Department of Psychology and Cognitive Science

Doctoral School in Cognitive Science

XXXIV Cycle

Fine-Grained Analyses of Early Autism-related Social Behavior in Real-World Scenarios by Machine Learning

Advisors:

Prof. Paola Venuti

Prof. Cesare Furlanello

Ph.D. candidate:

Gianpaolo Alvari

Academic Year 2020/2021

Acknowledgements

The work presented in this dissertation was the result of invaluable teamwork and mutual help from colleagues and friends. The first people I would like to acknowledge are my supervisors, Prof. Paola Venuti and Prof. Cesare Furlanello. They have guided and supported me throughout all of my research, even in the most challenging moments. I mostly owe them the skills and the experiences I have matured during these valuable years. I sincerely thank them for sharing their knowledge and mentorship, as well as for all the opportunities. I want to thank all the members of ODFLab (University of Trento) who allowed me to obtain the data used in my studies. In particular, I thank Dr. Arianna Bentenuto for her support, professionalism, and friendship. I also thank the members of the DSH research unit (FBK, Trento) for the chance to learn new computational skills. A special thanks to Luca Coviello for his dedication in helping me develop my ideas, and his precious contribution, especially in the third chapter. I want to thank all my colleagues from the doctoral school with whom I shared moments of fun and help. I would like to extend my gratitude to Prof. Marco Dondi for his advice and for sharing part of his research experience with me. Last but not least, I would like to sincerely thank all my family, friends, and Anna who have supported me and with whom I share all the joys and troubles of this journey.

Abstract

Autism Spectrum Disorder (ASD) is a condition that carries high costs for families and the healthcare system, requiring extensive management both in terms of diagnosis and treatment. The implementation of AI-based systems in clinical practice represents a possible supportive solution that can help clinicians by providing more systematic methods to monitor child behavior. The main advantage over more traditional observational approaches is to offer quantitative and refined analysis solutions that can be ecological at the same time. The relevance of AI in clinical applications can have a role both in the challenge of early detection and in designing intervention programs better tailored to the specific functioning of children with ASD. The research project presented in this dissertation focused on developing AI-based systems for fine-grained analysis of autism-related social behaviors and their validation in concrete clinical environments. Specifically, in Chapter 2, our first study is presented, which targets on implementing a computational phenotyping system to address the need for new early markers of the condition. Through fine-grained analytics of facial dynamics in videos, we identified a set of features that distinguished young (6-12 months) infants with ASD (18 ASD, 15 non-ASD) during unconstrained at-home interactions. In Chapters 3 and 4, we introduce EYE-C, a Behavior Imaging model for robust analysis of eye contact episodes in ecological therapist-child interactions. The system was validated in the clinical setting for personalized early intervention. First, we investigated the influence of extracted features in categorizing spectrum heterogeneity across a sample of 62 preschool (<6 years) children with ASD. Further, we tested our metrics as predictors of early intensive treatment outcomes in a sub-sample of 18 subjects with ASD. The project aims to demonstrate the feasibility of effective computational systems that are robust to the high variability of unstructured interactions, with emphasis on the applicative value in real-world scenarios. Even though based on limited sample sizes, the work presented may offer interesting insights into the perspective of integrating AI into clinical practice.
The research project was funded by an FBK scholarship and developed in a double internship at ODFLab (University of Trento) and the FBK Data Science for Health (DSH) research unit.

Contents

List of Abbreviations

ADTree Alternating Decision Tree

AI Artificial Intelligence

AIP Italian Association of Psychology

ANCOVA Analysis of Covariance

ANOVA Analysis of Variance

ASD Autism Spectrum Disorders

AU Action Unit

AUC Area Under the ROC Curve

BI Behavior Imaging

C Openface self-Confidence score

CE-CLM Convolutional Experts Constrained Local Model

CNN Convolutional Neural Network

CV Computer Vision

DL Deep Learning

EYE-C Eye-contact robust detection system

FACS Facial Action Coding System

GMDS-ER Griffiths Mental Development Scales

GMM Gaussian mixture models

GQ Global Developmental Quotient

HDBSCAN Hierarchical Density-Based Spatial Clustering of Applications with Noise

HOGs Histograms of Oriented Gradients

HV Home Video

LOOCV Leave-One-Out cross-validation

LSTM Long Short-Term Memory

MANCOVA Multivariate Analysis of Covariance

MANOVA Multivariate Analysis of Variance

MCC Matthews Correlation Coefficient

ML Machine Learning

MLR Multiple Linear Regression

MSE Mean Squared Error

NDBI Naturalistic Developmental Behavioral Interventions

PAFs Part affinity fields

PCA Principal Component Analysis

PDD Pervasive Developmental Disorders

RBB Repetitive and Restricted Behaviour

RF Random Forest

SA Social Abilities

SC Silhouette Coefficient

SI Simple Smile

SO Social Smile

SVM Support Vector Machine

SVR Support Vector Regression

T0	First evaluation
T1	Re-evaluation after intervention
TD	Typically Developing
UMAP	Uniform Manifold Approximation and Projection
VIF	Variable Inflation Factor

Chapter 1

Introduction

1.1 Autism Spectrum

Before we proceed with the discussion of Autism Spectrum Disorders (ASD), it is mandatory to provide an explanation about why we should define the spectrum as a disorder within the neurodiversity framework and not a mere pathology. There is currently an open debate among experts about the most appropriate way to define the autism spectrum and how to address all facets of its complex nature (Baron-Cohen, 2017; W. W. Lai and Oei, 2014). In this introductory part of the dissertation, this topic will be briefly exposed with an emphasis on the reasons for this choice of definition.

Autism refers to a construct used to define individuals with a specific pattern of behaviorally defined impairments in social communication, interactions, and repetitive behaviors. It is a critically heritable and complex neurodevelopmental condition with underlying cognitive attributes that often overlap with other disorders (Lord et al., 2020). In the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (APA, 2013), the autism spectrum is classified as a disorder with multiple levels of severity. Over the last few years, an interesting debate has been opened in research about the criticism of the DSM-5 categorization (W. W. Lai and Oei, 2014). From research, we learn that it makes more sense to define autism based on differences rather than just dysfunctions (Baron-Cohen, 2017; W. W. Lai and Oei, 2014). Social interaction and communication areas indeed represent aspects of fragility and disadvantage for individuals with autism, who need support and environmental restructuring. The notion of neurodevelopmental pathology strongly implies the need for a cure or treatment mischaracterization if not defined in the proper background (Baron-Cohen, 2017). From this perspective, it is more meaningful to define autism in a neurodiversity framework than a mere disease. The notion of neurodiversity also allows us to understand better the intrinsic characteristic of autism as a dynamic spectrum characterized by extremes ranging from individuals with minor impairments who adapt more easily to the environment up to individuals with severe behavioral compromises, requiring a more structural intervention. Defining people with ASD as individuals with a different level of socio-cognitive functioning better represents the autism spectrum both from a biological and an ethical standpoint.

Epidemiological studies suggest that autism has a variable prevalence based on the method of ascertainment employed and the sociocultural context (Lord et al., 2020). According to the 2010 Global Burden of Disease study, there are an estimated more than 52 million people with autism worldwide, corresponding to a prevalence of approximately 1 in 132 individuals (Lord et al., 2020). However, morbidity appears to be significantly higher in high-income countries, where it ranges between 1-2% (Baio et al., 2018; Maenner et al., 2020; Narzisi et al., 2020).

The rates of individuals diagnosed with ASD are overgrowing each year (A. Jacob et al., 2015; Lyall et al., 2017; Rogge and Janssen, 2019). Despite this, it is still unclear whether this growth is due to an actual increase of the condition, a shift in the diagnostic definition, improved public and medical awareness, or a combination of these factors (A. Jacob et al., 2015; Rogge and Janssen, 2019). In a recent systematic review, total annual expenses were estimated to reach over $460 trillion in the United States by 2025

(Leigh and Du, 2015) and over $2.4 million in lifetime costs per individual (Rogge and Janssen, 2019). Further, a child impairment generates a spill-over impact, decreasing the overall quality of life for the family (McCarty and Frye, 2020).

This massive impact of the condition both socially and economically has prompted greater attention to global challenges in autism research (Lord et al., 2020). First and foremost is the need to offer concrete solutions that can effectively lower costs, particularly for families. In parallel, it is necessary to offer services that can improve diagnosis and treatment in low-income countries, where medical awareness seems to be limited. Finally, global issues of stigma and integration must be taken into account (Lord et al., 2020).

1.1.1 Early Diagnosis

The diagnosis of ASD has a behavioral origin, based mainly on observation and description due to the lack of reliable biomarkers of biological or genetic anomalies. The current guideline for screening children with suspected ASD includes follow-up pediatric consultations with caregivers to gather a detailed medical history. When any suspect pattern of neurodevelopmental irregularities emerges, the child is referred to a specialist for in-depth evaluation through interviews, questionnaires, and psychological testing (related to both general cognitive development and social skills). Due to the complexity of acquiring reliable data from a subjective observational assessment, diagnostic training demands an extensive background of knowledge about child development and several years of training by specialized clinicians (Dawson and Sapiro, 2019).

Diagnosing autism in childhood is still an ongoing concern for the clinical setting as it requires long-term monitoring and detailed examination of the patient symptoms, which is costly and time-consuming. As the medical resource deficiency stands, it is very unfavorable for the clinical management of the condition (K. Campbell et al., 2019). Furthermore, through studies of developmental trajectories and intervention outcomes, it has become evident that early detection is a major challenge in autism research (Rogers and Talbott, 2016; Vivanti et al., 2019). From a pragmatic standpoint, being able to recognize the condition earlier in the development gives the chance to gain increased impact from specialized interventions and thus enhance the prospect for cognitive improvement and overall quality of life more in general (Dawson, 2008; Mozolic-Staunton et al., 2020; Rogers and Talbott, 2016; Vivanti et al., 2019; Waddington et al., 2016). Intervening early during cognitive maturation exploits brain plasticity, anticipating the full phenotype expression, and hopefully containing in some degree the atypical trend of cognitive development (Elder et al., 2017). There is now a consensus among experts that the first *red flags* of atypical socio-communicative development emerge rather evident from the second year of life (Barbaro and Dissanayake, 2013). Despite this, the median age of clinical diagnosis in high-income countries is still close to about 4 years of age (Brett et al., 2016). Nonetheless, several studies have reported that parents typically raise concerns about specific behaviors much earlier at 2 years of age (Lord et al., 2020). Reasons for this gap fall mainly into two domains. The primary cause is likely to be related to the condition inherent characteristics of being extremely heterogeneous, especially in its early manifestations. Within childhood, patterns of atypical developmental trajectories seem to vary between individuals and emerge at different stages of maturation (Lombardo et al., 2019; Zwaigenbaum et al., 2015). In addition, they frequently manifest through subtle signs that are difficult to detect by non-experts (Dawson et al., 2018). Second, socioeconomic barriers to early detection have emerged in healthcare systems worldwide, especially in disadvantaged environments where awareness of the condition is limited and families have restricted access to public health services (Eapen et al., 2013; Mozolic-Staunton et al., 2020). The challenge for autism research is to address these weaknesses by identifying reliable early markers and offering scalable, low-cost solutions.

Overall, early detection of prodromal symptoms has proven challenging. Atypical developmental traits can be identified in the second year of life when children with ASD display more marked socio-communicative impairments and stereotyped behaviors (Barbaro and Halder, 2016; Shic et al., 2014; Varcin and Jeste, 2017; Zwaigenbaum et al., 2019). At 12 months of development, infants with ASD exhibit interaction-related deficits in their first social exchanges with their parents. These impairments include a general poor expression (Esposito and Venuti, 2009; Shic et al., 2014; Zwaigenbaum et al., 2015), reduced eye contact (Barbaro and Dissanayake, 2013; Rozga et al., 2011), odd object play (Ozonoff et al., 2008; Wilson et al., 2017), a decreased responsiveness in episodes of joint attention (Rozga et al., 2011) and reduced capacity of name orientation (Nadig et al., 2007; Varcin and Jeste, 2017). In addition, there is evidence of early deficits in motor control (Esposito et al., 2011; Flanagan et al., 2012; Gima et al., 2018).

Therefore, the first years of growth mark a critical stage for cognitive development within the social-communicative domain. Social skills mature to become a crucial basis for communication throughout this period. The infant will start to actively employ these socio-cognitive resources to interact with his/her environment and accumulate positive interactive experiences. From studies about early detection, we can assume that in the autism spectrum, the first markers emerge at this stage (E. J. Jones et al., 2016). It is critical to provide the toddler with adequate tools to engage with the social environment as soon as possible. Therefore a major goal is to lower the age threshold for screening and anticipate the onset of severe symptoms (Chevallier et al., 2012; Landa, 2018).

Moreover, surveying signs of atypical behaviors within the first years of development may uncover the primary components of later autistic phenotype (Zwaigenbaum et al., 2019). Studying behavior so early, when impairments are less marked, is challenging yet represents a necessary step for supporting the clinical setting. At such a low age, behaviors can be subtle and difficult to analyze, so we need alternative solutions that are more systematic and fine-grained. Overcoming this weakness would allow clinicians to establish a more effective screening and support for the families.

1.1.2 Early Intervention

So far, we have emphasized early detection. However, this is just the starting point. Typically, the diagnostic effort has no clinical relevance if not to promptly launch a new intervention pathway. Early intervention is the real priority for clinicians. Young children with autism struggle to communicate and interact with other individuals, severely constraining their opportunities to engage in pleasant social experiences (Lord et al., 2020). In parallel, parents may find their child behaviors distressing and difficult to regulate, frequently leading to the establishment of maladaptive relationship patterns. Therefore, beginning a rehabilitative program timely is critical. Interventions for young children on the autism spectrum may result in significant improvements in cognitive, interactive, and communicative skills and eventually an overall reduction in symptom severity (Elder et al., 2017). It is crucial to provide the child with ASD as early as possible with adequate compensatory strategies to interact most adaptively with the social environment and minimize the loss of experience.

To date, there is no universally accepted intervention framework for autism (Jaliaawala and Khan, 2020). Challenges in developing universally approved rehabilitative models relate, on the one hand, to the underlying nature of the spectrum, on the other to the lack of reliable tools to evaluate the outcome. There is a high degree of variability in the timing and manifestation of the first prodromal symptoms (Lombardo et al., 2018; Zwaigenbaum et al., 2015). For many individuals, the onset of the first clinical manifestations could be early recognizable, whereas, for others, the phenotype may occur after an apparent period of typical development (Lord et al., 2020). In the latter, the child appears to develop in a typical manner but afterward starts to lose communication and social competencies, leading to a *regression* or stasis in behaviors (Boterberg et al., 2019; Lord et al., 2020; Pearson et al., 2018). Further, this pattern of variability in the onset is often associated with co-occurring difficulties or disorders (Havdahl and Bishop, 2019;

M.-C. Lai et al., 2019; E. Stevens et al., 2019). Studies on spectrum stratification have also reported a high degree of variance in condition progression and symptom severity (Fountain et al., 2012; Gotham et al., 2012; Venker et al., 2013; Wolfers et al., 2019). Among the most impactful effects of this heterogeneity are also the irregular response to the same kind of intervention and the variance in the outcome (Eapen et al., 2013; Lord et al., 2020; Uddin et al., 2019). Studies about the effectiveness of early intervention have indeed found evidence of overall improvement, but also wide inter-individual variance in the response (Bentenuto et al., 2020; Magiati et al., 2011). Children who are high-responders seem to achieve significant progress, while other low-responders are unlikely to gain much improvement (Bentenuto et al., 2020; Eapen et al., 2013). Several prodromic factors related to the baseline pre-treatment conditions, i. e. age, cognitive functioning, the severity of symptoms, and social skills (Ben-Itzchak et al., 2014; Magiati et al., 2011; Swallows and Graupner, 2005), seem to play a role in the definition of developmental trajectories, but their specific role needs further exploration (Bentenuto et al., 2020).

For these reasons, it is challenging to develop a standard intervention model that is generalizable and that maximizes outcomes for all patients across the autism spectrum (Estes et al., 2019; Rogers and Talbott, 2016). From this perspective, the necessity to design individualized intervention programs that account for variability in developmental trajectories and individual response to intervention has grown. Interest in precision psychiatry is growing fast in cutting-edge clinical environments. The demand for more effective models of therapy stems from the awareness that tailoring interventions to the patient individual characteristics may optimize the outcome.

Also of concern is the lack of systematic, autism-specific measures of the treatment response. In clinical practice, standard diagnostic tools are used to monitor the long-term improvement of a child with ASD. However, such tools are only structured for diagnostic assessment and not sensitive enough to measure improvements in the outcome of an intervention pathway (Bentenuto et al., 2020; Washington, Leblanc, Dunlap, Penev, et al., 2020). To assist the clinical setting, research efforts should address the need for standardized and shared systems among professionals for the evaluation and decision-making of interventions. Systematic behavioral correlates are necessary to step forward and help clinicians tailor the intervention according to children characteristics.

1.2 Machine Learning applications in ASD research

As noted in the previous sections, the clinical evaluation process is costly and time-consuming. Traditional research has been able to provide a positive contribution in trying to facilitate this process, increasing awareness of the prodromal stages of the autism spectrum and facilitating the link between families and health services (Lord et al., 2020). Still, not all shortcomings have been addressed, and many challenges remain. The most controversial aspects are inherent in the evaluation process, which, being observation-based, is bound to a subjective bias. In addition, the call for highly qualified specialists still entails long waiting lists for families (especially in low-income communities) and significant problems in the exportation of these methodologies in more ecological frameworks, such as school or at home. Finally, the challenges of early detection and a shortage of uniform guidelines are leading to a lack of large datasets collection (Dawson and Sapiro, 2019).

Therefore, the development of computational approaches for systematically analyzing behaviors is an important goal in autism research. Artificial Intelligence (AI) offers novel solutions for automated analysis of human behaviors. In recent years, the implementation of AI has scaled up substantially in behavioral research, with promising results in the context of ASD as well. Advanced Machine Learning (ML) models offer the potential to overcome many of the limitations of traditional observational research, offering alternative methodologies to provide support in studying and monitoring symptoms (Dawson and Sapiro, 2019; de Belen et al., 2020; Hyde et al., 2019; Sapiro et al., 2018).

The applications of computational psychology have largely proceeded towards three separate goals, attempting to address the major open issues in autism research: providing lower-cost tools for screening through classification, seeking to analyze subgroups on the spectrum, and providing alternative systematic methods for measuring behavior. Many studies have focused on (1) the implementation of supervised ML intending to speed up the diagnostic process, either by reducing the dimensionality of tests and questionnaires or by predicting the diagnosis based on data collected during the assessment (Cavus et al., 2021; Hyde et al., 2019); (2) a far smaller corpus of papers (mainly due to the lack of enough numerous samples) has been concerned with the unsupervised analysis of larger datasets, to investigate subgroups and heterogeneity of ASD (Wolfers et al., 2019). Lastly, (3) the most promising trend is represented by Behavior Imaging (BI), which is a technique used in medical contexts for the analysis and monitoring of behavioral disorders through the automated analysis of images or videos, thanks to advanced Computer Vision (CV) algorithms (de Belen et al., 2020; Sapiro et al., 2018). The first two objectives (1,2) represent advanced methods of data analysis, generally based on clinical data (psychological testing or parental reports) from the evaluation procedure of children with ASD. In contrast, the latter trend (3) represents an alternative approach to data collection no longer based on traditional standardized instruments from the clinical evaluation but on systematic measurements of old/new behavioral markers through CV models. The following sections will discuss each of these approaches in more detail.

1.2.1 Supervised methods

The development of algorithms aimed at fast classifying ASD symptoms and reducing the overall length of the diagnostic process was among the first applications of supervised ML into the clinical setting. Supervised approaches allow classification and prediction of autistic traits based on ground truth labels. One of the pioneering research groups in this field is the lab headed by Dr. Dennis P. Wall, from Stanford University (California, USA). In one initial study, Wall, Dally and colleagues (2012) attempted to accelerate the assessment process by mining a subset of items from the ADI-R that would be enough to classify ASD correctly. They collected a large dataset of 891 individuals with ASD and 75 with typical development. Using only 7 of the 93 test items, an Alternating Decision Tree (ADTree) classifier achieved nearly 100% accuracy and over 90% specificity (Wall, Dally, et al., 2012). In a related study, they tested the same methodology for the ADOS in a smaller sample. They reached almost 100% accuracy and 94% specificity by combining 8 of the 25 items in Module 1 while reducing the overall test length by near 70% (Wall, Kosmicki, et al., 2012).

Duda and colleagues (2014) further pursued validation of the eight-item classifier from (Wall, Kosmicki, et al., 2012) in a larger sample of 2333 children with ASD and 283 controls at younger ages. The algorithm showed significant correlations with both ADOS and ADOS-2 scoring (Lord et al., 2012), achieving nearly 96%, 98% sensitivity, and 84% specificity (Duda et al., 2014). A few years later, Duda and colleagues (2016) attempted to replicate the same results in distinguishing children with ASD (n=2775) and ADHD (n=150). Authors tested six different ML models and found that 5 of 65 Social Responsiveness Scale (SRS) scores were sufficient to distinguish between groups with an AUC=0.97. The authors then expanded the results to include a novel crowdsourced dataset to improve the classifier generalizability to real-world data. Pooling the two samples, they developed an algorithm that could correctly distinguish ASD from ADHD with only 15 items from the SRS, achieving an AUC score of nearly 0.9 (Duda et al., 2016).

In parallel, from the same research group, Levy and colleagues (2017) employed different ML models to diagnose children with ASD based on ADOS scores in two different modules: Module 2 (1319 ASD, 70 typical development) and Module 3 (2870 ASD, 273 typical development). By employing less than 10 features, they were able to achieve an AUC=0.95 for Module 3 and an AUC=0.93 for Module 2 (Levy et al., 2017).

More recently, the Wall Lab has been focused on designing low-cost systems for rapid screening of children on the autism spectrum, combining mobile-based technologies and psychological testing.

Tariq and colleagues (2018) developed a mobile web portal to collect 3-minute home videos of children with (n=116) and without (n=46) ASD. The videos were evaluated by several blinded non-expert raters based on 30 behavioral variables taken from the ADOS. The authors then employed several models for ASD classification over the 30 features, resulting in the highest performance with an AUC=0.92 (Tariq et al., 2018).

Abbas et al. (2018) merged two stand-alone classifiers to generate an ASD screening tool based on parental questionnaires and home videos. ADI-R and ADOS scores were collected to train 2 Random Forest classifiers for questionnaires (2299 ASD, 100 typical development, 287 other conditions) and videos (3310 ASD, 585 typical development, 364 with other conditions) respectively for predicting clinical diagnosis. Pooling the scores from both models outperformed traditional screening tools, including the Modified Checklist for Autism in Toddlers (M-CHAT; Robins et al., 2014) and the Child Behavior Checklist (CBCL; Achenbach and Dumenci, 2001) (Abbas et al., 2018).

Likewise, in a large recent study, Washington and colleagues (2021) further explored a similar system for early screening (<8 years of age) of ASD starting with non-expert raters. A novel system was implemented for crowdsourcing and identifying a pool of reliable evaluators to extract features (based on ADOS items) from 50 naturalistic videos (25 ASD, 25 controls). Two logistic regression classifiers were employed to assess the accuracy of the non-expert raters, yielding an AUC score above 0.98. The authors then compared performances across the same videos manipulated (by drawing a box over children faces, via OpenCV toolkit) for privacy-preserving, achieving similar results (Washington et al., 2021).

Overall, very promising outcomes emerged from this research stream. The contribution of ML models has major potential in lowering the high dimensionality of delayed psychological testing while still preserving a good degree of accuracy in discriminating children on the autism spectrum (Duda et al., 2014; Hyde et al., 2019; Kosmicki et al., 2015; Levy et al., 2017; Wall, Dally, et al., 2012; Wall, Kosmicki, et al., 2012). Moreover, when combined with questionnaires or checklists, ML also holds a strong contribution in developing more effective screening strategies (Abbas et al., 2018; Tariq et al., 2018). Within the domain of early detection, more recent studies have also shed light on the possibility of designing mobile-based frameworks that rely on the involvement of untrained raters, paving the way for translational at-home support solutions (Tariq et al., 2018; Washington et al., 2021). It seems that supervised ML actually offers the possibility to lighten the workload of lengthy clinical evaluations by making processes shorter and based on fewer features (Hyde et al., 2019). However, despite promising results, the impact is limited. The real objective is not only to facilitate diagnosis by reducing the dimensionality of psychological tests but rather to find new behavioral indicators to anticipate the screening process. Retaining data collection through traditional methods (psychological testing for socio-cognitive functioning and parental reports) is unlikely to lower the screening threshold for children on the autism spectrum, representing one of the most critical objectives in the diagnostic framework.

1.2.2 Unsupervised methods & heterogeneity

The second important AI approach in the context of autism research is unsupervised ML. In contrast to supervised, unsupervised approaches aims at analyzing data without the need for labels, but only based on the data structure. This method is perfect for facing heterogeneity within the autism spectrum, which is a persistent challenge for ASD research. With the term heterogeneity is intended the occurrence of subgroups within the condition that differ possibly functionally from each other. Various studies attempted to tackle this issue, trying to identify homogeneous and clinically relevant subgroups. Yet, the findings are more irregular and less supported among the experts (van Rentergem et al., 2021). It has been argued that stratification of the autism spectrum might actually be self-inflicted along with the progressive tendency to define the condition based on a spectrum of autistic-traits rather than well-defined and specific symptoms (Mottron and Bzdok, 2020). The inclusion of attenuated phenotypes (especially in the DSM-5) would have made delineation of the condition more difficult rather than easier. Symptoms that appeared more specific years ago would have faded away due to the “grinding” of autism into a spectrum of traits (Mottron and Bzdok, 2020). Debate is still open, nonetheless, since though broadening the spectrum has led to increased heterogeneity, it also led to improved overall clinical management of the condition (both in diagnosis and treatment) by increasing awareness and yielding a more valuable definition of the autism spectrum (Baron-Cohen, 2017; W. W. Lai and Oei, 2014; Lombardo et al., 2019). Beyond all the discussion, it is important to focus on an applicative perspective to offer concrete assistance to the clinical setting. For many reasons, it is key to investigate subtypes. Primarily, if we can recognize individuals in terms of multiple (stratification), we might better understand prognostic implications and provide ad-hoc support. Also, tracking down consistent sub-phenotypes might facilitate the detection of behavioral markers, which in turn could aid in early detection and intervention (Mottron and Bzdok, 2020). Unsupervised ML techniques offer optimal solutions for handling this challenge. ASD stratification via clustering algorithms provides the opportunity to identify behavioral patterns through a data-driven approach, in the absence of explicit labels. Typically, the procedure is based on the analysis of behavioral traits and the recognition of multiple clusters, which are then externally validated through measures of clinical presentation (Lombardo et al., 2019; van Rentergem et al., 2021; Wolfers et al., 2019).

Along with this perspective, Hu and Steinberg (2009) used several clustering algorithms to analyze a large dataset of individuals with ASD (n=1954) based on ADI-R scores.

The combined analysis yielded 4 distinct phenotypic clusters with different levels of symptom severity: severe language deficits, milder deficits in other domains (e.g., social development, interests and behaviors, nonverbal communication, and play skills), higher frequency of savant abilities, intermediate severity in general (Hu and Steinberg, 2009). In work by Lombardo and colleagues (2016), an analysis of a mixed dataset (694 ASD; 249 controls) revealed 5 discrete subgroups of ASD that differed based on performance on an emotion comprehension task (Reading the Mind in the Eyes Test, RMET). Notably, 3 of the 5 subgroups included most adults with ASD who showed a major impairment on the task, whereas the other subgroups were generally without any impairments (Lombardo et al., 2016).

A further alternative way to assess autism spectrum stratification is through developmental trajectories (E. Stevens et al., 2019). Fein and colleagues (1999) conducted a clustering analysis on a sample of preschool-aged children (n=194) with Pervasive Developmental Disorders (PDD). Two major subsets of high and low functioning emerged from the analysis, distinguished by dissimilar levels of functioning in the socio-cognitive domain. Participants were then re-evaluated at school-age while maintaining stable subgroup membership. Upon assessing developmental trajectories, high-functioning children showed greater overall improvement in the socio-cognitive domain, while low-functioning children showed little to no progress (Fein et al., 1999).

Similarly, Stevens and colleagues (2000) employed hierarchical clustering to explore a sample of children with ASD (n=138). Based on social, language, and nonverbal skills measurements, 2 low- and high-functioning clusters again emerged. The developmental trajectory of low-functioning children remained stable over time, while the high-functioning group diverged (M. C. Stevens et al., 2000).

More recently, Pickles and colleagues (2014) in a longitudinal study investigated the progression of language abilities in a sample (n=192) of individuals with ASD over 17 years. Patterns of development between 2 and 6 years proved to be strongly heterogeneous with significant variance in progress, with some groups showing improvement or decline in verbal skill progress. Conversely, after 6 years of age, the trajectory appeared uniform (Pickles et al., 2014).

In addition, Lord and colleagues (2015) furthered this line of research by following the developmental trajectories of a sample of children with ASD (n=85) from age 2 through adulthood. The authors analyzed parent-reported social impairments, adaptive social functioning, and stereotypical behaviors. Three groups emerged by the outcome, which included young adults with cognitive delays and persistent autistic traits, individuals with borderline or average cognitive abilities who continued to have socio-cognitive difficulties related to the autism spectrum, and a small group of young adults who adapted socially in later life despite being diagnosed with ASD in childhood (Lord et al., 2015). Overall, although several studies tackled the heterogeneity of the autism spectrum, no emerged subgroups have been well replicated. Unfortunately, as noted above, the lack of large datasets is suboptimal to investigate heterogeneity (Lombardo et al., 2016). Subtypes of ASD may contribute to variation in response to intervention and early onset of symptoms. This requires clustering analyses that are strongly geared toward providing support rather than identifying subgroups for their own sake.

To the best of our knowledge, there is only one example in research that combined unsupervised ML to identify subgroups and the evaluation of the intervention response. Stevens and colleagues (2019) employed Gaussian mixture models (GMM) and hierarchical clustering to identify behavioral phenotypes in a large sample (n=2400) of children with ASD. The sample was explored based on a comprehensive measure of over 3000 abilities in 8 domains: including social, language, adaptive, academic, cognitive, executive functioning, play, and motor skills. The initial application of GMM revealed 16 subgroups. Afterward, the subgroups were further analyzed by hierarchical clustering, which revealed 2 overlying behavioral phenotypes (high-/low-functioning) with unique impairment patterns, each comprising multiple in-groups that differed across deficits severity. In addition, differentiated response to treatment was found across subtypes, which showed distinct treatment response profiles when modeled with linear regression (E. Stevens et al., 2019). Participants included in clusters within the highest-functioning subgroup capitalized more rapidly across the treatment compared to participants included in the low-functioning subgroup (E. Stevens et al., 2019).

Altogether, the results of this study seem to support the idea that there are two main groups of functioning in the autism spectrum, in line with what has been found in previous studies on smaller samples (Fein et al., 1999; M. C. Stevens et al., 2000). Further exploration is needed to pursue this applicability perspective and to try as much as possible to offer more concrete solutions (Lombardo et al., 2019; E. Stevens et al., 2019).

Implementing data-driven systems has proven to be the right way to address the challenge of ASD heterogeneity. However, there is still a lack of consistency in the results. Challenges may arise not only from the shortage of large representative datasets but also from the lack of adequate behavioral indicators that are sensitive to differences between subgroups. Expanding the supervised case, the availability of new stratification methods based on unsupervised ML can help change the data collection structure, look for new, more systematic indicators, and detach from psychological testing.

1.2.3 Towards fine-grained analysis using Behavior Imaging

Behavior Imaging probably represents the AI application with the potential to contribute the most to the clinical setting by providing alternative ways to quantify behavior. Several of the studies discussed in Sections 1.2.1/2 already implemented ML models, which rely on providing data-driven methodologies to parse clinical measures or features collected through observation of behavior and make predictions (supervised) or identify new subgroups (unsupervised). The problem with these approaches is that they depend on variables collected during the diagnostic process or otherwise within the framework of psychological testing. As mentioned in previous sections, assessment is largely based on observation and thus suffers from limitations of subjective methodologies of analysis (Dawson and Sapiro, 2019).

In addition, such metrics are frequently not sensitive enough for detecting subtle irregularities on the autism spectrum (Dawson et al., 2018; Dawson and Sapiro, 2019; Sapiro et al., 2018). As compelling and certainly valuable as analyzing large data collections through ML algorithms, however, similar approaches inherit the limitations of the methodology employed in data collection. ML offers new insights into the inter-data relationships that may be missed, yet the benefit may remain constrained if data is collected through outdated methodologies. The importance of developing alternative solutions for behavior analysis has been emphasized multiple times in the literature (Dawson et al., 2018; Dawson and Sapiro, 2019; de Belen et al., 2020; Sapiro et al., 2018).

Despite standardized methods for clinical observation still being widely used, they suffer from significant limitations. These include the use of subjective rather than objective methods, poor precision of metrics, the requirement for highly trained professionals, and difficulties in translating these practices into naturalistic settings, such as at-home (Dawson and Sapiro, 2019).

Behavior Imaging relies on CV algorithms, which have the potential to solve each of these challenges by offering systematic solutions for behavior analysis through video, which are low-cost (no need for trained experts), automated, non-intrusive, fine-grained, and translational. In recent years the implementation of CV models has increased exponentially, becoming the best practice for several research frameworks, in particular within the healthcare sector (de Belen et al., 2020). Computational approaches provide quantitative metrics, letting clinicians represent behaviors by a dynamic pattern that ranges on a scale of a few milliseconds (Dawson and Sapiro, 2019).

Within psychiatry, Behavior Imaging implementation is growing but still maturing. In autism research, the emphasis of deployment is both for early detection, CV techniques have been effectively implemented to detect and analyze early behavioral markers automatically, and discover new ones, as well as in the intervention framework through symptom monitoring and assistive technology, even if to a lesser extent (Dawson and Sapiro, 2019; de Belen et al., 2020; Jaliaawala and Khan, 2020; Sapiro et al., 2018; Washington, Park, et al., 2020). Multiple categories of behavioral markers referring to both social-communicative and motor aspects have been considered: including eye gaze, attentional patterns, affective/facial expressions analysis, motor stereotypies, head movement, and movement patterns (de Belen et al., 2020; Sapiro et al., 2018).

The collaborative efforts of Dr. Geraldine Dawson and Dr. Guillermo Sapiro laboratories from Duke University (North Carolina, USA) are one of the most successful actions towards developing application-oriented systems (Sapiro et al., 2018).

In one of the first joint projects, Hashemi and colleagues (2014) attempted to develop a non-invasive system for early screening of children with ASD to aid in early risk recognition. They implemented a system based on CV and the Autism Observation Scale for Infants (AOSI, Bryson et al., 2008) to measure response in visual attention tasks by tracking facial dynamics. Exploring behavior in a small sample of at-risk children between 5 and 18 months (n=12) showed promising results. The authors examined frame-by-frame videos of playful clinician-child interaction, focusing on periods of attentional disengagement and visual tracking performance. The high temporal resolution (30 frames per second) pulled out possible latent patterns of head movements. Children with ASD displayed a delay in attentional disengagement and tracking compared to AOSI scores, consistent with ratings by experienced clinicians. In addition, participants showed a pattern of piecewise constant lateral head movement and less smooth movement that was almost impossible to recognize with the naked eye (Hashemi et al., 2014).

Within the same project, Egger and colleagues (2018) further developed a mobile-based technology to collect videos of young children with ASD in a non-clinical environment. The goal was to analyze children facial behavior while watching a video designed to elicit social autistic traits. The authors collected a large sample of toddlers between 12 and 72 months of age (n=1756) with over 4000 videos in a naturalistic environment. The sample was composed of toddlers at high risk of autism, detected based on parental reports and screening results (M-CHAT scores), and low-risk controls. Facial dynamics were automatically encoded by an AI-based tool developed by the authors. For each video, the percentage of neutral, positive, and negative expressions was quantified. High-risk children produced significantly more neutral expressions and a lower mean percentage of positive ones, controlling for age and sex. Findings demonstrated the relevance of an app-based framework to collect at-home ecological data and the potential of computational analysis from quantifying expressive and attentional features (Egger et al., 2018). Furthermore, this study confirmed the importance of studying child emotions in naturalistic settings.

Similarly, Campbell and colleagues (2019) compared typical and atypical attentional patterns in an answer-to-name task by tracking head movements from video. The authors collected a sample of more than 100 children between 16 and 31 months of age (22 ASD, 82 controls) while watching video stimuli on a tablet and built-in camera recording their motions. A CV-based algorithm was designed to examine the response to name calls, and the validity of the algorithm was tested against human coding with excellent results. Only 8% of children with ASD directed attention in response to name calls during the task, compared to 63% in the typically developing group. Differences emerged in latency, which was significantly longer in children on the autism spectrum; in addition, older children with ASD showed less attention to video in general (K. Campbell et al., 2019).

Carpenter and colleagues (2021) recently furthered their analysis of the same sample from (K. Campbell et al., 2019), investigating expressive facial dynamics. The authors designed a computational model to automatically detect and track facial landmarks, employed to estimate head pose and capture facial expressions. Specific sequences were identified throughout the videos, differentiating the two groups according to facial dynamics (AUC ranging from 0.62 to 0.73). Overall, children with ASD displayed more frequently neutral expressions, compared to typically developing controls (Carpenter et al., 2021).

Furthermore, the same research group provided interesting new applications within the motor domain (Sapiro et al., 2018). Martin and colleagues (2018) conducted a comparison study among children with (n=21) and without (n=21) ASD on head movement dynamics in relation to stimulus sociability. Children were frontally recorded while watching a 16-min movie, including social and non-social stimuli. To analyze head movements, 3 variables were evaluated: head nods, head turns, lateral head inclinations. For each video frame, the variables were coded via a CV algorithm that can capture a dense 3D shape in real-time (Jeni and Cohn, 2016). Children with ASD exhibited a greater degree of head-turning and a greater speed of inclination. In addition, these differences were found to be condition-specific to social stimuli (Martin et al., 2018).

Dawson and colleagues furthered these analyses by testing the sample used in (K. Campbell et al., 2019). Rates of spontaneous head movements during active attention states were coded using a CV-based system capable of detecting facial landmarks and estimating head pose angles relative to the camera. Consistent with findings from previous studies (Martin et al., 2018), toddlers with ASD displayed a greater degree of head movements, expressing difficulties in maintaining the head in midline position while engaged in attentional tasks (Dawson et al., 2018).

Literature related to the implementation of behavior imaging in autism research has grown steadily in recent years. Studies have contributed valuable insights into various aspects of atypical behavior. Promoting evidence has emerged regarding the classification of emotions and facial expressions as an important aspect of screening the monitoring of ASD, which commonly highlights a reduced expressiveness in children on the autism spectrum (de Belen et al., 2020; Drimalla et al., 2021; Guha et al., 2016; Kalantarian et al., 2018; Leo et al., 2018; Li et al., 2019; Manfredonia et al., 2019; Samad et al., 2019; Tang et al., 2019). Also, CV models have been employed to predict the social engagement intensity (Rudovic et al., 2018) and affect coordination (Zampella et al., 2020) of children with ASD. Further applications are also provided for eye gaze and attentional skills more in general, which showed a difference in patterns between children with and without ASD by computational video analysis and digital technologies (i.e smart-glasses) (Chong et al., 2017; de Belen et al., 2020; Higuchi et al., 2018; Jiang and Zhao, 2017; Liaqat et al., 2021; Liu et al., 2016; Liu et al., 2015; Xu et al., 2020). Results also emerged in the domain of motor abilities, such as differences in grasping (Zunino et al., 2020), body pose (Vyas et al., 2019), and detection of stereotyped movements (Negin et al., 2021; Rad et al., 2015).

Overall, the literature in this domain has demonstrated the enormous potential of Behavior Imaging in analyzing the dynamic nature of childhood behaviors during social interactions by providing a high resolution and continuous measure (Sapiro et al., 2018). The major benefit comes from the potential to deliver alternative metrics for data collection upon which supervised or unsupervised ML models can be implemented for early classification or stratification.

Concerning early detection, even though there is still no truly concrete solution, computational approaches have managed to gain popularity. However, within the intervention area, the progress of AI-based applications in autism research is still minimal (Jaliaawala and Khan, 2020). Some studies have tested the use of CV-based technologies to support intervention, in particular by giving feedback real-time on expressiveness to improve the recognition and production of facial expressions of children on the autism spectrum (Cockburn et al., 2008; de Belen et al., 2020; Golan et al., 2010; Hopkins et al., 2011; Tsai and Lin, 2011).

For example, Pan and colleagues (2015) developed an automatic emotion recognition system implemented in a robot-child interaction for the treatment of ASD. Their results suggested that computational analyses could help to improve the effectiveness of behavioral analyses during the intervention. Similarly, Harrold and colleagues (2014) developed a mobile-based framework that allowed children to learn to produce emotions with real-time feedback on performance (Harrold et al., 2014). Jain and colleagues (Jain et al., 2020) proposed an interactive game to support autism intervention for non-verbal abilities.

In addition, Rudovic and colleagues (2018) designed a multimodal system for recognizing a child affective states in a multimodal robot-assisted therapy setting. To improve the system, the authors additionally considered each child contextual information to tailor the framework outperforming non-personalized methods in recognition of engagement and affect (Rudovic et al., 2018).

Despite the increasing appeal of such applications, research has focused primarily on diagnosis (de Belen et al., 2020; Uddin et al., 2019). The implementation of AI-based methods in the treatment of ASD is much recent and still at an experimental stage. The integration of BI solutions in the therapeutic context could be a valuable step in the effort to achieve a joint approach among professionals. Yet, we are missing variables, perhaps subtle, that can be extracted automatically and have the power to predict the intervention outcomes and support the therapists in monitoring the progression of the symptoms. Thus although AI-based systems have considerable clinical relevance, they continue to suffer from limited translational applications (Jaliaawala and Khan, 2020). The analysis of intervention sessions is a very challenging domain that requires much flexibility, as maintaining a non-invasive approach is essential not to compromise the therapist-child interaction.

Further, fine-grained visual analysis (i.e., body or gaze tracking, facial expression recognition) is limited by real-life highly dynamic situations that often include occlusions or low signal quality, as with ASD intervention. In this regard, recent AI applications in the treatment scenario are still driven by heavy interaction structuring (de Belen et al., 2020;

Voss et al., 2019). As much as these approaches offer more robust performance, they do not scale to more naturalistic contexts that allow for generalizable results (Leclère et al., 2016).

1.3 Motivations for AI-based fine-grained analysis in ecological contexts

The autism spectrum is a very challenging, multifaceted condition. The spectrum is subtly and unevenly manifested in the initial stages, making early detection highly difficult, even when using the well-known standardized screening tools (Dawson Sapiro, 2019). It is a disorder with large individual variability and heterogeneity that prevents the further development of a uniform standardized management system. As a final point, there is a lack of predictive tools that can support the therapists in monitoring the intervention and maximizing the progress of children on the autism spectrum.

AI-based models are well suited to tackle the challenges of autism and redefine the diagnostic and intervention process. Data-driven analyses can provide interesting insights into aspects of heterogeneity (Lombardo et al., 2019; van Rentergem et al., 2021; Wolfers et al., 2019) and computational phenotyping offers alternative, more systematic solutions to extract behavioral variables that can serve as early markers or for monitoring ASD (Dawson and Sapiro, 2019; de Belen et al., 2020; Jaliaawala and Khan, 2020; Washington, Park, et al., 2020). A very interesting perspective is the rule-based classification paradigm, which brings the two aspects together, involving the use of CV models to extract relevant features and then apply a set of quantitative criteria for classification or clustering (Washington, Park, et al., 2020).

Another critical dimension is the feasibility of obtaining systematic measurements in a non-intrusive way, allowing to analyze behavioral aspects in ecological and naturalistic contexts, which is an essential aspect in the autism spectrum research (Dawson and

Sapiro, 2019; Hashemi et al., 2018; Sapiro et al., 2018).

Combined, all these aspects can upgrade the clinical infrastructure and provide a valuable support tool for both clinicians and families. On the diagnostic side, they have the potential to provide high-precision digital measurements to detect subtle early indicators that may be missed by the naked eye, and to support non-intrusive, perhaps remote, screening systems. In terms of intervention, they could help therapists identify precise measures to quantify the severity of symptoms and deliver care at the right time and frequency, perhaps through online feedback during intervention sessions. Moreover, AI-based solutions could help integrate aspects of heterogeneity paving the way for personalized treatment based on individual differences in the response (Lord et al., 2020; Rudovic et al., 2018; E. Stevens et al., 2019; Uddin et al., 2019). This may assist clinicians in optimally delivering treatment based on (1) quantitative data, (2) fine-grained analysis impossible to be carried out by humans involved in the social interplay, and (3) real-time feedback. From a clinical standpoint, this could allow for the implementation of personalized interventions aimed at maximizing efficacy in terms of developmental outcomes, symptom severity reduction, and adaptive functioning (Jaliaawala and Khan, 2020; Rudovic et al., 2018; E. Stevens et al., 2019). Finally, in a larger perspective, digital interventions and automated screening tools would allow the development of more objective assessment systems, which can reduce the bias for subjectivity and provide low-cost, non-invasive services that have the potential to lower health system expenditures and access families also in low-income countries (de Belen et al., 2020; Lord et al., 2020).

Finally, AI-assisted clinical practice is the new frontier. Unfortunately, clinicians and families are still generally unaware of the supportive potential of these technologies. Indeed there is a critical lack of systems that can be adopted for real-world scenarios. The shortage of effective concrete applications still prevents generalizability and filling the gap between the research community and the medical and psychological staff (Jaliaawala and Khan, 2020). Future research will have to push for empirical validation of these methods and develop more wide-scale concrete solutions. Therefore to address the problem, researchers need to design methods for AI-based systems that can amplify, augment, and enhance human performance in ways that make systems reliable, safe, and trustworthy (Jaliaawala and Khan, 2020).

This doctoral project aims precisely at attempting to move the first steps towards a more homogeneous integration of clinic and research. We tried to develop more translational applications by employing state-of-the-art algorithms for computational behavior analysis and a tight collaboration with stakeholders. Across two core studies, we have focused on implementing Behavior Imaging analysis to provide support from both screening and intervention perspectives through the detection of new markers by CV-based algorithms for data collection and the implementation of supervised and unsupervised ML models for data analysis. The following chapters will be devoted to a detailed description of the two studies and their implications.

Chapter 2

Is Smiling the Key? Machine Learning Analytics Detect Subtle Patterns in *Micro*-Expressions of Infants with ASD [1]

[1]The following chapter is based on:
Alvari, G., Furlanello, C., Venuti, P. (2021). Is Smiling the Key? Machine Learning Analytics Detect Subtle Patterns in *Micro*-Expressions of Infants with ASD. *Journal of clinical medicine*, *10*(8), 1776.

Abstract

The first study we conducted focused on addressing the primary challenge of autism research: early detection. From a clinical perspective, identifying the condition as early as possible is crucial for successful treatment. Despite advances in the literature, it is still difficult to identify early markers that can reliably predict the onset of symptoms. We developed a CV-based system to analyze early social behavior to uncover novel behavioral markers able to anticipate the diagnosis of ASD. Comparing a sample of infants with ASD (n=18) and a control sample (n=15), we examined the facial behavior during their first ecological interactions with the caregiver, between 6 and 12 months of age. We combined Openface, an AI-based software designed to encode frame-by-frame facial micro-movements systematically, and robust post-processing analysis to extract the subtle dynamics of social smiles in wild Home Videos. Infants with ASD displayed a reduced frequency and intensity of social smiles already within the first year of life. Also, we implemented a Random Forest classifier with cross-validation to predict the diagnosis and obtained an AUC=.90. The application of CV models allowed us to map facial behavior consistently, exposing early differences that are hardly detectable by the naked eye. The results of this study contribute to emphasize the potential of AI as a supporting tool for the clinical setting and offer excellent insights into an early screening of the autism spectrum.

2.1 Introduction

Early recognition of the autism condition is one of the most important challenges. From studies on early treatment, we know that age is a key factor in the success of the intervention program. Starting treatment earlier significantly increases the rate of improvement and eventually maximizes the outcome for a child with ASD (Elder et al., 2017; Rogers and Talbott, 2016; Vivanti et al., 2019). Identifying early markers has proved very complex, and to date, there are still difficulties in lowering the age of diagnosis. The obstacles extend both into the shortage of effective screening tools and into the complexity of disease manifestation in the early stages of development (Lombardo et al., 2018; Zwaigenbaum et al., 2015).

As highlighted in the second chapter of this dissertation, many people agree that it is possible to identify the first *red flags* as early as the second year of life, in which different studies have raised a variety of warning signs linked mainly to socio-interactive aspects (Barbaro and Dissanayake, 2013; Esposito and Venuti, 2009; Nadig et al., 2007; Ozonoff et al., 2018; Rozga et al., 2011; Shic et al., 2014; Varcin and Jeste, 2017; Wilson et al., 2017; Zwaigenbaum et al., 2015). In particular, social-attentive components apparently hold a central role in shaping the early onset of the condition while suggesting that these impairments may manifest very early in infants with ASD, leading to a cascade of developmental losses in social cognition. Problems in processing incoming social stimuli may render the interactive environment stepwise less rewarding, eventually resulting in a deprivation of core experiences and a collapse of the child interest in the surrounding social environment (Chevallier et al., 2012; E. J. Jones et al., 2014; E. J. Jones et al., 2016).

Behaviour Imaging techniques have proven to be the best choice, offering systematic measurements and being suitable to be robust against real-world scenarios. As a result of these properties, the implementation of state-of-the-art CV models in the study of behavior may be capable of exposing covert markers by disentangling behavior into more refined low-level components (K. Campbell et al., 2019; Dawson et al., 2018; Dawson and Sapiro, 2019; Sapiro et al., 2018). Accordingly, AI-based methodologies might be employed to discover new *red flags* for the autism spectrum through a richer characterization of existing markers from the literature by exploring primary components. This is a major concern when dealing with infantile behavior, whose diversity often results in a more detailed breakdown (Lombardo et al., 2019; Zwaigenbaum et al., 2015).

2.1.1 What marker? Facial expressions production in ASD

Facial expressions are an important channel to convey information and social experiences (Frith, 2009). Because of their importance in the socio-communicative development of the individual, they have often been taken into account within the autism spectrum. Nevertheless, it is challenging to infer significant findings concerning expressivity in ASD from the literature. This is partly due to conflicting results, partly due to outdated analysis methodologies. Many issues remain unresolved, but due to its nature, the production of facial expressions still is a key to investigation within atypical social development. In the population, atypical facial dynamics are related to lower social competence (Nowicki and Duke, 1994; Trevisan et al., 2018).

In the first years of life, newborns increase their expressiveness and begin to share feelings with caregivers (Begeer et al., 2008). At this stage, facial expressions are the main channel that toddlers exploit to interact with the social environment (Nichols et al., 2014). Moreover, newborns are biologically prepared to express emotional states to others (Nuske et al., 2013). This behavioral pattern elicits a series of physiological reactions in adults related to specific neural activations (Azhari et al., 2020; Esposito and Venuti, 2009; Nguyen et al., 2020). The dyadic coordination of discrete social signals provides the basis for the maturation of the child self-regulation and understanding of interactive dynamics (Azhari et al., 2020; Bornstein, 2013; Bornstein et al., 2016; Feldman et al., 2012). Consequently, this mechanism implies innate patterns of communicative behavior that subserve the development of a secure caregiver-child bond and the development of a typical social cognition (Azhari et al., 2020).

Overall findings in the literature suggest that there is variability in how children with ASD display facial expressions. This variety depends on many factors, including the type of analysis method employed and the experimental conditions. As reported in several studies, facial production in ASD appears to be more compromised in naturalistic contexts, when facial dynamics are spontaneous (Czapinski and Bryson, 2003; Stagg et al., 2014; Yirmiya et al., 1989). In these settings, the expressiveness of children with ASD results to be flatter, awkward, and less socially congruous (Nuske et al., 2013; Trevisan et al., 2018). Although facial expressions are generally perceived as more mechanical and irregular in appearance, the results are not always consistent with each other and sometimes show contradictions (Trevisan et al., 2018). By comparing studies that have used alternative measurement techniques, referring to more systematic analyses, the evidence becomes more explicit.

The exploitation of the AI framework allows employing more refined behavioral analysis by reducing the high coding costs. In the field of facial expressions, AI has allowed many researchers to extract subtle variables of facial dynamics avoiding time-consuming manual annotations (Dawson et al., 2018; Dawson and Sapiro, 2019; Manfredonia et al., 2019). This way, by shifting the focus to fine features, more consistent results seem to have emerged regarding facial production in children with ASD. Taking into account facial movements, clear differences in general expressivity emerged (Egger et al., 2018; Ekman and Friesen, 1978; Samad et al., 2019). The facial expressions of children with ASD appear to be less intense in terms of the magnitude of activation (Samad et al., 2019).

Regarding frequency, infants with ASD showed more neutral expressions than typically developing controls (Egger et al., 2018). In addition, analyzing the complexity, children with ASD showed lower coordination and variability between facial regions (Del Coco et al., 2017; Guha et al., 2016). Findings support what emerged in the more traditional literature about how parents perceive expressivity (Grossman, 2015; Trevisan et al., 2018). Micro-expressions seem to be an important subtle feature to consider, especially in childhood when the most prominent features of the condition are not yet easily recognizable. These observations are also demonstrated by the effectiveness of AI models in classification tasks.

The role of facial production seems to be salient in categorizing individuals with ASD (Drimalla et al., 2018; Grossard et al., 2020; Leo et al., 2018; Li et al., 2019). These results raise important assumptions in terms of social functioning in ASD. In literature, the ability to convey appropriate emotional signals is emphasized (Trevisan et al., 2018). Non-verbal skills are important to modulate social interactions, especially in the early stages of development. For this reason, it is relevant to consider how expressivity in ASD may contribute to the broader social phenotype. The atypical appearance of facial expressions could significantly undermine the establishment of an effective exchange with the caregiver (Faso et al., 2015; Trevisan et al., 2018). Impairments that are hard to recognize with the naked eye (Dawson et al., 2018).

Interestingly, this deficit seems to be related to the general level of functioning. Children with high-functioning ASD have proven to be less compromised in facial production (Grossard et al., 2020). One possible explanation for this pattern is that with the increase of cognitive skills, individuals with ASD may develop compensatory strategies such as reducing the atypical features of facial dynamics (Trevisan et al., 2018).

In support of these assumptions, studies on facial production training have proven its effectiveness in enhancing the appearance and fidelity of expressions in ASD (Gordon et al., 2014; Grossard et al., 2020). In this perspective, intervention programs may improve the expressiveness of children with ASD to help compensate for the deficit and raise the quality of social interactions. Correcting these irregularities could give children with

ASD the opportunity to experience adequate social exchanges and contain the atypical curve in socio-communicative development.

2.1.2 Aim and hypothesis

To summarise, the pursuit of novel early predictors is a priority in autism research. In pragmatic terms, the detection of precursor warnings for the condition should be (i) based on quantifiable and condition-specific markers, and (ii) based on ground-breaking methodologies that are feasible in naturalistic environments and robust to real-world scenarios (Alvari, Furlanello, et al., 2021).

The present study aims to test the applicative potential of CV models to explore novel markers of atypical behavior through video analysis. We referenced findings from the established literature about early impairments in socio-communicative and attentional components and their integration into initial manifestations of the autistic condition (E. J. Jones et al., 2016; W. Jones and Klin, 2013; Moore et al., 2018; Trevisan et al., 2018; Zwaigenbaum et al., 2015). We studied the characteristics of social smiles (Messinger and Fogel, 2007), which are characterized by being intrinsically linked to socio-attentive aspects, in the first home interactions between young toddlers with ASD and caregivers.

We collected retrospective 3-min Home Video (HV) recordings of infants between 6 and 12 months of development for a total of over 3000 usable frames per subject. Facial dynamics were extracted by embedding a robust coding system relying on a CV model for refined frame-by-frame analysis (Baltrušaitis et al., 2018), and a post-processing parsing (Bizzego et al., 2019) to derive salient features and structure a robust encoding of ecological HVs (Alvari, Furlanello, et al., 2021).

The morphological patterns of positive expressions have been further explored for the early identification of young infants who later develop ASD. Thus, we hypothesized that (1) it would be possible to detect early differences between young children with ASD and without through a refined analysis of facial dynamics by Deep Learning (DL), (2) that it would be possible to design ML models for the classification and prediction of infants on the autism spectrum based on the same expressive features (Alvari, Furlanello, et al., 2021).

Building a system that is capable of tagging a new marker at a very early age, within the first year of development, and robust to unstructured wild video data, would be an exciting step towards structuring new solutions with translational value and which may provide concrete support to the clinical setting.

2.2 Methods

2.2.1 Participants

All the analysis procedures and data collection were in accordance with the ethical standards of the Italian Association of Psychology (AIP) and with the ethical standards of the Ethics Committee of the APSS (Azienda Provinciale per i Servizi Sanitari, Trento, Italy) and the up-to-date Declaration of Helsinki (WMA, 2013).

The research sample was recruited at Observation, Diagnosis, and Formation Laboratory (ODFLab, University of Trento), a clinical and research center specialized in the functional diagnosis of neurodevelopmental disorders (Alvari, Furlanello, et al., 2021).

Population characteristics are summarized in Table 2.1. The study involved 18 children with ASD, with an average age of 8 (SD=1) months. The children had followed a full clinical assessment performed by specialists from ODFLab. The diagnosis was confirmed by Autism Spectrum Disorder (ASD) as described according to the criteria outlined in DSM IV/V (APA, 2013).

As a control sample, 15 children with typical development (TD) with a mean age of 9 (SD=2) months were also included. The TD group was collected outside the ODFLab by advertising on social networks and the laboratory website. The control sample in-

cluded children without any identifiable ASD condition according to DSM IV/V criteria (APA, 2013) and whose parents reported no significant abnormalities in socio-cognitive development (Alvari, Furlanello, et al., 2021).
The majority of the study population was male (ASD 94%, TD 75%), consistent with the asymmetrical ratio of gender within the condition (Loomes et al., 2017). All the children involved in the study were Italian. When the subjects were recruited, the mean age of the participants was 74 (SD=42) months for the ASD group and 85 (SD=30) months for the control group (Alvari, Furlanello, et al., 2021).
During the clinical evaluation of children on the autism spectrum, psychological tests to estimate general cognitive functioning and tests for social skills were administered. For social abilities, the ADOS-2 (Lord et al., 2012) was used as a criterion, and the average score was 7 (SD=1.8) for the ASD sample. Also, a measure of cognitive development was determined by administering two different tests (based on chronological age and level of development): the Griffiths Mental Development Scales (GMDS-ER, Luiz et al., 2006) to 8 children and the Wechsler Intelligence Scale for Children (Grizzle et al., 2011) for the rest of the sample. The average IQ=76 (SD=23) was evaluated over 16 subjects in the ASD group; for two subjects from the ASD group, an IQ score was missing (Alvari, Furlanello, et al., 2021).
In the control group, a measure of cognitive development was not available for most, but all children were reported by their parents as properly functioning for their chronological age. The average IQ=96 (SD=6) was assessed over 4 subjects within the TD group (Alvari, Furlanello, et al., 2021).

Table 2.1: Population characteristics of TD and ASD groups (Alvari, Furlanello, et al., 2021)

	ASD	**TD**	t/x^2	**p**
	n = 18	n = 15		
Gender, N(%)			0.533	0.465
Male	17 (94.4)	13 (86.7)		
Female	1 (5.6)	2 (13.3)		
Age (months), mean (SD)	74.4 (41.5)	84.5 (29.5)	0.544	0.593
Average Video Age (months), mean (SD)	8.3 (1.2)	8.8 (1.7)	1.153	0.258
ADOS CSS Total Score, mean (SD)	7 (1.8)	-	-	-
IQ Composite Score, mean (SD)	76.2 (22.5)[a]	95.8 (16.1)[b]	-	-
Video Length (sec.), mean (SD)	121.7 (5.9)	123.3 (6.3)	0.723	0.475
Average Interactions number, mean (SD)	3.6 (0.7)	3.3 (0.8)	0.839	0.408
Confidence Score, mean (SD)	0.94 (0.01)	0.95 (0.14)	1.933	0.062
Confidence Percentage (%), mean (SD)	90.5 (5.8)	88.5 (8.3)	0.811	0.424

Note: ADOS CSS: Autism Diagnostic Observation Schedule, 2nd edition, comparison score; IQ Intelligence Quotient; ASD Autism Spectrum Conditions; TD Typically Developing.
Average Video Age refers to the average age of subjects in the videos. Age refers to the age at the moment of evaluation. Average interaction number refers to the average number of interactions for each subject. Confidence Score refers to the average confidence (expressed from 0 to 1) of frames analyzed by Openface (Baltrušaitis et al., 2018). Confidence Percentage refers to the proportion of frames with high Confidence Score (>0.75), over the whole interaction video. IQ and ADOS scores were available only for the ASD group.
[a] The average IQ score is calculated over 16 subjects from the ASD group.
[b] The average IQ score is calculated over 4 subjects from the TD group.

Parents were asked to provide video recordings of at-home interactions with their children during the first year of age during data collection. Multiple HVs were collected for each subject at different ages ranging from 6 to 12 months of development. Interactive continuous play sequences were extracted from the HVs at different ages (6-12 months). To obtain suitable interactive sequences and to optimize the CV-based analysis of the children facial expressiveness, the HVs were edited (using video editing software, Movavi) by blurring other faces appearing in the frames and cutting the sequences in which the child face was covered or not frontally visible.

Subsequently, video interactions were filtered based on several selection criteria: must be continuous (without interruptions of more than 4 seconds), must be positive (that include play sequences and are free of manifestations of distress or discomfort), must involve the mother, cannot be shorter than 20 seconds each, and the child face must be frontally visible (Alvari, Furlanello, et al., 2021). Once selected by the criteria, all remaining video data were kept. As a result, segments of positive interactions were collected for each subject. The average number of interactions selected for each subject was 4 (SD=1) in the ASD group and 3 (SD=1) in the TD group. Finally, for each subject, the interactive segments were merged into 2-minutes composite videos, with a total average duration of 122 seconds (SD=6) for the ASD group and 123 seconds (SD=6) for the TD group (Alvari, Furlanello, et al., 2021).

2.2.2 Methodological considerations

In the context of the autism spectrum, research has shifted from studies based primarily on retrospective data (through the administration of parental reports or the analysis of HVs) to analysis designs built on prospective longitudinal studies of children considered to be at high risk of later being classified on the autism spectrum, most frequently younger siblings of individuals with ASD (Zwaigenbaum et al., 2013).

In terms of methodology, retrospective surveys measure events that occurred in a period before the study design and therefore fall victim to various biases. In particular, parental questionnaires are severely limited by both the reliability of recall and subjective judgment. The use of pre-diagnosis HVs, on the other hand, somewhat manages to correct some of the biases mentioned above, providing more objective information and the opportunity to use more standardized coding methods.

Nevertheless, analyzing HVs yields data subject to other types of errors related more to sampling, including the difficulty in containing the internal variability in the content of the different videos (such as the quantity and quality of recorded and unrecorded behaviors); in addition, the experimenter cannot have control over either the influence of the environmental context on response or the intrinsic characteristics of the footage, such as the quality (orientation and graphic quality) or the duration of the videos (Buxbaum and Hof, 2012; Zwaigenbaum et al., 2015; Zwaigenbaum et al., 2013). Analyses depend on previously collected data, perhaps in unreliable ways, so the researcher cannot personally assess the quality of the sampling.

Prospective studies do not fall victim to these systematic errors in the data structure, but on the other hand, they require high investments in time and resources, and collecting a representative field is a great obstacle, in fact, siblings of subjects with ASD have a risk rate of around 20% (Ozonoff et al., 2010). Further, while there is a problem of poor structuring and control in retrospective designs, there is often a bias from over-structuring in longitudinal studies. The major advantage of HVs is that they offer raw data with a very high ecological value. To structure translational frameworks is crucial to maintain the data as naturalistic as possible (Dawson et al., 2018; Dawson and Sapiro, 2019; Sapiro et al., 2018). The real challenge is to design analysis systems that will compensate for the lack of structure through systematic and sufficiently standardized measurements. For these reasons, in this study, we have chosen this approach to transform data variability from a limitation into a resource for designating more robust and generalizable models.

2.2.3 Facial *micro*-expressions analysis

The systematic analysis of facial behavior was based on the manual Facial Action Coding System (FACS). Developed by Ekman and Friesen (1978; Ekman et al., 2002), the FACS is a thorough taxonomy for coding facial expressions systematically and quantitatively. Facial dynamics are described based on the combination of over 70 observable fine movements of different parts of the face. Since these movements are the expression of the contraction of the underlying muscles, FACS provides a comprehensive set of features that are labeled Action Units (AUs) to represent the basic components of facial expressions, allowing each possible display to be decoded and measured without any interpretation (Ekman and Friesen, 1978; Ekman et al., 2002; Manfredonia et al., 2019). This way, the analysis is no longer dependent on the definition of discrete emotions and is structured based on objective and quantifiable motions (Fasel and Luettin, 2003).

FACS has been taken as a standardized reference in many studies for a systematic analysis of facial expressions and is increasingly used in autism research (Grossard et al., 2020; Trevisan et al., 2018). The major issue is that to minimize the bias of suggestiveness in expression ratings, the FACS requires expert graders with extensive and specific training. Furthermore, being very accurate, even if fully systematic, requires time-consuming human coding.

Several types of positive facial expressions (smiles) are described in FACS, which have a specific anatomical representation and functional correlation (Ekman and Friesen, 1978; Ekman et al., 2002). All smiling patterns take shape starting from a shared core AU12 referred to as the *lip corner puller*, which is formed by the activation of the *zygomaticus major* muscle around the corners of the lips (Ekman and Friesen, 1978; Ekman et al., 2002).

For the purposes of this study, we considered a particular type of smile, the Duchenne Smile, which is distinguished by the additional activation of AU06, the so-called cheek raiser, which corresponds to the contraction of the lateral part of the orbicularis oculi muscle (Ekman and Friesen, 1978; Ekman et al., 2002). This particular morphological combination has been well studied in the literature and has often been associated with the functional value of the communicative smile (social smile) in infants, perceived as more expressive (Bolzani et al., 2002; Dondi et al., 2007; Ekman et al., 1990; Mattson et al., 2013; Messinger and Fogel, 2007; Messinger et al., 2012).

During the early stages of development, the social smile (combination of AU12 and AU06) produced by the infant is characterized by being perceived by the caregiver as a more intense expression than simpler smiles (involving only AU12 activation) (Ekman and Friesen, 1978; Ekman et al., 2002; Mattson et al., 2013). Further, in the first dyadic relationships, the social smile occurs more frequently in salient interactive periods when the infant attention is focused on the smiling mother face (Fogel et al., 2000; Messinger and Fogel, 2007; Messinger et al., 2001). From a developmental point of view, this type of expression emerges very early, from the first weeks of life, and matures during the first semester of growth in combination with acquiring new patterns of attention to social stimuli (Lavelli and Fogel, 2005; Messinger and Fogel, 2007). In this critical period, marked by a shift in attentive preferences, the social smile appears in schemes involving a sequence of infant smiles preceded by states of focusing on the caregiver face (Lavelli and Fogel, 2005). Compared to other simpler smile typologies, the social smile seems to have a strong socio-attentive value and a very important protective role as one of the first communicative tools.

Considering the rich literature on early impairments in attentional (Barbaro and Dissanayake, 2013; Bryson et al., 2008; W. Jones and Klin, 2013; Maestro et al., 2002; Rozga et al., 2011) components and reduced expressivity (Czapinski and Bryson, 2003; Esposito and Venuti, 2009; Shic et al., 2014; Stagg et al., 2014; Trevisan et al., 2018; Zwaigenbaum et al., 2015) within infants on the autism spectrum, we have considered these expressive dynamics as a possible novel early marker (Alvari, Furlanello, et al., 2021).

2.2.4 Automated Action Units encoding

As outlined in the previous section, AUs have largely been used in the literature for more systematic coding of facial behavior. The quality of the results benefits greatly from the employment of systematic analyses; more refined coding of facial dynamics can bring out hidden features (Ekman and Friesen, 1978; Ekman et al., 2002; Manfredonia et al., 2019). This aspect is particularly relevant in ASD research, in which atypical behaviors in early development may be subtle and difficult to detect (Dawson et al., 2018; Dawson and Sapiro, 2019). However, the bottleneck in facial dynamics analysis is the cumbersome encoding of mimics through coding systems (Bangerter et al., 2020). Recently, the development of new AI-based models has allowed the automatic measurement of AUs effectively, highly simplifying the annotation effort. Training automated CV algorithms solves the cost of manual coding while maintaining objective measurements (Leo et al., 2018; Manfredonia et al., 2019). So far, affect analysis results have been very promising (Bangerter et al., 2020; Leo et al., 2018; Samad et al., 2019; Sapiro et al., 2018). The number of studies that have adopted these methodologies in ASD research is growing, although there are not yet many contributions regarding development age. Overall, findings go in the direction of confirming the presence of atypical features in facial expressions production. By analyzing refined components of facial muscles, the expressiveness emerged flatter and less complex in ASD children (Egger et al., 2018; Samad et al., 2019).

In the present study to encode facial behavior, we referred to Openface: a CV-based open-source software designed to estimate the real-time activation intensity of AUs from images and videos (Baltrušaitis et al., 2018). The architecture was based on recent state-of-the-art models in the AU recognition framework (Baltrušaitis et al., 2015; Gudi et al., 2015; Valstar et al., 2015; Zhao et al., 2016). In addition, it includes an additional implementation to make the algorithm more robust to in-the-wild video data (Baltrušaitis et al., 2018). Openface has been trained on several publicly available datasets (SEMAINE, McKeown et al., 2010; BP4D, X. Zhang et al., 2014; DISFA, Mavadati et al., 2013) to predict both the presence and intensity of 17 different AUs.

The model structure is divided into an initial face detection module, followed by a face landmark detection analysis module. Ultimately, the first two steps are combined for estimating the intensity and presence of AUs frame-by-frame (Baltrušaitis et al., 2018). The first step is implemented using a multi-task cascaded Convolutional Neural Network (CNN) to detect and align faces in the image (K. Zhang et al., 2016). Next, a Convolutional Experts Constrained Local Model (CE-CLM) is employed for in-the-wild face landmark detection and tracking (Zadeh et al., 2017). Histograms of Oriented Gradients (HOGs) are applied to the extracted aligned images to extract facial appearance features from the images (Felzenszwalb et al., 2009). At this stage, a Principal Component Analysis (PCA) model was further applied to reduce the dimensionality of the features (Baltrušaitis et al., 2018). As a final step, a linear kernel Support Vector Machine (SVM) was employed for the prediction of AUs presence, and a linear kernel Support Vector Regression (SVR) was used for AUs intensity estimation, thus providing an analysis of their presence, co-occurrence, and dynamics in videos (Baltrušaitis et al., 2018). The model was then validated on the DISFA challenging dataset (Mavadati et al., 2013) against a set of baselines outperforming the most recent and complex models (CCNF, Baltrušaitis et al., 2014; IRKR, Nicolle et al., 2016; LT, Kaltwang et al., 2015; CNN, Gudi et al., 2015; D-CNN, K. Zhang et al., 2016).

Openface has been successfully adopted in several third-party studies, proving its reliability in examining AUs patterns from videos (Kawulok et al., 2021; McDuff and Girard, 2019). Further, it has been tested to map facial dynamics on different clinical samples including autism (Drimalla et al., 2018; Drimalla et al., 2020; Owada et al., 2018; Rudovic et al., 2018).

In our study, by working with Openface we have been able to output an intensity vector for each of the AU in any frame of our video sequences from the HV dataset (over 3500 frames per subject). Afterward, we extracted AU12 and AU06 for the quality control and signal processing step (Alvari, Furlanello, et al., 2021).

2.2.5 Signal *post*-processing

After employing Openface, we jumped to a post-processing phase of the resulting matrix to extract the occurrence of Social (SO) and non-Social Smiles (SI). The regression of AU intensity consisted of a value that varied frame-by-frame over time expressed on a continuous scale between 0 and 5, in which 1 represented the activation threshold of the AU. Consequently, we considered an AU to be active when its activation intensity was greater than the threshold. We thus initially considered for further processing all the frames of the video sequences in which AU12 or 06 were >1. The data post-processing stage is organized into four distinct steps (Figure 2.1).

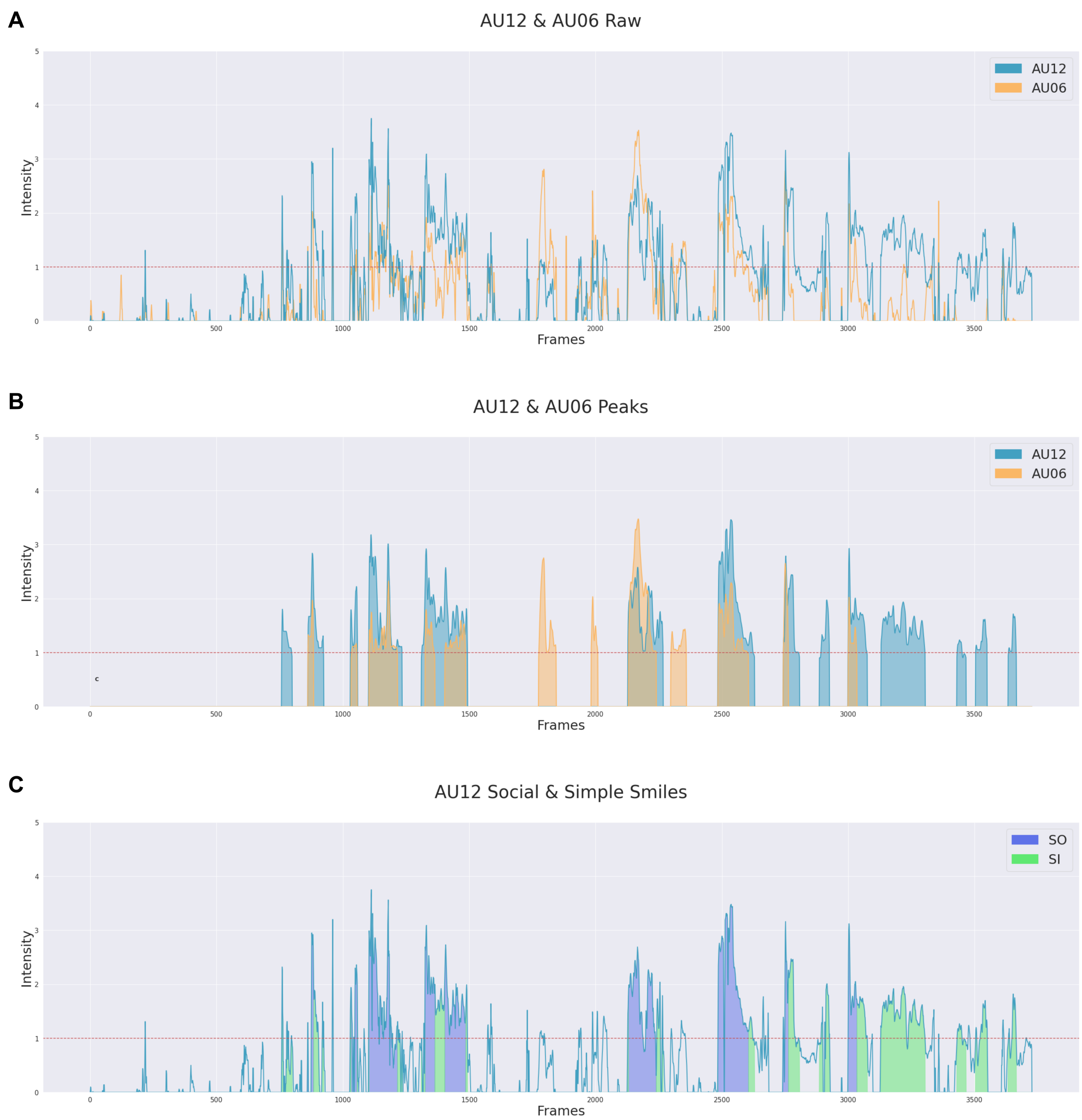

Figure 2.1: Signal Processing for AU12 and AU06 intensities over the interaction (data represent a subject from TD group). A) AU12/06 raw data; B) AU12/06 peaks extraction step; C) Social (SO) and Simple (SI) Smiles periods extraction.

By implementing a three-layer CNN trained to predict the detection error of facial landmarks, Openface provides a self-confidence score (C) of the estimates frame-by-frame

(Baltrušaitis et al., 2018). As an initial step, (1) we filtered the matrix by including only frames with high internal reliability ($C > 0.75$) in the analyses (Baltrušaitis et al., 2018). As a result, the average percentage of frames with high confidence exceeded 88%, with an average C of over 0.9 in both the sample with and without ASD (Table 2.1) (Alvari, Furlanello, et al., 2021).

Subsequently, (2) the matrix was processed using the python library `pyphisio`, structured to analyze physiological signals in time-series sequences and built on well-established libraries for scientific computing (Bizzego et al., 2019). The library includes functions that provide filtering by removing noise and improving the overall quality of the signal. We first built EvenlySignal instances that allow storing signal values with a fixed sample frequency. Then, a 4-frame convolutional filter was applied to smooth the signal and maximize follow-up analysis (Figure 2.1) (Alvari, Furlanello, et al., 2021; Bizzego et al., 2019).

As a third step, (3) we filtered the signal based on a minimum activation threshold of the AUs intensities. Originally from (Baltrušaitis et al., 2018), the activation threshold value was set to 1, but as a result of the smoothing process, the signal values were overall lowered, and accordingly, we recalibrated the minimum activation threshold to 0.8 (Alvari, Furlanello, et al., 2021).

Finally, to extract the activation peaks of the AUs, (4) we then determined a minimum duration for the expressive events. Consistent with the literature concerning the duration of facial dynamics, we excluded spikes of AUs intensities too brief to be classified as fully expressive events (<1 sec.), trying to avoid false positives as much as possible (Figure 2.1) (Alvari, Furlanello, et al., 2021; Davison et al., 2018; Dondi et al., 2007; Merghani, Davison, et al., 2018; Merghani, Davison, et al., 2018).

The filtering and smoothing process resulted in a clearer signal, from which we were able to extract activation peaks corresponding to the contraction of AU12 and AU06 (Alvari, Furlanello, et al., 2021). Combining the data from the two units, we thus pulled out

variables corresponding to (i) Social Smiles (SO, AU12 AU06) when the two AUs were active simultaneously, and (ii) Simple Smiles (SI, AU12) when only AU12 was distinctly active (with AU06 inactive).
As output, 3 measures were computed based on the peaks extracted for the two smile types: mean duration, mean intensity of the AUs, and mean frequency (defined by the mean number of peaks in 120 sec.).

2.2.6 Statistical analysis

The 3 extracted variables (duration, frequency, and intensity) were taken into account for the analysis of the two smile categories: Social (SO) and Simple (SI). The analysis phase was divided into two distinct parts.
The first stage (1) focused on testing between-group differences based on the extracted smile features. Firstly, two series of Spearman's Rank Order Correlations (rho) controlling for IQ scores and sex were applied within the ASD group to ascertain that gender and the level of general cognitive functioning were unrelated to the smiling features and that the results were not biased (Alvari, Furlanello, et al., 2021). A one-way Multivariate Analysis of Variance (MANOVA) was performed with groups (ASD vs. TD) as an independent factor and the 3 measures (frequency, intensity, and duration) as dependent variables. Follow-up analysis included one-way Analyses of Variance (ANOVAs) for the dependent variables. Also, a Bonferroni correction test was included to adjust probability (p) values due to the increased risk of a type I error when performing multiple statistical tests (Alvari, Furlanello, et al., 2021).
In the second part of the analysis, (2) we explored the effectiveness of predicting diagnosis with supervised ML. A Random Forest (RF) classifier was designed to verify the validity of the social smiling features for ASD classification. Starting from the results of the previous analyses, to train the model we reduced the dimensionality of the measures by selecting for training only those features of the social smile that differed

most between the two groups. We randomly split the dataset into training, validation, and testing, reserving 30% for the latter. We employed 5-fold cross-validation on the training and validation datasets to tune hyperparameters by performing a grid search. Based on the resulting AUC, we selected the top-performing hyperparameters. Finally, we trained, validated, and tested the RF to classify ASD and TD groups based on social smiling features. We evaluated the model performance by considering the AUC score and the Matthews Correlation Coefficient (MCC), which produces a more informative and truthful score in evaluating binary classifications than accuracy and F1 score (Chicco and Jurman, 2020).

2.3 Results

2.3.1 Between-group analysis

In the first part of the analysis, the distributions of the two groups were compared by analyzing the variance. The smiling features extracted were included as dependent variables: average AUs activation intensity, average duration, and frequency of smiles. Both Social Smiles (SO, AU12 AU06) and Simple Smiles (SI, AU12) were considered.

Two series of Spearman rank-order correlations were conducted to compute the relationship between IQ, gender, and each of the dependent variables. No statistically significant correlation was found (Table S1 in Supplementary Materials) (Alvari, Furlanello, et al., 2021). A one-way MANOVA was applied to test for statistically significant discrimination between TD and ASD groups based on features of two smile types. Before proceeding with testing, we checked the assumptions of the MANOVA. Preliminary checking revealed that residuals were not normally distributed for all of the dependent variables, as assessed by Shapiro-Wilk test ($p<.01$). Despite the violation of the normality distribution assumption, the MANOVA still stands as an optimal statistical model thanks to its robustness against deviations from normality. No multivariate outliers were found in data distribution, as assessed by Mahalanobis distance. There was no multicollinearity, as assessed by Pearson correlation (r=0.91, $p<.0001$). There was homogeneity of variances, as assessed by Levene's test ($p>.05$) (Alvari, Furlanello, et al., 2021).

From MANOVA testing, a statistically significant difference emerged between the group on the combined dependent variables (ASD vs. TD), ($F(6, 22)=14.513$, $p<.0001$). As follow-up analyses, univariate One-way ANOVAs were performed for each of the dependent variables, with a Bonferroni correction resulting in statistical significance with an adjusted p-value ($p<.007$) (Table S2 in Supplementary Materials).

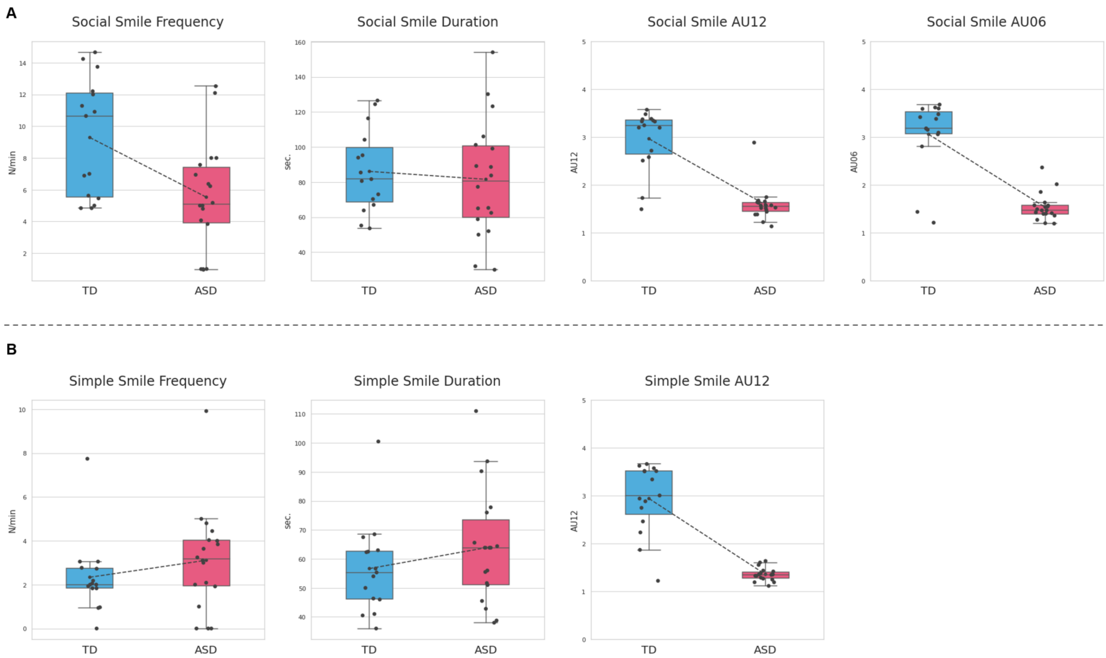

Figure 2.2: A) Social Smile (SO) features (frequency, duration, AU12/AU06 intensity) Boxplots; B) Simple Smile (SI) features (frequency, duration, AU12 intensity) Boxplots.

There were no statistically significant differences between the ASD and TD groups in average duration for both Social ($F(1, 27)=.219$, $p=.644$, partial $\eta^2=.008$) and Simple smiles (F(1, 27)=.992, p =.328, partial η^2=.035) (Figure 2.2).

A statistically significant difference was found in average AUs activation intensity. In Simple smiles, the activation intensity of AU12 was significantly lower in the ASD group (F(1, 27)=70.528, p<.0001, partial η^2=.723); in the Social smiles, both the AU06 (F(1, 27)=80.293, p<.0001, partial η^2=.748) and the AU12 (F(1, 27)=66.925, p<.0001, partial η^2=.713) were hypoactive in the ASD group (Figure 2.2).

About smiling frequency, statistically significant differences emerged only in the case of Social smiles (F(1, 27)=11.526, p<.00714, partial η^2=.299) (Figure 2.2 A). In contrast, the Simple smiles frequency did not reveal a significant difference between the two groups (F(1, 27)=3.133, p=.088, partial η^2=.104) (Figure 2.2 B).

Ultimately, pairwise comparisons with Bonferroni-adjusted p-values were made for all dependent variables (Table S3 in Supplementary Materials). Adjusted means were significantly different between groups regarding AUs activation intensity (p<.0001) for both Simple and Social smiles. Besides, a significant difference in the adjusted mean for the frequency of Social smiles was also found (p<.01). All other pairwise comparisons were not statistically significant (Alvari, Furlanello, et al., 2021).

2.3.2 Supervised ML

In the second part of the analysis, we implemented an RF classifier to categorize the two samples according to the social smiling parameters. The dimensionality of the variables was reduced by taking into account only those features that were found to be statistically significant in discriminating the two groups from the previous testing: AUs average intensity and frequency of Social smiles (Figure 2.3).

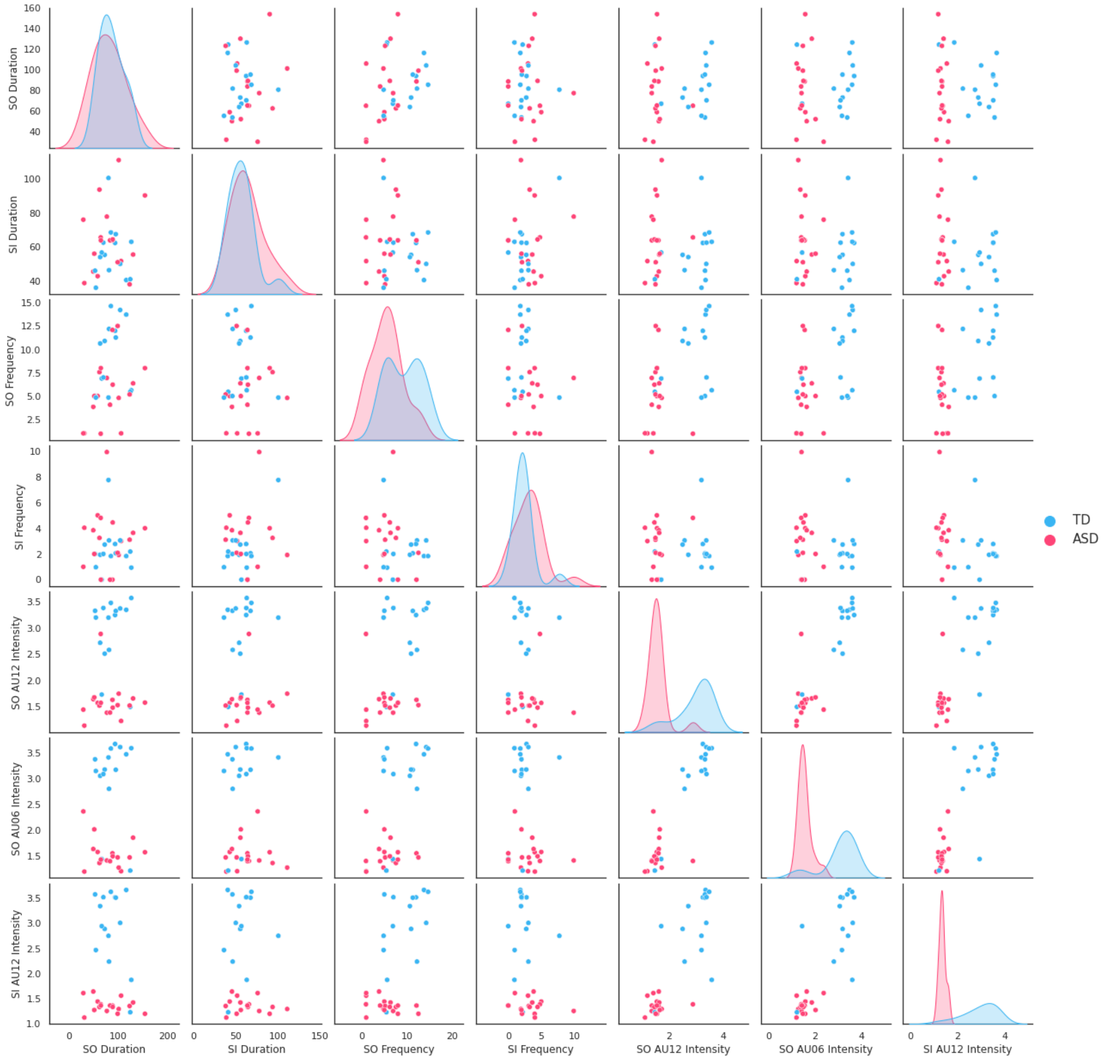

Figure 2.3: Social (SO) and Simple (SI) Smiles features pairplot.

For hyperparameter tuning, we employed 5-fold cross-validation with grid-search on the train-val datasets. A number of estimators =100 and a maximum depth =9 maximized the model performance in validation, resulting in an AUC=1.0 and an MCC=1.0.
The RF classifier was then tested on the test dataset by initializing the hyperparameters that emerged, resulting in an AUC=0.9 and an MCC=0.82 for classifying the two groups

(ASD vs. TD) based on AUs intensity and frequency of Social smiling (Figure 2.4).

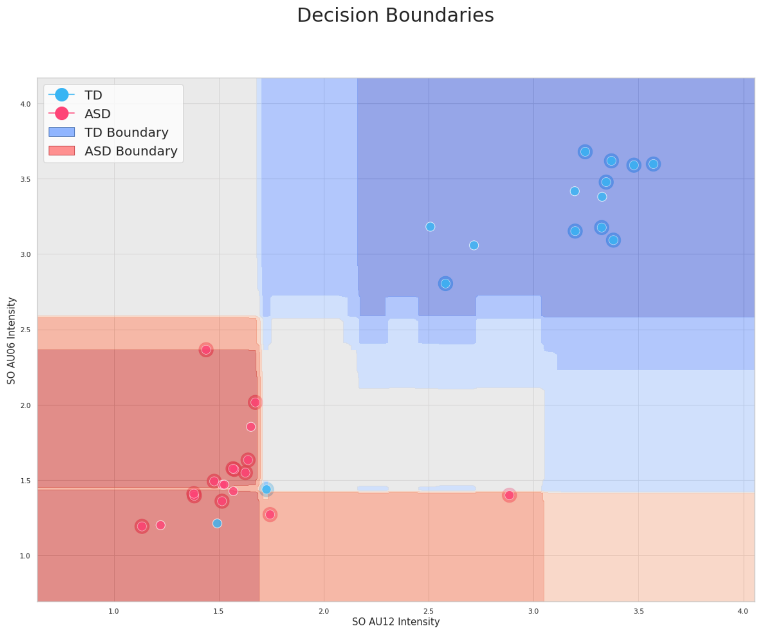

Figure 2.4: Random Forest classifier decision boundaries. SO: Social Smiles; SI: Simple Smiles.

Furthermore, the importance of the variables in the performance of the RF model was calculated using SHAP (SHapley Additive exPlanations) python library to evaluate the output of ML models. The average intensities of AU12 and AU06 of Social smiles were the most relevant features for classifying infants with and without ASD (Figure 2.5).

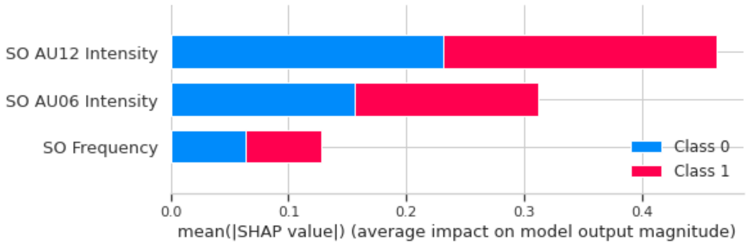

Figure 2.5: RF classifier feature importance (SHAP). Class 1: TD group; Class 2: ASD group.

2.4 Discussion

In this initial exploratory study, we attempted to tackle the challenge of early detection in autism research. The study end goal was to explore novel quantitative markers of atypical social-interactive development in young infants with ASD by developing computational systems for human behavior analysis. In this attempt, we implemented a combination of a CV-based model and an advanced signal processing system for a fine-grained and robust analysis of facial dynamics in at-home interactions through video.

Accent on attentive components in the processing of social stimuli has been highlighted in defining the early markers of atypical neurodevelopment in the autism spectrum (Alvari, Furlanello, et al., 2021; Chawarska et al., 2013; E. J. Jones et al., 2016; W. Jones and Klin, 2013; Moore et al., 2018; Shic et al., 2014). Social smiling involves the dynamic combination of an expression (smiling) and an attentive (gazing) component in terms of behavioral display. This facial display is recognized within early infancy by having a robust temporal link with the maturation of new orientation patterns towards the mother face (Lavelli and Fogel, 2005; Messinger and Fogel, 2007; Messinger et al., 2001).

The social smile has been suggested as a possible early marker of later manifestation of the autism phenotype (Nichols et al., 2014; Ozonoff et al., 2008). Employing computational methods, we analyzed fine-grained components of social smiles in early play interactions among caregivers and infants between 6 and 12 months of life. We compared both social and non-social smiling in young toddlers with and without ASD by computing the average duration, average intensity, and frequency for each of the two smiling types (Alvari, Furlanello, et al., 2021). We parsed the variables both through inferential analysis and using supervised ML.

From the analysis of variance of the two samples, findings seem to suggest that subjects with ASD smile with significantly lower intensity within the first year of development. This hypoactivation of facial movements, measured by AUs (Ekman and Friesen, 1978; Ekman et al., 2002), occurs both from smiles with and without a gazing component, suggesting a general poor positive expressive production during the first social exchanges in the autism condition, regardless of the socio-attentive valence of the expression (Shic et al., 2014; Varcin and Jeste, 2017; Zwaigenbaum et al., 2015).

A difference was found in the number of expressions displayed over the exchange between the two types of smiles from our data. Indeed, while there was a reduced frequency of social smiling in infants with ASD, simple smiles, which have less communicative value, have a similar frequency between ASD and TD groups. This result seems to replicate previous findings of lower rates for social smiling but similar rates for non-social smiling in children with ASD compared to controls (Alvari, Furlanello, et al., 2021; Nichols et al., 2014. Furthermore, the expressive deficit did not appear to correlate with the level of general cognitive functioning in our ASD sample. This finding might be consistent with previous studies that have suggested a relationship between IQ and compensatory strategies to reduce expressive deficits in children with ASD (Grossard et al., 2020; Trevisan et al., 2018). It is reasonable to assume that this compensatory pattern is not yet evident in infancy when cognitive functioning does not fully support the maturation of compensating mechanisms (Alvari, Furlanello, et al., 2021). There was no significant difference in both social and simple smiles between subjects with and without ASD concerning average duration.

Finally, by further developing a supervised ML model, it was possible to accurately predict the diagnosis of ASD based on social smiling features, considering both accuracy and specificity (Chicco and Jurman, 2020). In particular, according to an analysis of the weight of features in the classification, expressive intensity emerged as the predominant factor in discriminating infants with ASD.

Overall, the relevance of the results emerging from this study is significant in two research areas: a context more related to clinical aspects of the condition and the study of atypical development; and a context related to the contribution of new methodological solutions for human behavior analysis (Alvari, Furlanello, et al., 2021).

In the clinical scenario, findings seem to suggest that it could be interesting to include the quantification of expressive dynamics as a potential feature for the early detection of young infants with ASD. To our knowledge, this study was the first to highlight potential differences in positive expressive dynamics already within the first year of development in the autism spectrum. Thus, our findings seem to suggest that fine-grained analysis of social smiling yields salient cues for early detection of impairments in social development as early as 6 months of age (Alvari, Furlanello, et al., 2021). In addition, the differences that emerged in terms of the smiling frequency help underline the importance of the attentional components in identifying the condition in the early stages of development. Compared to simpler smiles, which are less charged with a communicative social value, social smiles are displayed differently in intensity and frequency. This discrepancy between social and non-social smile dynamics seems to corroborate previous research emphasizing the importance of the primary attentive components (Chawarska et al., 2013; E. J. Jones et al., 2016; Lord et al., 2020; Moore et al., 2018; Shic et al., 2014; Zwaigenbaum et al., 2015; Zwaigenbaum et al., 2019).

Although the nature of the collected data is not adequate to define conclusions about the specific mechanisms underlying atypical socio-cognitive development, it may have a heuristic value to include speculation. Possibly, one hypothesis is that the lack of access to a powerful communication tool (social smile) may cause the child with ASD to miss out on the experience of effective interactive exchanges during the first months of development. This atypical conduct may lead to a deprivation of enough social experiences during a critical period (Messinger and Fogel, 2007) essential to develop adequate social cognition and lead to the subsequent onset of the autism phenotype (Chevallier et al., 2012; Dawson, 2008; Landa, 2018; Messinger and Fogel, 2007; Zwaigenbaum et al., 2015).

On the other side, the innovative contribution of this study is also related to methodological aspects. Computational approaches provided high-quality analysis based on quantitative and objective measurements to categorize behavior, which could be non-invasive and automated (D. J. Campbell et al., 2014; Dawson and Sapiro, 2019; S. Jacob et al., 2019). These features make the application of such models extremely promising in autism research, in which discrete ecological measurements are essential. In the literature, such technologies have been proposed as alternative tools for measuring behavior in infants with ASD, targeting many of the markers that were previously identified through more traditional measures, such as attentional and motor deficits (Barbaro and Dissanayake, 2013; Chawarska et al., 2013; Esposito and Venuti, 2009; Gima et al., 2018; W. Jones and Klin, 2013; Moore et al., 2018; Ouss et al., 2018; Purpura et al., 2017; Rozga et al., 2011; Shic et al., 2014; Varcin and Jeste, 2017). In the present study, developing a CV-based system allowed the analysis of infant behavior in highly naturalistic scenarios, offering a systematic and fine-grained measurement (Alvari, Furlanello, et al., 2021). The analysis of interactions through this approach allowed novel markers to be identified as had been suggested in the literature. Furthermore, the combination of BI algorithms to extract subtle variables from the videos and the application of supervised ML models led to the development of a system capable of early predicting the occurrence of socio-cognitive developmental impairments in young children on the autistic spectrum.

Finally, the success in structuring a solution suitable for a naturalistic scenario offers an interesting contribution towards the definition of more translational approaches.
The present research also carries an important methodological limitation. Home Videos provide the enormous advantage of early ecological data, but at the cost of data control. This study intentionally focused on retrospective material, intending to analyze features in an at-home and natural environment. The literature further suggests investigating facial expressions in settings as spontaneously as possible (Trevisan et al., 2018). Nevertheless, the more unconstrained the interactions are, the less control there is over the data. As mentioned before, to develop applicative solutions, it is important to keep the setup natural, yet there are significant constraints. To manage these limitations, we adopted strict selection criteria for interactions. However, a certain degree of variability in the data is still unavoidable, in particular concerning the diversity of parental behavior (Alvari, Furlanello, et al., 2021).
To summarise, the objective of this work was to test AI in eliciting new traits in the frame of early autism detection. To our knowledge, this was the first reported evidence to explore novel markers during early dyadic social interactions based on a refined analysis of smiling. A generally poor positive display in infants with ASD has often been reported in the research literature. In particular, social smiles have already been raised as a potential early marker of the ASD phenotype, but only after the second year of life when impairments related to the condition phenotype emerge more consistently (Nichols et al., 2014; Zwaigenbaum et al., 2015; Zwaigenbaum and Penner, 2018). We hypothesize that the limitation of these approaches may lie in the analysis of facial dynamics through generic and unrefined systems, focusing on macro components of smiles, i.e. as the integration between expression and gaze, without addressing subtle facial elements. The employment of automated systems for analyzing behavior through images (Baltrušaitis et al., 2018) allowed us to effectively extract fine-grained features of facial movements (Action Units), which typically require challenging and time-consuming manual

coding by experts (Ekman and Friesen, 1978; Ekman et al., 2002). Leveraging these technologies, we were able to compute a quantitative and systematic measurement of smiling in very early interactions within the first year of development. In addition, the structuring of unsupervised ML models has led to these markers also being tested as predictors for diagnosis, achieving promising results. Beyond the methodological innovation aspects, the results are consistent with past research that had reported an important role in early abnormalities in primary social-attentional components in young infants on the autism spectrum (E. J. Jones et al., 2016; Moore et al., 2018).

From a more comprehensive perspective, this study provides an opportunity to highlight once again the promising contribution that AI-based methods can bring to clinical practice. The extensive support of automated techniques enabled to capture of fine features and potentially uncover subtle developmental patterns that are hardly detectable by the naked eye (Dawson et al., 2018; Dawson and Sapiro, 2019; Sapiro et al., 2018). In autism research, significant effort has been placed on seeking reliable early indicators that may predict the risk of later diagnosis. The benefit is to start the intervention earlier and hopefully improve the outcome (Estes et al., 2019; Lord et al., 2020; Vivanti et al., 2019; Zwaigenbaum and Penner, 2018). In this perspective, CV-based solutions may offer excellent opportunities to explore new subtle features, taking advantage of extracting systematic measurements non-intrusively in naturalistic at-home scenarios.

Chapter 3

EYE-C: Eye-Contact Robust Detection and Analysis during Unconstrained Child-Therapist Interactions in the Clinical Setting of Autism Spectrum Disorders [2]

[2]The following chapter is based on:
Alvari, G., Coviello, L., Furlanello, C., (2021). EYE-C: Eye-Contact Robust Detection and Analysis during Unconstrained Child-Therapist Interactions in the Clinical Setting of Autism Spectrum Disorders. *Brain Sciences*, *11*(12), 1555.

Abstract

In our second study, we focused on tackling spectrum stratification. The high level of heterogeneity in ASD and the lack of systematic measurements complicate predicting outcomes of early intervention and the identification of better-tailored treatment programs. Computational phenotyping may assist therapists in monitoring child behavior through quantitative measures and personalizing the intervention based on individual characteristics, yet real-world behavior analysis is still an ongoing challenge. For this purpose, we designed EYE-C, a system based on Openpose and Gaze360 for fine-grained analysis of eye-contact episodes in unconstrained therapist-child interactions via a single video camera. The model was validated on video data varying in resolution and setting, achieving promising performance. We further tested EYE-C on a clinical sample of 62 preschoolers with ASD for spectrum stratification based on eye-contact features and age. By unsupervised clustering, three distinct sub-groups were identified, distinguished by eye-contact dynamics and a distinctive clinical phenotype. Overall, this study highlights the potential of AI in categorizing atypical behavior and providing translational tools that might assist clinical practice.

3.1 Introduction

Designing early intervention programs that can reduce symptom severity may significantly enhance the well-being and quality of life of individuals on the autism spectrum and their families (Bishop-Fitzpatrick et al., 2016; Estes et al., 2019; Lord et al., 2020; Rodgers et al., 2021). In this respect, the detection of prodromal markers certainly holds a crucial role in enabling the *early* treatment. However, we strongly need more effective tools to identify the best pathway and capitalize on potential improvements (Baldwin et al., 2021; Fuller and Kaiser, 2020; E. Stevens et al., 2019). Systematic behavioral indicators that help predict the outcome of the intervention are still lacking. Additionally, studies on the effectiveness of early intervention treatment have shown that, despite a general improvement, children with autism conditions showed a high rate of individual variability in their response to the treatment pathway (Bentenuto et al., 2020; Eapen et al., 2013; Fuller and Kaiser, 2020; Lord et al., 2020; Uddin et al., 2019).

There is still a shortage of a shared framework or systematic behavioral indicators to analyze and evaluate the intervention. A major cause of this problem is the high heterogeneity of the spectrum, both in onset and developmental trajectories (Lombardo et al., 2019; Lord et al., 2020; Pearson et al., 2018; Wolfers et al., 2019; Zwaigenbaum et al., 2015). This wide variability complicates the implementation of shared frameworks and recalls the need to adapt the intervention pathway to individual child characteristics (Bentenuto et al., 2020; Rudovic et al., 2018; Rudovic et al., 2019). Some prodromal factors have been considered in the literature as moderators of the impact of treatment, but the results are mixed, and their specific importance still needs to be investigated (Bentenuto et al., 2020).

Possible concerns may arise from using behavioral characteristics designated for clinical diagnosis to measure the impact of treatment and monitor long-term improvements (Bentenuto et al., 2020; Washington, Leblanc, Dunlap, Penev, et al., 2020). Likewise, the same approach has been applied in most stratification studies, where data collection has remained bound to variables derived from the administration of standard psychological testing. As already pointed out, such metrics may not be sensitive enough to measure subtle differences in either improvement or sub-group differentiation (Washington, Park, et al., 2020). The research effort should be directed towards structuring systematic methodologies based on fine-grained descriptors suitable to measure more specific behavioral variables. Furthermore, considering the strong connection between spectrum heterogeneity and varying response to the treatment, a promising strategy should be by examining behavioral traits that are sensitive to the inter-individual variability of the autism spectrum and carry applicative value in predicting intervention effectiveness. Behavior Imaging has proven to have great potential within the context of refined behavioral analysis, especially for early recognition. However, the application of AI-based systems on intervention is still underdeveloped and at an experimental step. Biobehavioral correlates underlying specific aspects of emotional regulation (Cabibihan et al., 2018; Jarraya et al., 2020; Masmoudi et al., 2019), joint attention (Alnajjar et al., 2021), social engagement (Rudovic et al., 2017; Rudovic et al., 2019) and social interaction 2020) have been studied for treatment response (Baldwin et al., 2021; E. Stevens et al., 2019) and implementation (Washington, Leblanc, Dunlap, Penev, et al., 2020; Wetherby et al., 2018). Nonetheless, the literature is still limited, and further investigation is needed.

The primary difficulty with real-world data from intervention sessions is that analysis is complex and needs much versatility. Within the context of autism, it is well-known that it is crucial to assess behavior in settings as naturalistic as possible. This becomes particularly significant in the framework of assessment and intervention sessions in which spontaneity is already inevitably partly compromised, both by the non-familiar environment and by the involvement of a therapist in the interaction. Maintaining a non-invasive approach is essential to prevent affecting the therapist-child interplay. In addition, fine-grained analyses have to deal with the high level of dynamics of the setting, in which both child and adult are constantly moving and which includes periods of interrupted or low-quality signals. For these reasons, in the literature, the application of AI models in the intervention scenario is based on heavy interaction and setting restructuring, which often compromises the applicative value (Jaliaawala and Khan, 2020; Voss et al., 2019).

The more the interaction and the environment are structured, the better the quality of the data collected, but at the high cost of limited flexibility (Leclère et al., 2016). In most studies, the trade-off has been weighted in favor of more robust model performances, resulting in a lack of translational solutions (Jaliaawala and Khan, 2020). Research needs to move towards designing more balanced computational methods that account for data quality yet emphasize the ecology of interactions. Thus, it will be feasible to deliver effective AI-based systems that can be scaled to real-world scenarios and provide support for clinicians and therapists working across the autism spectrum. The novel contribution of this study in this field is the implementation of a robust system for eye contact analysis and its validation in concrete clinical environments.

3.1.1 Gaze patterns

Lack of eye contact is an iconic trait on the autism spectrum (Fonger and Malott, 2019; Ninci et al., 2013). Maintenance of sustained eye contact may significantly enhance the quality of the social experience as well as increase the likelihood of success in responding properly to stimuli and prompts as well as potentially improve the acquisition of adaptive social competencies (Carbone et al., 2013; Cook et al., 2017; Fonger and Malott, 2019). As already pointed out, children with autism, even at an early age, show marked difficulties in gaze integration and an atypical response to adult gaze (Barbaro and Dissanayake, 2013; Chawarska et al., 2013; W. Jones and Klin, 2013; Madipakkam et al., 2017; Rozga et al., 2011). Therefore, gaze integration is already a prominent goal in early intervention programs (Fonger and Malott, 2019). Learning appropriate gaze modulation early in social interaction may enhance success in many domains and potentially improve intervention outcomes of young children with ASD (Fabiano et al., 2020; Miller et al., 2017).

In this research area, AI has found many applications in both symptom monitoring (Hashemi et al., 2014; Sapiro et al., 2018) and treatment, especially through robot-mediated therapy (Chung, 2019; Rudovic et al., 2018; Yun et al., 2017). Notably, computational approaches introduce the possibility of collecting quantitative and fine-grained measures with high temporal sensitivity. Most of these approaches were designed upon employing advanced eye-tracking technologies through wearable devices (i.e., smart-glasses) (Hurwitz et al., 2020; Thorup et al., 2018; Ye et al., 2012), frontal cameras (Bovery et al., 2019; K. Campbell et al., 2019; Chang et al., 2021; Hashemi et al., 2018; Li et al., 2019; Yun et al., 2017) or strong interaction structuring (Hashemi et al., 2018; Sapiro et al., 2018). The major drawback of implementing these techniques, as highlighted above, remains the constraint of operating in not-so-naturalistic environments (Jaliaawala and Khan, 2020; Voss et al., 2019). Despite the advances and the considerable appeal of this area of study, there is still a lack of examples for eye contact detection with sufficiently ecological methodologies in autism research. Given the relevance of integrating eye gaze into intervention programs, this issue is an important goal. Additionally, examples of the role of gaze patterns in defining different shadings of the spectrum also are lacking in terms of a functional perspective. Most studies focused on discriminating between those diagnosed with ASD and typically developing peers (Fabiano et al., 2020; Georgescu et al., 2019; Ye et al., 2012). To the best of our knowledge, very few studies have investigated the role of gaze in stratifying the autism spectrum.

Campbell and colleagues (D. J. Campbell et al., 2014) investigated the role of variability in attention to direct gaze in differentiating the autism spectrum. They employed unsupervised clustering on young children with ASD based on visual response to dyadic stimuli from videos. The analysis identified three different sub-groups that were compared for verbal, social, and adaptive functioning skills (D. J. Campbell et al., 2014). The cluster that exhibited limited attention to social scenes subsequently demonstrated a poor outcome; conversely, the sub-group with good attentional abilities was verbal and high functioning at three years. The results of this work confirm that gaze analysis may have an interesting clinical role both in addressing spectrum heterogeneity and as a predictor of outcome (D. J. Campbell et al., 2014).

In a different approach, Fabiano and colleagues (2020) used a combination of handcrafted and raw gaze variables with demographic characteristics, such as age and gender, to classify multiple levels of ASD risk. They constructed several classifiers showing that the different classes, low, medium, high, and ASD correspond to different patterns that can be used to classify risk. The results confirmed the potential of gaze as an indicator that needs to be further investigated. In addition, the analyses showed that age is an important factor in classifying ASD risk, resulting in an overall accuracy of 93.45% (Fabiano et al., 2020).

More recently, Latreche and colleagues (2021) investigated the role of social orientation in modulating treatment outcomes in preschool children with ASD. They employed eye-tracking technologies to measure subjects attentional patterns while watching videos of an adult engaging in child-directed speech. The results confirmed that the degree of attention to the adult face strongly correlated with the severity of autistic symptoms at baseline and also predicted improvement after treatment. Children with ASD who stared less at the actress's face and avoided eye contact suffered more impairment in the socio-communicative domain and showed less after-treatment improvement, particularly in the verbal domain (Latrèche et al., 2021).

Overall, the results are very promising and underline the importance of social-attentive skills within intervention programs, but further work is needed. Behavior Imaging approaches may be well suited to address this need by offering interesting alternatives for the ecological measurement of child behavior.

3.1.2 Current study

The present work is focused on developing an AI-based method for the ecological analysis of therapist-child interactions through video capable of extracting dyadic gaze coordination episodes. The purpose of the present work is to test the validity of computational solutions to systematically analyze the socio-attentional components of the interactions and identify behavioral indicators that allow for the identification of sub-groups within the spectrum.

We considered eye contact coding for analysis as it is a major impairment in autism and for its crucial role in the early intervention (Latrèche et al., 2021; Miller et al., 2017). We collected video recordings (around 60 min each) of ADOS administration sessions of children with ASD in preschool age. We included videos at different resolutions, from low to high, and in different rooms of the same laboratory, from small to large, to ensure a robust and translational framework.

For gaze analysis, we developed a combined AI-based approach based on a module to extract multi-person body and head pose keypoints (Cao et al., 2019) and a module to derive a 3D vector of gaze direction frame-by-frame from wild videos (Kellnhofer et al., 2019). We further developed a system for robust derivation of eye contact periods experienced between therapist and child during unconstrained interactions.

The model was validated by matching the output with hand-coded features. Continuous interactive sequences of about 10 minutes were extracted from 4 different videos (with different video resolutions and in different lab rooms) for a total of more than 70000 frames. The sequences were hand-coded frame-by-frame, and results were compared with the model's output to evaluate the performance and understand under which conditions the data quality stayed too low.

Finally, we tested the gaze features for autism spectrum stratification and for predicting the intervention outcome based on machine learning methods. We hypothesized that based on our metrics, it would be possible to identify sub-groups within the sample through unsupervised clustering and validation on clinical variables.

3.2 Methods

3.2.1 Data collection

All analyses and data collection were carried out in accordance with the ethical standards of AIP and the Ethics Committee of the APSS (Trento, Italy). The study involved 85 (11F, 74M) preschool children (< 6 years of age) with a confirmed diagnosis of Autism Spectrum Disorders (ASD). All the participants were Italian and recruited within ODFLab patients. All families involved in this study were well informed about the procedure and agreed to written informed consent. They also were aware of the possibility of abandoning the procedure at any time.

The diagnosis of ASD was confirmed through a comprehensive assessment of the children functional profile and validated through a clinical judgment by an independent clinician based on DSM-V criteria (APA, 2013) and through the administration of the ADOS-2 (Lord et al., 2012). Population characteristics are summarized in Table 3.1.

Table 3.1: Population characteristics (Alvari, Coviello, et al., 2021).

	ASD Sample
	n = 85
Age (months), mean (SD)	56.32 (13.8)
GQ, mean (SD)	71.54 (17.4)
ADOS, mean (SD)	14.82 (1.4)
Gender, N (%)	
Male	74 (87.1)
Female	11 (12.9)

Note: ADOS: Autism Diagnostic Observation Schedule, 2nd edition, raw score; GQ: Global Developmental Quotient (GMDS-ER); ASD: Autism Spectrum Disorders.

Inclusion criteria required that the subjects were diagnosed as having ASD without displaying other medical conditions and that they had been assessed within 6 years of age. During the clinical evaluation, psychological tests were administered to assess general cognitive functioning and social skills. All assessment meetings were video-recorded. In particular, videos of the ADOS-2 administration sessions were collected. Clinical variables collected in the study included ADOS-2 raw scores for social abilities and symptom severity and the Griffiths Developmental Scales (GMDS-ER) for an overall assessment of cognitive development quotient and related subscales for all the participants.

The ADOS-2 is a golden standard for the diagnosis of autism and is carried out by an experienced, trained specialist. The administration procedure consists of a sustained semi-structured play interaction between the clinician and the child to elicit different socio-cognitive skills. The instrument is structured in 4 different modules according to the child chronological age and level of expressive language. Each module is divided into Social Abilities (SA) and Repetitive and Restricted Behaviour (RRB) subscales, combined into an overall comparison score to classify the severity of the child autistic symptoms. In the present study, the raw scores have been included, as for the Toddler module (suitable for younger children) it is not possible to compute a standardized score (Lord et al., 2012).

The GMDS-ER are developmental scales (also normalized in an Italian sample) administered to children in a laboratory setting through semi-structured activities to assess different domains of mental development in young children. The testing provides a global developmental quotient (GQ) and specific scoring on six different subscales of cognitive functioning, including gross motor, hand-eye coordination, communication, social, performance, and practical reasoning abilities. In the present study, the GQ and the subscale scores were taken into consideration (the practical reasoning scale was excluded because it is not administered to young children) (Luiz et al., 2006).

3.2.2 Videos specifics

The video recordings considered for the attentional pattern analysis included play interactions between therapists and children during the administration of the ADOS-2. The average duration of the recordings was approximately 1 hour, but it varied with the child responsiveness and the quality of the interplay (duration M=63.91 minutes, SD=26.2). The play activity with the child is kept spontaneous by the therapist, although using standardized materials and a predefined sequence.

The videos were all recorded in the same laboratory, the ODFLab, but in 4 different rooms, two of which were larger (around 28 m^2) and two were smaller (around 12 m^2). Acquisition of the recording was carried out by using a single environmental camera in the corner of the room. The location and resolution of the cameras varied based on the room, which ranged across 384/640/720/1280/1920 pixels of width. Details regarding video resolution in the sample are shown in Table 3.2.

Table 3.2: Video resolution for the ASD sample.

Room Size	**Video resolution (px width)**				
	384	640	720	1280	1920
Small, N (%)	10 (11.8)	5 (5.9)	43 (50.6)	1 (1.2)	0
Large, N (%)	3 (3.5)	10 (11.8)	11 (12.9)	0	2 (2.4)

The resulting data collection were recordings with a high variability of the content, both in terms of the interaction and the video quality of the material. The data can vary from high-resolution videos shot in a relatively small room to low-resolution videos taken in a larger room, where the subjects were more distant and less clearly visible. Such variability represents both a handicap and a resource. While it complicates testing and weakens video analysis performance, it also forces the design of a highly robust system. The study's primary objective was to define a resilient framework to extract attentional patterns from our real-world clinical data automatically. Details about the model development are described in Section 3.2.4.

3.2.3 Related work

The application of AI-based models for the analysis of attentional patterns has recently advanced, with promising results (Sapiro et al., 2018). The potential again relies on the opportunity to automatically measure attentional behavior through video and extract quantitative parameters in a systematic way. However, predicting gaze direction in real-world scenarios has been proven challenging. The strong variability of the environment, the occlusion of the image, and the dynamism of the interaction remain difficult variables to manage.

Most systems have been integrated to analyze gaze with a frontal camera through eye recognition and geometrical segmentation (Ghosh et al., 2021; Huang et al., 2017). However effective, the strong limitation of these approaches is that they are based on heavy interaction structuring and rely on a fixed light source. They are not suitable for unconstrained environments nor for analyzing dynamic interactions within the clinical setting (Kellnhofer et al., 2019).

An alternative is appearance-based methods that learn more direct gaze mapping using large annotated datasets (X. Zhang et al., 2015). These methods for gaze estimation work well in everyday settings, yet most of the state-of-the-art models are still being developed and evaluated based on datasets collected under controlled conditions in the laboratory, often acquired with a frontal camera. These conditions are constrained by limited variability in appearance and little change in head pose (Fischer et al., 2018; X. Zhang et al., 2015, 2017a; Zhu and Deng, 2017).

There are no benchmark designs regarding the specific analysis of eye contact episodes in dynamic interactions. This is because eye contact recognition not only requires an accurate estimation of gaze direction and information about the position and orientation of the target. A few examples attempted to address this issue by offering advanced solutions also using standard cameras in literature (Müller et al., 2018).

Smith and colleagues (2013) employed a classification approach to determine eye contact from a camera video. Yet, their methodology required a priori knowledge about the size and pose of the target (Smith et al., 2013). Similarly, Parekh and colleagues (2017) developed a CNN architecture that recognized eye contact. Their method performed well but demanded the subject to be stationary in front of a camera (Parekh et al., 2017).

Muller and colleagues (2018) developed a novel approach to recognizing eye contact in multi-person interactions to address this issue. The setting consisted of a setup of 8 different environmental cameras installed around 4 adults interacting while sitting. The model combined both gaze direction information and speech (determined by analysis of facial Action Units), assuming that people tend to look at the person who is talking during conversations (Ho et al., 2015; Müller et al., 2018). The model was further evaluated on datasets of natural group interactions and performed better against more standard approaches (Müller et al., 2018; X. Zhang et al., 2017a).

Preliminary interesting examples are available also in the context of multi-person interactions (Müller et al., 2018; Parekh et al., 2017). Although robust, these solutions stay constrained to highly structured environments and are not suitable for naturalistic clinical settings, where children and therapists often rapidly change both position and orientation. Designing a system suitable for dynamic interactions and real-world scenarios is the primary goal of the present study.

3.2.4 Model design

We attempted to develop a robust eye contact detection system (EYE-C) well suited to analyze collected clinical videos. The first objective was to implement a computational solution for extracting multi-person gaze directions in naturalistic videos. To address this problem, we designed a system based on state-of-the-art pre-trained algorithms composed of (1) a module for extracting the head position of targets in the image (Cao et al., 2019) and (2) a module for estimating a frame-by-frame gaze direction vector (Kellnhofer et al., 2019).

For step (1), we used Openpose, which is a CV model that can do real-time multi-person 2D pose estimation from in-the-wild videos (Cao et al., 2019). The model takes as input the colored image and produces the 2D coordinates of the anatomical keypoints for each person in the image. The Openpose pipeline consists of a first step in which the input RGB image is fed to a multi-stage CNN architecture, initialized with the VGG-19 model, and then fine-tuned (Simonyan and Zisserman, 2014). In the first set of stages, a feedforward network predicts the 2D confidence map of the body keypoints. In the second stage, part affinity fields (PAFs) are predicted, representing a degree of association between the keypoints and enabling body parts to integrate into a full-body pose (Cao et al., 2019; Cao et al., 2017). In the end, the confidence map and PAFs are parsed through inference to produce 2D keypoints of all people in the image (Cao et al., 2019).

The model was evaluated on multiple datasets (Andriluka et al., 2014; Lin et al., 2014) and compared against Mask R-CNN (He et al., 2017) and Alphapose (Fang et al., 2017), achieving the best performance considering the trade-off between speed and accuracy in the COCO Challenge 2017. The output of the model consists of a JSON file of 135 landmarks of different body parts divided into 3 blocks: body+foot, hand, and face detection.

We employed Openpose to extract the features from the first main block (body+foot) and then computed the head bounding boxes of the targets by using the keypoints of ears, eyes, nose, and neck. Once we extracted the therapist and child head coordinates in the video frames, we can apply the gaze estimation module.

In the second module, (2) we used Gaze360, an appearance-based model capable of extracting a 3D gaze vector from 2D videos in-the-wild (Kellnhofer et al., 2019). Given the absence of real-world datasets to estimate gaze, the authors first collected a large-scale dataset for gaze-tracking in unconstrained images. The dataset is the largest publicly available set and consists of 238 subjects in both indoor and outdoor environments with labels of 3D gaze coordinates in many head poses and distances (Kellnhofer et al., 2019). Based on the dataset, the authors further implemented a model for gaze direction estimation. The architecture of Gaze360 is based on bidirectional LSTM capsules, which provide an average of the modeling sequences in which the output depends on both previous and future inputs (Kellnhofer et al., 2019).

Thus, a window of 7 consecutive frames of head crops is used as input (centered around the target frame) to predict gaze. In the first stage, the head crop of each frame is processed individually through a CNN, which produces 256-dimensional features. The features are fed to the bidirectional LSTMs to produce compact representation vectors in the second step. Finally, vectors are concatenated into fully connected layers to predict both 3D gaze coordinates and a quantile error estimate (Kellnhofer et al., 2019). The architecture was evaluated cross-dataset using several benchmark datasets of high- and low-resolution 3D gaze (Fischer et al., 2018; Kellnhofer et al., 2019; Smith et al., 2013; X. Zhang et al., 2017b). The model was further fine-tuned into new domains using a self-supervised approach and improved performance across all datasets.

The large variability of the Gaze360 dataset and the cross-domain adaptation of the model allowed for excellent performance even in unseen videos from uncurated online media sources, such as Youtube videos, demonstrating flexibility and robustness (Kellnhofer et al., 2019). The final output of the model is represented by a coordinate matrix of the gaze vector $\boldsymbol{g}$ for any head crop in each frame of the video. The coordinates in Gaze360 are computed in a spherical system and expressed in observing the camera Cartesian perspective system $\boldsymbol{g} = [\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}]$. The origin of the vector represents the center of the head (based on the coordinates of the eyes, mouth, and nose), and the coordinates (expressed between 1 and -1) define its direction. For example, if $\boldsymbol{g} = [\mathbf{0}, \mathbf{0}, -\mathbf{1}]$ the target is looking directly at the camera, regardless of its position. In this manner, the estimation of gaze vectors is based only on head crop appearance and without any other global information from the environment (Kellnhofer et al., 2019).

In summary, in our study we first combined the two modules using (1) Openpose to extract head crops and then fed them to (2) Gaze360 to compute therapist and child gaze vectors in our dataset. We rendered all clinical videos by drawing the headboxes and gaze vectors to double-check the result. From video inspection, it was evident that the model's performance dropped during periods of high interaction dynamism, i.e. when the child moved around the room and frequently changed distance and head orientation relative to the camera. In these cases, the head recognition module failed, producing head boxes that were generally smaller and varied a lot in size during short sequences.

The results of Gaze360 are based upon the information of multiple consecutive frames (Kellnhofer et al., 2019). Thus, headboxes that vary a lot in size over a few seconds compromised gaze direction estimation, often resulting in faulty vector predictions. In addition, this effect was more noticeable in videos recorded in larger rooms, where the distance to the camera was higher and the headboxes smaller. Overall, the performance suffered heavily in the most dynamic periods of the session, both for the size variability and the reduced dimension of the head crops.

We proceeded in two directions to try to solve these problems. To better handle larger settings, (i) we increased the size of the headboxes by 50%, providing bigger input images for gaze estimation. In addition, to cope with moments of high mobility of the subjects, we (ii) forced a matched dimension of the headboxes for continuous sequences from a single target, normalizing the shape of the head crops according to the largest size recorded in short consecutive frame sequences. In practice, we increased the overall size of the head crops fed to the gaze estimation module, and we normalized the headbox shape in consecutive frames keeping the size constant in video sequences to manage the headbox variability in dynamic contexts. Model validation is discussed in more detail in Sections 3.2.4.1 and 3.2.4.2.

3.2.4.1 Eye-contact detection

Once the 3D gaze coordinates of the child with ASD and the therapist were extracted, the challenge was to successfully build a function to extract the periods of eye contact between targets in wild 2D clinical videos. The system needed to be robust to variance in predictions and flexible to different setting conditions to accomplish this task. Eye contact periods were defined based on the relationship between therapist and child frame-by-frame gaze estimations. One subject was looking at the other if the gaze vector was directed toward the other head. If both subjects were looking at each other for a certain amount of time, then eye contact was present.

To operationalize this dynamic, first, we computed the 2D coordinates $[\boldsymbol{x}, \boldsymbol{y}]$ of the intersection p between the line passing through the coordinates of the gaze vector $\boldsymbol{g}$, and the line passing through the center $\boldsymbol{g0}$ of the other target headbox and perpendicular to the x-axis. Namely, we were able to establish the point $\boldsymbol{p}$ where the gaze of subject

A crossed the position of the head of subject B on the ordinate (Figure 3.1 B). We then calculated the distance $\boldsymbol{d}$ (in pixels) between the intersection point $\boldsymbol{p}$ and the origin $\boldsymbol{g0}$ to understand the proximity of a subject gaze to the target head, as follows 3.1:

$$\boldsymbol{d_A} = |\boldsymbol{p_B} - \boldsymbol{g0_B}| \tag{3.1}$$

The smaller the distance $\boldsymbol{d}$ with respect to the therapist head, the more the child gaze will be oriented towards the face. To understand whether the child was looking at the adult face, we established a maximum distance threshold $\boldsymbol{T_d}$ (in pixels). We adopted a threshold rather than the precise center to attempt to contain slight inaccuracies in gaze prediction.

When both distances $\boldsymbol{d_A}$ and $\boldsymbol{d_B}$ were below threshold $\boldsymbol{T_d}$ we would potentially get eye contact. However, this first step is constrained to a bidimensional representation of data. The outputs of Openpose are two-dimensional coordinates of the landmarks (Cao et al., 2019). On the other hand, Gaze360 provides a three-dimensional vector (Kellnhofer et al., 2019). An issue of considering only the 2D coordinates is to recognize as episodes of eye contact some moments without such coordination, for instance, situations when the heads of the subjects are located at very different depths or more often when they are very close to each other. Neglecting depth may result in many scenarios where gaze directions appear to cross but only from a 2D perspective. This approach would lead to include several false positives in the analysis and compromise the quality of the data as well as lose information.

A

B

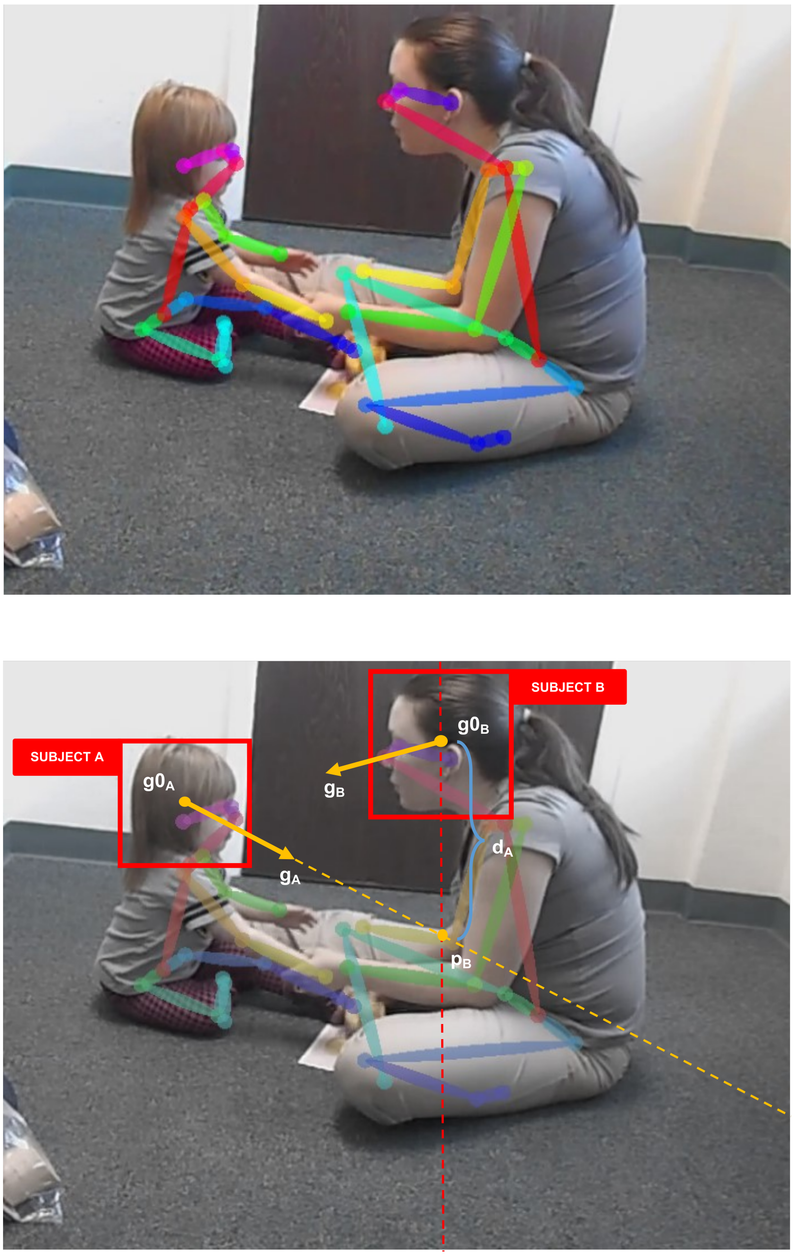

Figure 3.1: Eye-contact detection model (EYE-C). The images represent the output of the model run on a video example from YouTube (the video is licensed under a CC licence, and was kindly offered by White, R. [Good Behavior Beginnings]. (2015, May 15). How to Redirect Escape Behavior in 2 year olds [Video]. YouTube. https://www.youtube.com/watch?v=GzGLF8GlPmo); A) Openpose body keypoints output (Cao et al., 2019); B) Gaze360 gaze vectors output (Kellnhofer et al., 2019) and eye-contact detection system; $\boldsymbol{g0_A}/\boldsymbol{g0_B}$: headbox center of subject A/B; $\boldsymbol{g_A}/\boldsymbol{g_B}$: gaze vector of subject A/B; $\boldsymbol{p_B}$: intersection point between gaze of subject A and headbox x-axis coordinates of subject B; $\boldsymbol{d_A}$: distance (pixels) between $\boldsymbol{p_B}$ and $\boldsymbol{g0_B}$.

To address this problem, we used a simple but effective approach. The output of Gaze360 is a 3D vector in which depth is expressed through a $\boldsymbol{z}$ value that varies between -1 and 1 (Kellnhofer et al., 2019). When $\boldsymbol{z}$ assumes a negative value, the subject is looking toward the camera, conversely, it assumes a positive value when the subject looks away from the camera. Whenever child and therapist look at each other and are on the same depth level in the room, their gaze vectors will have a value of $\boldsymbol{z}$ close to 0. This means the therapist gaze will be fully oriented towards the right or left side of the room, vice-versa for the child. Conversely, when subjects look at each other from two different depths of the room, the $\boldsymbol{z}$ value of each of the two gaze vectors will be different from 0 and with opposite signs. Therefore, the $\boldsymbol{z}$ value of the therapist gaze vector will start to increase (if the direction of the vector points away from the camera) or decrease (if the direction of the vector points toward the camera) according to its direction. Similarly, the $\boldsymbol{z}$ value of the child gaze vector will change, but with an opposite sign. If $\boldsymbol{z}$ of subject A increases, then $\boldsymbol{z}$ of subject B decreases below 0. This is because if vectors are aligned at different depths, they will always have opposite signs. In this way, to recognize eye contact between child and therapist, both gazes need to be close enough to each other head and need opposite depth direction.

Lastly, a rarer situation to consider is when both subjects stay at the same depth position, and $\boldsymbol{z}$ is close to 0. In such circumstances, little fluctuations and errors in gaze prediction might vary vector orientation, compromising the analysis and including possible false negatives. To solve this problem, we again established a threshold $\boldsymbol{T_z}$ by setting a maximum degree of tolerance for the absolute value of $\boldsymbol{z}$. When $\boldsymbol{z}$ was close enough to 0, and therefore the gaze directions had nearly no depth, it was unnecessary for the two gaze vectors to have opposite signs. In this way, we were able to control all cases in which subjects were looking at each other closely and at the same depth level of the room.

In summary, conditions for discriminating eye contact periods included that (1) the gaze vectors were both oriented toward the headbox of the other within a threshold distance, (2) the eye contact period had a minimum duration, and (3) that the vectors had opposite directions when the absolute value of $\boldsymbol{z}$ exceeded a certain value.

This pipeline enables a dyadic eye contact detection system resilient to common variations in terms of video resolution and ecological clinical setting, with enough flexibility to handle interactions with high levels of dynamism.

3.2.4.2 Model evaluation

Following the design and method definition part, we evaluated the performance of EYE-C using hand-crafted features. We picked five videos from the sample matching different video resolutions and room settings. A continuous interactive dyadic sequence of about 10 minutes was extracted from each selected video, for a total of more than 70000 frames (M=14173.4, SD=261.7 frames for video) for the model testing (a total of 4360 positive frames, labelled with eye-contact).

The 10-minute videos were subsequently hand-coded frame-by-frame using a software for observational video coding (BORIS, https://github.com/olivierfriard/BORIS). The interactive periods in which there was eye contact between therapist and child were noted for each frame. In parallel, the same sequences were further encoded using the eye-contact detector according to the pipeline described in the previous sections.

Finally, the output of the model was evaluated using the hand-crafted features as ground truth (eye-contact present/absent) for each frame. The parameters described in Section 3.2.4.1 were fine-tuned to maximize matching using the Matthews Correlation Coefficient (MCC) to evaluate the model's performance (Chicco and Jurman, 2020).

Overall, the best performance was achieved by using a maximum distance threshold $\boldsymbol{T_d}$ corresponding to 80% of the target headbox size, a depth threshold $\boldsymbol{T_z}$=.3 , and a minimum length of 30 consecutive frames second) for eye-contact sequences. The results are described in Table 3.3, for increasing size of the videos.

Table 3.3: Model evaluation results.

Video	Frames (N)	Res (px)	Room	Acc	Pre	Rec	MCC
1	13786	640	Small	0.96	0.65	0.80	0.70
2	13955	1280	Small	0.95	0.53	0.76	0.61
3	14317	720	Large	0.96	0.79	0.65	0.69
4	14497	720	Small	0.99	0.94	0.94	0.93
5	14312	384	Small	0.93	0.34	0.71	0.46

Note: Res: Resolution; Acc: Accuracy; Pre: Precision; Rec: Recall; MCC: Matthews Correlation Coefficient.

As expected, the EYE-C performance decreases in conditions where at the same time the video had lower quality and the interaction occurred in larger rooms achieving an MCC=.46, which corresponds to high accuracy but low precision. In these borderline cases, the shapes of people are too small in the frame sequences preventing a correct analysis of body landmarks and gaze direction. By excluding the video with the latter characteristics, the model performed well across all the other conditions with an average MCC=.74.

For these reasons, videos with lower quality and recorded in larger rooms were filtered out. Based on the analysis of e gaze vectors and subject headboxes, we removed 23 subjects among those with interactions recorded in larger rooms at lower quality. The sample was accordingly downsized from 85 to 62 subjects. As a result, the sub-sample with re-evaluation after the intervention was reduced from 25 to 18 subjects.

Finally, EYE-C was run on the filtered dataset after the evaluation to extract fine-grained features of the eye contact periods between therapists and children with ASD. The metrics mined and employed in subsequent analyses included: average duration, average child distance $\boldsymbol{d}$, number, and frequency of the eye-contact episodes.

3.2.5 Data analysis plan

For dataset analysis, we considered the eye-contact features as independent variables: average duration (dur), total number (num) and frequency (freq, number per minute) of eye-contact episodes, and the overall average distance $\boldsymbol{d}$ of the child gaze from the therapist face during the interaction.

As dependent variables, we included the scores of psychological testing from the clinical evaluation. Concerning cognitive functioning, we considered the general developmental quotient (GQ) and 5 relative subscales: gross motor (Motor), hand-eye coordination (Coordination), communication (Language), social (Social), and performance (Perform) abilities. Also, we included the ADOS-2 total raw score (ADOS) and the sub-score of the social abilities subscale (SA) for the socio-communicative dimensions

We did not consider the single item scores of the ADOS-2 as they are rated on a qualitative scale with little variance (0-2). Indeed, in our sample, 78 (91.8%) of the 85 subjects received the maximum score (=2) in the item related to gaze modulation during the first evaluation. Therefore we chose to keep only the overall raw score of the ADOS-2.

Accordingly, we considered developmental quotient (GQ) for cognitive functioning and raw ADOS-2 scores (ADOS) for social-communication skills within the re-evaluation phase of the intervention sub-sample.

The analysis procedure was further divided into two separate parts: correlation and stratification.

Correlation - As a first step, we explored the correlation between our eye contact features and the clinical variables. We first converted all variables into z-scores to normalize the standard deviation to 1 (using Standard Scaler from the scikit-learn library). Subsequently, we employed Multiple Linear Regressions (MLRs) and checked the assumptions to analyze the relationship between the independent variables jointly against each dependent variable.

Stratification - To investigate spectrum heterogeneity, we then employed unsupervised clustering based on eye-contact features. We first standardized the variables and then employed Uniform Manifold Approximation and Projection (UMAP) for manifold learning and dimensionality reduction Allaoui et al., 2020. Further, for clustering, we employed Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), which assumes clusters based on density regions and leaves scattered background classified as noise. HDBSCAN is a suitable algorithm for data-driven approaches because it does not need to determine the number of clusters a priori and thus is more robust in exploratory analysis design than deterministic partitioning algorithms such as K-Means (Schubert et al., 2017). Finally, we externally validated the clusters by testing for differences in clinical variables based on the resulting sub-groups. We applied a One-way MANOVA using cluster membership as an independent variable. Then we applied One-way ANCOVAs, with video length and resolution as covariates, and Tukey Tests for post-hoc analysis, pairwise comparisons, and an adjusted p-value. Finally, we further analyzed the duration and number of eye-contact episodes over time in a mixed design to see if there were any differences between the sub-groups over the course of the interaction.

3.3 Results

3.3.1 Correlations

Before proceeding with the definition of MLRs, we checked for the assumptions. Multicollinearity occurs when you have two or more independent variables that are highly correlated with each other. We computed the Variable Inflation Factor (VIF) to determine the correlation between eye-contact features by obtaining a score for each variable of how well it is explained by the others.

A VIF score above 5 indicates high multicollinearity. As expected, a strong correlation between frequency and number of eye-contact episodes was found (Table 3.4). For these reasons, we decided to eliminate frequency (freq) and keep number (num) in the following analysis.

Table 3.4: Model evaluation results.

Variable	VIF
freq	5.6
num	5.42
d	1.07
dur	1.34

Note: freq: eye-contact periods frequency; num: eye-contact periods total number; d: average child gaze distance d; dur: average eye-contact periods duration.

In addition, to have more control over the independent variables, we applied Pearson Correlation Coefficient to control the association between the eye-contact features and the length of the videos to avoid a bias due to the duration of the interactions. No significant correlation emerged.

Then we checked the distributions of the variables to check that they followed a normal distribution. We initially used Q-Q Plots to test the distribution of the variables. From a visual inspection of the diagnostic plots, the distance (dist) and number (num) of eye contacts, and the communication abilities quotient (Language) did not follow a normal distribution (Figure S1 in Supplementary Materials).

We converted the two variables with a logarithmic transformation. Then we conducted Shapiro Wilk tests to check for normality of distributions. All variables resulted normally distributed after the conversion (Figure S2 in Supplementary Materials). Next, we standardized all the measurements and computed an MLR for each dependent variable, using our eye-contact features as independent variables (excluding frequency) (Figure S3 in Supplementary Materials).

The assumption of homoscedasticity is that the residuals are equal for all values of the predicted dependent variable (i.e., the variances along the line of best fit remain similar as you move along the line). We checked for homoscedasticity by controlling the plots of studentized residuals versus unstandardized predicted values and by performing the Breusch-Pagan test for heteroscedasticity (Figure S4 in Supplementary Materials). For each of the MLRs, homoscedasticity of the residuals was confirmed by visual inspection and non-significant test results. Finally, we checked the distribution of residuals by again using the Shapiro-Wilk test and found the normality of the error distributions for all MLRs (Figure S5 in Supplementary Materials). The results of all MLRs for each of the dependent variables are summarized in Table S4 in Supplementary Materials.

A significant regression equation with a non-robust negative correlation was found for the ADOS-2 total score ($F(3,58)=2.718$, $p<.05$), with an $R^2=.123$, and the related Social Abilities subscale ($F(3,58)=2.866$, $p<.05$), with an $R^2=.129$. No significant regression was found for the general developmental quotient, but a significant regression equation with a non-robust positive correlation was found for the related subscales of communication ($F(3,58)=2.795$, $p<.05$), with an $R^2=.126$, and hand-eye coordination ($F(3,58)=2.783$, $p<.05$), with an $R^2=0.126$ abilities. All remaining MLRs for the subscales of the GQ were non-significant.

The average duration (dur) and distance (d) of eye contact episodes were not significant predictors for all regression models. Otherwise, the total number (num) of eye contacts during child-therapist interactions was a significant predictor of ADOS ($p<.01$), SA ($p<.01$), Language ($p<.05$), and Coordination ($p<.05$) scores.

3.3.2 Stratification

The second part of the analysis design further explored the findings by investigating the effectiveness of gaze patterns (num, dur, freq, d) for spectrum stratification and identifying the possible occurrence of sub-groups within the sample. Given recent findings regarding the importance (Fabiano et al., 2020), we also considered age as a factor for clustering . We initially normalized the variables by converting them into z-scores.

We employed the Uniform Manifold Approximation and Projection (UMAP) for non-linear dimensionality reduction and improved data visualization (Allaoui et al., 2020; McInnes et al., 2018). A 2-component UMAP was applied on the 4 scaled eye-contact features and age jointly to reduce data structure using 5 nearest neighbors and a minimum distance of 0 as hyperparameters of the algorithm. The output is a 2D projection of the data structure into low-dimensional space, based on a transformation (embedding) of the selected features (Figure 3.2 A). The resulting data embedding was further processed for unsupervised clustering. We used HDBSCAN algorithm by setting a minimum cluster size of 10, equal to twice the number of features used (2*5 dim), as suggested in the literature (Schubert et al., 2017).

In the dataset, three different clusters were identified (sub-groups 0/1/2), and two single data points were classified as noise (sub-group -1) (Figure 3.2 B).

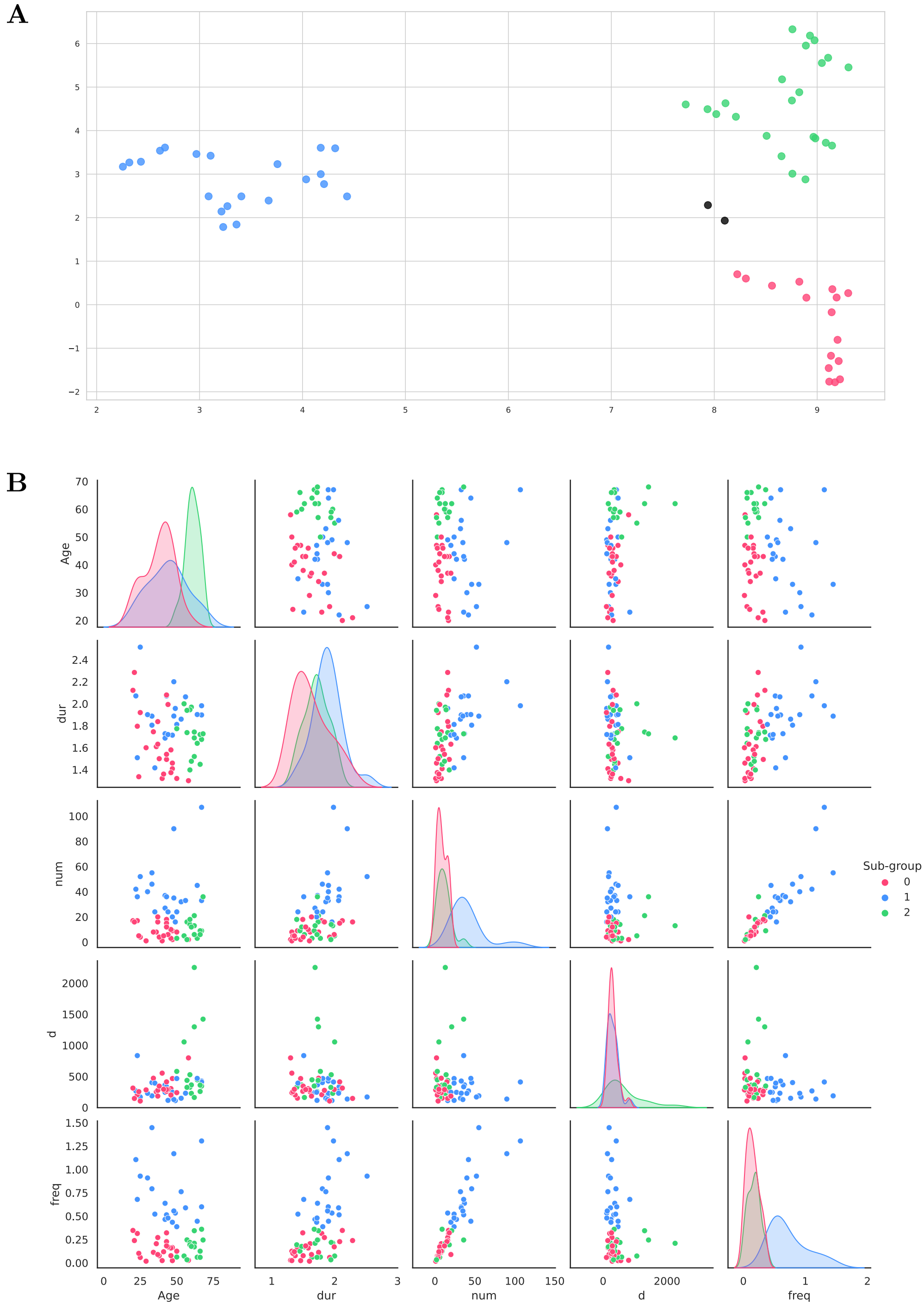

Figure 3.2: A) HDBSCAN clusters on 2-components UMAP output; B) Eye-contact metrics and age pairplot; freq: frequency of eye-contact episodes; num: total number of eye-contact episodes; dur: average duration of eye-contact episodes; d: average distance of children gaze vectors from therapists headboxes during interaction; age: children age at first assessment.

We computed the Silhouette Coefficient (SC) to assess the consistency and homogeneity of the resulting clusters and achieved a score of SC=0.56. The characteristics and population size of the clusters are summarised in Table 3.5.

Table 3.5: Sub-groups characteristics.

	Sub-group 0	**Sub-group 1**	**Sub-group 2**	**F / x^2**	**p**
	n = 23	n = 21	n = 16		
Gender, N (%)				3.572	0.734
Male	22 (95.7)	17 (80.9)	12 (75)		
Female	1 (4.3)	4 (19.1)	4 (25)		
Video Resolution (px)	-	-	-	11.069	0.748
Age (months), mean (SD)	37.9 (10.2)	43.9 (10.1)	60.9 (10.1)		
Eye-contact num, mean (SD)	9 (6.2)	40.7 (6.2)	11.6 (6.1)		
Eye-contact freq (N/min), mean (SD)	0.2 (0.1)	0.7 (0.1)	0.2 (0.1)		
Eye-contact dur (sec), mean (SD)	1.6 (0.3)	1.9 (0.3)	1.7 (0.3)		
Eye-contact d (px), mean (SD)	308.9 (152.1)	301 (152.1)	649.4 (152.1)		
GQ, mean (SD)a	69.7 (15.4)	77.1 (15.4)	65 (15.4)		
Coordination, mean (SD)	69.6 (14.5)	82.2 (14.5)	64.8 (14.5)		
Language, mean (SD)	55.7 (26.6)	67.8 (26.6)	60.4 (26.6)		
Motor, mean (SD)	78.4 (13.8)	79.1 (13.8)	71.5 (13.8)		
Social, mean (SD)	64.6 (16.6)	75.8 (16.6)	58.7 (16.6)		
Perform, mean (SD)	89.9 (18)	88.1(18)	72.6 (18)		
ADOS, mean (SD)a	15.6 (3.5)	13.1 (3.5)	15.4 (3.5)		
SA, mean (SD)	12.2 (3.5)	10.2 (3.5)	11.8 (3.5)		

Note: freq: eye-contact periods frequency; num: eye-contact periods total number; d: average child gaze distance d; dur; average eye-contact periods duration. GQ: Global Developmental Quotient; SA: Social Abilities subscale.

Comparison of sub-groups characteristics showed no significant differences in gender, video resolution, and duration of therapist-child interaction.

Finally, clustering was evaluated by comparing the differences between the groups, considering both clinical and eye-contact features (Figure 3.3). A One-way MANOVA was applied with group membership as the independent variable and clinical metrics as the dependent variables. A statistically significant difference emerged between the sub-groups on the combined dependent variables, ($F(13,46)=6.393$, $p<.0001$).

As follow-up analyses, univariate One-way ANCOVAs and Tukey's Tests were performed for post-hoc pairwise comparisons of each dependent variable.

Figure 3.3: Boxplots of each dependent and independent variable for resulting sub-groups; num: eye-contact periods total number; freq: eye-contact periods frequency (N/min); dur: eye-contact periods average duration (sec); d: average distance of children gaze vectors from therapists headboxes during interaction (px); age: children age at first assessment (months); length: video total duration (min).

Concerning our eye-contact features, we found a significant difference in total number ($F(2,57)=31.82$, $p<.0001$), frequency ($F(2,57)=55.577$, $p<.0001$), average duration ($F(2,57)=5.815$, $p<.01$), and the overall average distance ($F(2,57)=6.618$, $p<.01$) between the sub-groups. A significant difference between the sub-groups was also found in terms of the children age ($F(2,57)=23.597$, $p<.0001$).

Comparing the sub-groups based on clinical variables revealed significant differences concerning the ADOS ($F(2,57)=3.549$, $p<.05$) total score and the cognitive development subscales of social ($F(2,57)=3.207$, $p<.05$), hand-eye coordination ($F(2,57)=5.803$, $p<.01$), and performance ($F(2,57)=3.238$, $p<.05$) abilities.

We further tested the significant results achieved through pairwise comparisons with an adjusted p-value (Table S5 in Supplementary Materials). Tukey post-hoc tests showed that sub-group 1 showed a significantly higher number and frequency of eye contact than both sub-groups 0 (num $p<.001$, freq $p<.001$) and 2 (num $p<.001$, freq $p<.001$). Conversely, sub-groups 0 and 2 did not differ either in number ($p=.82$) or frequency ($p=.9$). In terms of duration, sub-group 1 showed significantly longer episodes of eye contact than sub-group 0 ($p<.01$), but no difference emerged either between sub-groups 1 and 2 ($p=.096$) or 0 and 2 ($p=.595$). In addition, sub-group 2 showed a significantly higher overall distance d than both sub-groups 1 ($p<.01$) and 0 ($p<.01$); the latter showed no difference in distance ($p=.9$).

Regarding the clinical variables, a difference emerged between the ADOS-2 total scores of sub-groups 0 and 1 ($p<.05$), the latter having significantly lower scores. In contrast, there were no significant differences between the other sub-groups (2 vs. 0 $p=.9$, 2 vs. 1 $p=.1$).

In terms of hand-eye coordination abilities, sub-group 1 showed a significantly higher mean quotient compared to both sub-groups 0 ($p<.05$) and 2 ($p<.01$), whereas sub-groups 1 and 2 did not differ ($p=.063$). Sub-group 1 showed a significantly higher score in the social abilities subscale than sub-group 2 ($p<.05$), but not sub-group 0 ($p=.195$); there was no difference between sub-groups 0 and 2 ($p=.653$). Concerning the performance subscale, pairwise comparisons showed no significant differences among sub-groups. When comparing age, sub-group 2 was significantly older than sub-groups

1 ($p<.001$) and 0 ($p<.001$), which also did not differ in age ($p=.152$).

We also compared the three sub-groups based on the resolution of the videos, the duration of the video, the room where the interaction occurred, and the gender of the subjects as controls and found no significant differences.

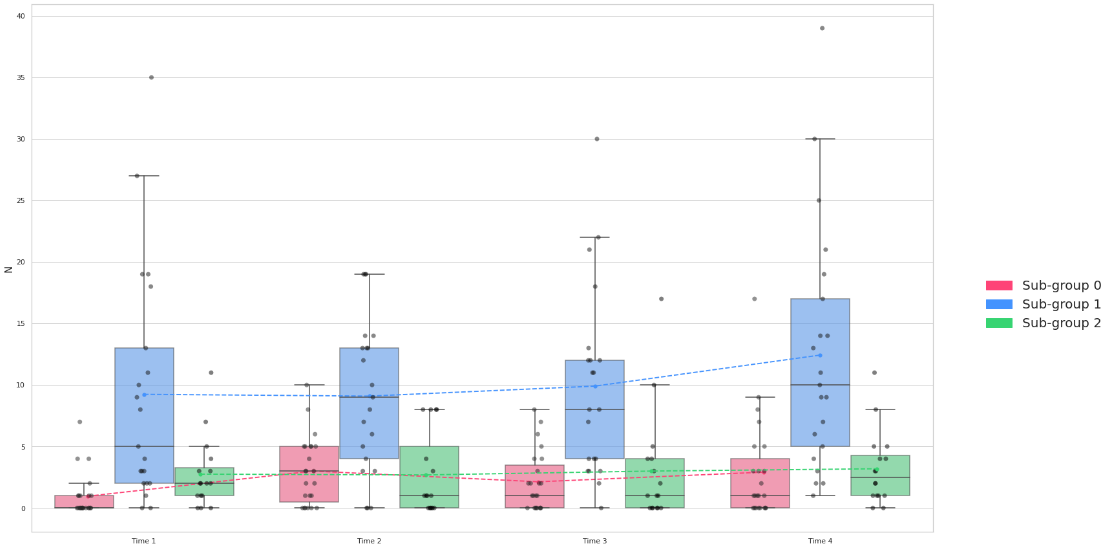

Figure 3.4: Eye-contact periods total number (num) over time.

As the last step of clustering exploration, we divided interactions into 4 equivalent time-points to check whether there was any difference in gaze patterns over time between the sub-groups. We measured both the number and duration of eye contact periods at 4 consecutive time points. Mixed Two-way ANOVAs were performed with number and duration over time as the within factors and sub-group membership as the between factor (Figure 3.4).

As expected from previous analyses, there was an overall significant difference between the sub-groups in both number ($p<.0001$) and duration ($p<.0001$). There was no significant difference within the number ($p=.195$) and duration ($p=.246$) of eye-contact episodes over periods of the interaction. Finally, there was no significant interaction between the sub-groups and the duration (p=.41) or number (p=.725) over time.

3.4 Discussion

The study aimed to develop and test an efficient computational phenotyping method to study the interactive behavior of young children with ASD for ecological exploration of gaze patterns during therapy. Identifying a marker sensitive to individual differences is an important goal in the perspective of personalized treatment. Despite promising results in the literature, the major bottleneck is the development of generalizable and flexible methods into real-world scenarios. For this reason, we implemented a method that is resilient to variability in data structure with added applicative value. Our approach combined unsupervised machine learning analysis with a data collection based on fine-grained features acquired by Behavior Imaging solutions.

Eye contact was studied both for its central role in the diagnostic framework as well as its value as an indicator of the severity of social-communicative symptoms in the autism spectrum (D. J. Campbell et al., 2014; Carbone et al., 2013; Cook et al., 2017; Fabiano et al., 2020; Fonger and Malott, 2019; Latrèche et al., 2021; Madipakkam et al., 2017). Indeed, it has recently been suggested in the literature that attentional patterns should be further investigated in the context of outcome prediction (Latrèche et al., 2021e) and stratification of the condition (D. J. Campbell et al., 2014; Fabiano et al., 2020).

The first part of our study addressed the development of EYE-C, the eye-contact detection model. Our main goal was to offer a more robust solution trying to overcome the translational limitations of previous implementations of CV-based systems in the clinical setting (Jaliaawala and Khan, 2020; Voss et al., 2019). Towards this end, we implemented a state-of-the-art model-based design for behavior analysis in wild videos (Cao et al., 2019; Kellnhofer et al., 2019). Our system performed well during validation in clinical scenarios with an average MCC=.74 across different interaction videos, proving to be robust to variability in both setting and video quality. We were able to identify with good precision and accuracy episodes of eye contact in highly dynamic interactions between child and therapist. To the best of our knowledge, ours is the first solution for eye contact detection in non-structured clinical settings. We implemented a method to deliver more reliable and quantifiable measurements of a behavioral feature that is very important in clinical practice on the autism spectrum. Previously developed solutions were based on a heavy structuring of interaction, which often limited the value of subsequent analysis and application aspects (Bovery et al., 2019; K. Campbell et al., 2019; Chang et al., 2021; Hashemi et al., 2018; Hurwitz et al., 2020; Li et al., 2019; Sapiro et al., 2018; Thorup et al., 2018; Yun et al., 2017). Considering the importance of integrating eye contact into intervention programs (Fonger and Malott, 2019), our method offers a solution with high potential to support the real clinical context of ASD. Our system can better decompose the dynamics of gaze and eye contact than commonly used testing techniques, i. e. ADOS-2 scores, which are not suitable to accurately quantify behavior.

In the second part of the study, we employed EYE-C to explore the dynamics of dyadic gaze coordination in child-therapist interactions. The need for more refined measurements of behavior has been highlighted in the literature to address major challenges such as stratification and outcome prediction (Leblanc et al., 2020; Washington, Park, et al., 2020). Traditional psychological testing, which is generally used in this area of research, is rather designed to diagnose and detect differences from typical development. Therefore, it is not well suited to recognize the subtle variability within the spectrum (Bentenuto et al., 2020; Washington, Leblanc, Dunlap, Penev, et al., 2020; Washington, Park, et al., 2020). In our study, more than 90% of the sample had the same maximum score in the item of ADOS-2 concerning eye contact abnormalities. Behavior Imaging offers an excellent opportunity in this perspective by allowing quantitative and refined measurements to study behavior in a more systematic way (Sapiro et al., 2018). Consistently, in the present work, we have employed eye-contact features collected in this manner combined with data-driven analysis of unsupervised and supervised ML.

The first step of the analysis covered a preliminary exploration of the gaze features extracted. We examined the correlation between our metrics and the clinical variables collected during the assessment of preschool children with ASD. The results of the regression models confirmed the presence of associations, although not robust, between the eye-contact features and the rates of symptom severity, and some subscales of cognitive functioning. As hypothesized, a negative correlation emerged between ADOS scores together with the relative subscale of social impairments (SA) and the number of eye contact episodes. Children with a higher degree of interactive deficits and higher severity of social symptoms displayed less eye-contact coordination with the therapist. These data are consistent with findings in the literature regarding the association between the degree of attention to the adult face and the severity of autistic symptoms in preschool children (Latrèche et al., 2021). No correlation was found concerning general cognitive functioning, suggesting a stronger association between attentional patterns and socio-interactive rather than cognitive aspects. However, taking into account the individual subscales, positive correlations emerged within the domains of hand-eye coordination and communication abilities. These findings are also coherent if we consider the importance of motor coordination aspects in the integration of attentive schemas and the critical role of eye contact in the later development of socio-communicative skills (Crippa et al., 2013; Fonger and Malott, 2019; Miller et al., 2017; Nebel et al., 2016). Contrary to expectation, there was no significant association with the social abilities subscale. This might be explained in part by better understanding that the social subscale of the GMDS-ER includes both items related to interactive skills and items related to the child level of autonomy, which is less related to social abilities, yet this will need to be further investigated.

In the second stage of data analysis, we stepped forward to address the challenge of autism heterogeneity. We employed unsupervised clustering based on eye-contact features to check whether sub-groups would emerge within the spectrum. Unsupervised approaches applied to computational phenotyping can also facilitate the development of fine-grained instruments and the identification of novel specifiers that may help detect reliable subtypes. Along with attentional metrics, we also considered age as a factor given the findings regarding its importance in stratification (Fabiano et al., 2020). Three different homogeneous clusters were found, which differ in gaze coordination and age. Subgroup 1 (high-coordination) is characterized by including toddlers who showed improved gaze coordination, including a higher number, frequency, and duration of eye contact episodes. In contrast, the other two sub-groups (low-coordination, 0 and 2) were characterized by lower and similar eye contact features. Considering age, the low-coordination cluster 2 (old-low-coordination) was distinguished by including children with significantly higher age than the low-coordination cluster 0 and the high-coordination cluster 1.

To summarize, two low- and high-coordination sub-groups (0 and 1) of age-matched toddlers were found, which differed significantly in terms of number, frequency, and duration of eye-contact episodes with the therapist during the interaction. In addition, a smaller third old-low-coordination sub-group (2) was identified, which was characterized by a quality of gaze coordination comparable to the other low-coordination sub-group, but a higher age. Interestingly, the children in the old-low-coordination cluster also differed from the others in displaying a higher overall distance $\boldsymbol{d}$, which measures the distance of the child's gaze from the face of the therapist. This may be explained by the fact that also in typical development, older children generally tend to be less focused on dyadic interaction and explore the environment more by gazing at objects and paying less attention to adults in general.

Afterward, the clustering was validated by taking into account the clinical variables collected during children diagnostic assessment, which included both symptom severity (ADOS-2, SA), and level of functioning scores (GQ, Coordination, Social, Motor, Language, Perform). When comparing the clinical characteristics of the sub-groups, the high-low distinction remained consistent, with high-coordination children showing higher levels of general cognitive functioning, social abilities, and hand-eye coordination along with lower scores on the symptom severity and social impairment. In comparison, the two low-coordination clusters showed comparable clinical features, including a higher degree of symptom severity and social impairments and lower cognitive functioning, social abilities, and coordination.

In a statistical analysis of the distributions, the two age-matched sub-groups of low- and high-coordination were significantly distinguished for symptom severity and hand-eye coordination abilities. The older low-coordination cluster did not differ in any clinical variable from the other low-coordination sub-group but showed differences in hand-eye coordination and social abilities compared to the high-coordination sub-group.

Altogether, some interesting data emerged from stratification. Findings seem to support the hypothesis that the autism spectrum could be stratified into two major levels of functioning, consistent with what was found in previous studies (D. J. Campbell et al., 2014; M. C. Stevens et al., 2000; Wolfers et al., 2019). Two core age-matched sub-groups emerged, one cluster consisting of autistic children with a milder symptom phenotype, better hand-eye coordination skills, and showing a higher number and duration of eye-contact episodes with the therapist, while the other cluster included autistic children with lower eye-contact features, lower hand-eye coordination abilities and a higher degree of symptom severity.

When observing data distributions, high and low functioning profiles remained stable in the three sub-groups also across the social impairments subscale of the ADOS-2. Nonetheless, no significant differences emerged as we expected, but further investigation is necessary. In addition, differences in hand-eye coordination abilities were also found significant between the high-coordination and the old-low-coordination sub-

groups. These persistent differences across clusters corroborate the results of prior analyses and are consistent with previous studies that supported a strong link between fine-motor coordination and social competencies (Crippa et al., 2013; Flori and Angeli, 2021; Johnson et al., 2016; Nebel et al., 2016; Sumner et al., 2016).
Overall, this work highlights once again the major potential of Behavior Imaging for the analysis of behavior in clinical practice (Dawson et al., 2018; Dawson and Sapiro, 2019; Sapiro et al., 2018). To provide concrete support, it is necessary to develop robust, translational approaches that are flexible to the dynamics of interaction and that take into account the variability of settings (Jaliaawala and Khan, 2020). In our study, structuring an effective system to measure gaze patterns in a refined and ecological way yielded interesting results in the field of stratification within the autism spectrum. A flexible analytical system was developed by employing advanced AI-based models with high potential for translational applications to real-world clinical scenarios (Cao et al., 2019; Kellnhofer et al., 2019). The limited number of data collected constrained the impact of the results achieved, yet it provides a promising starting point for further studies. As outlined in the literature, with increased sample size the reported number of clusters in the spectrum tends to grow (Fonger and Malott, 2019; Wolfers et al., 2019). It would be interesting to include new data to examine whether new subgroups emerge from the patterns of eye contact and to further analyze stratification in the context of personalized early intervention to investigate the role of these variables in outcome prediction.

Chapter 4

Machine Learning analysis of Eye-Contact features for predicting Early Intervention Outcomes in ASD [3]

[3]The following chapter describes analyses based on the clinical sample from (Alvari, Coviello, et al., 2021).

Abstract

This brief chapter provides an analytical extension of the data collected in (Alvari, Coviello, et al., 2021). We further explored eye-contact features in the framework of outcome prediction after early intervention. From the second study sample, we selected a sub-sample of 25 preschool children with ASD undergoing an early intensive intervention to predict symptom severity and general cognitive functioning after 1 year. Random Forest (RF) regressors were employed to predict GQ and ADOS scores at re-evaluation. Adding eye-contact features to the clinical measures as predictors improved the performance of the models in forecasting the outcome of the intervention. We further explored eye-contact features in the framework of outcome prediction after early intervention. The preliminary results of these analyses are only exploratory but represent a promising avenue that leads that prompts more in-depth investigations.

4.1 Introduction

As emphasized in the previous chapter, there is still a lack of adequate tools to measure and predict the outcome of the intervention (Baldwin et al., 2021; Bentenuto et al., 2020; Fuller and Kaiser, 2020; E. Stevens et al., 2019). While certain factors have been addressed as moderators of treatment impact, the results are not consistent (Bentenuto et al., 2020). In the clinical practice context, there is an urgent need to identify more reliable behavioral markers that can be easily monitored and can help clinicians to maximize the effectiveness of treatment.

Already widely highlighted, the application of AI-based solutions offers promising alternatives in the systematic investigation of interactive behavior (Alvari, Coviello, et al., 2021; Alvari, Furlanello, et al., 2021; Jaliaawala and Khan, 2020; Sapiro et al., 2018).

In this brief exploratory study, we tested the validity of computational solutions in identifying markers that have a predictive value for intervention outcomes. We extended the analyses of the previous study by taking into account the patterns of eye contact, which plays a very important role in the intervention and can be predictive of the acquisition of subsequent social-communication skills (Cook et al., 2017; Fabiano et al., 2020; Fonger and Malott, 2019; Miller et al., 2017).

Therefore, we hypothesized that predicting cognitive functioning and symptom severity after 12 months of intervention would be more accurate by combining gaze features and clinical variables compared with employing clinical variables alone.

4.2 Methods

4.2.1 Data collection

From the previous study sample (Chapter 3), we selected a sub-sample of 25 (3F, 22M) preschoolers with ASD. Some of the families included in the sample from our second

work agreed to undertake an intervention pathway for their child at ODFLab.
The Laboratory currently delivers early intensive intervention with a developmental approach in the local community in line with Naturalistic Developmental Behavioral Interventions (NDBI) (Tiede and Walton, 2019) as well as in accordance with the Italian sanitary system (Venuti and Bentenuto, 2017; Venuti, 2012). Intervention is comprehensive and integrates behavioral, developmental, and relational principles. The program is intensive (4-6 hours per week), including child-specific activities (speech therapy, music therapy, cognitive activities, and social play) and parental involvement. The therapists may shift during the course of therapy, but they are all trained and specialized in the same protocols and goals. Child improvement is constantly monitored and supervised.
After about 1 year of treatment, the children were reassessed to measure their progress by collecting data on cognitive functioning (GQ) and social skills (ADOS-2 raw score, Lord et al., 2012). Inclusion criteria required that the subjects had a diagnosis of ASD without presenting other medical conditions, assessed within 6 years of age, and that subjects started an intensive intervention program with a re-evaluation at least 10 months later (M duration=14.96 months, SD=5).
For each of the subjects, we collected the eye-contact features (freq, num, dur, d) and clinical variables (GQ, Coordination, Motor, Social, Perform, Language, ADOS, SA) during the first assessment from the previous study (Alvari, Coviello, et al., 2021), and in addition, the clinical variables (GQ, ADOS) measured after the intervention (T1). Following the data filtering according to the model evaluation as described in the second study (Section 3.2.4), the sub-sample was reduced from 25 to 18 subjects.
Population characteristics are summarized in Table 4.1.

Table 4.1: Population characteristics..

	T0	**T1**
	n = 18	n = 18
Age (months), mean (SD)	41.1 (12.8)	54.5 (13.3)
GQ, mean (SD)	75.39 (15.4)	82.29 (18.6)
ADOS, mean (SD)	14.1 (3.1)	12.9 (3.8)
Intervention duration (months), mean (SD)		13.5 (3.5)

Note: ADOS: Autism Diagnostic Observation Schedule, 2nd edition, raw score; GQ: Global Developmental Quotient; ASD: Autism Spectrum Disorders; T0: first assessment; T1: re-evaluation after the intervention.

4.2.2 Data analysis

To analyze the predictive power of eye-contact features, we employed supervised ML RF regressors. We randomly split the dataset into training, validation, and testing, leaving 30% for the latter. Leave-One-Out cross-validation (LOOCV) was then applied to the train-val dataset for hyperparameter tuning by grid search. The best-performing model according to Mean Squared Error (MSE) score was then selected. We first employed the RF to predict ADOS and GQ scores at re-evaluation (T1) based on clinical variables and age at first evaluation (T0). Subsequently, we re-run the procedure by adding eye-contact features as predictors to check for improvement in the performance of models.

4.3 Results

The predictive power of eye-contact features was tested in the intervention sub-sample to assess the prediction of ADOS and GQ at re-evaluation after treatment (T1) (Figure 4.1).

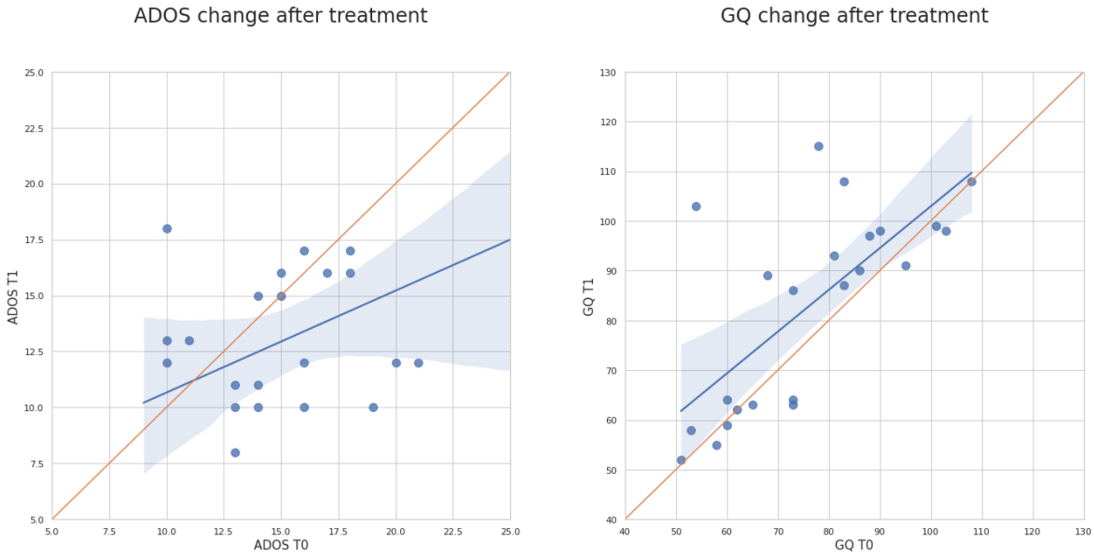

Figure 4.1: ADOS and GQ change after treatment (T0 vs. T1). ADOS T0/T1: ADOS raw score at T0/T1; GQ T0/T1: general developmental quotient at T0/T1.

We employed RF regressors using first the clinical variables with age at T0 as predictors and then adding eye-contact metrics to check for changes in performance.
Regarding the GQ in T1, the model with the best performance (16 trees, 2 maximum depths) obtained an MSE=60.08 in validation and MSE=229.2 in testing. By further exploring the weight of the features in the prediction, the Motor subscale achieved the highest importance score (Figure 4.2 A1). When adding eye-contact features to the predictors, the model (16 trees, 2 maximum depth) obtained an MSE=64.86 in validation and an MSE=127.93 in testing. The Motor subscale maintained a high importance score along with the frequency of eye-contact periods (Figure 4.2 A2).
Concerning the ADOS-2 score in T1, the RF with the best performance (16 trees, 2 maximum depth) obtained an MSE=4.23 in validation and an MSE=22.44 in testing. The predictors with a higher importance score were the subscale of motor and social abilities (Figure 4.2 B1). When we added eye-contact features to the predictors, the best model (16 trees, 3 maximum depth) achieved an MSE=3.88 in validation and an MSE=16.23

in testing. Components with a higher importance in prediction included the Motor subscale along with the frequency of eye-contact periods during interactions (Figure 4.2 B2).

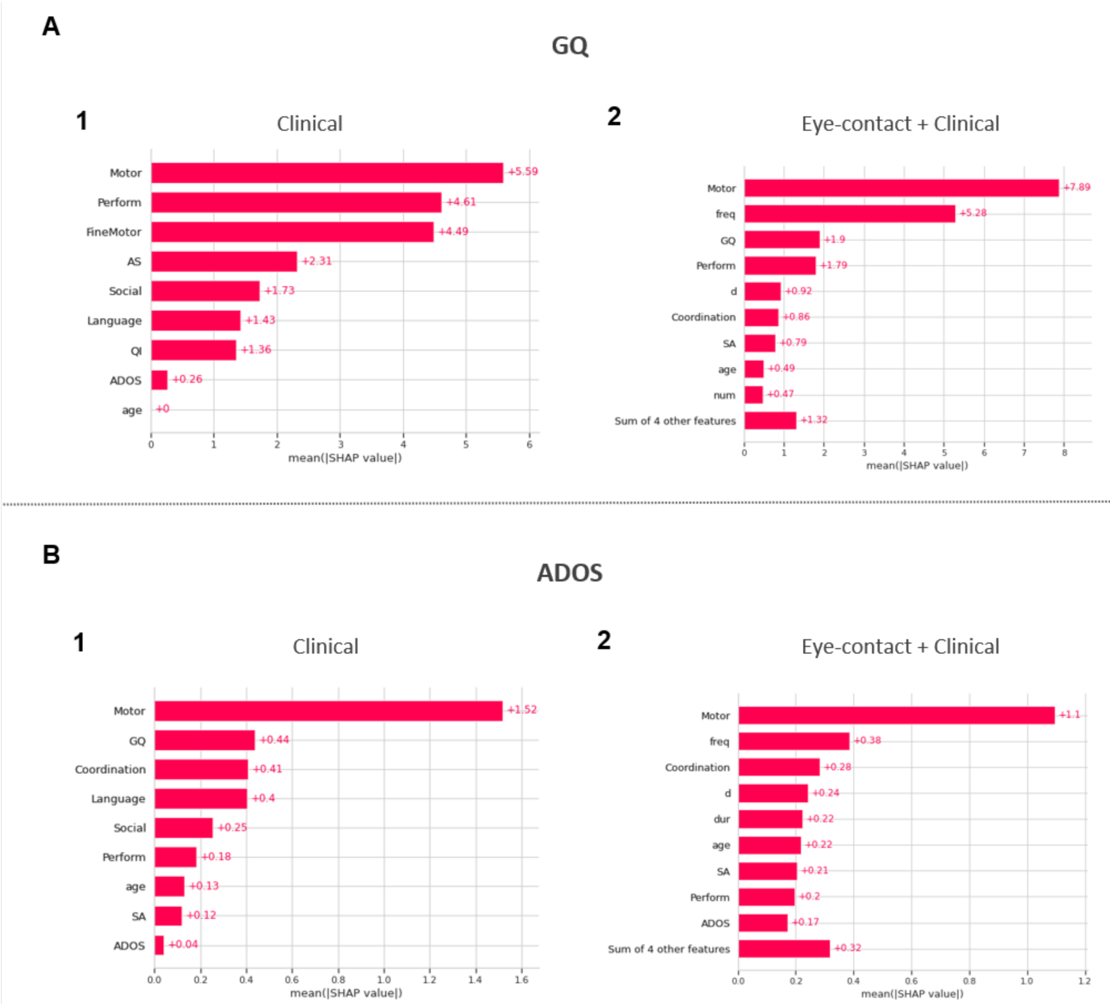

Figure 4.2: Random Forest regressor feature importance in the prediction of GQ (A) and ADOS (B) at T1; A1) results for GQ with only clinical variables; A2) results for GQ with both clinical variables and eye-contact features; B1) results for ADOS with only clinical variables; B2) results for ADOS with both clinical variables and eye contact features.

Finally, we further analyzed the predictive power of our gaze features by applying them alone in predicting both GQ and ADOS after the intervention (T1). Concerning GQ, the RF with the best performance (64 trees, 2 maximum depth) obtained an MSE=87.92 in validation and an MSE=249.74 in testing. Frequency was the feature with the highest weight in the prediction (Figure 4.3 A). As for ADOS, the best model (16 trees, 2 maximum depth) obtained an MSE=4.49 in validation and an MSE=7.41 in testing. Both frequency of eye-contact periods and the overall distance of the child gaze during interaction scored high in feature importance (Figure 4.3 B).

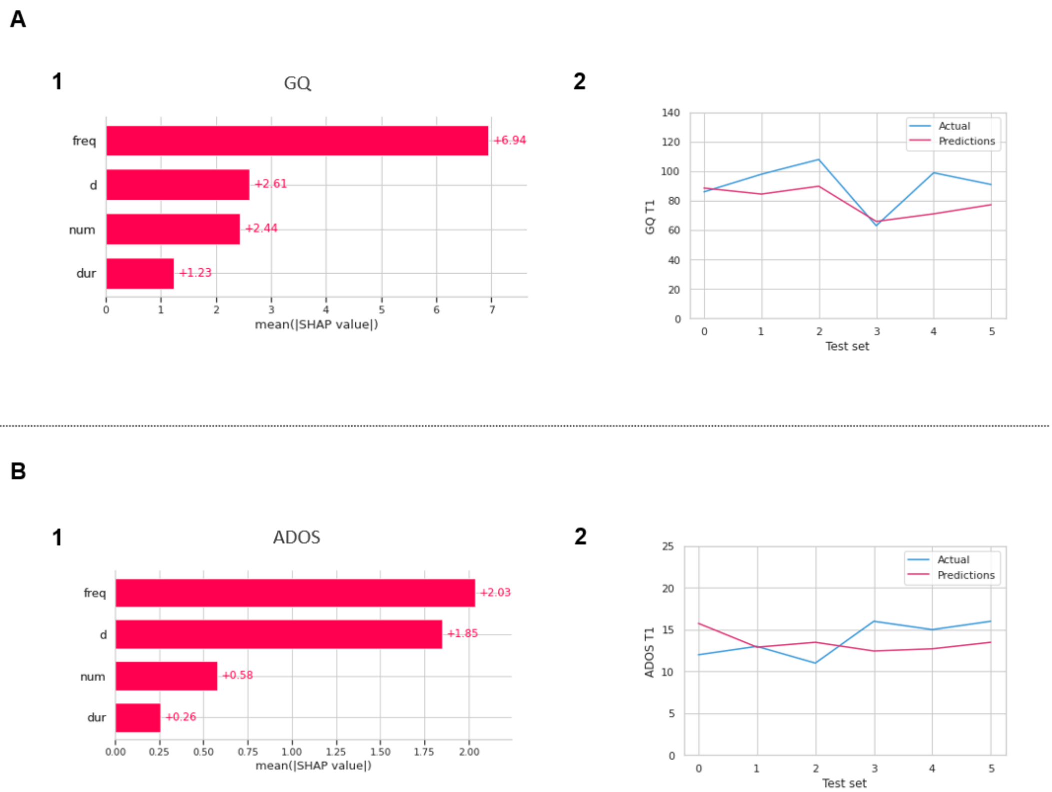

Figure 4.3: Random Forest regressors (1) feature importance and (2) prediction results for A) GQ; B) ADOS.

4.4 Discussion

In this latter exploratory study, we extended the analyses of previous work (Alvari, Coviello, et al., 2021) by testing our eye-gaze features to predict early treatment outcomes on 18 cases after 12 months. Including gaze metrics among the predictors improved the performance of the ML models for both general cognitive functioning and symptom severity. The combination of eye-contact features and clinical measurements resulted in improved MSE scores of around 30% in predicting ADOS-2 and 40% in predicting the GQ at re-evaluation, respectively.

Compared to clinical variables, we did not achieve better results in predicting general cognitive functioning (GQ) when using eye contact metrics alone. In this respect, the highest scores are still obtained by combining clinical and eye-contact data. However, we performed much better (near 70%) in predicting the symptom severity (ADOS) after the intervention when using eye contact metrics versus clinical variables.

Given the limited sample size, the results of this study should only be interpreted from an exploratory perspective. Nonetheless, fine-grained eye contact measurements appear to have a predictive value concerning treatment outcomes, especially in the degree of condition impairments. The results seemed to confirm what already emerged in the literature regarding the role of socio-attentive patterns in the impact of treatment and acquisition of social competencies (D. J. Campbell et al., 2014; Latrèche et al., 2021), highlighting the importance of sustained eye contact in early intervention programs (Fonger and Malott, 2019). Nonetheless, fine-grained eye contact measurements appear to have a predictive value concerning treatment outcomes, especially in the degree of condition impairments.

Along with findings from the previous study (Alvari, Coviello, et al., 2021), these analyses corroborate the potential of computational phenotyping for providing systematic analyses that may have an important role in the clinical framework of ASD. The application of AI-based methodologies may deliver proper solutions to address the challenges of personalized intervention (Jaliaawala and Khan, 2020). Designing methods capable of monitoring behaviors sensitive to spectrum heterogeneity and have predictive value for treatment outcomes may support clinicians in tailoring intervention programs to children characteristics and maximizing the impact.

Chapter 5

Conclusions & Future Directions

This project aims to highlight the potential of advanced computational solutions for addressing the ongoing challenges in clinical practice of the autism spectrum. Our first study focused on developing methods to elicit new early behavioral markers for early detection of the condition within a diagnostic framework (Alvari, Furlanello, et al., 2021). Other works focused on early intervention, trying to address the issues of heterogeneity and lack of behavioral predictors of outcome related to the design of personalized intervention programs (Alvari, Coviello, et al., 2021). The twofold approach shared by all studies was on the one hand to develop advanced methods that could systematically and ecologically analyze atypical child behavior, and on the other hand to show the feasibility of such tools in a real clinical environment, addressing the needs of clinicians for more practical solutions with applicative value.

Behavioral Imaging and computational phenotyping now represent possible solutions to support clinical practice (Dawson et al., 2018; Dawson and Sapiro, 2019; de Belen et al., 2020; Sapiro et al., 2018). Although related literature is still experimental in the psychiatric domain, the findings are very promising. Our studies represent further evidence of the potential of AI-based methods (Alvari, Coviello, et al., 2021; Alvari, Furlanello, et al., 2021).

The autism spectrum is a very complex but widely studied condition. Perhaps the underlying nature of the issues persisting in ASD research is less a matter of asking the wrong questions and more a matter of using the wrong methods. This project explored ways to address clinicians needs for systematic and refined behavioral measures (Lord et al., 2012). The objective was to develop effective solutions with translational value, as this is a leading concern between clinical and research environments (Jaliaawala and Khan, 2020).

Therefore, we prioritized efforts to provide methods that were quantitative, refined, ecological, and robust to the high variability of dynamic interactive scenarios. As well as offering automatic alternatives that are cheaper in terms of cost and time, these approaches can allow clinicians and researchers to observe otherwise inaccessible subtle behaviors (i.e., infrequent, hard-to-detect, or triggered at-home behaviors) in more natural scenarios (Dawson et al., 2018; Nazneen et al., 2015). Further, our approaches relied on paradigms involving rule-based Behavior Imaging models to extract a quantitative set of features from images and then leverage supervised or unsupervised ML for classification or clustering (Alvari, Coviello, et al., 2021; Alvari, Furlanello, et al., 2021; Washington, Park, et al., 2020). Combining these techniques allowed us to categorize atypical infant behavior according to more systematic and fine-grained metrics and overcome the limitations of more traditional methods of observation (Bentenuto et al., 2020; Dawson and Sapiro, 2019; de Belen et al., 2020; Hyde et al., 2019; Sapiro et al., 2018).

The results we achieved through this project may offer novel insights both from a methodological standpoint, highlighting the application potential of state-of-the-art technologies, and from a more clinical perspective in terms of the relevance of specific social behaviors (i.e., facial expressions, eye contact) for early diagnosis and intervention (Alvari, Coviello, et al., 2021; Alvari, Furlanello, et al., 2021).

The interpretation of the findings from the studies we conducted is restricted by the limited size of the clinical samples, which in turn complicates drawing reliable theoretical inferences about the underlying behavioral processes. Nonetheless, the major contribution of our work is related to the approach and methods employed. Another critical limitation refers to the scarce control over data, i. e. the lack of structuring of interactions between child and therapist/caregiver or the high variability in settings and quality of the videos. This wide variance of data may compromise the quality of the analysis and the reliability of the methods employed yet promote the generalizability of the results. The gap between research and stakeholders is mainly due to a preference for more structured data favoring better performing models (Jaliaawala and Khan, 2020; Voss et al., 2019). The nature of our approach is oriented towards developing more applicative solutions and greater attention to the needs of the clinical practice. Hence, the results of the studies must also be interpreted by taking into account this reasoning and the inherent limitations of adopting this strategy.

Implementing AI-based models in the clinical framework represents the new frontier and can give valuable support to children with ASD and their environment. An interesting outlook derives from the opportunity to offer telehealth solutions that may allow clinicians to monitor behavior remotely through parent-recorded materials (Nazneen et al., 2015). This viewpoint becomes even more relevant if we consider socio-cultural scenarios where access to healthcare services is limited, such as during a pandemic.

Overall, the studies we have carried out in this doctoral project yield some encouraging insights (Alvari, Coviello, et al., 2021; Alvari, Furlanello, et al., 2021). Future perspectives of our research are manifold and include both a more in-depth analysis of the results as well as the implementation of these solutions in more concrete systems. The models we developed represent an initial step towards structuring technologies capable of analyzing social behavior more comprehensively. It will be interesting to design more advanced models that can capture more complex interactive dynamics, i.e., by including elements of object-based attentional triangulation, which is another significant factor in the diagnosis and intervention of toddlers with ASD (Bentenuto et al., 2020; Rodgers

et al., 2021; Rogers and Talbott, 2016; Venuti and Bentenuto, 2017; Venuti, 2012).
A further step is to combine multiple behavioral features by acquiring a multimodal perspective about child behavior, which has proven to be a successful strategy in Behavior Imaging (Rudovic et al., 2018). From the other side, it will also be important to evaluate how advanced AI-based solutions can be embedded into interactive systems which can be employed within clinical practice. In this perspective, the focus will be to design concrete applications solutions in cooperation with families and clinicians, leading to increasingly adaptive user interfaces.

Bibliography

Abbas, H., Garberson, F., Glover, E., & Wall, D. P. (2018). Machine learning approach for early detection of autism by combining questionnaire and home video screening. *Journal of the American Medical Informatics Association*, *25*(8), 1000–1007.

Achenbach, T. M., & Dumenci, L. (2001). Advances in empirically based assessment: Revised cross-informant syndromes and new dsm-oriented scales for the cbcl, ysr, and trf: Comment on lengua, sadowski, friedrich, and fisher (2001).

Allaoui, M., Kherfi, M. L., & Cheriet, A. (2020). Considerably improving clustering algorithms using umap dimensionality reduction technique: A comparative study. *International Conference on Image and Signal Processing*, 317–325.

Alnajjar, F., Cappuccio, M., Renawi, A., Mubin, O., & Loo, C. K. (2021). Personalized robot interventions for autistic children: An automated methodology for attention assessment. *International Journal of Social Robotics*, *13*(1), 67–82.

Alvari, G., Coviello, L., & Furlanello, C. (2021). Eye-c: Eye-contact robust detection and analysis during unconstrained child-therapist interactions in the clinical setting of autism spectrum disorders. *Brain Sciences*, *11*(12), 1555.

Alvari, G., Furlanello, C., & Venuti, P. (2021). Is smiling the key? machine learning analytics detect subtle patterns in micro-expressions of infants with asd. *Journal of clinical medicine*, *10*(8), 1776.

Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, 3686–3693.

APA. (2013). Diagnostic and statistical manual of mental disorders: Dsm-5. *Arlington, VA*.

Association, W. M. et al. (2013). World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. *Jama*, *310*(20), 2191–2194.

Azhari, A., Gabrieli, G., Bizzego, A., Bornstein, M. H., & Esposito, G. (2020). Probing the association between maternal anxious attachment style and mother-child brain-to-brain coupling during passive co-viewing of visual stimuli. *Attachment & Human Development*, 1–16.

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Rosenberg, C. R., White, T., et al. (2018). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2014. *MMWR Surveillance Summaries*, *67*(6), 1.

Baldwin, M. M., Xiao, Z., & Murray, A. (2021). Temporal synchrony in autism: A systematic review. *Review Journal of Autism and Developmental Disorders*, 1–22.

Baltrušaitis, T., Mahmoud, M., & Robinson, P. (2015). Cross-dataset learning and person-specific normalisation for automatic action unit detection. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, *6*, 1–6.

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2014). Continuous conditional neural fields for structured regression. *European conference on computer vision*, 593–608.

Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, 59–66.

Bangerter, A., Chatterjee, M., Manfredonia, J., Manyakov, N. V., Ness, S., Boice, M. A., Skalkin, A., Goodwin, M. S., Dawson, G., Hendren, R., et al. (2020). Automated recognition of spontaneous facial expression in individuals with autism spectrum disorder: Parsing response variability. *Molecular autism*, *11*(1), 1–15.

Barbaro, J., & Dissanayake, C. (2013). Early markers of autism spectrum disorders in infants and toddlers prospectively identified in the social attention and communication study. *Autism*, *17*(1), 64–86.

Barbaro, J., & Halder, S. (2016). Early identification of autism spectrum disorder: Current challenges and future global directions. *Current Developmental Disorders Reports*, *3*(1), 67–74.

Baron-Cohen, S. (2017). Editorial perspective: Neurodiversity–a revolutionary concept for autism and psychiatry.

Begeer, S., Koot, H. M., Rieffe, C., Terwogt, M. M., & Stegge, H. (2008). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. *Developmental Review*, *28*(3), 342–369.

Ben-Itzchak, E., Watson, L. R., & Zachor, D. A. (2014). Cognitive ability is associated with different outcome trajectories in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(9), 2221–2229.

Bentenuto, A., Bertamini, G., Perzolli, S., & Venuti, P. (2020). Changes in developmental trajectories of preschool children with autism spectrum disorder during parental based intensive intervention. *Brain sciences*, *10*(5), 289.

Bishop-Fitzpatrick, L., Hong, J., Smith, L. E., Makuch, R. A., Greenberg, J. S., & Mailick, M. R. (2016). Characterizing objective quality of life and normative outcomes in adults with autism spectrum disorder: An exploratory latent class analysis. *Journal of autism and developmental disorders*, *46*(8), 2707–2719.

Bizzego, A., Battisti, A., Gabrieli, G., Esposito, G., & Furlanello, C. (2019). Pyphysio: A physiological signal processing library for data science approaches in physiology. *SoftwareX*, *10*, 100287.

Bolzani, L., Messinger, D. S., Yale, M. E., & Dondi, M. (2002). Smiling in infancy. *An empirical reflection on the smile*, 111–135.

Bornstein, M. H. (2013). Parenting and child mental health: A cross-cultural perspective. *World Psychiatry*, *12*(3), 258–265.

Bornstein, M. H., Hahn, C.-S., & Putnick, D. L. (2016). Stability of core language skill across the first decade of life in children at biological and social risk. *Journal of Child Psychology and Psychiatry*, *57*(12), 1434–1443.

Boterberg, S., Van Coster, R., & Roeyers, H. (2019). Characteristics, early development and outcome of parent-reported regression in autism spectrum disorder. *Journal of autism and developmental disorders*, *49*(11), 4603–4625.

Bovery, M. D. M. J., Dawson, G., Hashemi, J., & Sapiro, G. (2019). A scalable off-the-shelf framework for measuring patterns of attention in young children and its application in autism spectrum disorder. *IEEE Transactions on Affective Computing*.

Brett, D., Warnell, F., McConachie, H., & Parr, J. R. (2016). Factors affecting age at asd diagnosis in uk: No evidence that diagnosis age has decreased between 2004 and 2014. *Journal of autism and developmental disorders*, *46*(6), 1974–1984.

Bryson, S. E., Zwaigenbaum, L., McDermott, C., Rombough, V., & Brian, J. (2008). The autism observation scale for infants: Scale development and reliability data. *Journal of autism and developmental disorders*, *38*(4), 731–738.

Buxbaum, J. D., & Hof, P. R. (2012). *The neuroscience of autism spectrum disorders*. Academic Press.

Cabibihan, J.-J., Chellali, R., So, C. W. C., Aldosari, M., Connor, O., Alhaddad, A. Y., & Javed, H. (2018). Social robots and wearable sensors for mitigating meltdowns in autism-a pilot test. *International Conference on Social Robotics*, 103–114.

Campbell, D. J., Shic, F., Macari, S., & Chawarska, K. (2014). Gaze response to dyadic bids at 2 years related to outcomes at 3 years in autism spectrum disorders: A subtyping analysis. *Journal of autism and developmental disorders*, *44*(2), 431–442.

Campbell, K., Carpenter, K. L., Hashemi, J., Espinosa, S., Marsan, S., Borg, J. S., Chang, Z., Qiu, Q., Vermeer, S., Adler, E., et al. (2019). Computer vision analysis captures atypical attention in toddlers with autism. *Autism*, *23*(3), 619–628.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE transactions on pattern analysis and machine intelligence*, *43*(1), 172–186.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7291–7299.

Carbone, V. J., O'Brien, L., Sweeney-Kerwin, E. J., & Albert, K. M. (2013). Teaching eye contact to children with autism: A conceptual analysis and single case study. *Education and Treatment of Children*, *36*(2), 139–159.

Carpenter, K. L., Hahemi, J., Campbell, K., Lippmann, S. J., Baker, J. P., Egger, H. L., Espinosa, S., Vermeer, S., Sapiro, G., & Dawson, G. (2021). Digital behavioral phenotyping detects atypical pattern of facial expression in toddlers with autism. *Autism Research*, *14*(3), 488–499.

Cavus, N., Lawan, A. A., Ibrahim, Z., Dahiru, A., Tahir, S., Abdulrazak, U. I., & Hussaini, A. (2021). A systematic literature review on the application of machine-learning models in behavioral assessment of autism spectrum disorder. *Journal of Personalized Medicine*, *11*(4), 299.

Chang, Z., Di Martino, J. M., Aiello, R., Baker, J., Carpenter, K., Compton, S., Davis, N., Eichner, B., Espinosa, S., Flowers, J., et al. (2021). Computational methods to measure patterns of gaze in toddlers with autism spectrum disorder. *JAMA pediatrics*.

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological psychiatry*, *74*(3), 195–203.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in cognitive sciences*, *16*(4), 231–239.

Chicco, D., & Jurman, G. (2020). The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*(1), 1–13.

Chong, E., Chanda, K., Ye, Z., Southerland, A., Ruiz, N., Jones, R. M., Rozga, A., & Rehg, J. M. (2017). Detecting gaze towards eyes in natural social interactions and its use in child assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–20.

Chung, E. Y.-h. (2019). Robotic intervention program for enhancement of social engagement among children with autism spectrum disorder. *Journal of Developmental and Physical Disabilities*, *31*(4), 419–434.

Cockburn, J., Bartlett, M., Tanaka, J., Movellan, J., Pierce, M., & Schultz, R. (2008). Smilemaze: A tutoring system in real-time facial expression perception and production in children with autism spectrum disorder. *ECAG 2008 workshop facial and bodily expressions for control and adaptation of games*, *3*.

Cook, J. L., Rapp, J. T., Mann, K. R., McHugh, C., Burji, C., & Nuta, R. (2017). A practitioner model for increasing eye contact in children with autism. *Behavior Modification*, *41*(3), 382–404.

Crippa, A., Forti, S., Perego, P., & Molteni, M. (2013). Eye-hand coordination in children with high functioning autism and asperger's disorder using a gap-overlap paradigm. *Journal of autism and developmental disorders*, *43*(4), 841–850.

Czapinski, P., & Bryson, S. (2003). Reduced facial muscle movements in autism: Evidence for dysfunction in the neuromuscular pathway? *Brain and Cognition*, *51*(2), 177–179.

Davison, A., Merghani, W., Lansley, C., Ng, C.-C., & Yap, M. H. (2018). Objective micro-facial movement detection using facs-based regions and baseline evaluation. *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, 642–649.

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and psychopathology*, *20*(3), 775–803.

Dawson, G., Campbell, K., Hashemi, J., Lippmann, S. J., Smith, V., Carpenter, K., Egger, H., Espinosa, S., Vermeer, S., Baker, J., et al. (2018). Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder. *Scientific reports*, *8*(1), 1–7.

Dawson, G., & Sapiro, G. (2019). Potential for digital behavioral measurement tools to transform the detection and diagnosis of autism spectrum disorder. *JAMA pediatrics*, *173*(4), 305–306.

de Belen, R. A. J., Bednarz, T., Sowmya, A., & Del Favero, D. (2020). Computer vision in autism spectrum disorder research: A systematic review of published studies from 2009 to 2019. *Translational psychiatry*, *10*(1), 1–20.

Del Coco, M., Leo, M., Carcagni, P., Spagnolo, P., Luigi Mazzeo, P., Bernava, M., Marino, F., Pioggia, G., & Distante, C. (2017). A computer vision based approach for understanding emotional involvements in children with autism spectrum disorders. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 1401–1407.

Dondi, M., Messinger, D. S., Colle, M., Tabasso, A., Simion, F., Barba, B. D., & Fogel, A. (2007). A new perspective on neonatal smiling: Differences between the judgments of expert coders and naive observers. *Infancy*, *12*(3), 235–255.

Drimalla, H., Baskow, I., Behnia, B., Roepke, S., & Dziobek, I. (2021). Imitation and recognition of facial emotions in autism: A computer vision approach. *Molecular autism*, *12*(1), 1–15.

Drimalla, H., Landwehr, N., Baskow, I., Behnia, B., Roepke, S., Dziobek, I., & Scheffer, T. (2018). Detecting autism by analyzing a simulated social interaction. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 193–208.

Drimalla, H., Scheffer, T., Landwehr, N., Baskow, I., Roepke, S., Behnia, B., & Dziobek, I. (2020). Towards the automatic detection of social biomarkers in autism spectrum disorder: Introducing the simulated interaction task (sit). *NPJ digital medicine*, *3*(1), 1–10.

Duda, M., Kosmicki, J., & Wall, D. P. (2014). Testing the accuracy of an observation-based classifier for rapid detection of autism risk. *Translational psychiatry*, *4*(8), e424–e424.

Duda, M., Daniels, J., & Wall, D. P. (2016). Clinical evaluation of a novel and mobile autism risk assessment. *Journal of autism and developmental disorders*, *46*(6), 1953–1961.

Eapen, V., Črnčec, R., & Walter, A. (2013). Clinical outcomes of an early intervention program for preschool children with autism spectrum disorder in a community group setting. *BMC pediatrics*, *13*(1), 1–9.

Egger, H. L., Dawson, G., Hashemi, J., Carpenter, K. L., Espinosa, S., Campbell, K., Brotkin, S., Schaich-Borg, J., Qiu, Q., Tepper, M., et al. (2018). Automatic emotion and attention analysis of young children at home: A researchkit autism feasibility study. *NPJ digital medicine*, *1*(1), 1–10.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The duchenne smile: Emotional expression and brain physiology: Ii. *Journal of personality and social psychology*, *58*(2), 342.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding systems*. Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system: Facial action coding system: The manual: On cd-rom*. Research Nexus.

Elder, J. H., Kreider, C. M., Brasher, S. N., & Ansell, M. (2017). Clinical impact of early diagnosis of autism on the prognosis and parent–child relationships. *Psychology research and behavior management*.

Esposito, G., & Venuti, P. (2009). Comparative analysis of crying in children with autism, developmental delays, and typical development. *Focus on Autism and Other Developmental Disabilities*, *24*(4), 240–247.

Esposito, G., Venuti, P., Apicella, F., & Muratori, F. (2011). Analysis of unsupported gait in toddlers with autism. *Brain and development*, *33*(5), 367–373.

Estes, A., Swain, D. M., & MacDuffie, K. E. (2019). The effects of early autism intervention on parents and family adaptive functioning. *Pediatric medicine (Hong Kong, China)*, *2*.

Fabiano, D., Canavan, S., Agazzi, H., Hinduja, S., & Goldgof, D. (2020). Gaze-based classification of autism spectrum disorder. *Pattern Recognition Letters*, *135*, 204–212.

Fang, H.-S., Xie, S., Tai, Y.-W., & Lu, C. (2017). Rmpe: Regional multi-person pose estimation. *Proceedings of the IEEE international conference on computer vision*, 2334–2343.

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: A survey. *Pattern recognition*, *36*(1), 259–275.

Faso, D. J., Sasson, N. J., & Pinkham, A. E. (2015). Evaluating posed and evoked facial expressions of emotion from adults with autism spectrum disorder. *Journal of autism and developmental disorders*, *45*(1), 75–89.

Fein, D., Stevens, M., Dunn, M., Waterhouse, L., Allen, D., Rapin, I., & Feinstein, C. (1999). Subtypes of pervasive developmental disorder: Clinical characteristics. *Child Neuropsychology*, *5*(1), 1–23.

Feldman, H. M., Lee, E. S., Yeatman, J. D., & Yeom, K. W. (2012). Language and reading skills in school-aged children and adolescents born preterm are associated with white matter properties on diffusion tensor imaging. *Neuropsychologia*, *50*(14), 3348–3362.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, *32*(9), 1627–1645.

Fischer, T., Chang, H. J., & Demiris, Y. (2018). Rt-gene: Real-time eye gaze estimation in natural environments. *Proceedings of the European Conference on Computer Vision (ECCV)*, 334–352.

Flanagan, J. E., Landa, R., Bhat, A., & Bauman, M. (2012). Head lag in infants at risk for autism: A preliminary study. *American Journal of Occupational Therapy*, *66*(5), 577–585.

Flori, V., & Angeli, B. (2021). Motor intervention and autism: One way, several domains of effectiveness. *Journal of Advanced Health Care*, *3*(2), 52–61.

Fogel, A., Nelson-Goens, G. C., & Hsu, H.-C. (2000). Do different infant smiles reflect different positive emotions? *Social Development*, *9*(4), 497–520.

Fonger, A. M., & Malott, R. W. (2019). Using shaping to teach eye contact to children with autism spectrum disorder. *Behavior analysis in practice*, *12*(1), 216–221.

Fountain, C., Winter, A. S., & Bearman, P. S. (2012). Six developmental trajectories characterize children with autism. *Pediatrics*, *129*(5), e1112–e1120.

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3453–3458.

Fuller, E. A., & Kaiser, A. P. (2020). The effects of early intervention on social communication outcomes for children with autism spectrum disorder: A meta-analysis. *Journal of Autism and Developmental Disorders*, *50*(5), 1683–1700.

Georgescu, A. L., Koehler, J. C., Weiske, J., Vogeley, K., Koutsouleris, N., & Falter-Wagner, C. (2019). Machine learning to study social interaction difficulties in asd. *Frontiers in Robotics and AI*, *6*, 132.

Ghosh, S., Dhall, A., Hayat, M., Knibbe, J., & Ji, Q. (2021). Automatic gaze analysis: A survey of deeplearning based approaches. *arXiv preprint arXiv:2108.05479*.

Gima, H., Kihara, H., Watanabe, H., Nakano, H., Nakano, J., Konishi, Y., Nakamura, T., & Taga, G. (2018). Early motor signs of autism spectrum disorder in spontaneous position and movement of the head. *Experimental brain research*, *236*(4), 1139–1148.

Golan, O., Ashwin, E., Granader, Y., McClintock, S., Day, K., Leggett, V., & Baron-Cohen, S. (2010). Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces. *Journal of autism and developmental disorders*, *40*(3), 269–279.

Gordon, I., Pierce, M. D., Bartlett, M. S., & Tanaka, J. W. (2014). Training facial expression production in children on the autism spectrum. *Journal of autism and developmental disorders*, *44*(10), 2486–2498.

Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ados scores. *Pediatrics*, *130*(5), e1278–e1284.

Grizzle, R., Goldstein, S., & Naglieri, J. (2011). Wechsler intelligence scale for children. *Encyclopaedia of Child Behaviour and Development*, 1553–1555.

Grossard, C., Dapogny, A., Cohen, D., Bernheim, S., Juillet, E., Hamel, F., Hun, S., Bourgeois, J., Pellerin, H., Serret, S., et al. (2020). Children with autism spectrum disorder produce more ambiguous and less socially meaningful facial expressions: An experimental study using random forest classifiers. *Molecular autism*, *11*(1), 1–14.

Grossman, R. B. (2015). Judgments of social awkwardness from brief exposure to children with and without high-functioning autism. *Autism*, *19*(5), 580–587.

Gudi, A., Tasli, H. E., Den Uyl, T. M., & Maroulis, A. (2015). Deep learning based facs action unit occurrence and intensity estimation. *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, *6*, 1–5.

Guha, T., Yang, Z., Grossman, R. B., & Narayanan, S. S. (2016). A computational study of expressive facial dynamics in children with autism. *IEEE transactions on affective computing*, *9*(1), 14–20.

Harrold, N., Tan, C. T., Rosser, D., & Leong, T. W. (2014). Copyme: A portable real-time feedback expression recognition game for children. *Chi'14 extended abstracts on human factors in computing systems* (pp. 1195–1200).

Hashemi, J., Dawson, G., Carpenter, K. L., Campbell, K., Qiu, Q., Espinosa, S., Marsan, S., Baker, J. P., Egger, H. L., & Sapiro, G. (2018). Computer vision analysis for quantification of autism risk behaviors. *IEEE Transactions on Affective Computing*.

Hashemi, J., Tepper, M., Vallin Spina, T., Esler, A., Morellas, V., Papanikolopoulos, N., Egger, H., Dawson, G., & Sapiro, G. (2014). Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism research and treatment*, *2014*.

Havdahl, A., & Bishop, S. (2019). Heterogeneity in prevalence of co-occurring psychiatric conditions in autism. *The lancet. Psychiatry*, *6*(10), 794–795.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

Higuchi, K., Matsuda, S., Kamikubo, R., Enomoto, T., Sugano, Y., Yamamoto, J., & Sato, Y. (2018). Visualizing gaze direction to support video coding of social attention for children with autism spectrum disorder. *23rd International Conference on Intelligent User Interfaces*, 571–582.

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PloS one*, *10*(8), e0136905.

Hopkins, I. M., Gower, M. W., Perez, T. A., Smith, D. S., Amthor, F. R., Wimsatt, F. C., & Biasini, F. J. (2011). Avatar assistant: Improving social skills in students with an asd through a computer-based intervention. *Journal of autism and developmental disorders*, *41*(11), 1543–1555.

Hu, V. W., & Steinberg, M. E. (2009). Novel clustering of items from the autism diagnostic interview-revised to define phenotypes within autism spectrum disorders. *Autism Research*, *2*(2), 67–77.

Huang, Q., Veeraraghavan, A., & Sabharwal, A. (2017). Tabletgaze: Dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets. *Machine Vision and Applications*, *28*(5), 445–461.

Hurwitz, S., Ryan, T., & Kennedy, D. P. (2020). Developing social communication skills using dual first-person video recording glasses: A novel intervention for adolescents with autism. *Journal of autism and developmental disorders*, *50*(3), 904–915.

Hyde, K. K., Novack, M. N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D. R., & Linstead, E. (2019). Applications of supervised machine learning in autism spectrum disorder research: A review. *Review Journal of Autism and Developmental Disorders*, *6*(2), 128–146.

Jacob, A., Scott, M., Falkmer, M., & Falkmer, T. (2015). The costs and benefits of employing an adult with autism spectrum disorder: A systematic review. *PloS one*, *10*(10), e0139896.

Jacob, S., Wolff, J. J., Steinbach, M. S., Doyle, C. B., Kumar, V., & Elison, J. T. (2019). Neurodevelopmental heterogeneity and computational approaches for understanding autism. *Translational psychiatry*, *9*(1), 1–12.

Jain, S., Thiagarajan, B., Shi, Z., Clabaugh, C., & Matarić, M. J. (2020). Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Science Robotics*, *5*(39).

Jaliaawala, M. S., & Khan, R. A. (2020). Can autism be catered with artificial intelligence-assisted intervention technology? a comprehensive survey. *Artificial Intelligence Review*, *53*(2), 1039–1069.

Jarraya, S. K., Masmoudi, M., & Hammami, M. (2020). Compound emotion recognition of autistic children during meltdown crisis based on deep spatio-temporal analysis of facial geometric features. *IEEE Access*, *8*, 69311–69326.

Jeni, L. A., & Cohn, J. F. (2016). Person-independent 3d gaze estimation using face frontalization. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 87–95.

Jiang, M., & Zhao, Q. (2017). Learning visual attention to identify people with autism spectrum disorder. *Proceedings of the ieee international conference on computer vision*, 3267–3276.

Johnson, B. P., Lum, J. A., Rinehart, N. J., & Fielding, J. (2016). Ocular motor disturbances in autism spectrum disorders: Systematic review and comprehensive meta-analysis. *Neuroscience & Biobehavioral Reviews*, *69*, 260–279.

Jones, E. J., Gliga, T., Bedford, R., Charman, T., & Johnson, M. H. (2014). Developmental pathways to autism: A review of prospective studies of infants at risk. *Neuroscience & Biobehavioral Reviews*, *39*, 1–33.

Jones, E. J., Venema, K., Earl, R., Lowy, R., Barnes, K., Estes, A., Dawson, G., & Webb, S. (2016). Reduced engagement with social stimuli in 6-month-old infants with later autism spectrum disorder: A longitudinal prospective study of infants at high familial risk. *Journal of neurodevelopmental disorders*, *8*(1), 1–20.

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431.

Kalantarian, H., Washington, P., Schwartz, J., Daniels, J., Haber, N., & Wall, D. P. (2018). A gamified mobile system for crowdsourcing video for autism research. *2018 IEEE international conference on healthcare informatics (ICHI)*, 350–352.

Kaltwang, S., Todorovic, S., & Pantic, M. (2015). Latent trees for estimating intensity of facial action units. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 296–304.

Kawulok, M., Nalepa, J., Kawulok, J., & Smolka, B. (2021). Dynamics of facial actions for assessing smile genuineness. *Plos one*, *16*(1), e0244647.

Kellnhofer, P., Recasens, A., Stent, S., Matusik, W., & Torralba, A. (2019). Gaze360: Physically unconstrained gaze estimation in the wild. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6912–6921.

Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational psychiatry*, *5*(2), e514–e514.

Lai, M.-C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., Szatmari, P., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *The Lancet Psychiatry*, *6*(10), 819–829.

Lai, W. W., & Oei, T. P. S. (2014). Coping in parents and caregivers of children with autism spectrum disorders (asd): A review. *Review Journal of Autism and Developmental Disorders*, *1*(3), 207–224.

Landa, R. J. (2018). Efficacy of early interventions for infants and young children with, and at risk for, autism spectrum disorders. *International Review of Psychiatry*, *30*(1), 25–39.

Latrèche, K., Kojovic, N., Franchini, M., & Schaer, M. (2021). Attention to face as a predictor of developmental change and treatment outcome in young children with autism spectrum disorder. *Biomedicines*, *9*(8), 942.

Lavelli, M., & Fogel, A. (2005). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: The 2-month transition. *Developmental psychology*, *41*(1), 265.

Leblanc, E., Washington, P., Varma, M., Dunlap, K., Penev, Y., Kline, A., & Wall, D. P. (2020). Feature replacement methods enable reliable home video analysis for machine learning detection of autism. *Scientific reports*, *10*(1), 1–11.

Leclère, C., Avril, M., Viaux-Savelon, S., Bodeau, N., Achard, C., Missonnier, S., Keren, M., Feldman, R., Chetouani, M., & Cohen, D. (2016). Interaction and behaviour imaging: A novel method to measure mother–infant interaction using video 3d reconstruction. *Translational Psychiatry*, *6*(5), e816–e816.

Leigh, J. P., & Du, J. (2015). Brief report: Forecasting the economic burden of autism in 2015 and 2025 in the united states. *Journal of autism and developmental disorders*, *45*(12), 4135–4139.

Leo, M., Carcagnı, P., Distante, C., Spagnolo, P., Mazzeo, P. L., Rosato, A. C., Petrocchi, S., Pellegrino, C., Levante, A., De Lumè, F., et al. (2018). Computational assessment of facial expression production in asd children. *Sensors*, *18*(11), 3993.

Levy, S., Duda, M., Haber, N., & Wall, D. P. (2017). Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Molecular autism*, *8*(1), 1–17.

Li, B., Mehta, S., Aneja, D., Foster, C., Ventola, P., Shic, F., & Shapiro, L. (2019). A facial affect analysis system for autism spectrum disorder. *2019 IEEE International Conference on Image Processing (ICIP)*, 4549–4553.

Liaqat, S., Wu, C., Duggirala, P. R., Cheung, S.-c. S., Chuah, C.-N., Ozonoff, S., & Young, G. (2021). Predicting asd diagnosis in children with synthetic and image-based eye gaze data. *Signal Processing: Image Communication*, *94*, 116198.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *European conference on computer vision*, 740–755.

Liu, W., Li, M., & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, *9*(8), 888–898.

Liu, W., Yu, X., Raj, B., Yi, L., Zou, X., & Li, M. (2015). Efficient autism spectrum disorder prediction with eye movement: A machine learning framework. *2015 International conference on affective computing and intelligent interaction (ACII)*, 649–655.

Lombardo, M. V., Lai, M.-C., Auyeung, B., Holt, R. J., Allison, C., Smith, P., Chakrabarti, B., Ruigrok, A. N., Suckling, J., Bullmore, E. T., et al. (2016). Unsupervised data-driven stratification of mentalizing heterogeneity in autism. *Scientific Reports*, *6*(1), 1–15.

Lombardo, M. V., Lai, M.-C., & Baron-Cohen, S. (2019). Big data approaches to decomposing heterogeneity across the autism spectrum. *Molecular psychiatry*, *24*(10), 1435–1450.

Lombardo, M. V., Pramparo, T., Gazestani, V., Warrier, V., Bethlehem, R. A., Barnes, C. C., Lopez, L., Lewis, N. E., Eyler, L., Pierce, K., et al. (2018). Large-scale associations between the leukocyte transcriptome and bold responses to speech differ in autism early language outcome subtypes. *Nature neuroscience*, *21*(12), 1680–1688.

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(6), 466–474.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., Bishop, S., et al. (2012). Autism diagnostic observation schedule–2nd edition (ados-2). *Los Angeles, CA: Western Psychological Corporation*, *284*.

Lord, C., Bishop, S., & Anderson, D. (2015). Developmental trajectories as autism phenotypes. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *169*(2), 198–208.

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J., Jones, R. M., Pickles, A., State, M. W., et al. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, *6*(1), 1–23.

Luiz, D., Barnard, A., Knosen, N., Kotras, N., Horrocks, S., McAlinden, P., & O'Connell, R. (2006). Gmds-er 2-8. griffith mental devel-opmental scales-extended revised: 2 to 8 years. the test agency.

Lyall, K., Croen, L., Daniels, J., Fallin, M. D., Ladd-Acosta, C., Lee, B. K., Park, B. Y., Snyder, N. W., Schendel, D., Volk, H., et al. (2017). The changing epidemiology of autism spectrum disorders. *Annual review of public health*, *38*, 81–102.

Madipakkam, A. R., Rothkirch, M., Dziobek, I., & Sterzer, P. (2017). Unconscious avoidance of eye contact in autism spectrum disorder. *Scientific reports*, *7*(1), 1–6.

Maenner, M. J., Shaw, K. A., Baio, J., et al. (2020). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2016. *MMWR Surveillance Summaries*, *69*(4), 1.

Maestro, S., Muratori, F., Cavallaro, M. C., Pei, F., Stern, D., Golse, B., & Palacio-Espasa, F. (2002). Attentional skills during the first 6 months of age in autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(10), 1239–1245.

Magiati, I., Moss, J., Charman, T., & Howlin, P. (2011). Patterns of change in children with autism spectrum disorders who received community based comprehensive interventions in their pre-school years: A seven year follow-up study. *Research in Autism Spectrum Disorders*, *5*(3), 1016–1027.

Manfredonia, J., Bangerter, A., Manyakov, N. V., Ness, S., Lewin, D., Skalkin, A., Boice, M., Goodwin, M. S., Dawson, G., Hendren, R., et al. (2019). Automatic recognition of posed facial expression of emotion in individuals with autism spectrum disorder. *Journal of autism and developmental disorders*, *49*(1), 279–293.

Martin, K. B., Hammal, Z., Ren, G., Cohn, J. F., Cassell, J., Ogihara, M., Britton, J. C., Gutierrez, A., & Messinger, D. S. (2018). Objective measurement of head movement differences in children with and without autism spectrum disorder. *Molecular autism*, *9*(1), 1–10.

Masmoudi, M., Jarraya, S. K., & Hammami, M. (2019). Meltdowncrisis: Dataset of autistic children during meltdown crisis. *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 239–246.

Mattson, W. I., Cohn, J. F., Mahoor, M. H., Gangi, D. N., & Messinger, D. S. (2013). Darwin's duchenne: Eye constriction during infant joy and distress. *PloS one*, *8*(11), e80161.

Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, *4*(2), 151–160.

McCarty, P., & Frye, R. E. (2020). Early detection and diagnosis of autism spectrum disorder: Why is it so difficult? *Seminars in Pediatric Neurology*, *35*, 100831.

McDuff, D., & Girard, J. M. (2019). Democratizing psychological insights from analysis of nonverbal behavior. *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 220–226.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

McKeown, G., Valstar, M. F., Cowie, R., & Pantic, M. (2010). The semaine corpus of emotionally coloured character interactions. *2010 IEEE International Conference on Multimedia and Expo*, 1079–1084.

McNaughton, K. A., & Redcay, E. (2020). Interpersonal synchrony in autism. *Current psychiatry reports*, *22*(3), 1–11.

Merghani, W., Davison, A., & Yap, M. (2018). Facial micro-expressions grand challenge 2018: Evaluating spatio-temporal features for classification of objective classes. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 662–666.

Merghani, W., Davison, A. K., & Yap, M. H. (2018). A review on facial micro-expressions analysis: Datasets, features and metrics. *arXiv preprint arXiv:1805.02397*.

Messinger, D. S., & Fogel, A. (2007). The interactive development of social smiling.

Messinger, D. S., Fogel, A., & Dickson, K. L. (2001). All smiles are positive, but some smiles are more positive than others. *Developmental psychology*, *37*(5), 642.

Messinger, D. S., Mattson, W. I., Mahoor, M. H., & Cohn, J. F. (2012). The eyes have it: Making positive expressions more positive and negative expressions more negative. *Emotion*, *12*(3), 430.

Miller, M., Iosif, A.-M., Hill, M., Young, G. S., Schwichtenberg, A., & Ozonoff, S. (2017). Response to name in infants developing autism spectrum disorder: A prospective study. *The Journal of pediatrics*, *183*, 141–146.

Moore, A., Wozniak, M., Yousef, A., Barnes, C. C., Cha, D., Courchesne, E., & Pierce, K. (2018). The geometric preference subtype in asd: Identifying a consistent, early-emerging phenomenon through eye tracking. *Molecular autism*, *9*(1), 1–13.

Mottron, L., & Bzdok, D. (2020). Autism spectrum heterogeneity: Fact or artifact? *Molecular Psychiatry*, *25*(12), 3178–3185.

Mozolic-Staunton, B., Donelly, M., Yoxall, J., & Barbaro, J. (2020). Early detection for better outcomes: Universal developmental surveillance for autism across health and early childhood education settings. *Research in Autism Spectrum Disorders*, *71*, 101496.

Müller, P., Huang, M. X., Zhang, X., & Bulling, A. (2018). Robust eye contact detection in natural multi-person interactions using gaze and speaking behaviour. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 1–10.

Nadig, A. S., Ozonoff, S., Young, G. S., Rozga, A., Sigman, M., & Rogers, S. J. (2007). A prospective study of response to name in infants at risk for autism. *Archives of pediatrics & adolescent medicine*, *161*(4), 378–383.

Narzisi, A., Posada, M., Barbieri, F., Chericoni, N., Ciuffolini, D., Pinzino, M., Romano, R., Scattoni, M., Tancredi, R., Calderoni, S., et al. (2020). Prevalence of autism spectrum disorder in a large italian catchment area: A school-based population study within the asdeu project. *Epidemiology and psychiatric sciences*, *29*.

Nazneen, N., Matthews, N., Smith, C. J., Rozga, A., Abowd, G. D., Oberleitner, R., Reischl, U., & Arriaga, R. I. (2015). Use of a novel imaging technology for remote autism diagnosis: A reflection on experience of stakeholders. *Procedia Manufacturing*, *3*, 293–300.

Nebel, M. B., Eloyan, A., Nettles, C. A., Sweeney, K. L., Ament, K., Ward, R. E., Choe, A. S., Barber, A. D., Pekar, J. J., & Mostofsky, S. H. (2016). Intrinsic visual-motor synchrony correlates with social deficits in autism. *Biological psychiatry*, *79*(8), 633–641.

Negin, F., Ozyer, B., Agahian, S., Kacdioglu, S., & Ozyer, G. T. (2021). Vision-assisted recognition of stereotype behaviors for early diagnosis of autism spectrum disorders. *Neurocomputing*, *446*, 145–155.

Nguyen, T.-A., Coursaris, C. K., Léger, P.-M., Sénécal, S., & Fredette, M. (2020). Effectiveness of banner ads: An eye tracking and facial expression analysis. *International Conference on Human-Computer Interaction*, 445–455.

Nichols, C. M., Ibañez, L. V., Foss-Feig, J. H., & Stone, W. L. (2014). Social smiling and its components in high-risk infant siblings without later asd symptomatology. *Journal of autism and developmental disorders*, *44*(4), 894–902.

Nicolle, J., Bailly, K., & Chetouani, M. (2016). Real-time facial action unit intensity prediction with regularized metric learning. *Image and Vision Computing*, *52*, 1–14.

Ninci, J., Lang, R., Davenport, K., Lee, A., Garner, J., Moore, M., Boutot, A., Rispoli, M., & Lancioni, G. (2013). An analysis of the generalization and maintenance of eye contact taught during play. *Developmental neurorehabilitation*, *16*(5), 301–307.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal behavior*, *18*(1), 9–35.

Nuske, H. J., Vivanti, G., & Dissanayake, C. (2013). Are emotion impairments unique to, universal, or specific in autism spectrum disorder? a comprehensive review. *Cognition & emotion*, *27*(6), 1042–1061.

Ouss, L., Le Normand, M.-T., Bailly, K., Leitgel Gille, M., Gosme, C., Simas, R., Wenke, J., Jeudon, X., Thepot, S., Da Silva, T., et al. (2018). Developmental trajectories of hand movements in typical infants and those at risk of developmental disorders: An observational study of kinematics during the first year of life. *Frontiers in psychology*, *9*, 83.

Owada, K., Kojima, M., Yassin, W., Kuroda, M., Kawakubo, Y., Kuwabara, H., Kano, Y., & Yamasue, H. (2018). Computer-analyzed facial expression as a surrogate marker for autism spectrum social core symptoms. *PloS one*, *13*(1), e0190442.

Ozonoff, S., Gangi, D., Hanzel, E. P., Hill, A., Hill, M. M., Miller, M., Schwichtenberg, A., Steinfeld, M. B., Parikh, C., & Iosif, A.-M. (2018). Onset patterns in autism: Variation across informants, methods, and timing. *Autism Research*, *11*(5), 788–797.

Ozonoff, S., Heung, K., Byrd, R., Hansen, R., & Hertz-Picciotto, I. (2008). The onset of autism: Patterns of symptom emergence in the first years of life. *Autism research*, *1*(6), 320–328.

Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., Sigman, M., et al. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(3), 256–266.

Pan, Y., Hirokawa, M., & Suzuki, K. (2015). Measuring k-degree facial interaction between robot and children with autism spectrum disorders. *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 48–53.

Parekh, V., Subramanian, R., & Jawahar, C. (2017). Eye contact detection via deep neural networks. *International Conference on Human-Computer Interaction*, 366–374.

Pearson, N., Charman, T., Happé, F., Bolton, P. F., & McEwen, F. S. (2018). Regression in autism spectrum disorder: Reconciling findings from retrospective and prospective research. *Autism Research*, *11*(12), 1602–1620.

Pickles, A., Anderson, D. K., & Lord, C. (2014). Heterogeneity and plasticity in the development of language: A 17-year follow-up of children referred early for possible autism. *Journal of Child Psychology and Psychiatry*, *55*(12), 1354–1362.

Purpura, G., Costanzo, V., Chericoni, N., Puopolo, M., Scattoni, M. L., Muratori, F., & Apicella, F. (2017). Bilateral patterns of repetitive movements in 6-to 12-month-old infants with autism spectrum disorders. *Frontiers in psychology*, *8*, 1168.

Rad, N. M., Bizzego, A., Kia, S. M., Jurman, G., Venuti, P., & Furlanello, C. (2015). Convolutional neural network for stereotypical motor movement detection in autism. *arXiv preprint arXiv:1511.01865*.

Robins, D. L., Casagrande, K., Barton, M., Chen, C.-M. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (m-chat-r/f). *Pediatrics*, *133*(1), 37–45.

Rodgers, M., Simmonds, M., Marshall, D., Hodgson, R., Stewart, L. A., Rai, D., Wright, K., Ben-Itzchak, E., Eikeseth, S., Eldevik, S., et al. (2021). Intensive behavioural interventions based on applied behaviour analysis for young children with autism: An international collaborative individual participant data meta-analysis. *Autism*, *25*(4), 1137–1153.

Rogers, S. J., & Talbott, M. (2016). Early identification and early treatment of autism spectrum disorder. *International review of research in developmental disabilities* (pp. 233–275). Elsevier.

Rogge, N., & Janssen, J. (2019). The economic costs of autism spectrum disorder: A literature review. *Journal of Autism and Developmental Disorders*, *49*(7), 2873–2900.

Rozga, A., Hutman, T., Young, G. S., Rogers, S. J., Ozonoff, S., Dapretto, M., & Sigman, M. (2011). Behavioral profiles of affected and unaffected siblings of children with autism: Contribution of measures of mother–infant interaction and nonverbal communication. *Journal of autism and developmental disorders*, *41*(3), 287–301.

Rudovic, O., Lee, J., Dai, M., Schuller, B., & Picard, R. W. (2018). Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, *3*(19).

Rudovic, O., Lee, J., Mascarell-Maricic, L., Schuller, B. W., & Picard, R. W. (2017). Measuring engagement in robot-assisted autism therapy: A cross-cultural study. *Frontiers in Robotics and AI*, *4*, 36.

Rudovic, O., Park, H. W., Busche, J., Schuller, B., Breazeal, C., & Picard, R. W. (2019). Personalized estimation of engagement from videos using active learning with deep reinforcement learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 217–226.

Samad, M. D., Diawara, N., Bobzien, J. L., Taylor, C. M., Harrington, J. W., & Iftekharuddin, K. M. (2019). A pilot study to identify autism related traits in spontaneous facial actions using computer vision. *Research in Autism Spectrum Disorders*, *65*, 14–24.

Sapiro, G., Hashemi, J., & Dawson, G. (2018). Computer vision applications to computational behavioral phenotyping: An autism spectrum disorder case study. *Curr. Opin. Biomed. Eng.*

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). Dbscan revisited, revisited: Why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, *42*(3), 1–21.

Shic, F., Macari, S., & Chawarska, K. (2014). Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. *Biological psychiatry*, *75*(3), 231–237.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Smith, B. A., Yin, Q., Feiner, S. K., & Nayar, S. K. (2013). Gaze locking: Passive eye contact detection for human-object interaction. *Proceedings of the 26th annual ACM symposium on User interface software and technology*, 271–280.

Stagg, S. D., Slavny, R., Hand, C., Cardoso, A., & Smith, P. (2014). Does facial expressivity count? how typically developing children respond initially to children with autism. *Autism*, *18*(6), 704–711.

Stevens, E., Dixon, D. R., Novack, M. N., Granpeesheh, D., Smith, T., & Linstead, E. (2019). Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervised machine learning. *International journal of medical informatics*, *129*, 29–36.

Stevens, M. C., Fein, D. A., Dunn, M., Allen, D., Waterhouse, L. H., Feinstein, C., & Rapin, I. (2000). Subgroups of children with autism by cluster analysis: A longitudinal examination. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(3), 346–352.

Sumner, E., Leonard, H. C., & Hill, E. L. (2016). Overlapping phenotypes in autism spectrum disorder and developmental coordination disorder: A cross-syndrome comparison of motor and social skills. *Journal of Autism and Developmental Disorders*, *46*(8), 2609–2620.

Swallows, G., & Graupner, T. (2005). Intensive behavioral treatment for children with autism: A research synthesis. *Journal of autism and developmental disorders*, *32*(5), 423–446.

Tang, J. S., Chen, N. T., Falkmer, M., Blte, S., & Girdler, S. (2019). Atypical visual processing but comparable levels of emotion recognition in adults with autism during the processing of social scenes. *Journal of autism and developmental disorders*, *49*(10), 4009–4018.

Tariq, Q., Daniels, J., Schwartz, J. N., Washington, P., Kalantarian, H., & Wall, D. P. (2018). Mobile detection of autism through machine learning on home video: A development and prospective validation study. *PLoS medicine*, *15*(11), e1002705.

Thorup, E., Nyström, P., Gredebäck, G., Bölte, S., & Falck-Ytter, T. (2018). Reduced alternating gaze during social interaction in infancy is associated with elevated symptoms of autism in toddlerhood. *Journal of Abnormal Child Psychology*, *46*(7), 1547–1561.

Tiede, G., & Walton, K. M. (2019). Meta-analysis of naturalistic developmental behavioral interventions for young children with autism spectrum disorder. *Autism*, *23*(8), 2080–2095.

Trevisan, D. A., Hoskyn, M., & Birmingham, E. (2018). Facial expression production in autism: A meta-analysis. *Autism Research*, *11*(12), 1586–1601.

Tsai, T.-W., & Lin, M.-Y. (2011). An application of interactive game for facial expression of the autisms. *International Conference on Technologies for E-Learning and Digital Entertainment*, 204–211.

Uddin, M., Wang, Y., & Woodbury-Smith, M. (2019). Artificial intelligence for precision medicine in neurodevelopmental disorders. *NPJ digital medicine*, *2*(1), 1–10.

Valstar, M. F., Almaev, T., Girard, J. M., McKeown, G., Mehu, M., Yin, L., Pantic, M., & Cohn, J. F. (2015). Fera 2015-second facial expression recognition and analysis challenge. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, *6*, 1–8.

van Rentergem, J. A. A., Deserno, M. K., & Geurts, H. M. (2021). Validation strategies for subtypes in psychiatry: A systematic review of research on autism spectrum disorder. *Clinical Psychology Review*, 102033.

Varcin, K. J., & Jeste, S. S. (2017). The emergence of autism spectrum disorder (asd): Insights gained from studies of brain and behaviour in high-risk infants. *Current opinion in psychiatry*, *30*(2), 85.

Venker, C. E., Eernisse, E. R., Saffran, J. R., & Weismer, S. E. (2013). Individual differences in the real-time comprehension of children with asd. *Autism Research*, *6*(5), 417–432.

Venuti, P., & Bentenuto, A. (2017). Studi di caso. disturbi dello spettro autistico. *Quaderni di psicologia*.

Venuti, P. (2012). *Intervento e riabilitazione nei disturbi dello spettro autistico*. Carocci.

Vivanti, G., Dissanayake, C., Duncan, E., Feary, J., Capes, K., Upson, S., Bent, C. A., Rogers, S. J., & Hudry, K. (2019). Outcomes of children receiving group-early start denver model in an inclusive versus autism-specific setting: A pilot randomized controlled trial. *Autism*, *23*(5), 1165–1175.

Voss, C., Schwartz, J., Daniels, J., Kline, A., Haber, N., Washington, P., Tariq, Q., Robinson, T. N., Desai, M., Phillips, J. M., et al. (2019). Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: A randomized clinical trial. *JAMA pediatrics*, *173*(5), 446–454.

Vyas, K., Ma, R., Rezaei, B., Liu, S., Neubauer, M., Ploetz, T., Oberleitner, R., & Ostadabbas, S. (2019). Recognition of atypical behavior in autism diagnosis from video using pose estimation over time. *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6.

Waddington, H., van der Meer, L., & Sigafoos, J. (2016). Effectiveness of the early start denver model: A systematic review. *Review Journal of Autism and Developmental Disorders*, *3*(2), 93–106.

Wall, D. P., Dally, R., Luyster, R., Jung, J.-Y., & DeLuca, T. F. (2012). Use of artificial intelligence to shorten the behavioral diagnosis of autism.

Wall, D. P., Kosmicki, J., Deluca, T., Harstad, E., & Fusaro, V. A. (2012). Use of machine learning to shorten observation-based screening and diagnosis of autism. *Translational psychiatry*, *2*(4), e100–e100.

Washington, P., Leblanc, E., Dunlap, K., Kline, A., Mutlu, C., Chrisman, B., Stockham, N., Paskov, K., & Wall, D. P. (2021). Crowd annotations can approximate clinical autism impressions from short home videos with privacy protections. *medRxiv*.

Washington, P., Leblanc, E., Dunlap, K., Penev, Y., Kline, A., Paskov, K., Sun, M. W., Chrisman, B., Stockham, N., Varma, M., et al. (2020). Precision telemedicine through crowdsourced machine learning: Testing variability of crowd workers for video-based autism feature recognition. *Journal of personalized medicine*, *10*(3), 86.

Washington, P., Park, N., Srivastava, P., Voss, C., Kline, A., Varma, M., Tariq, Q., Kalantarian, H., Schwartz, J., Patnaik, R., et al. (2020). Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(8), 759–769.

Wetherby, A. M., Woods, J., Guthrie, W., Delehanty, A., Brown, J. A., Morgan, L., Holland, R. D., Schatschneider, C., & Lord, C. (2018). Changing developmental trajectories of toddlers with autism spectrum disorder: Strategies for bridging research to community practice. *Journal of Speech, Language, and Hearing Research*, *61*(11), 2615–2628.

Wilson, K. P., Carter, M. W., Wiener, H. L., DeRamus, M. L., Bulluck, J. C., Watson, L. R., Crais, E. R., & Baranek, G. T. (2017). Object play in infants with autism spectrum disorder: A longitudinal retrospective video analysis. *Autism & developmental language impairments*, *2*, 2396941517713186.

Wolfers, T., Floris, D. L., Dinga, R., van Rooij, D., Isakoglou, C., Kia, S. M., Zabihi, M., Llera, A., Chowdanayaka, R., Kumar, V. J., et al. (2019). From pattern classification to stratification: Towards conceptualizing the heterogeneity of autism spectrum disorder. *Neuroscience & Biobehavioral Reviews*, *104*, 240–254.

Xu, K., Ji, B., Wang, Z., Liu, J., & Liu, H. (2020). An auxiliary screening system for autism spectrum disorder based on emotion and attention analysis. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2299–2304.

Ye, Z., Li, Y., Fathi, A., Han, Y., Rozga, A., Abowd, G. D., & Rehg, J. M. (2012). Detecting eye contact using wearable eye-tracking glasses. *Proceedings of the 2012 ACM conference on ubiquitous computing*, 699–704.

Yirmiya, N., Kasari, C., Sigman, M., & Mundy, P. (1989). Facial expressions of affect in autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry*, *30*(5), 725–735.

Yun, S.-S., Choi, J., Park, S.-K., Bong, G.-Y., & Yoo, H. (2017). Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system. *Autism Research*, *10*(7), 1306–1323.

Zadeh, A., Chong Lim, Y., Baltrušaitis, T., & Morency, L.-P. (2017). Convolutional experts constrained local model for 3d facial landmark detection. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2519–2528.

Zampella, C. J., Csumitta, K. D., Simon, E., & Bennetto, L. (2020). Interactional synchrony and its association with social and communication ability in children with and without autism spectrum disorder. *Journal of Autism & Developmental Disorders*, *50*(9).

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, *23*(10), 1499–1503.

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., Liu, P., & Girard, J. M. (2014). Bp4d-spontaneous: A high-resolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing*, *32*(10), 692–706.

Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2015). Appearance-based gaze estimation in the wild. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4511–4520.

Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2017a). It's written all over your face: Full-face appearance-based gaze estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 51–60.

Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2017b). Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE transactions on pattern analysis and machine intelligence*, *41*(1), 162–175.

Zhao, K., Chu, W.-S., & Zhang, H. (2016). Deep region and multi-label learning for facial action unit detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3391–3399.

Zhu, W., & Deng, H. (2017). Monocular free-head 3d gaze tracking with deep learning and geometry constraints. *Proceedings of the IEEE International Conference on Computer Vision*, 3143–3152.

Zunino, A., Cavazza, J., Volpi, R., Morerio, P., Cavallo, A., Becchio, C., & Murino, V. (2020). Predicting intentions from motion: The subject-adversarial adaptation approach. *International Journal of Computer Vision*, *128*(1), 220–239.

Zwaigenbaum, L., Bauman, M. L., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., McPartland, J. C., Natowicz, M. R., Choueiri, R., Fein, D., et al. (2015). Early identification of autism spectrum disorder: Recommendations for practice and research. *Pediatrics*, *136*(Supplement 1), S10–S40.

Zwaigenbaum, L., Brian, J. A., & Ip, A. (2019). Early detection for autism spectrum disorder in young children. *Paediatrics & child health*, *24*(7), 424–432.

Zwaigenbaum, L., Bryson, S., & Garon, N. (2013). Early identification of autism spectrum disorders. *Behavioural brain research*, *251*, 133–146.

Zwaigenbaum, L., & Penner, M. (2018). Autism spectrum disorder: Advances in diagnosis and evaluation. *Bmj*, *361.*