AUS DER

KLINIK UND POLIKLINIK FÜR PSYCHIATRIE UND PSYCHOTHERAPIE

KLINIKUM DER LUDWIG-MAXIMILIANS- UNIVERSITÄT MÜNCHEN

Objectifying social interaction in autism: Digitally assisting the diagnostic process using behavioral parameters and machine learning

DISSERTATION

ZUM ERWERB DES DOCTOR OF PHILOSOPHY (PH.D.)

AN DER MEDIZINISCHEN FAKULTÄT DER

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

VORGELEGT VON

JANA CHRISTINA KÖHLER

AUS

HILDEN, DEUTSCHLAND

2024

Mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

1. Gutachten: Prof. Dr. Christine Falter-Wagner
2. Gutachten: Prof. Dr. Nikolaos Koutsouleris
3. Gutachten: Prof. Dr. Sophia Stöcklein
4. Gutachten: Prof. Dr. Dr. Kai Vogeley

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 04.11.2024

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Dean's Office Medical Faculty
Faculty of Medicine

Affidavit

Köhler, Jana

Surname, first name

Nußbaumstraße 7, 80336 München

Address

I hereby declare, that the submitted thesis entitled

Objectifying social interaction in autism: Digitally assisting the diagnostic process using behavioral parameters and machine learning

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Place, Date

Signature doctoral candidate

Confirmation of congruency between printed and electronic version of the doctoral thesis

Doctoral candidate: Jana Köhler

Address: Nußbaumstraße 7, 80336 München

I hereby declare that the electronic version of the submitted thesis, entitled

Objectifying social interaction in autism: Digitally assisting the diagnostic process using behavioral parameters and machine learning

is congruent with the printed version both in content and format.

Place, Date

Signature doctoral candidate

Abstract

The rising prevalence of autism spectrum disorder (ASD), a neurodevelopmental disorder characterized by severe impairments in social interaction and communication persisting across the lifespan, constitutes a pressing issue in comprehensive healthcare provision. The heterogeneous phenotype hampers the identification of solid diagnostic markers across the entire spectrum, resulting in subjective and cumbersome diagnostic tools in praxis. Diagnostic assessment so far is conducted entirely at a behavioral level, requiring a skilled diagnostician and a lengthy diagnostic process, resulting in long waiting periods. Hence, updating our understanding of the complexity of autistic social interaction difficulties is crucial for the optimization of the diagnostic process.

This dissertation project aimed at applying the multivariate methodology of machine learning to social interaction characteristics of autism which were digitally extracted from video and audio in order to build classification models that could potentially assist the diagnostic process.

A first perspective on how to embed ML in social interaction research was proposed in a proof-of-concept study (Perspective Article, Appendix). In the main part of this dissertation, diagnostic classification of ASD is explored using support vector machine (SVM) models trained on objectively and automatedly extracted features from dyadic conversations between autistic and non-autistic study participants, tapping into the core symptomatology in the domains of social interaction (Original Study I) and communication (Original Study II) and finding promising classification accuracy. Further, a proof-of-concept application using social interaction data from a real-life clinical sample, including a control group of patients with differential and co-morbid disorders, is provided (Original Study III).

The research conducted within this dissertation project provides an important step towards objectifying social interaction difficulties in autism and can serve as a framework to potentially augment diagnostic decision-making in the future. Ultimately, finding the best combination of diagnostic markers will be crucial for improving patient care in the long run.

Contents

Abbreviations

ADHD	Attention Deficit Hyperactivity Disorder
ADOS-2	Autism Diagnostic Observation Schedule 2nd Edition
ASD	Autism Spectrum Disorder
BAC	Balanced Accuracy
DSM-5	Diagnostic and Statistical Manual of Mental Disorder
EEG	Electroencephalography
MEA	Motion Energy Analysis
ML	Machine Learning
MRI	Magnetic Resonance Imaging
SVM	Support Vector Machine
TD	Typically Developing

Full list of publications

Bierlich, A.M., Scheel, N.T., Keeser, D., Tepest, R., Georgescu, A.L., **Koehler, J.C.**, Plank, I.S.*, & Falter-Wagner, C.M.* Interpersonal synchrony modulates neural processing in the Action Observation Network and is spared in autism. *Under review at Human Brain Mapping.*

Koehler, J.C.*, Dong, M.S.*, Song, D.-Y., Bong, G., Koutsouleris, N., Yoo, H., & Falter-Wagner, C.M. (2024). Classifying autism in a clinical population based on motion synchrony: a proof-of-concept study using real-life diagnostic interviews. *Sci Rep,* 14, 5663. https://doi.org/10.1038/s41598-024-56098-y (*shared first-authorship)

Koehler, J.C., Dong, M.S., Bierlich, A.M., Fischer, S., Späth, J., Plank, I.S., Koutsouleris, N., & Falter-Wagner, C.M. (2024). Machine learning classification of autism spectrum disorder based on reciprocity in naturalistic social interactions. *Translational Psychiatry*, 14(1), 76. https://doi.org/10.1038/s41398-024-02802-5

Plank, I.S., **Koehler, J.C.,** Nelson, A.M., Koutsouleris, N., & Falter-Wagner, C.M. (2023). Automated extraction of speech and turn-taking parameters in autism allows for diagnostic classification using a multivariable prediction model. *Frontiers in Psychiatry*, 14. https://doi.org/10.3389%2Ffpsyt.2023.1257569.

Plank, I.S., Traiger, L.S., Nelson, A.M., **Koehler, J.C.**, Lang, S.F., Tepest, R., Vogeley, K., Georgescu, A.L., Falter-Wagner, C.M. (2023). The role of interpersonal synchrony in forming impressions of autistic and non-autistic adults. *Scientific Reports,* 13, 15306. https://doi.org/10.1038/s41598-023-42006-3.

Koehler, J.C., & Falter-Wagner, C.M. (2023). Digitally assisted diagnostics of autism spectrum disorder. *Frontiers in Psychiatry*, 14:1066284, doi: 10.3389/fpsyt.2023.1066284

Koehler, J.C.*, Georgescu, A.L.*, Weiske, J., Spangemacher, M., Burghof, L., Falkai, P., Koutsouleris, N., Tschacher, W., Vogeley, K., & Falter-Wagner, C.M. (2022). Brief report: Specificity of interpersonal synchrony deficits to autism spectrum disorder and its potential for digitally assisted diagnostics. *Journal of autism and developmental disorders*, 52(8), 3718-3726. https://doi.org/10.1007/s10803-021-05194-3 (*shared first-authorship)

Falter-Wagner, C.M. & **Koehler, J.C.** (2021). Digital unterstützte Diagnostik bei Autismus-Spektrum-Störung. Pädiatrische Praxis.

Georgescu, A.L.*, **Koehler, J.C.***, Weiske, J., Vogeley, K., Koutsouleris, N., & Falter-Wagner, C.M. (2019). Machine Learning to Study Social Interaction Difficulties in ASD. Frontiers in Robotics and AI, 6:132. https://doi.org/10.3389/frobt.2019.00132 (*shared first-authorship)

1 Introductory summary

1.1 Challenges in Autism Diagnostics

Autism spectrum disorder (ASD) is a neurodevelopmental disorder encompassing a spectrum of persistent impairments in social interaction and communication as well as the presence of stereotypical behaviors [1]. With an estimated prevalence of 1% [1], ASD is typically diagnosed in early childhood, though an increasing number of diagnoses are made in adulthood [2,3]. The highly heterogeneous phenotype with its diverse manifestations of autistic symptoms challenges diagnostic decision-making and research on behavioral or biological markers with sufficient universality across the entire spectrum. Though a growing list of around 100 candidate genes strongly linked to ASD has been proposed [4], no common variant has been identified yet. Several emerging diagnostic markers on the neurophysiological (e.g., altered gamma-band activity) and biological (e.g., extra-axial fluid volume) level have limited reproducibility [5].

Therefore, the diagnostic process for ASD poses a significant bottleneck: Currently, diagnostics are conducted entirely at the behavioral level. Best practice procedures include a series of diagnostic interviews and behavioral observations conducted by a multidisciplinary team, i.a., expert ratings of a patient's social communication impairments and stereotypical and repetitive behaviors, testing procedures concerning neurocognitive functioning, and an assessment of the developmental history with a caretaker [6,7]. The involved process results in months of waiting time from referral until final diagnosis (median length around 7 months in a Canadian sample according to [8]), leading to a backlog in clinical institutions. This is particularly crucial, not least because early detection and intervention may facilitate improved developmental trajectories [9]. The lack of objective diagnostic markers on the behavioral level is also problematic in other ways: Due to the varied manifestations of social interaction impairments, detecting and correctly classifying autistic symptoms and behaviors is particularly challenging for a clinician. Hence, interrater reliability is heavily influenced by the expertise of the diagnostician [10,11]. Further, although standard screening measures for ASD show sufficient sensitivity in the general population, they nevertheless lack specificity in clinical settings [12,13] and challenge differential diagnostic decision-making of conditions with overlapping phenotypes such as attention deficit hyperactivity disorder (ADHD) [14], schizophrenia [15], intellectual disability [16], or other disorders that manifest in social interaction difficulties (e.g., Borderline personality disorder [17]). Differential diagnostic decision-making is further complicated by the high rate of co-occurring mental health conditions present in autism [18,19]. The reduced specificity, in turn, may increase false diagnostic decisions and hinder accurate treatment.

Taken together and considering the inconsistent findings on diagnostic markers across the entire autism spectrum, updating our understanding of social interaction impairments in the context of ASD diagnostics is a pressing issue. Novel and complex multivariate methods that can account for the interdependencies in the highly heterogeneous autistic phenotype need to be applied.

1.2 Diagnostic Classification of Autism using ML: Advances and Limitations

Methods of artificial intelligence like machine learning (ML) are increasingly studied to aid diagnostic decision-making for psychiatric and neurodevelopmental conditions in a data-driven way [20,21]. ML allows a computer to learn patterns across large amounts of multivariate data using a variety of computational algorithms and, in turn, apply those to unseen data, thereby providing an estimate of predictiveness of the data at hand. This method offers significant benefits for diagnostic decision-making, as it enables predictions on the single-subject level instead of the mere comparison of groups, as well as detecting predictive patterns and unknown interdependencies in high-dimensional data sets [22]. A common approach is the application of supervised learning techniques, whereby an algorithm is trained on labeled data sets, hence, learning the relationships between input variables and diagnostic label, allowing in turn for the prediction of diagnostic status in unseen data. Considering these benefits in light of the complexity of the autistic phenotype, advances have been made in multiple fields using ML for establishing the classification value of potential diagnostic markers.

One line of research focuses on extracting features from the Autism Diagnostic Observation Schedule (ADOS-2) [23] to train machine learning algorithms. The ADOS-2 is considered the current gold-standard instrument in childhood diagnostics of autism, though in Germany explicitly not recommended for adulthood [24]. Usually, it is conducted by a licensed clinical expert, who qualitatively assesses a patient regarding their reciprocal communication and social interaction skills through a mixture of semi-structured interview and observation. Because of its widespread application, the data sets available are usually rather large, allowing for comprehensive data analyses. Several studies have used ML techniques aiming to reduce the number of ADOS-2 interview items and to identify the most predictive core domains [25–29]. However, as the final diagnostic decision is usually heavily reliant on the outcome of the ADOS-2, the features used for prediction and the diagnostic label within these models are often not independent of each other [30], resulting in a circularity problem. Moreover, its administration in more heterogeneous populations including individuals with other disorders affecting social interactions, as well as differences in the clinical experience of the diagnosticians, ultimately lead to decreased interrater reliability in diagnostic classification [31]. Conclusively, while data-driven ML analyses of ADOS-2 may contribute to the optimization of existing diagnostic tools, their inherent weaknesses remain.

ML methods have been used in the broader field of personalized psychiatry based on neuroimaging data, aiming to build predictive models for early recognition, diagnosis, and prognosis of a wide range of psychiatric disorders. In the domain of autism research, neuroimaging data accounts for the majority of diagnostic classification studies [22], based on identified abnormalities regarding the volume of grey and white matter, as well as atypical functional and structural connectivity [32]. Considering that a key strength of ML is the ability to test for a model's generalizability to unseen data, large data quantities are usually required, reducing the scalability of neuroimaging methods such as magnetic resonance imaging (MRI). Most importantly, we are faced with a unique group of individuals known for their sensory sensitivities in the field of autism, further hampering the feasibility of using sensory intensive techniques on a large-scale basis.

1.3 Objective Assessment of Autistic Social Interaction and Communication

Peculiarities of social communication in autism include, for instance, poorly integrated verbal and nonverbal communication [33], impaired use of gestures [34], reduced facial expressiveness [35], reduced conversational turn-taking [36], and aberrated [37] or complete lack of speech [1]. As marked differences in social interaction and communication style constitute the core features of ASD, these symptoms naturally lend themselves to be further explored for diagnostic prediction. Importantly, as social interactions account for the majority of daily life, this kind of data is easily obtainable. Due to the mentioned drawbacks of observational coding, automated and objective assessment methods of autistic social interaction and communication provide an exciting avenue for further investigation within a diagnostic context.

Study findings have supported clinical anecdotal evidence suggesting a certain “oddness” about the first impression of autistic individuals: They have been judged more socially awkward [38], or have been rated less favorably regarding their character traits [39] than their non-autistic peers based on brief video snippets and still images. Various motion tracking techniques have attempted to capture this awkwardness of nonverbal behaviors: Wearable motion tracking sensors can be applied with little effort, resulting in high resolution recordings of kinematic features. Studies have found promising classification results for ML models for both infants (e.g., based on ball playing motion [40]) and adults (e.g., based on hand motion in an imitation task [41]). However, those highly controlled study set ups require a distinct laboratory infrastructure, resulting in limited sample sizes and therefore warranting cautious interpretation. A more scalable alternative to quantify motion aberrances can be achieved using advanced computer vision techniques, where two- or even three-dimensional motion is extracted from video excerpts using frame-differencing algorithms [42] or artificial intelligence [43], hence, requiring no complex setups. For instance, head motion features have been extracted using computer vision to classify ASD in children [e.g., 44,45]. Importantly, such techniques allow for motion analyses in more naturalistic settings. This is crucial because autism is regarded as a disorder of social interaction, i.e., it is based on the manifestation of difficulties when communicating with other individuals. In fact, reduced interpersonal synchrony with an interactional counterpart has been identified as a driver of negative first impressions of autistic individuals [46]. Interpersonal synchrony, meaning the adaptation of verbal and/or nonverbal behaviors during a social interaction, has frequently been investigated and found to be reduced in autism [47,48]. Further, the coordination of individual social behaviors across different modalities, *intra*personal synchrony, is suggested to be impaired in autism [33,49], providing a potential constituent of reduced *inter*personal synchrony [50]. In a proof-of-concept study (Perspective Article, Appendix A), we computed intrapersonal head and body synchrony of autistic, non-autistic and mixed (i.e., including an autistic and a non-autistic individual) conversational dyads based on objective motion extraction using Motion Energy Analysis (MEA) [42], a frame-differencing algorithm. The resulting synchrony time series were subsequently used as features in a SVM classification model to predict diagnosis with an accuracy of 75.9% [51]. Hence, combining objectively extracted behavioral data stemming from actual social interactions could aid in capturing subliminal signals contributing to autistic social interaction difficulties and, therefore, provide valuable markers for diagnosis.

Another pillar constituting the autistic phenotype is atypical communication. While an estimated 25% of individuals on the spectrum are non- or minimally verbal [52,53], the majority have sufficient expressive language skills. The ADOS-2 [23] requires the rating of communication on multiple domains regarding content (echolalia, stereotypical use of language), prosody (aberrant speech rhythm, intonation, pitch, volume), as well as reciprocity within the conversation. While content-related verbal peculiarities such as echolalia or stereotypical language are sufficiently straightforward to detect for a trained clinician, more subtle speech features might be harder to assess. Hence, automated speech analysis has the advantage of not requiring phonetic experts for manual coding to operationalize the prosodic “awkwardness” in autism. This approach allows for the exploration of the robustness of potential vocal markers. In fact, a promising body of research is directed towards the classification of a diverse range of psychiatric disorders using speech recognition techniques [54]. For autism, classification has been attempted based on automated speech analysis in tasks requiring conversations with an avatar [55], story retelling [56], or specific word repetitions [57]; the latter study reaching an accuracy of 91%. In real-life settings, two studies have used individual speech as well as turn-taking features extracted from ADOS interviews to predict diagnosis with accuracies ranging between 59-89% [36,58]. However, as autistic individuals have pronounced difficulties with small talk and building connections with others, an interview context encompassing a clear conversational pattern of question and answer might not adequately reflect the communication impairments present in autism. Therefore, the automated and objective extraction of both individual and interactional speech features during naturalistic reciprocal verbal exchanges poses an intriguing concept in the context of vocal marker research for diagnostic classification.

One aim of the present thesis was to comprehensively capture both non-verbal (Original Study I) and verbal (Original Study II) social interaction and communication features on multiple modalities from naturalistic social interactions between autistic and typically-developing (TD) individuals using largely automated analysis pipelines and, subsequently, explore their predictive power for assisting diagnosis. To reduce the sensory load on participants, the experimental setup contained minimal use of wearable devices. To maximize scalability, feature extraction was performed largely automatically from videos of the interactions, using established computer vision methods for motion and facial expression, as well as automated speech analysis. Instead of predicting individual diagnostic status, the developed ML models predict dyad membership to either mixed (ASD-TD) or non-autistic (TD-TD) dyad, therefore, putting the dyad in the focus of analysis and underlining interactional dependencies.

In the first study (Original Study I), synchrony in facial expressions, head and body motion was used as the base for multiple SVM classification models. The best performance was obtained based on the reciprocal interpersonal synchrony of facial motion (balanced accuracy/BAC of 79.5%), followed by models based on the total amount of head and body motion and facial expressiveness (BAC = 68.8%), and head motion (BAC = 62.1%). The *intra*personal coordination of head and body movement did not yield significant predictive performance. Combining modalities within a stacking model [59] did not improve overall classification accuracy.

In the second study (Original Study II), the predictiveness of automatically extracted speech features was explored for a subsample of the participants. Importantly, we included both individual and interactional aspects of vocal interactions and were able to predict dyad membership using a SVM algorithm with a BAC of 76.2%. Further, in an attempt to objectify the perceived “awkwardness” in autistic social interaction, distinct group differences between autistic and non-autistic participants were analyzed regarding prosodic and interactional features. We found slower and more monotonous speech in autistic participants, as well as an increased amount of silent intervals in ASD-TD interactions.

Though employing carefully cross-validated ML algorithms, the arguably limited sample size and lack of external validation cannot rule out the possibility of overfitting, i.e., a lack of generalization to other, unseen samples. However, in using a minimally invasive setup as well as objective and scalable analysis methods, we were able to capture social interaction in autism across multiple modalities and, hence, provide a framework potentially to facilitate future validation in more translational settings.

1.4 Scalable translation to clinical settings

The continuing developments regarding the remote assessment of behaviors using computer vision and wearables provide the opportunity to overcome scalability issues and deliver insights into subtle mechanisms of autistic social interaction. The classification accuracy for both verbal and non-verbal aspects of social communication in autism explored in Original Study I and II shows that the automatic extraction of these behaviors is a promising avenue in the quest for behavioral marker research in autism. Importantly, shifting the dyad into focus of the analysis and assessing features of an interaction along with features describing the individual represents a viable route for more fine-grained investigations of social interaction difficulties in autism. The scalable setup facilitates implementation into clinical practice, potentially augmenting diagnostic decision-making in the long run.

Therefore, another aim of the present thesis was to classify autism using a SVM model based on synchronized behaviors extracted via computer vision techniques between patient and clinician. Hence, in the third study (Original Study III), we analyzed videos of real ADOS interviews from a representative clinical sample of an autism outpatient unit in Korea with MEA and used the objectively extracted features to train a SVM model. Importantly, instead of TD individuals, patients with a diverse range of differential psychiatric diagnoses served as a control group, providing the possibility to address the issue of specificity. Additionally, while the previous studies were comprised of adults, a larger age range of children and young adults (5.5 – 28.7 years) was considered. Although the videos were not initially recorded for the purpose of this study – therefore constituting a rather uncontrolled setup – we did find above-chance classification accuracy (BAC = 63.4%) in support of interpersonal synchrony as a marker specific to autism.

1.5 Outlook

The research presented in this work provides new insights into the autism spectrum and is an important step towards the goal of finding the best combination of markers to predict autism. In the future, it will be important to enhance diagnostic classification models with multiple modalities [60], i.e., combining behavioral with biological measures. The increasing availability of wearable devices able to measure physiological (e.g., medical wristbands), neurological (e.g., EEG headbands) or neuropsychological (e.g., mobile eye tracking) signals allows for capturing detailed aspects of social interaction without being too invasive or too restrictive regarding clinical accessibility. Further, in the endeavor of translational research, it will be important to target specificity and validation in large-scale diagnostic contexts. This includes ethical considerations concerning data availability as such recordings are often bound by patient-doctor confidentiality [61], as well as mitigating bias in data collection and model development to ensure that minority groups are sufficiently represented [62]. Moreover, to fully overcome the circularity problem and address heterogeneity issues, unsupervised clustering approaches, i.e., identifying subclusters in a data-driven way, would be an interesting approach to investigate social interaction differences in autism. In the latest revisions of the common diagnostic classification systems, the categorization of autism into subtypes (e.g., Asperger's syndrome or childhood autism) has been abandoned in favor of regarding autism as a spectrum of social interaction difficulties and repetitive behaviors [1,63]. However, a recent meta-analysis [64] across 11 meta-analyses of autism-control group comparisons revealed a decrease in effect sizes over time, concluding that the substantial heterogeneity in autism symptomatology resulting from broadened diagnostic criteria may hamper the identification of solid diagnostic markers. Therefore, with the increased availability of larger data sets, the data-driven development of behavioral subclusters has the potential for gaining insight into the underpinnings of psychopathology and, ultimately, improve treatment options [65]. Finally, to shed further light on social interaction difficulties and their complex interdependencies in autism, the growing application of interpretable ML techniques can assist in increasing model transparency and, therefore, facilitate our understanding of *why* a certain diagnostic prediction is made for an individual [66].

When diagnosing ASD, clinical expertise remains of utmost importance and potential digital enhancements should be viewed as augmentation to and not replacement of a thorough diagnostic assessment conducted by a clinical expert. Yet, the objective assessment of autistic social interaction dynamics offers the opportunity for a scalable translation into clinical practice and could inform diagnostic decision-making in the future.

2 Original Study I – Machine learning classification of autism spectrum disorder based on reciprocity in naturalistic social interactions

Contributions and reference

In this study, the aim was to extract non-verbal behavioral features from social interactions between autistic and non-autistic participants and build diagnostic prediction models thereof. I oversaw the overarching project administration which involved gaining approval from the ethics board of LMU and study insurance matters. I programmed the paradigm, including the script-based synchronization of the hardware via PsychoPy [67] to ensure fully synchronized data collection. I presented the study design and collected feedback from members of the autism community during an outreach event. I was directly involved in and supervised recruitment and data collection of the autistic and non-autistic participants. I curated the data and developed the preprocessing pipelines using DaVinci Resolve, OpenFace [68], MEA [42] and RStudio [69]. I analyzed the final data set using NeuroMiner [70]. I presented the results to the scientific community during talks and poster presentations at different conferences. In compliance with the endeavor of reproducibility and open science, I published the analysis scripts on GitHub (https://github.com/jckoe/MLASS-study) and made a preprint available on MedRXiv. I wrote the manuscript and was primarily responsible for the revision process and finalizing the published article.

Koehler, J.C., Dong, M.S., Bierlich, A.M., Fischer, S., Späth, J., Plank, I.S., Koutsouleris, N., Falter-Wagner, C.M. (2024). **Machine learning classification of autism spectrum disorder based on reciprocity in naturalistic social interactions.** *Translational Psychiatry,* 14(1), 76. https://doi.org/10.1038/s41398-024-02802-5

Note: For better readability, the supplementary material of this publication can be accessed online.

Journal Citation Report:

Impact Factor 2022: 6.8

5 Year Impact Factor 2022: 7.0

Ranked 24/264 in Psychiatry Journals (Journal Citation Indicator 2022)

Translational Psychiatry www.nature.com/tp

ARTICLE OPEN

Check for updates

Machine learning classification of autism spectrum disorder based on reciprocity in naturalistic social interactions

Jana Christina Koehler [1]✉, Mark Sen Dong [1], Afton M. Bierlich [1], Stefanie Fischer [1,2], Johanna Späth [1], Irene Sophia Plank [1], Nikolaos Koutsouleris [1,3,4] and Christine M. Falter-Wagner [1]✉

Autism spectrum disorder is characterized by impaired social communication and interaction. As a neurodevelopmental disorder typically diagnosed during childhood, diagnosis in adulthood is preceded by a resource-heavy clinical assessment period. The ongoing developments in digital phenotyping give rise to novel opportunities within the screening and diagnostic process. Our aim was to quantify multiple non-verbal social interaction characteristics in autism and build diagnostic classification models independent of clinical ratings. We analyzed videos of naturalistic social interactions in a sample including 28 autistic and 60 non-autistic adults paired in dyads and engaging in two conversational tasks. We used existing open-source computer vision algorithms for objective annotation to extract information based on the synchrony of movement and facial expression. These were subsequently used as features in a support vector machine learning model to predict whether an individual was part of an autistic or non-autistic interaction dyad. The two prediction models based on reciprocal adaptation in facial movements, as well as individual amounts of head and body motion and facial expressiveness showed the highest precision (balanced accuracies: 79.5% and 68.8%, respectively), followed by models based on reciprocal coordination of head (balanced accuracy: 62.1%) and body (balanced accuracy: 56.7%) motion, as well as intrapersonal coordination processes (balanced accuracy: 44.2%). Combinations of these models did not increase overall predictive performance. Our work highlights the distinctive nature of non-verbal behavior in autism and its utility for digital phenotyping-based classification. Future research needs to both explore the performance of different prediction algorithms to reveal underlying mechanisms and interactions, as well as investigate the prospective generalizability and robustness of these algorithms in routine clinical care.

Translational Psychiatry (2024)14:76 ; https://doi.org/10.1038/s41398-024-02802-5

INTRODUCTION

Background

The diagnosis of autism spectrum disorder (ASD) encompasses a range of symptoms in reciprocal social interaction and communication as well as restricted, repetitive behaviors and interests [1]. The currently rising prevalence [2] exacerbates waiting times for an already long and demanding diagnostic process, increasing psychological stress on seeking diagnostic clarification [3]. Gold-standard recommendations include assessment with semi-structured diagnostic interviews or observational tools conducted by multidisciplinary teams, along with neuropsychological assessments and an anamnesis of developmental history by a caregiver [3]. With the increasing number of patients seeking diagnosis in adulthood, the lack of recommended diagnostic instruments for this population [4] poses an additional challenge. Therefore, the improvement of the diagnostic process of autism in adulthood has been named one of the top priorities in autism research [5].

Digitalized methods have high potential to improve screening and diagnostic procedures, such as assessing home videos [6] or interactions with virtual characters [7, 8]. While promising, these findings often rely on time-consuming manual behavioral coding or, more importantly, may not adequately reflect real-time social interactions, which are especially relevant for judging symptom strength [9]. Additionally, the increased use of artificial intelligence methods, such as machine learning (ML), has furthered research on increasing the efficiency of existing diagnostic tools, e.g., by identifying subsets of the most important items for diagnosis [10, 11], pointing to areas of impairments most indicative for diagnosis. These include aberrances in, e.g., gesturing, facial expressions and reciprocal social communication [10], traits which seemingly influence first impressions of people with ASD, who are judged as interacting more awkwardly by typically developing (TD) peers [12]. This suggests that non-verbal behavior also represents an important pillar of clinical impression formation.

Several computer vision approaches have been investigated to capture this different non-verbal behavior and explored its use for autism diagnosis, underlining its usefulness for the quantification of behavioral markers [13]. For example, using motion tracking, the degree of imitation of isolated hand movements could identify autistic and non-autistic adults with an accuracy of 73% [14]. A recent deep learning approach analyzing videos that depict

[1]Department of Psychiatry and Psychotherapy, Medical Faculty, LMU, Munich, Germany. [2]Goethe University Frankfurt, University Hospital, Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, Frankfurt am Main, Germany. [3]Max Planck Institute of Psychiatry, Munich, Germany. [4]Institute of Psychiatry, Psychology and Neuroscience, King's College, London, UK. ✉email: jana.koehler@med.uni-muenchen.de; christine.falter@med.uni-muenchen.de

Received: 13 January 2023 Revised: 19 January 2024 Accepted: 23 January 2024
Published online: 03 February 2024

clinical interviews of autistic and non-autistic children revealed a classification accuracy of 80.9% based on pose estimation [15].

Reciprocal interaction in ASD

A way to quantify this aberrant interaction style is through closely examining the way two interacting partners temporally adjust their behavior with each other, or, in other words, how well they are "in sync". Interpersonal synchrony or coordination can not only be achieved through mutual, bilateral matching, but also by establishing leader-follower relationships through unilaterally adapting to the behavior of the interactant [16]. Interpersonal synchrony has repeatedly been associated with rapport, affiliation, and perception [17, 18], emphasizing its importance for social cognition. In ASD, reduced interpersonal synchrony or coordination has in fact been described on multiple modalities and across the lifespan [19]. For instance, reduced coordination of emotional facial expressions has been found in autistic youth in conversation with a partner [20]. Interpersonal synchrony in head motion has been found to be reduced in diagnostic interviews with patients subsequently diagnosed with autism as compared to those who were not [21]. Another study investigating head and body motion synchrony in autistic and non-autistic adults found both to be reduced when an autistic person was part of the conversation [22], once again reflecting the importance of the interactional perspective. Further, synchrony and coordination differences in autism have also been found within the individual (intrapersonally), with reduced or differing coordination of simultaneous movements [23, 24] or across communication modalities [25]. Lastly, movement atypicalities, apart from coordination, appear to be pronounced in autism, including reduced facial expressiveness in autistic children [26], as well as a unique kinematic profile of biological motion and motor control [27]. A recent meta-analysis found a significant correlation between gross motor and social skills in autism [28], underlining the significance of movement differences for the core symptomatic profile of ASD.

Aims

In summary, the mere definition of ASD as a disorder of social interaction implies an interdependency and calls for shifting to the dyad as unit of analysis [29]. However, feasible measures are lacking due to their reliance on extensive manual coding, experimental paradigms appearing staged or unnaturalistic, or investigating only isolated aspects of social interaction. Hence, the aim of this proof-of-concept study was to build upon existing knowledge of adaptation difficulties in autism and use the richness of non-verbal social interaction data in an efficient way to build an objective (i.e. independent of self- or clinician-ratings) classification model of autistic social interaction. To this end, we trained several Support Vector Machine (SVM) classification models to optimally differentiate between members of autistic vs. non-autistic interactional dyads. To increase objectivity and feasibility for potential further development in clinical practice, we used existing open-source algorithms that maximized automation in the annotation and analysis process.

METHODS

Sample

We recruited 35 participants with ASD from a clinical database, as well as local autism networks. The diagnosis (F84.0 or F84.5) had to have been given by a qualified clinical psychologist or psychiatrist according to ICD-10 criteria as confirmed by a full diagnostic report. Inclusion criteria were an age between 18-60 years, normal intelligence (IQ > 70, as measured by an IQ score based on a verbal and non-verbal IQ test [30]) and no current neurological disorder. Additionally, 69 typically developing (TD) participants with no current or history of psychiatric or neurological disorders or psychotropic medication were recruited. Two ASD participants had to be excluded from the final sample because their diagnosis could not be verified on the basis of an incomplete diagnostic report. An additional five ASD participants were excluded during the analysis due to data loss from imprecise facial tracking. Due to the dyadic nature of the study, their interactional partners had to be excluded as well. Another TD-TD dyad was excluded due to technical issues during script loading, leading to a final sample of 88 participants. Groups were matched with respect to age and IQ. A chi-square-test of independence revealed no significant association between group membership and sex, $\chi^2(1,\ N = 88) = 2.6$, $p = 0.11$. A description of the final sample can be found in Table 1. All participants gave written informed consent before study participation and were compensated monetarily afterwards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects/patients were approved by the ethics committee of the medical faculty of the LMU Munich (number 19-702).

Table 1. Sample description.

	ASD ($n = 28$, 18 female)	**TD ($n = 60$, 26 female)**	$p_{adjusted}$
Age	37.18 (13.14)	31.48 (10.78)	0.101
Crystalline IQ	113.68 (16.90)	113.98 (16.69)	0.879
Non-verbal IQ	119.75 (23.44)	117.03 (17.37)	0.382

Mean parameter values (SD in parentheses) for each of the IQ tests for the ASD and TD participants, as well as the results of Wilcoxon tests (assuming unequal variances). Verbal IQ as measured by the Mehrfach-Wortschatz-Test. Non-verbal IQ as measured by the Culture-Fair-Test 20-Revised. Participants with ASD either had a diagnosis of F84.0 [3] or F84.5 [25]. p values adjusted for multiple comparisons using the false discovery rate (FDR) [69].

Study setup

Participants were randomly paired resulting in 28 ASD-TD (mixed) and 16 TD-TD (non-autistic control) dyads. All were naïve to the diagnosis of their interactional partner. They were seated approximately 190 cm across from each other in fixed chairs. Two cameras (Logitech C922) were installed on a tripod on a table in front of the participants, recording their respective facial expression at 30 frames per second. A third camera was mounted at a wide angle on a tripod at a distance of approximately 240 cm (Fig. 1). All recordings were operated from a single computer using custom PsychoPy [31] scripts, allowing for maximal synchronization of the three video input streams. To control for any biases in subsequent video analyses caused by lighting change [32], measurements were taken in stable artificial light. To maximize hygienic safety measures during the Covid-19 pandemic, slight changes to the setup were required after the first nine participants were assessed (see Supplementary Information S4.3).

Participants engaged in two ten-minute conversation tasks for which they were instructed beforehand by the study personnel. After giving a starting cue with a clapping board, all study personnel left the room. Participants were asked to engage in a conversation about their hobbies, as well as to plan a fictional five course meal with dishes they both disliked. The mealplanning task has been used in previous synchrony studies (e.g., [22, 33]), with the rationale that a collaborative and funny task increases affiliation and synchrony respectively. As the overlap between dishes two people dislike tends to be smaller than finding common food preferences, this task requires more collaboration from both interactants. In contrast, a conversation about their hobbies was introduced. Restricted interests are a key diagnostic criterion of autism according to DSM-5 [1], whereby autistic individuals tend to switch to monologue style when talking about their interests [34]—a unique behavior, which we aimed to capture. The order of the tasks was counterbalanced among participants.

Additionally, participants completed a series of questionnaires to assess their level of self-reported autistic traits (Autism Quotient; AQ [35]), empathy (Saarbrücker Persönlichkeitsfragebogen; SPF [36], the German version of the Interpersonal Reactivity Index; IRI [37]), alexithymia (Toronto Alexithymia Scale; TAS20 [38]), depressiveness (Beck Depression Inventory; BDI [39]), self-monitoring (self-monitoring scale; SMS [40]), and movement difficulties (a German translation of the Adult Dyspraxia Checklist; ADC [41]). To obtain a best estimate of both their crystalline (Mehrfachwahl-Wortschatz-Intelligenztest; MWT [42]) and non-verbal (Culture-Fair Test;

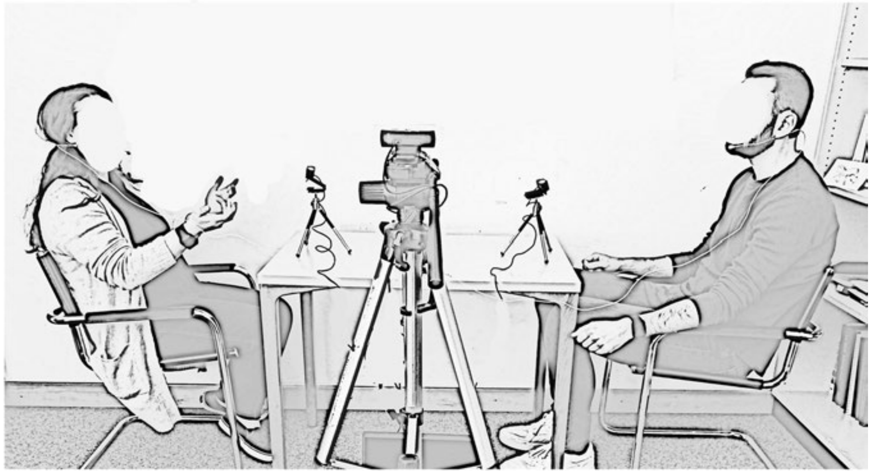

Fig. 1 Experimental setup. Participants were seated across from each other and asked to conduct two conversational tasks. For additional setup info see Supplementary Material.

CFT 20-R [43]) IQ, two IQ assessments were undertaken, and their results averaged. Since difficulties in recognizing emotional facial expressions could potentially cause a bias in the investigation of synchrony in facial expressions, participants additionally completed a computer task for facial expression recognition (Berlin Emotion Recognition Test; BERT [44]).

Data preparation and feature extraction

Videos were cut to a duration of ten minutes in DaVinci Resolve (Version 16.2.0054). Facial expression was analyzed with the open-source algorithm *Openface 2.0* [45], identifying action units (AUs) and three head pose parameters (pitch, yaw, roll) and extracting a time series of their presence and intensity for every frame. Motion Energy analysis (MEA [32]) was used to analyze head and upper body movement captured with the scenic camera. MEA extracts time series of grayscale pixel changes for every frame in pre-specified regions of interest (ROI). Due to the constant lighting conditions and a stable camera, pixel changes within each ROI indicate movement.

Prior to the final analyses, the behavioral time series from both tasks were synchronized between, and (in case of intrapersonal coordination) within, participants in the respective modalities. For this purpose, windowed cross-lagged correlations were computed in R. The size of the respective windows and lags for each modality were carefully chosen, relying on previous research wherever applicable [20, 46], to ensure maximum standardization. For the estimation of intrapersonal coordination, head movement, as derived from OpenFace, was cross-correlated with the body motion energy times series derived from MEA. Finally, summary scores (mean, median, standard deviation, minimum, maximum, skewness, and kurtosis) of the maximum synchrony instances from both tasks for each person were extracted. The extent as to which each person was synchronizing within the dyadic interaction was defined as their degree of imitating (following) their partners movements. For further details on the cross-correlation and feature extraction procedures, refer to Supplementary Information S2.

Facial emotion recognition capabilities were operationalized as mean accuracy (in %) and response time (in ms) (see Supplementary Information S3.4).

A full list of features can be found in Supplementary Table S13.

Classification models

Separate SVM classification models were trained using features grouped according to the interaction modalities. The feature vectors for each participant combined the values from both the mealplanning and hobbies task. In each base model, the SVM algorithm independently modeled linear relationships between features and classification label. To account for the interactional nature of the underlying feature set for classification, participants were labeled as belonging to either a mixed (ASD-TD) or non-autistic control (TD-TD) dyad, resulting in groups of 56 and 32 individuals respectively. Consequently, both interactants within one dyad received the same label, regardless of their individual diagnosis. This labeling procedure was modeled closely to a diagnostic setting in clinical reality, in which only one interactant's diagnostic status would be at question whereas the other interactant would represent the clinical rater. Linear SVM optimized a linear hyperplane in a high-dimensional data space that maximized separability between individuals belonging to either of the two dyad types (i.e., the support vectors). Based on the trained hyperplane, the data was subsequently projected into the linear kernel space and their geometric distance to the decision boundary was measured, therefore, predicting each participant's classification. Every participant was assigned a decision score and a predicted classification label.

We built separate models for the synchrony of facial action units (FACEsync; 168 features per individual), head movement (HEADsync; global head movement, as well as pitch, yaw and roll; 56 features per individual), and body movement (BODYsync; 14 features per individual), as well as intrapersonal head-body movement coordination (INTRAsync; 14 features per individual), and individual movement parameters (MovEx; total head and body movement, and facial expressiveness; 6 features per individual). The decision scores of all our base models, as well as the model covering the head region (FACEsync + HEADsync), were subsequently combined in a stacking-based data fusion framework [47] to assess whether a combination of the modalities would result in superior prediction results than the unimodal classifiers themselves.

We additionally conducted supplementary analyses using individual diagnosis as classification label. Results of these analyses can be found in the Supplementary Information S3.6.

Support vector machine learning analysis

Machine learning analyses were conducted with the toolbox NeuroMiner (Version 1.1; https://github.com/neurominer-git/NeuroMiner_1.1) [48] in MATLAB (Version 2022b) [49]. A repeated, nested, stratified cross-validation (CV) structure was implemented with 11 outer CV folds and ten permutations (CV2) and ten inner CV folds with one permutation (CV1). At the CV2 level, we iteratively held back participants from four dyads as test samples (approx. 9% of data), while the rest of the data (approx. 81%) entered the CV1 cycle, where the data were again split into validation and training sets. Both interactants from a dyad would always remain in the same fold. This nested stratified CV allows for a strict separation between training and testing data, with hyper-parameter tuning happening entirely within the CV1 loop while the CV2 loop exclusively measured the model's generalizability to unseen data. Additionally, the stratified design ensured that proportion of dyad type in every fold would adequately reflect the proportion of dyad type in the full sample in order to avoid training bias. The five base models were pre-processed and trained separately using LIBLINEAR Support Vector L2-regularized L2-loss classification algorithms (see Supplementary Information S3.1 and S3.2). Given that the current dataset contains a rather high feature-to-sample ratio, this specific algorithm was chosen because of its similarity to LIBSVM but without implementing complex kernels which could potentially result in over-fitting. All models were corrected for class imbalance by hyperplane weighting. Balanced Accuracy (BAC = (sensitivity + specificity)/2) was used as the performance criterion for parameter optimization. Statistical

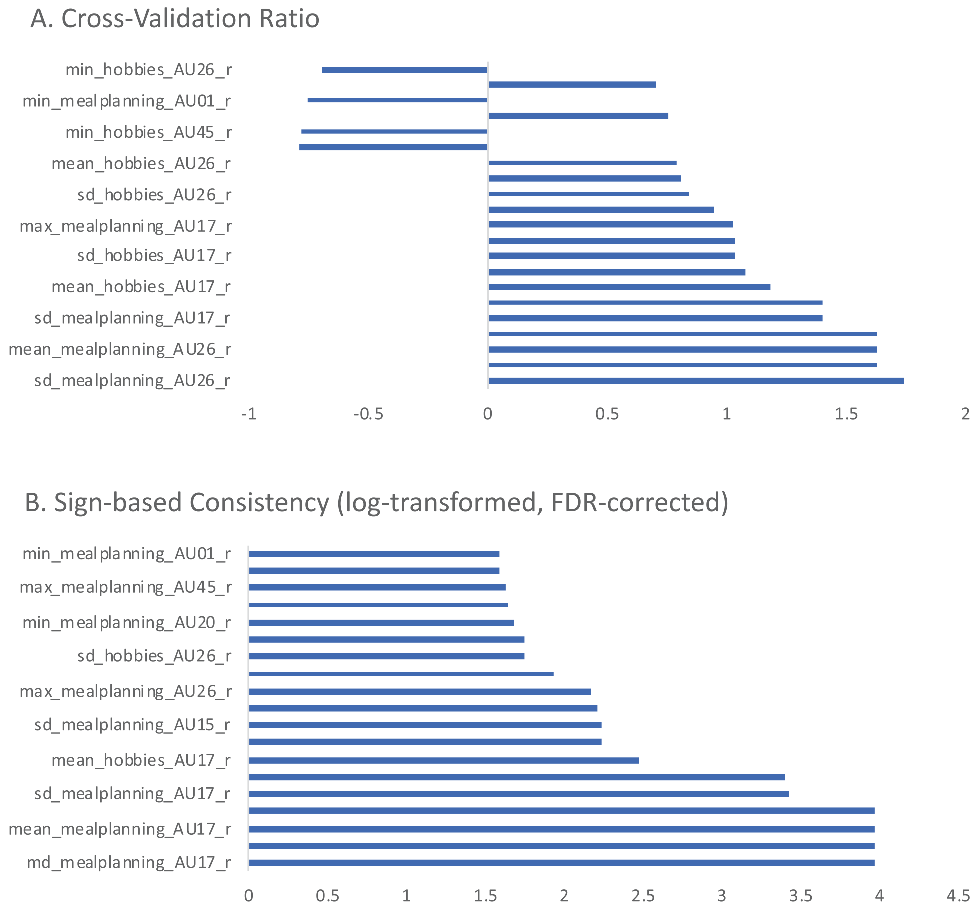

Fig. 2 Contribution of features in FACEsync model. Cross-validation ratio of feature weights (**A**) and sign-based consistency (**B**) for the FACEsync model. The features depicted correspond to the person-specific adaptation of intensity of a participant to their dyadic counterpart in the respective facial action units (AU) for either hobbies or mealplanning task (min minimum, sd standard deviation, max maximum).

significance of the base classifiers was assessed through permutation testing [50]. The permutation testing procedure determines how statistically significant is the model's performances (i.e., BAC) using the current data compare to models trained on the dataset but with the labels randomly permuted. The permutation test was repeated 1000 times. The significance level was set to $\alpha = 0.05$. In current discussions, an alpha level of .005 has been proposed [51], though the appropriateness of this approach has been called into question [52]. Hence, to reassure statistically rigorous results, we additionally annotated when a significant model's permutation test result would fail significance using the stricter alpha level. To control for potential bias of the dyadic nature of the data on each model's significance, each permutation analysis was conducted with both participants of each dyad permuted in pairs according to their dyadic structure. For further details on the permutation testing procedure, see Supplementary Information S3.3. The two stacking models [53] were trained on the resulting decision scores (all base models, facial expression + head motion synchrony) by wrapping them in the identical cross-validation framework as the base models. A L1-loss LIBSVM algorithm with Gaussian kernel was employed to find a parsimonious combination of decision scores which maximized BAC across the C parameter range. For details, see Supplementary Information S3.

RESULTS

Base model performances

Using facial action unit (AU) synchrony data, the repeated nested stratified cross-validation FACEsync model yielded a balanced accuracy (BAC) of 79.5%, and an area under the receiver operating curve (AUC) of .82 ($p < 0.001$, also see Supplementary Fig. S9). The contribution of the different features to classification group (Fig. 2) was calculated by feature weights (see Supplementary Information S3.4) and cross-validation ratio. Additionally, the sign-based consistency was explored as an indicator of the feature classification reliability. Assignment to the ASD-TD dyads was mainly driven by features describing an elevated and highly varied extent of adaptation in AU17 (chin raiser) and AU26 (jaw drop). Minimized adaptation in AU01 (inner brow raiser), AU20 (lip stretcher) and AU45 (blink) were indicative of belonging to the TD-TD interaction type. In order to investigate any associations of facial emotion recognition abilities and adaptation behaviors of the different facial AUs, correlation analyses were performed between the decision scores derived from the FACEsync model and accuracy and response time (rt) from the Berlin Emotion Recognition Test (BERT [44]). No significant associations were found ($r_{accuracy}(86) = -0.16$, $r_{rt}(86) = 0.13$; both $p = 0.23$ after FDR correction).

The model using only head motion coordination data (HEADsync) achieved a BAC of 62.1% and an AUC of 0.64 ($p = 0.002$). Assignment to the TD-TD group was driven by higher values in minimum adaptation of global head movement whereas higher maximum and more variant values for head movement adaptation predicted the ASD-TD group.

The classification model based on upper body movement coordination (BODYsync) predicted dyad origin with a BAC of 56.7% and an AUC of 0.55 ($p = 0.009$). Using a stricter alpha level of 0.005, this model would not be judged as performing significantly better than chance.

Our classification model based on intrapersonal head-body coordination (INTRAsync) performed around chance level with a BAC of 44.2% and an AUC of 0.44 ($p = 0.994$).

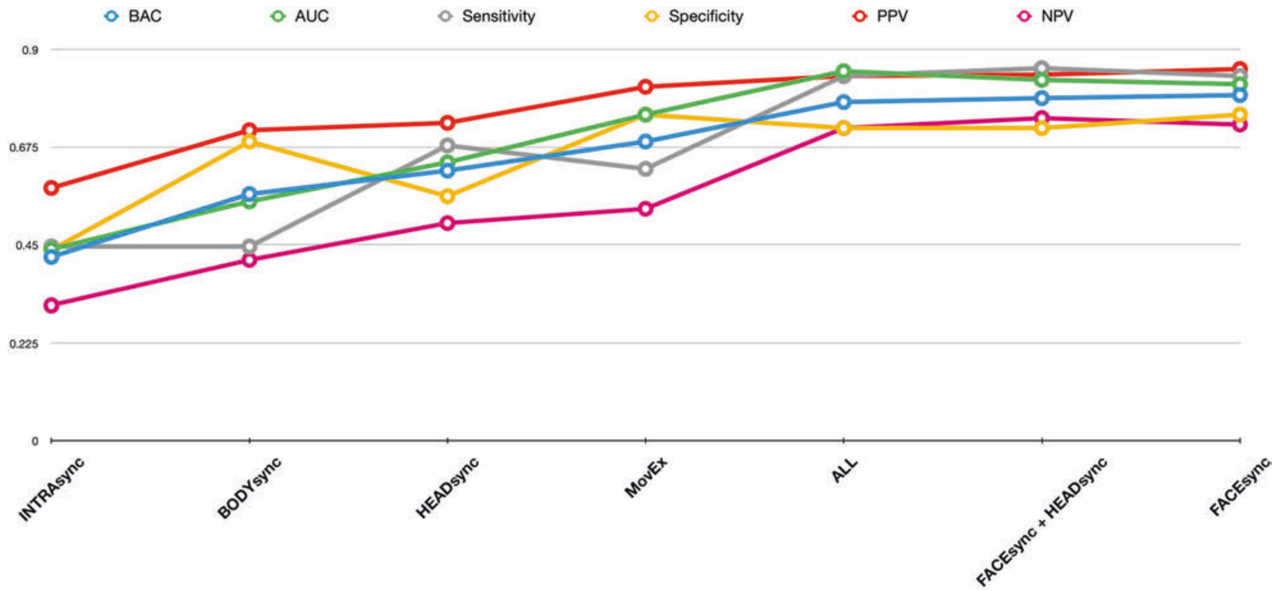

Fig. 3 Classification metrics for all base and stacking models. BAC balanced accuracy, AUC area under the curve, PPV positive predictive value, NPV negative predictive value. Models are depicted in the order of lowest to highest performing BAC.

The SVM classification model based on features of total head and body movement and general facial expressiveness (MovEx) predicted dyad origin with a BAC of 68.8% and an AUC of 0.75 ($p < 0.001$).

Additional classification metrics for all models can be found in Supplementary Table S6.

Stacking model

All base model decision scores were extracted and combined into a hierarchical stacking-based fusion framework to assess potential prediction improvements. Combinations of only the head region (FACEsync + HEADsync; BAC = 78.8%, AUC = 0.83), as well as of all modalities (BAC = 77.9%, AUC = 0.85) did not outperform the most predictive base model (FACEsync) with 79.5%.

Additional classification metrics of all models are depicted in Fig. 3.

Classification based on diagnostic group

We repeated all SVM analyses using different labels based on diagnostic groups while ignoring interaction type. These additional analyses were conducted in order to investigate if our collected social interaction data was specific enough to identify an autistic individual, regardless of interaction dyad origin. All models generated inferior prediction accuracies compared to the dyad labeling approach (3.1). Detailed results can be found in Supplementary Information S3.6.

DISCUSSION

The aim of the current study was to quantify social interaction in ASD for the purpose of automatized diagnostic classification. In this proof-of-concept study, we set out to utilize a dyadic setting for classification of autistic vs. non-autistic interaction based on reciprocity. Participants were filmed conducting two brief conversations about pre-set topics. Using repeated nested cross-validation techniques, we could show that SVM classification models based on different modalities of behavioral reciprocity were sufficient to predict dyad membership to a high degree. Contrary to our hypothesis, combining different non-verbal modalities did not improve overall predictive accuracy. Classification into individual diagnostic groups (ASD vs. TD) based on social interaction data performed worse on all modalities, as well as the model classifying on individual measures of full body movement and general facial expressiveness. This highlights the importance of the social context to capture the manifestation of autistic symptoms.

A model based on reciprocity of facial action units within the interactions showed the best classification accuracy (79.5%) within our sample. When looking more closely at individual feature importance in the facial region, we found heightened and more varied scores for reciprocal adaptation in the AUs chin raiser, jaw drop and lip corner depressor in both tasks to be indicative for classification into the autistic interaction type. This was especially pronounced for the mealplanning task, suggesting higher and more varied synchrony in this task in the ASD-TD interactions. While elevated synchrony in ASD might seem counterintuitive at first glance, especially in light of findings on reduced mimicry in autism [54, 55], taking a closer look at feature importance for the TD-TD group provides a differentiated picture. Participants with higher values for minimum adaptation across all features had an increased likelihood to be classified into the TD-TD group, suggesting a potential floor effect for facial synchrony in this group. Thus, their synchrony did not subceed a certain lower threshold. This was especially pronounced in the action units for inner brow raiser (AU1), lip stretcher (AU20) and blinking (AU45). Additionally, motor synchrony in autistic interactions has previously been found to vary along with the level of autistic traits, social-communicative functioning, and context [19]. The same mechanisms may hold true for mimicry. For example, in a study investigating mimicry in the BERT emotion recognition task, Drimalla and colleagues [56] found significantly more variance in the intensity of facial expressions in autistic participants. Importantly, since machine learning analyses factor in countless interdependencies between features, interpretations based on feature weights should be considered with caution. Nevertheless, the rather high classification accuracy based solely on facial synchrony features found in our study provides valuable implications for future research on classification based on social interactions in an even more ecological setting (e.g., diagnostic assessments via video conferencing).

Interestingly, our model based on measures of individual amount of full body movement and general facial expressiveness (MovEx) was the second-best of the base learners, supporting findings of a characteristic motor signature in autism. For example, Zhao and colleagues [57] investigated head movements in autistic children during live interactions and found aberrances on all three axes. Notably though, our classification model factoring in dyad type, thus, data that included the TD interaction partners, showed superior performance compared to classification based on diagnosis. Hence, interactional aspects also seem to have an association with individual movement features, supporting the hypothesis that intra- and interpersonal adjustment processes are not entirely independent of each other [58].

Contrary to previous findings of high classification accuracy for head and body coordination [24], our model based on this modality performed at a below-chance level, showing low specificity of head-body coordination for autistic vs. non-autistic interaction. However, interpretation should be considered cautiously given the specifications of our experimental setup. Due to our data being collected as part of a larger setup, participants wore wristbands on their non-dominant hand measuring physiological data (see Supplementary Information S1). In order to reduce artefacts in physiological data acquisition, participants were instructed to relax their non-dominant hand in their lap. Arguably, this instruction and setup difference with regards to the previous study could well account for the lack of classification power by intrapersonal coordination in the current study. This is supported by the absence of a significant difference between body synchrony found between our participants' motion time series and randomly matched time series (see Supplementary Information S2.3).

While our results support previous findings on head motion synchrony as a distinguishing feature of autistic communication [21], combining it with facial expression synchrony did not yield a higher prediction accuracy in a stacking model. This was also the case for our overall stacking model. However, stacking may be able to improve predictive performance of any problem primarily in cases where the underlying data is not well represented by a single model [59], which is not the case in the current study. Furthermore, combining several models with significantly different predictive accuracies might in fact harm overall performance of the stacker. Additionally, if the underlying base models are highly correlated, combining them does not necessarily lead to improved performance [59]. In fact, we did find significant associations of our MovEx model (total head and body movement and general facial expressiveness) with HEADsync for the ASD-TD group ($r = 0.55$, $p < 0.001$), as well with INTRAsync for the TD-TD group ($r = 0.52$, $p > 0.05$; for further details see Supplementary Information S4.5). In our study, we aimed to combine different modalities in a hypothesis-driven way to retain a certain amount of interpretability. We found no added benefit for increasing model complexity. However, it is possible that in order to improve predictive performance of social interactions features, non-verbal aspects of social interaction could be complemented by different modalities in the future, such as speech, eye-movements, physiological or neurological measures. For example, in a recent study conducted by Liao et al. [60], simultaneous measures of EEG, eye tracking and facial expression were assessed of autistic children viewing social and non-social stimuli. The authors found superior prediction accuracies for the combination of behavioral and physiological classifiers.

Notably, there are several limitations within the scope of the present study.

First, the sample size in the current study is limited. To counter this, we implemented a repeated nested cross-validation structure as well as careful feature reduction methods. Nevertheless, our findings should be considered as proof-of-concept and will have to be validated in a larger and external sample, possibly including adults with different psychiatric diagnoses, including comorbidities, to examine specificity within a clinical context more closely and, hence, strengthen the translational aspect [61]. Additionally, regarding the differing incidence rates and possibly phenotypical presentation in males and females with autism, larger samples will allow for thorough analyses of sex and gender effects on social interaction in autism. In any case, we believe that automatic extraction and classification algorithms of social behaviors can support human observation, as they have the possibility to extract behavioral subtleties reliably (e.g., subtle facial expressions [62]), and, thus, could augment diagnostic decision making [63] over and above potential biases. We are convinced that the high scalability of our largely automatized setup can facilitate a simplified data collection process within clinical settings, ideally allowing for cross-site validation approaches which are crucial to the development of reliable clinical prediction models [64]. Second, though interpersonal synchrony has been found to be reduced in interactional dyadic settings independent of partner diagnosis [22], a preference for interactions within purely autistic dyads as compared to mixed interactions has been suggested [12]. This is reflected in theoretical frameworks, such as the "double empathy problem" [65] as well as "dialectic misattunement" [66], specifying autistic impairments to be especially pronounced between people with fundamentally different ways of information processing and interacting. While this underlines the notion of ASD as a social interaction disorder, in a real-world and especially clinical setting this homogenous combination is rarely to be found, which is why this dyad composition was not assessed in this study.

Third, though highly scalable, we relied on different existing computer vision algorithms for our study. On the one hand, this means that the direct comparison of the base models' accuracies has to be interpreted with caution, as both computer vision algorithms used employ different methods of movement extraction. On the other hand, these algorithms are associated with certain limitations themselves. For example, Motion Energy Analysis (MEA) as a video analysis method has constraints regarding the dimensionality of movement. Because MEA only outputs changes in motion, no specifications regarding direction or magnitude of movement can be made. However, while more distinct investigations of these factors in ASD are certainly desirable, they nevertheless add another layer of complexity to already highly dimensional prediction models. With an increasing feature-to-sample ratio, the ability of ML classifiers to learn more complex relationships may be restricted. Therefore, this was not a focus in our study. Regarding facial expression, a range of AUs and participants had to be excluded due to their extent of missing values within their resulting time series. This was partially due to the participants moving out of the camera frame. Though OpenFace employs person-specific normalization by subtracting a "neutral" face from all other frames of a person, the algorithm is nevertheless reported as potentially less accurate if a face does not show a lot of movement dynamics [67]. Further, within this study the AUs were extracted in a completely automated fashion, without external validation by human coders. While performance accuracy measures for OpenFace are generally favorable compared to other automatic facial expression detection algorithms [45] and this fully automatic approach furthers scalability, nevertheless, it cannot be ruled out that the AUs were not measured accurately, limiting direct interpretations in terms of specific AUs. However, even considering those technical drawbacks, our FACEsync model achieved high classification accuracy. We believe that with the continuing technological developments within computer vision methodology this limitation will likely be overcome in the future.

Lastly, the application of machine learning in clinical psychology and psychiatry is providing novel possibilities for increased precision in individualized diagnosis, prognosis and treatment [68]. However, with increasing model complexity, interpretation of findings and their implications become more challenging. While our findings point to the predictive accuracy of reciprocity in social interactions for autism, future research should aim to gain a greater understanding about the underlying mechanisms of those features. For instance, while we have found high predictive accuracy for an overall estimate of autistic reciprocal interaction within a conversation, a more fine-grained analysis of behavioral synchrony at different time points could shed light on possibly fluctuating interaction dynamics. In addition, while this study mainly explored the use and performance of one of the most widely used machine learning algorithms in psychiatric research [68], there exist a range of other supervised and unsupervised machine learning algorithms that, given a careful cross-validation procedure, tend to perform well with small sample sizes. An additional exploratory analysis using Random Forest Classification is included in the Supplementary Material (Section S4.7). However, to gain deeper understanding of underlying interactions and mechanisms in autistic social interaction, future research should compare the performance indices and feature spaces selected by different algorithms across different samples. Furthermore, interpretable machine learning models could be used in future studies to take feature analysis to the individual level and, thus, study the heterogeneity of ASD in more detail as well as develop more personalized psychosocial interventions.

In this study, we tested adults with autism with a diverse range of cognitive functioning levels, autistic traits and ages. Nevertheless, our SVM models managed to identify participants of an autistic social interaction with high accuracy. While this approach prevents disclosure of the diagnostic status of each individual within a dyad, thus, preserves anonymity, the continuing developments in computer vision prediction models may raise concerns of those affected over the risk of unwilling identification. Hence, it should be emphasized that a professional clinician's rating is essential for diagnostic decision making in psychiatric care. Consequently, diagnostic prediction models should be viewed as augmenting, rather than replacing diagnostic assessments made by trained clinicians [63]. However, a shift of data collection from traditional questionnaire-based or behavior observation diagnostic tools to objective digital markers will produce sensitive data that needs to be continuously treated with greatest care and data protection standards need to be abided by. Here, automated coding of behaviors is especially beneficial as opposed to manual approaches, allowing for instant anonymization of extracted time series.

Conclusively, using carefully cross-validated ML algorithms, we were able to classify members of autistic and non-autistic dyads based on multiple objective non-verbal measures of reciprocity in naturalistic social interactions. Facial synchrony within the dyad as unit of analysis [29] proved to be the most valuable marker for diagnostic classification of ASD. We are confident that with the growing interconnectedness in psychiatric and computational research, the complexity of social interaction difficulties in autism can be optimally captured.

DATA AVAILABILITY

The dataset generated and analyzed during the current study contains clinical information and is therefore not publicly available. They are available from the first author upon reasonable request pending the approval of the coauthors. The preprocessing scripts used during this study are available under https://github.com/jckoe/MLASS-study.

REFERENCES

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). 5th ed. Arlington, VA: American Psychiatric Pub; 2013.
2. Matson JL, Kozlowski AM. The increasing prevalence of autism spectrum disorders. Res Autism Spectr Disord. 2011;5:418–25.
3. Zwaigenbaum L, Penner M. Autism spectrum disorder: advances in diagnosis and evaluation. BMJ. 2018;361:k1674.
4. AWMF. Autismus-Spektrum-Störungen im Kindes-, Jugend- und Erwachsenenalter, Teil 1: Diagnostik: Interdisziplinäre S3-Leitlinie der DGKJP und der DGPPN sowie der beteiligten Fachgesellschaften, Berufsverbände und Patientenorganisationen. 2016. https://www.awmf.org/uploads/tx_szleitlinien/028-018l_S3_Autismus-Spektrum-Stoerungen_ASS-Diagnostik_2016-05.pdf.
5. Autistica. Your questions: shaping future autism research. 2016. https://www.autistica.org.uk/downloads/files/Autism-Top-10-Your-Priorities-for-Autism-Research.pdf.
6. Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: A development and prospective validation study. PLoS Med. 2018;15:1–20.
7. Drimalla H, Scheffer T, Landwehr N, Baskow I, Roepke S, Behnoush B, et al. Towards the automatic detection of social biomarkers in autism spectrum disorder: introducing the simulated interaction task (SIT). NPJ Digit Med. 2020;3:1–10. https://www.nature.com/articles/s41746-020-0227-5?utm_source researcher_app&utm_medium referral&utm_campaign RESR_MRKT_Research-er_inbound.
8. Robles M, Namdarian N, Otto J, Wassiljew E, Navab N, Falter-Wagner C, et al. A virtual reality based system for the screening and classification of autism. IEEE Trans Vis Comput Graph. 2022;28:2168–78.
9. Redcay E, Dodell-Feder D, Mavros PL, Kleiner M, Pearrow MJ, Triantafyllou C, et al. Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. 2013;34:2511–23. https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.22086.
10. Küpper C, Stroth S, Wolff N, Hauck F, Kliewer N, Schad-Hansjosten T, et al. Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. Sci Rep. 2020;10:1–11.
11. Kosmicki JA, Sochat V, Duda M, Wall DP. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Transl Psychiatry. 2015;5:e514–e514.
12. Morrison KE, DeBrabander KM, Jones DR, Faso DJ, Ackerman RA, Sasson NJ. Outcomes of real-world social interaction for autistic adults paired with autistic compared to typically developing partners. Autism. 2019;1362361319892701. https://doi.org/10.1177/1362361319892701.
13. de Belen RAJ, Bednarz T, Sowmya A, Del Favero D. Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019. Transl Psychiatry. 2020;10:333.
14. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Applying machine learning to kinematic and eye movement features of a movement imitation task to predict autism diagnosis. Sci Rep. 2020;10:1–13.
15. Kojovic N, Natraj S, Mohanty SP, Maillart T, Schaer M. Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. Sci Rep. 2021;11:1–10.
16. Koehne S, Hatri A, Cacioppo JT, Dziobek I. Perceived interpersonal synchrony increases empathy: Insights from autism spectrum disorder. Cognition. 2016;146:8–15. https://doi.org/10.1016/j.cognition.2015.09.007.
17. Hove MJ, Risen JL. It's all in the timing: Interpersonal synchrony increases affiliation. Soc Cogn. 2009;27:949–60.
18. Miles LK, Nind LK, Macrae CN. The rhythm of rapport: Interpersonal synchrony and social perception. J Exp Soc Psychol. 2009;45:585–9. May
19. McNaughton KA, Redcay E. Interpersonal synchrony in autism. Curr Psychiatry Rep. 2020;22:12 https://doi.org/10.1007/s11920-020-1135-8.
20. Zampella CJ, Bennetto L, Herrington JD. Computer vision analysis of reduced interpersonal affect coordination in youth with autism spectrum disorder. Autism Res. 2020;13:2133–42.
21. Koehler JC, Georgescu AL, Weiske J, Spangemacher M, Burghof L, Falkai P, et al. Brief report: specificity of interpersonal synchrony deficits to autism spectrum disorder and its potential for digitally assisted diagnostics. J Autism Dev Disord. 2021. https://doi.org/10.1007/s10803-021-05194-3.
22. Georgescu AL, Koeroglu S, De AF, Vogeley K, Falter-Wagner CM, Tschacher W. Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: a motion energy study. Mol Autism. 2020;11:1–14.
23. McAuliffe D, Pillai AS, Tiedemann A, Mostofsky SH, Ewen JB. Dyspraxia in ASD: impaired coordination of movement elements. Autism Res. 2017;10:648–52.
24. Georgescu AL, Koehler JC, Weiske J, Vogeley K, Koutsouleris N, Falter-Wagner C. Machine learning to study social interaction difficulties in ASD. Front Robot AI. 2019;6:1–7.
25. de Marchena A, Eigsti IM. Conversational gestures in autism spectrum disorders: asynchrony but not decreased frequency. Autism Res. 2010;3:311–22. 2010/12/25
26. Stagg SD, Slavny R, Hand C, Cardoso A, Smith P. Does facial expressivity count? How typically developing children respond initially to children with autism. Autism. 2014;18:704–11.
27. Cook JL, Blakemore SJ, Press C. Atypical basic movement kinematics in autism spectrum conditions. Brain. 2013;136:2816–24.
28. Wang LAL, Petrulla V, Zampella CJ, Waller R, Schultz RT. Gross motor impairment and its relation to social skills in autism spectrum disorder: a systematic review and two meta-analyses. Psychol Bull. 2022;148:273–300.

29. Redcay E, Schilbach L. Using second-person neuroscience to elucidate the mechanisms of social interaction. Nat Rev Neurosci. 2019;20:495–505. https://doi.org/10.1038/s41583-019-0179-4.
30. Schuwerk T, Vuori M, Sodian B. Implicit and explicit Theory of Mind reasoning in autism spectrum disorders: the impact of experience. Autism. 2015;19:459–68.
31. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: experiments in behavior made easy. Behav Res Methods. 2019;51:195–203.
32. Ramseyer FT. Motion energy analysis (MEA): A primer on the assessment of motion from video. J Couns Psychol. 2020;67:536–49.
33. Tschacher W, Rees GM, Ramseyer F. Nonverbal synchrony and affect in dyadic interactions. 2014;5. https://doi.org/10.3389/fpsyg.2014.01323.
34. Nadig A, Lee I, Singh L, Bosshart K, Ozonoff S. How does the topic of conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. Neuropsychologia. 2010;48:2730–9.
35. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. J Autism Dev Disord. 2001;31:5–17.
36. Paulus C. DER SAARBRÜCKER PERSÖNLICHKEITSFRAGEBOGEN SPF(IRI) ZUR MESSUNG VON EMPATHIE: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. https://doi.org/10.23668/psycharchives.9249.
37. Davis MH. A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology. 1980;10:85.
38. Bagby RM, Parker JDA, Taylor GJ. The twenty-item Toronto Alexithymia scale. Item selection and cross-validation of the factor structure. J Psychosom Res. 1994;38:23–32. https://doi.org/10.1016/0022-3999(94)90005-1.
39. Hautzinger M, Bailer M, Worall H, Keller F. BECK-DEPRESSIONS-INVENTAR - Beck Depression Inventory (BDI; Beck AT, Ward CH, Mendelson M, MockJ, Erbaugh J, 1961) - German version [Internet]. Prof. Dr. Martin Hautzinger bei PsychAuthors: https://www.psychauthors.de/psychauthors/index.php?wahl forschung&uwahl psychauthors&uuwahl p00492MH 1994.
40. Graf A. Eine deutschsprachige Version der Self-Monitoring-Skala. Z fur Arb- und Organisationspsychologie. 2004;48:109–21.
41. Kirby A, Edwards L, Sugden D, Rosenblum S. The development and standardization of the Adult Developmental Co-ordination Disorders/Dyspraxia Checklist (ADC). Res Dev Disabil. 2010;31:131–9. http://www.sciencedirect.com/science/article/pii/S0891422209001437.
42. Lehrl S, Triebig G, Fischer BANS. Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. Acta Neurologica Scandinavica. 1995;91:335–345.
43. Weiß RH. CFT 20-R: grundintelligenztest skala 2-revision. Hogrefe; 2006. Göttingen.
44. Drimalla H, Baskow I, Behnia B, Roepke S, Dziobek I. Imitation and recognition of facial emotions in autism: a computer vision approach. Molecular autism. 2021;12:1–15.
45. Baltrušaitis T, Zadeh A, Lim YC, Morency LP. Openface 2.0: facial behavior analysis toolkit. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). Xi'an, China: IEEE; 2018. pp. 59–66.
46. Schoenherr D, Paulick J, Worrack S, Strauss BM, Rubel JA, Schwartz B, et al. Quantification of nonverbal synchrony using linear time series analysis methods: Lack of convergent validity and evidence for facets of synchrony. Behav Res Methods. 2019;51:361–83.
47. Wolpert DH. Stacked generalization. Neural networks. 1992;5:241–259.
48. Koutsouleris, Vetter & Wiegand (2023). Neurominer [Computer software]. Retrieved from https://github.com/neurominer-git/NeuroMiner_1.2.
49. The MathWorks Inc. MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com.
50. Golland P, Fischl B. Permutation Tests for Classification: Towards Statistical Significance in Image-Based Studies. In: Taylor C, Noble JA, editors. Information Processing in Medical Imaging. IPMI 2003. Lecture Notes in Computer Science, vol 2732. Berlin, Heidelberg: Springer.
51. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. Nat Hum Behav. 2018;2:6–10.
52. Lakens D, Adolfi FG, Albers CJ, Anvari F, Apps MAJ, Argamon SE, et al. Justify your alpha. Nat Hum Behav. 2018;2:168–71.
53. Wolpert DH. Stacked generalization. Neural Netw. 1992;5:241–59. https://www.sciencedirect.com/science/article/pii/S0893608005800231.
54. Yoshimura S, Sato W, Uono S, Toichi M. Impaired overt facial mimicry in response to dynamic facial expressions in high-functioning autism spectrum disorders. J Autism Dev Disord. 2015;45:1318–28.
55. McIntosh DN, Reichmann-Decker A, Winkielman P, Wilbarger JL. When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. Dev Sci. 2006;9:295–302. 2006/05/04
56. Drimalla H, Baskow I, Behnia B, Roepke S, Dziobek I. Imitation and recognition of facial emotions in autism: a computer vision approach. Mol Autism. 2021;12:1–15.
57. Zhao Z, Zhu Z, Zhang X, Tang H, Xing J, Hu X, et al. Atypical Head Movement during face-to-face interaction in children with autism spectrum disorder. Autism Res. 2021;1–12.
58. Bloch C, Vogeley K, Georgescu AL, Falter-Wagner CM. INTRApersonal synchrony as constituent of INTERpersonal synchrony and its relevance for autism spectrum disorder. Front Robot AI. 2019;6:1–8.
59. Sagi O, Rokach L. Ensemble learning: A survey. WIREs Data Mining Knowl Discov. 2018;8:e1249.
60. Liao M, Duan H, Wang G. Application of machine learning techniques to detect the children with autism spectrum disorder. J Healthc Eng. 2022;2022:9340027.
61. Dwyer D, Krishnadas R. Five points to consider when reading a translational machine-learning paper. Br J Psychiatry. 2022;220:169–71.
62. Leo M, Carcagnì P, Distante C, Spagnolo P, Mazzeo PL, Rosato AC, et al. Computational assessment of facial expression production in ASD children. Sensors. 2018;18:1–25.
63. Koehler JC, Falter-Wagner CM. Digitally assisted diagnostics of autism spectrum disorder. Front Psychiatry. 2023;14,1066284.
64. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges. BMJ. 2016;353:i3140, https://doi.org/10.1136/bmj.i3140.
65. Milton DEM. On the ontological status of autism: the 'double empathy problem'. Disabil Soc. 2012;27:883–7. https://doi.org/10.1080/09687599.2012.710008.
66. Bolis D, Balsters J, Wenderoth N, Becchio C, Schilbach L. Beyond autism: introducing the dialectical misattunement hypothesis and a bayesian account of intersubjectivity. Psychopathology. 2018;50:355–72.
67. Baltrušaitis T, Robinson P, Morency LP. OpenFace: an open source facial behavior analysis toolkit. 2016. https://www.omron.com/ecb/products/mobile/.
68. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol. 2018;14:91–118. 2018/02/06.
69. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. Ser B (Methodological). 1995;57:289–300.

ACKNOWLEDGEMENTS

We would like to extend our gratitude to all participants as well as individuals from the autistic community who gave feedback on study design and interpretation during data collection and an outreach event. We would also like to acknowledge Mariia Seleznova for providing valuable advice on ML strategies.

AUTHOR CONTRIBUTIONS

CFW conceptualized the research program. JCK conceived and programmed the study design. JCK, AB and SF collected the data. JCK and JS pre-processed the data. JCK, MD, ISP, CFW, and NK conceptualized the analysis. MD provided the cross-validation pipeline scripts. JCK analyzed the data and wrote the manuscript. MD, ISP, CFW, and NK provided feedback and consulted on day-to-day issues with JCK. CFW and NK provided the resources. All authors have approved the submitted version.

FUNDING

This work was supported by Stiftung Irene (PhD scholarship awarded to JCK) and the German Research Council (grant numbers 876/3-1 and FA 876/5-1 awarded to CFW). Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-024-02802-5.

Correspondence and requests for materials should be addressed to Jana Christina Koehler or Christine M. Falter-Wagner.

3 Original Study II – Automated extraction of speech and turn-taking parameters in autism allows for diagnostic classification using a multivariable prediction model

Contributions and reference

In this study, objective verbal features from conversational dyads consisting of autistic and non-autistic participants were automatically extracted and used to build a diagnostic classification model. As the dataset featured in this study is a subset of the sample in Study I, all statements in 2.1 regarding project administration, experimental paradigm, recruitment and data collection, and data curation apply. I supervised a master's thesis exploring objective speech analysis approaches based on a preliminary data set. I was involved in data preprocessing and final machine learning analysis in NeuroMiner. I was involved in the literature search, gave extensive feedback on the manuscript drafts and was involved in revisions for the final published article.

Plank, I. S., **Koehler, J. C.**, Nelson, A. M., Koutsouleris, N., & Falter-Wagner, C. M. (2023). **Automated extraction of speech and turn-taking parameters in autism allows for diagnostic classification using a multivariable prediction model.** *Frontiers in Psychiatry*, 14. https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1257569

Note: For better readability, the supplementary material of this publication can be accessed online.

Journal Citation Report:
Impact Factor 2022: 4.7
5 Year Impact Factor 2022: 5.0
Ranked 76/264 in Psychiatry Journals (Journal Citation Indicator 2022)

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 06 November 2023
DOI 10.3389/fpsyt.2023.1257569

OPEN ACCESS

EDITED BY
Lawrence Fung,
Stanford University, United States

REVIEWED BY
Philippine Geelhand,
Université libre de Bruxelles, Belgium
Kenji J. Tsuchiya,
Hamamatsu University School of Medicine, Japan

*CORRESPONDENCE
I. S. Plank
irene.plank@med.uni-muenchen.de

RECEIVED 21 July 2023
ACCEPTED 20 October 2023
PUBLISHED 06 November 2023

CITATION
Plank IS, Koehler JC, Nelson AM, Koutsouleris N and Falter-Wagner CM (2023) Automated extraction of speech and turn-taking parameters in autism allows for diagnostic classification using a multivariable prediction model.
Front. Psychiatry 14:1257569.
doi: 10.3389/fpsyt.2023.1257569

Automated extraction of speech and turn-taking parameters in autism allows for diagnostic classification using a multivariable prediction model

I. S. Plank[1*], J. C. Koehler[1], A. M. Nelson[1], N. Koutsouleris[1,2,3] and C. M. Falter-Wagner[1]

[1]Department of Psychiatry and Psychotherapy, LMU University Hospital, LMU Munich, Munich, Germany, [2]Max Planck Institute of Psychiatry, Munich, Germany, [3]Institute of Psychiatry, Psychology and Neuroscience, King's College, London, United Kingdom

Autism spectrum disorder (ASD) is diagnosed on the basis of speech and communication differences, amongst other symptoms. Since conversations are essential for building connections with others, it is important to understand the exact nature of differences between autistic and non-autistic verbal behaviour and evaluate the potential of these differences for diagnostics. In this study, we recorded dyadic conversations and used automated extraction of speech and interactional turn-taking features of 54 non-autistic and 26 autistic participants. The extracted speech and turn-taking parameters showed high potential as a diagnostic marker. A linear support vector machine was able to predict the dyad type with 76.2% balanced accuracy (sensitivity: 73.8%, specificity: 78.6%), suggesting that digitally assisted diagnostics could significantly enhance the current clinical diagnostic process due to their objectivity and scalability. In group comparisons on the individual and dyadic level, we found that autistic interaction partners talked slower and in a more monotonous manner than non-autistic interaction partners and that mixed dyads consisting of an autistic and a non-autistic participant had increased periods of silence, and the intensity, i.e. loudness, of their speech was more synchronous.

KEYWORDS
speech, turn-taking, diagnostic classification, autism, prediction model, conversation

1. Introduction

Speech as a form of communication is unique to humans. According to Ferdinand de Saussure, it is based on signs combining acoustic forms (the signifier) with meaning (the signified) (1). All signifiers can vary in their production to add contextual meaning to their signified. Important speech features like pitch, referring to the tone of speech, intensity, referring to the volume of speech, and articulation rate, referring to the speed of speech, are all influenced by the affective and mental state of the speaker (2, 3). Therefore, they strongly influence how a certain utterance is perceived: Meaning is not only *what* we say, but *how* we say it.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that entails symptoms regarding communication, social behaviour and behavioural rigidity (4). Speech can be completely absent in autistic people. Even in verbal individuals, speech of autistic people differs to that of non-autistic people (5, 6). One of the most common diagnostic instruments for

ASD, the Autism Diagnostic Observation Schedule [ADOS(R); (7)], highlights that both changes in prosody and speech rate may indicate ASD, amongst other verbal behaviours.

A recent meta-analysis evaluated the alterations of speech features in ASD (5). The authors found that pitch differs between autistic and non-autistic people in terms of increased mean and variance. However, results concerning intensity and speech rate were more equivocal. For both domains, many included studies did not show differences between autistic and non-autistic people, while other studies found effects, though not all of them in the same direction. The meta-analysis did not include studies that investigated variance of intensity over the course of a conversation. In another systematic review (8), two studies investigating variance of intensity were mentioned, one of which did not find differences in intensity range (9), and the other found decreased standard deviation of intensity (10). It is important to note that both Fusaroli et al. (8) and Asghari et al. (5) included various modes of speech production, ranging from spontaneous production over narration to social interactions. Additionally, both included all age ranges, so it is possible that not all outcomes apply to adults.

In addition to the importance of speech differences, autistic people report having difficulties with small talk and are perceived as more awkward in conversations (11–14). Since small talk and conversations with strangers are essential for building connections with others, it is important to understand how autistic verbal behaviours differ from non-autistic verbal behaviours in these situations. Reciprocal communication is characterised by a to and fro of speaking and listening. Successful turn-taking not only requires mutual prediction of an upcoming transition point but also a minute concertation of behaviours between interaction partners allowing them to be in sync (15). The length of turn-taking gaps can be an estimate of how in sync interaction partners were and is associated with social connection (16). If two strangers lose their flow, they tend to feel awkward and try to fill the silence (17). A recent study by Ochi et al. (18) found increased turn-taking gaps and more silence vs. talking as measured by the silence-to-turn ratio (19). However, the sample consisted of only male autistic and non-autistic participants, and it is unclear whether the results generalise to people of other genders. Therefore, it is especially important to investigate turn structure in a more general sample to assess the quality of verbal communication.

Finally, the investigation of speech features should be extended to include the temporal fine-tuning within interaction dyads, given the increasing literature showing reduced interactional synchrony in dyads of one autistic and one non-autistic compared to two non-autistic interaction partners [e.g. (20); for a review, see (21)]. Behavioural synchrony is the product of coordination between interaction partners. This coordination can be achieved by the interaction partners adapting their behaviour to each other. Synchrony of speech features is well documented (22–25); however, research investigating speech synchrony in autistic people is scarce. Ochi et al. (18) found that non-autistic participants showed more synchrony between the ADOS interviewer's intensity and their own than autistic participants, but they found no differences regarding synchrony of pitch. Wynn et al. (26) altered the speed in trial prompts and found that non-autistic adults adapted the speed of their answer in the corresponding trial, while autistic adults and children did not. Both studies show that interpersonal coordination of speech features is a promising avenue to investigate differences in verbal interaction between autistic and non-autistic people.

Additionally, a recent study also used parts of ADOS interviews to investigate classification between autistic and non-autistic children based on synchrony of speech features (27). They extracted lexical features and calculated the similarity of the lexical content of the interviews. Machine learning classifiers were able to predict whether a child was diagnosed with ASD with better accuracy when the synchrony measures were added to the model as compared to a model that only included individual speech features. However, in that study, the ADOS was used both for creating the true labels and to extract features for the classification, risking circularity that might artificially inflate accuracies. Therefore, it is vital to assess the performance of classifiers with features extracted from data that is independent from the diagnostic process. In a recent study using automatically extracted interpersonal synchrony of motion quantity and facial expressions, we show that pursuing more naturalistic study designs can yield high classification accuracy of almost 80% (28). If these results can be extended to speech and interactional features of verbal communication in adults, this would provide a low-tech and scalable route to assist clinicians with the diagnosis of ASD.

This study design fills the outlined gaps in the literature by extracting speech parameters with an automated pipeline from naturalistic conversations that are independent of the diagnostic assessment to avoid any circularity in the classification procedure. The automated extraction of features increases objectivity, specificity and applicability of the pipeline to a variety of conversational paradigms. The main aim of the current study was (i) to determine the potential of speech coordination as a diagnostic marker for ASD. Additionally, we defined two secondary aims: (ii) to describe individual speech feature differences, and (iii) interactional speech differences that can help explain the classification power. Concerning our main aim (i), we expected that a multivariable prediction model would be able to classify dyad type based on individual speech and dyadic conversational features, thereby offering an exciting possibility for assisting diagnostics of ASD. On the individual level regarding our aim (ii), we expected that autistic and non-autistic individuals would differ in their pitch variance, intensity variance and articulation rate. Additionally, we computed turn-based adaptation of pitch, intensity and articulation rate and expected increased turn-based adaptation in non-autistic compared to autistic individuals. On the dyadic level regarding our aim (iii), we hypothesised that interactional differences would be found in silence-to-turn ratios, turn-taking gaps as well as time-course synchrony of pitch and intensity.

2. Materials and methods

This study is part of a larger project to find diagnostic markers for ASD. The preregistration of the hypotheses regarding aim (ii) and (iii) can be retrieved from OSF.[1] Preprocessing was performed using Praat *6.2.09* (29), the uhm-o-meter scripts provided by De Jong et al. (30, 31) and R *4.2.2* (32) in Rstudio *2022.12.0* (33). The Bayesian analysis was performed in R and JASP *0.16.4* (34). The machine learning analysis was conducted with the NeuroMiner toolbox *1.1* (35)

1 https://osf.io/jhetr

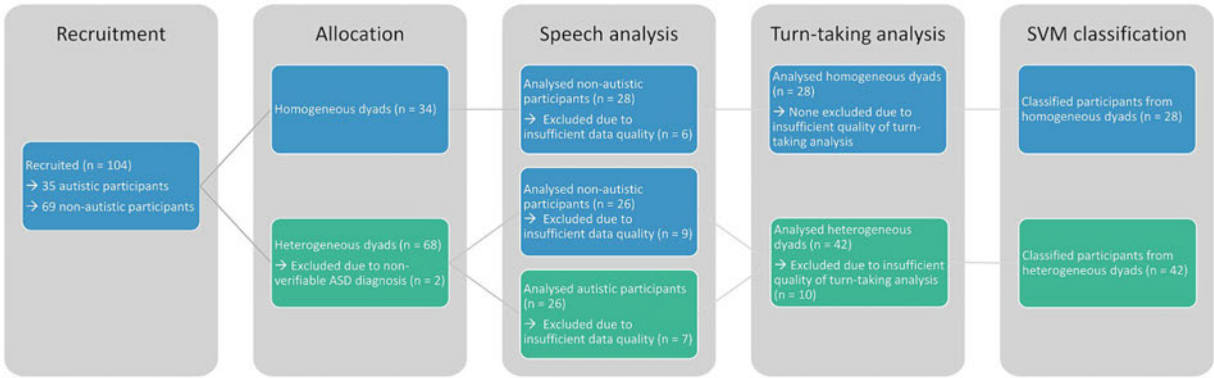

FIGURE 1
This consort chart shows the recruitment and exclusion of participants. All sample sizes are given per participant and not per dyad. Colours indicate the group affiliation at the respective analysis step.

implemented in MATLAB *R2022b* (36) and Python *3.9*.[2] All code used to preprocess and analyse the data can be found on GitHub.[3] We report our prediction model following the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines (37).

2.1. Participants

We recruited 35 autistic and 69 non-autistic participants from the general population and the outpatient clinic at the LMU University Hospital Munich by posting flyers at the university and at the hospital as well as distributing them online on social media and mailing lists. Of these participants, 26 autistic (*mean* age = 34.85 ± 12.01 years, 17 male) and 54 non-autistic (*mean* age = 30.80 ± 10.42 years, 21 male) participants were analysed (Figure 1). Non-autistic participants were recruited to match the overall gender and age distribution of the autistic sample. This sample is a subset of the sample analysed by Koehler et al. (28) containing all participants with sufficient audio data quality.

All participants were between 18 and 60 years old, had no current neurological disorder and had an IQ above 70 based on verbal and non-verbal IQ tests (38, 39). For each autistic participant, an ASD diagnosis (F84.0 or F84.5) according to the ICD-10 (40) was confirmed by evaluating the diagnostic report. All non-autistic participants had no current or previous psychiatric diagnosis and no intake of psychotropic medication. Autistic and non-autistic participants did not differ credibly in age, verbal [measured with the Mehrfachwahl-Wortschatz-Intelligenztest, MWT-B; (38)] or nonverbal IQ [measured with the Culture Fair Intelligence Test, CFT-20-R; (39)], but they differed credibly on the Adult Dyspraxia Checklist [ADC; (41)], the Autism Quotient [AQ; (42)], the Beck's Depression Inventory [BDI; (43)], Self-Monitoring Scale [SMS-short; (44)], the Saarbrückener Persönlichkeitsfragebogen [SPF; (45), German version of the Interpersonal Reactivity Index, IRI, (46)] and the Toronto Alexithymia Scale [TAS-20; (47); see Table 1]. Two autistic participants had a comorbid diagnosis of attention deficit hyperactivity disorder (ADHD), nine of an affective disorder and five of a neurotic stress-related or somatoform disorder. The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the medical faculty of the LMU. All participants provided written, informed consent and received a monetary compensation for their participation.

2.2. Experimental procedure

After giving informed consent, blood samples were taken, followed by demographics and the intelligence assessments. Throughout the session, participants completed the above listed questionnaires. They also performed a task assessing emotion recognition [BERT, (48)]. In addition, some of the participants took part in a separate study measuring endocrinology and effects of social ostracism.

We paired participants in either mixed dyads consisting of one autistic and one non-autistic participant or non-autistic dyads. Participants were paired based on availability regardless of age and gender. Dyads did not differ in average age or age difference between the interaction partners. However, there was strong evidence in favour of a difference in gender composition (mixed dyads: *mean* age = 33.15 ± 7.72, *mean* age difference = 12.69 ± 9.18 [1 to 32 years], 15% female, 35% male and 50% gender-mixed dyads; non-autistic dyads: *mean* age = 30.18 ± 8.22, *mean* age difference = 10.64 ± 11.15 [1 to 31 years], 50% female and 50% gender-mixed; for statistical values see Supplementary material S1.1). We did not disclose their interaction partner's diagnostic status to them. The dyads engaged in two 10-minute long conversations: one about their hobbies and one fun task in which they were asked to plan a menu consisting of food and drinks that they both disliked (49). On the one hand, we chose the *hobbies* task because special interests are a core symptom of ASD (4). On the other hand, the *meal planning* task facilitates a more collaborative interaction and has been shown to promote increased synchrony in non-autistic dyads (20). The experimenter left the room during the conversations. After both conversation tasks, participants were asked to rate the quality of their interactions. During the COVID-19 pandemic, testing had to be moved to a different room after nine dyads and a plexiglass was placed between the participants

2 https://github.com/neurominer-git/NeuroMiner_1.1
3 https://github.com/IreneSophia/MLSPE

TABLE 1 Mean and standard deviation of the autistic and non-autistic samples analysed in this study as well as group comparisons performed with Bayesian Mann–Whitney U tests based on 10,000 samples.

	Autistic	Non-autistic	Log(BF_{10})	W
Age	34.85 ± 12.01	30.80 ± 10.42	−1.028	564.50
IQ – nonverbal	115.35 ± 22.96	117.07 ± 15.21	−1.384	716.50
IQ – verbal	112.12 ± 15.01	113.96 ± 16.53	−1.297	743.00
ADC	50.12 ± 16.06	15.57 ± 8.99	9.339	61.50
AQ	33.00 ± 8.41	14.26 ± 4.55	8.863	65.50
BDI	18.35 ± 12.36	3.94 ± 3.96	8.137	139.00
SMS-short	5.92 ± 2.86	9.54 ± 2.98	5.080	1,136.00
SPF	36.77 ± 6.62	45.43 ± 5.35	6.498	1,214.00
TAS-20	61.27 ± 11.74	36.91 ± 7.61	8.359	77.00

Note. ADC, Adult Dyspraxia Scale; AQ, Autism Quotient; BDI, Beck's Depression Inventory; SMS-short, Self-Monitoring Scale; SPF, Saarbrückener Persönlichkeitsfragebogen; TAS-20, Toronto Alexithymia Scale.

as a health and safety measure. Participants did not wear masks during the conversations and the quality of interactions was rated equal before and after the measures had been put into place (28).

We captured participants' behaviour via multiple channels. The current study focuses on speech coordination captured with one recording device to which two separate microphones were connected (t.Bone earmic 500 with ZoomH4n recorder). The nonverbal communication parameters, body movement captured by a scene camera (Logitech C922), facial expressions captured by two face cameras (Logitech C922), heart rate and electrodermal activity captured by wearables (Empatica E4), as well as the analysis of the blood samples were outside of the scope of the current analysis and published elsewhere (28). For more details on the data collection procedure, please consult Supplementary material S1.2.

2.3. Preprocessing

We extracted individual phonetic features for each task and participant using praat (29) (for more details, see Supplementary material S1.3). We calculated pitch and intensity synchrony with rMEA's cross-correlation function to calculate windowed cross-lagged correlations (WCLC) using the same window length of 16 s, step size of 8 s and lag of 2 s as Ochi et al. (18). We used the uhm-o-meter (30, 31) to extract turns from conversations, with a turn defined as all speaking instances of one interactant until the end of the speaking instance preceding the next speaking instance of someone else (see Figure 2). For each turn, we calculated turn-taking gap, average pitch, average intensity and number of syllables to calculate articulation rate. Additionally, we used turn-based information to calculate how much each participant adapted their pitch, intensity and articulation rate to the pitch, intensity and articulation rate of the previous turn.

2.4. Comparison of synchrony with pseudosynchrony

We used segment shuffling as described by Moulder et al. (50) to determine whether synchrony and turn-based adaptation calculations are credibly different from their corresponding pseudo values (see also Supplementary material S3). These pseudo values are created by randomly shuffling one of the interactant's data and then computing synchrony between the shuffled and real data. For each synchrony and turn-based adaptation value, we computed the average of 100 pseudosynchrony or pseudoadaptation values. Then, we used a Bayesian paired t-test as implemented in the BayesFactor package to compare the values. There was evidence in favour of the hypotheses that pitch and intensity synchrony as well as turn-based adaptation of pitch, intensity and articulation rate were all credibly higher than the corresponding pseudo values (see Table 2). This indicates that the obtained synchrony values exceeded chance coordination.

2.5. Support vector machine for classification

We used a linear L2-regularised L2-loss support vector machine (SVM) as implemented by LIBLINEAR in NeuroMiner to predict each individual's participation in either a non-autistic or mixed dyad to address our main aim (i). SVMs have not only been applied to classify several psychiatric diagnoses (51, 52), but have been specifically applied to predict ASD based on interactional data (28, 53, 54). Therefore, we chose to use an SVM to allow for comparability with previous results and decided on a linear SVM for its computational speed (55). We combined the SVM with a L2 or ridge regression since this seems to perform better with correlated predictors (56). The SVM algorithm optimises a linear hyperplane to achieve maximum separability between non-autistic and mixed dyads in the high-dimensional feature space. Separability was assessed using balanced accuracy, which equally weighs sensitivity (ratio of true positives to the sum of the true positives and false negatives) and specificity (ratio of true negatives to the sum of the true negatives and false positives). We used hyperplane weighting to account for unbalanced sample sizes where the misclassification penalty for the smaller sample is increased (35). The algorithm optimises the hyperplane so that the geometric margin between most similar instances of opposite classes (i.e. the support vectors) is maximised, thus increasing generalisability to new observations following the principles of statistical learning theory (57–59). The SVM algorithm determines dyad membership by the position of an individual with respect to the optimally separating hyperplane (OSH), while the individual's decision score measures the geometric distance to the OSH with higher absolute values indicating a more pronounced expression of the given separating pattern. This decision score then determines the label assigned to an individual, specifically if the individual was part of a non-autistic or mixed dyad. All features (see Table 3) were scaled from −1 to 1 and pruned to exclude zero variance features as recommended by the NeuroMiner manual. We used a repeated, nested stratified cross-validation (CV) structure to account for the unbalanced sample sizes and the dyadic nature of the data, ensuring that two interactants of the same dyad were always in the same fold and that the ratio of interactants from mixed and non-autistic dyads was consistent in all folds. The CV structure consisted of two loops with the outer loop being implemented in seven folds and 10 permutations and the inner loop with 10-fold and one permutation. The outer loop iteratively held back five dyads to validate the algorithm on unseen data, while the rest of the dyads was included

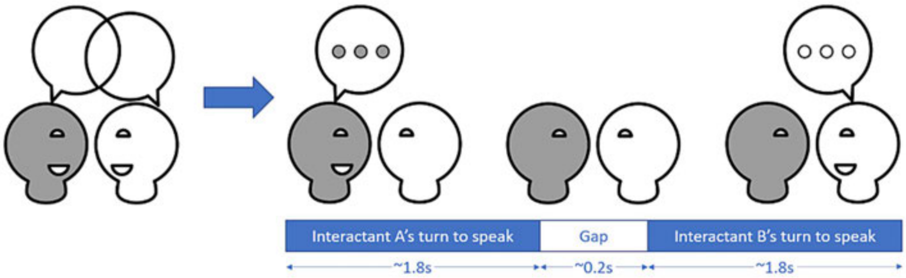

FIGURE 2
Conversations can be broken down into turns where one of the interactant is speaking and gaps between the turns. In a large-scale study, Templeton et al. (16) found that turns in an unstructured conversation between strangers have a median length of 1.8 s, while the median length of gaps was about 0.2 s.

TABLE 2 Comparison of synchrony and turn-based adaptation values with their corresponding pseudo values.

	Mean and SD of values	Mean and SD of pseudo values	$Log(BF_{10})$
Individual adaptation			
Turn-based pitch	0.121 ± 0.107	0.087 ± 0.020	4.561
Turn-based intensity	0.146 ± 0.086	0.091 ± 0.017	20.977
Turn-based articulation rate	0.138 ± 0.096	0.099 ± 0.019	7.017
Dyadic synchrony			
Pitch	0.197 ± 0.022	0.190 ± 0.003	1.423
Intensity	0.368 ± 0.048	0.164 ± 0.010	104.491

TABLE 3 List of individual and dyadic features.

Individual	Dyadic
Articulation rate	Number of turns
Number of pauses	Silence-to-turn ratio
Number of syllables	Speech rate
Phonation time	Synchrony of intensity
Turn-based adaptation of articulation rate	Synchrony of pitch
Turn-based adaptation of intensity	Turn-taking gap
Turn-based adaptation of pitch	
Variance of intensity	
Variance of pitch	

Note. All features were entered for the meal planning and the hobbies task separately, resulting in 30 features. Articulation rate refers to the number of syllables per phonation time, while speech rate refers to the number of syllables per total time (phonation time and silence).

in the inner loop. Here, three dyads were held back for validation. The decision scores of the test data of the inner loop were additionally post-hoc optimised according to the receiver operator function. Last, we used label permutation testing while keeping the cross-validation structure intact to assess whether the resulting SVM performance was above chance (5,000 permutations, $\alpha_{Bonferroni\text{-}corrected} = 0.007$).

2.6. Bayesian analysis

We tested our hypotheses regarding aims (ii) and (iii) using Bayesian repeated-measures ANOVAs as implemented in JASP. Each ANOVA included one within-subjects factor (task: *meal planning*, *hobbies*) and one between-subjects factor, either diagnostic status (autistic, non-autistic) or dyad type (mixed, non-autistic). We checked for equality of variance and visually inspected whether the residuals were normally distributed. In the case of violations of these assumptions, we computed a non-parametric alternative and compared the results. We used the Bayes Factor to assess the strength of evidence for or against a model or inclusion of a factor. The Bayes Factor is the ratio of marginal likelihoods, thereby quantifying how much more or less likely one model is than the other. We interpreted the logarithmic Bayes Factor according to Jeffrey's scheme (60). For example, if a model is more than 100 times as likely [$Log(BF) > \mathrm{Log}(100) = 4.6$], we consider this *decisive* evidence in favour of this model [very strong: $Log(BF) > 3.4$; strong: $Log(BF) > 2.3$; moderate: $Log(BF) > 1.1$; anecdotal: $Log(BF) > 0$]. We use the logarithm of the Bayes Factor because it leads to symmetric thresholds: a $Log(BF)$ of 4 signifies very strong evidence in favour of a model and a $Log(BF)$ of −4 the same strength of evidence against a model.

There was a credible difference between the gender composition of the non-autistic and the mixed dyads due to no non-autistic male dyads. Since studies have shown differences between genders with regard to language in ASD (61–63), we repeated all group comparisons on the dyad level excluding the male mixed dyads to ensure that possible differences are not driven by gender composition.

3. Results

3.1. Performance of support vector machine for classification

Our SVM algorithm was able to distinguish between individuals from a non-autistic and a mixed dyad with 76.2% balanced accuracy on the basis of both individual and dyadic speech and communication features. Specifically, 78.6% of the individuals from a non-autistic dyad were correctly labelled as such (specificity), while 73.8% of the individuals from a mixed dyad were assigned the correct label (sensitivity, see Figure 3). While this model performs significantly above chance levels ($p < 0.001$; area under the curve: 0.81 [CI 0.72–0.92]; please consult Supplementary material S2 for more details on

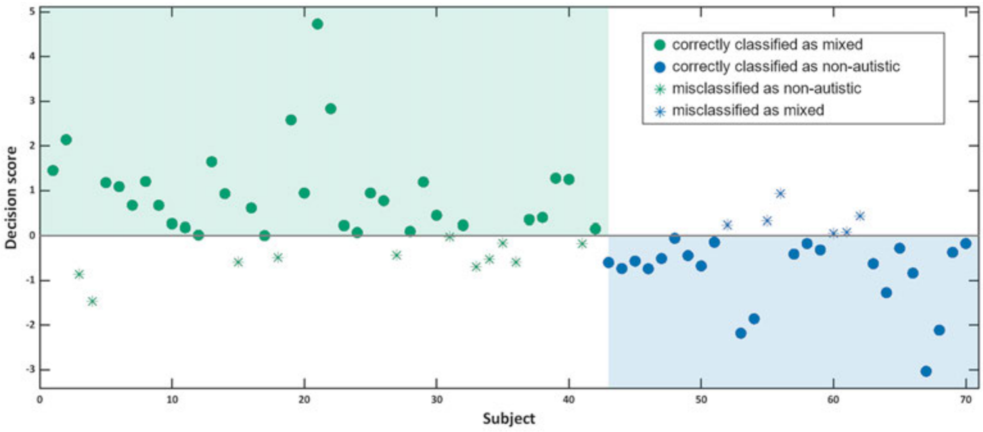

FIGURE 3
This graph shows for each participant the decision score calculated by the SVM classifier with participants from non-autistic dyads in blue and participants from mixed dyads in green. Filled circles were correctly categorised by the classifier, while empty circles were misclassified.

the SVM classifier), it does not outperform a model trained on synchrony of facial expressions with a balanced accuracy of 79.5% or a stacked model with a balanced accuracy of 77.9% including multiple movement parameters automatically extracted from video recordings of dyadic interactions (28).

3.2. Group comparisons on the individual and the dyad level

3.2.1. Individual differences between autistic and non-autistic participants

Autistic participants differed from non-autistic participants in their speech features as evidenced by the results of the Bayesian ANOVAs.

3.2.1.1. Pitch

Pitch variance was best explained by a model including task and diagnostic status but not the interaction of the two [$Log(BF_{10}) = 5.612$]. The analysis of effects across matched models revealed very strong evidence for the inclusion of task and anecdotal evidence for the inclusion of diagnostic status [task: $Log(BF_{incl}) = 4.455$; diagnostic status: $Log(BF_{incl}) = 1.030$]. There was anecdotal evidence against the inclusion of the interaction [task × diagnostic status: $Log(BF_{incl}) = -0.510$]. However, the Q-Q plot of the residuals revealed deviations from the normal distribution and the variances were not homogeneous. Therefore, we computed a Bayesian Mann–Whitney U test to determine whether the anecdotal evidence in favour of an effect of diagnostic status can be reproduced with a non-parametric test, which was the case [$Log(BF_{10}) = 0.888$, $W = 439.00$]. Pitch variance was increased in non-autistic compared to autistic participants.

3.2.1.2. Intensity

The best model describing intensity variance was the full model including the predictors task and diagnostic status as well as their interaction [$Log(BF_{10}) = 3.205$]. The analysis of effects across matched models revealed that this was mainly driven by the interaction with decisive evidence in favour of the interaction effect and moderate and anecdotal evidence against task and diagnostic status, respectively [task × diagnostic status: $Log(BF_{incl}) = 5.163$; task: $Log(BF_{incl}) = -1.544$; diagnostic status: $Log(BF_{incl}) = -0.436$]. Specifically, while intensity variance of autistic participants was increased in the *hobbies* condition, the reverse was true for non-autistic participants (see Figure 4).

3.2.1.3. Articulation rate

Articulation rate was again best described by the full model including task, diagnostic status and the interaction [$Log(BF_{10}) = 6.727$], with moderate evidence in favour of including diagnostic status as well as strong evidence in favour of including task and the interaction [task × diagnostic status: $Log(BF_{incl}) = 2.517$; task: $Log(BF_{incl}) = 2.656$; diagnostic status: $Log(BF_{incl}) = 1.517$]. Articulation rate was faster in non-autistic than autistic participants.

3.2.1.4. Turn-based adaptation

Last, the null model outperformed all alternative models with anecdotal evidence in favour of the null model for turn-based adaptation of pitch and intensity (see Supplementary material S4). In the case of adaptation of articulation rate, there was anecdotal evidence in favour of the model including task but no other predictor [$Log(BF_{10}) = 1.072$] with higher articulation rate in the *meal planning* condition. Since the residuals were not normally distributed, we performed non-parametric tests which confirmed no effect of diagnostic status on all three adaptation parameters (see Supplementary material S4).

3.2.2. Dyadic differences between non-autistic and mixed dyads

Some interactional features differed between non-autistic and mixed dyads; however, others were comparable in both dyad types (see Figure 5).

3.2.2.1. Silence-to-turn ratio

The silence-to-turn ratio was best predicted by the full model including both task and dyad type as well as the interaction [$Log(BF_{10}) = 4.141$]. A closer look at the analysis of effects across matched models revealed strong evidence in favour of the inclusion of task and moderate evidence in favour of the inclusion of the interaction as well as anecdotal evidence against the inclusion of dyad type as a predictor [task × dyad type: $Log(BF_{incl}) = 1.690$; task: $Log(BF_{incl}) = 2.449$; dyad type: $Log(BF_{incl}) = 0.049$]. This seems to

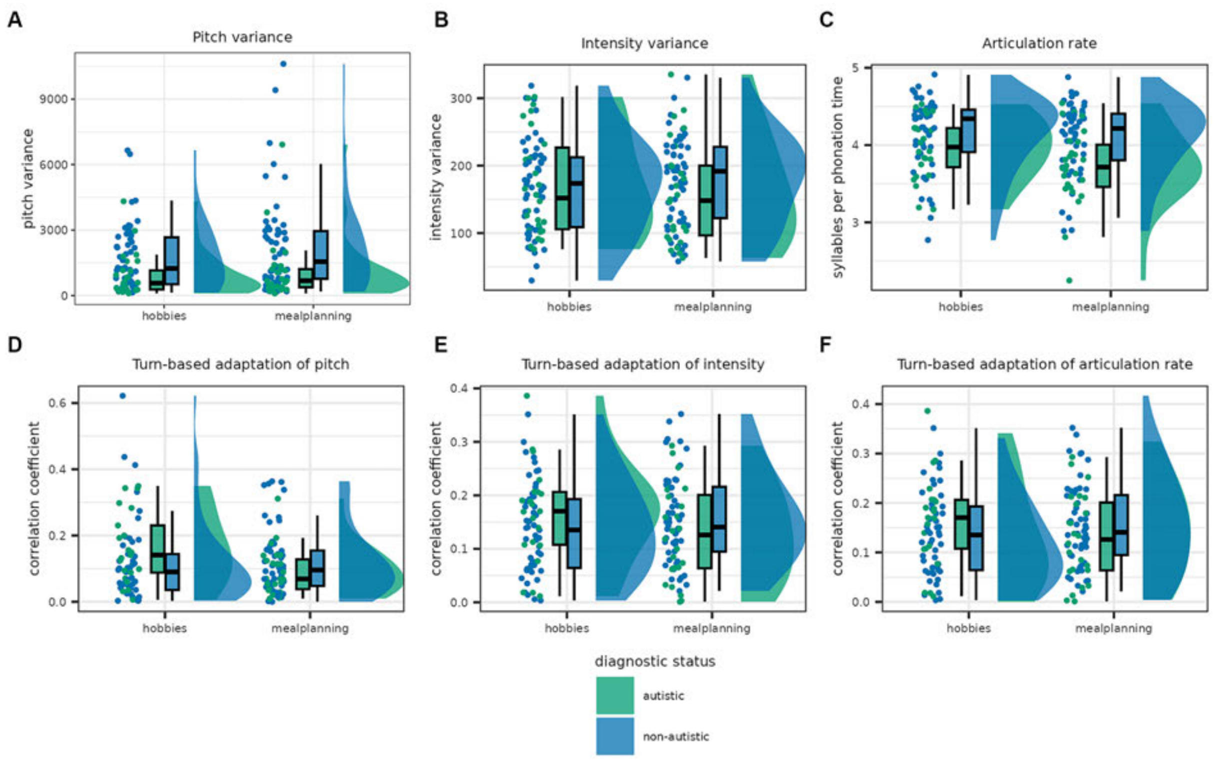

FIGURE 4
This graph shows the distribution of individual features in the autistic and non-autistic participants as scatterplots, density plots and box plots. The boxes show the interquartile range and the median, while the whiskers show 1.5 times the interquartile range added to the third and subtracted from the first quartile. In the first row, panel **(A)** shows pitch variance, panel **(B)** intensity variance and the panel **(C)** articulation rate. All three features were increased in non-autistic compared to autistic participants. The second row shows the amount participants adapted their pitch **(D)**, intensity **(E)** and articulation rate **(F)** to the previous turn. There were no significant differences in adaption of all three speech factors between autistic and non-autistic participants.

be driven by the increased difference between mixed and non-autistic dyads in the *meal planning* condition; although, in both conditions the silence-to-turn ratio was smaller in the case of non-autistic dyads.

3.2.2.2. Turn-taking gap

Turn-taking gap was best explained by the model only including the predictor dyad type for which there was anecdotal evidence [$Log(BF_{10}) = 0.267$]. Similarly, there was anecdotal evidence in favour of including dyad type as well as the interaction of dyad type and task but moderate evidence against including task [task × dyad type: $Log(BF_{incl}) = 0.877$; task: $Log(BF_{incl}) = -1.369$; dyad type: $Log(BF_{incl}) = 0.279$]. Turn-taking gaps tended to be slightly longer in the mixed dyads, especially in the *meal planning* task.

3.2.2.3. Pitch synchrony

Pitch synchrony, as calculated with WCLC, was best explained by the null model, suggesting that interactants of both dyad types adjusted their pitch to a similar extent to each other (see Supplementary material S5).

3.2.2.4. Intensity synchrony

Nonetheless, non-autistic and mixed dyads differed in their WCLC synchrony of intensity with the best model predicting WLCL synchrony of intensity including both task and dyad type but not the interaction [$Log(BF_{10}) = 7.150$]. Indeed, there is anecdotal evidence against the inclusion of the interaction, while there is decisive evidence for the inclusion of task and moderate evidence for the inclusion of dyad type [task × dyad type: $Log(BF_{incl}) = -0.813$; task: $Log(BF_{incl}) = 5.567$; dyad type: $Log(BF_{incl}) = 1.576$]. Mixed dyads adjusted their intensity more strongly, with more synchrony in the *hobbies* condition in both dyad types.

3.2.2.5. Comparison of dyads excluding male dyads

We repeated the analyses of silence-to-turn ratio, turn-taking gap, pitch synchrony and intensity synchrony in a limited sample excluding all male dyads to ensure that the found differences are not driven by differences in gender composition between mixed and non-autistic dyads. For all four parameters, the same model was supported by the evidence as the best model as for the full sample (see Supplementary material S6). Therefore, it is unlikely that the found differences were driven by gender composition.

4. Discussion

Differences in verbal communication are an important symptom of ASD (4, 7). We paired strangers and asked them to have two conversations, one about their hobbies and one where they collaboratively planned a meal with food and drinks that they both dislike. The dyads either consisted of two non-autistic adults or of one

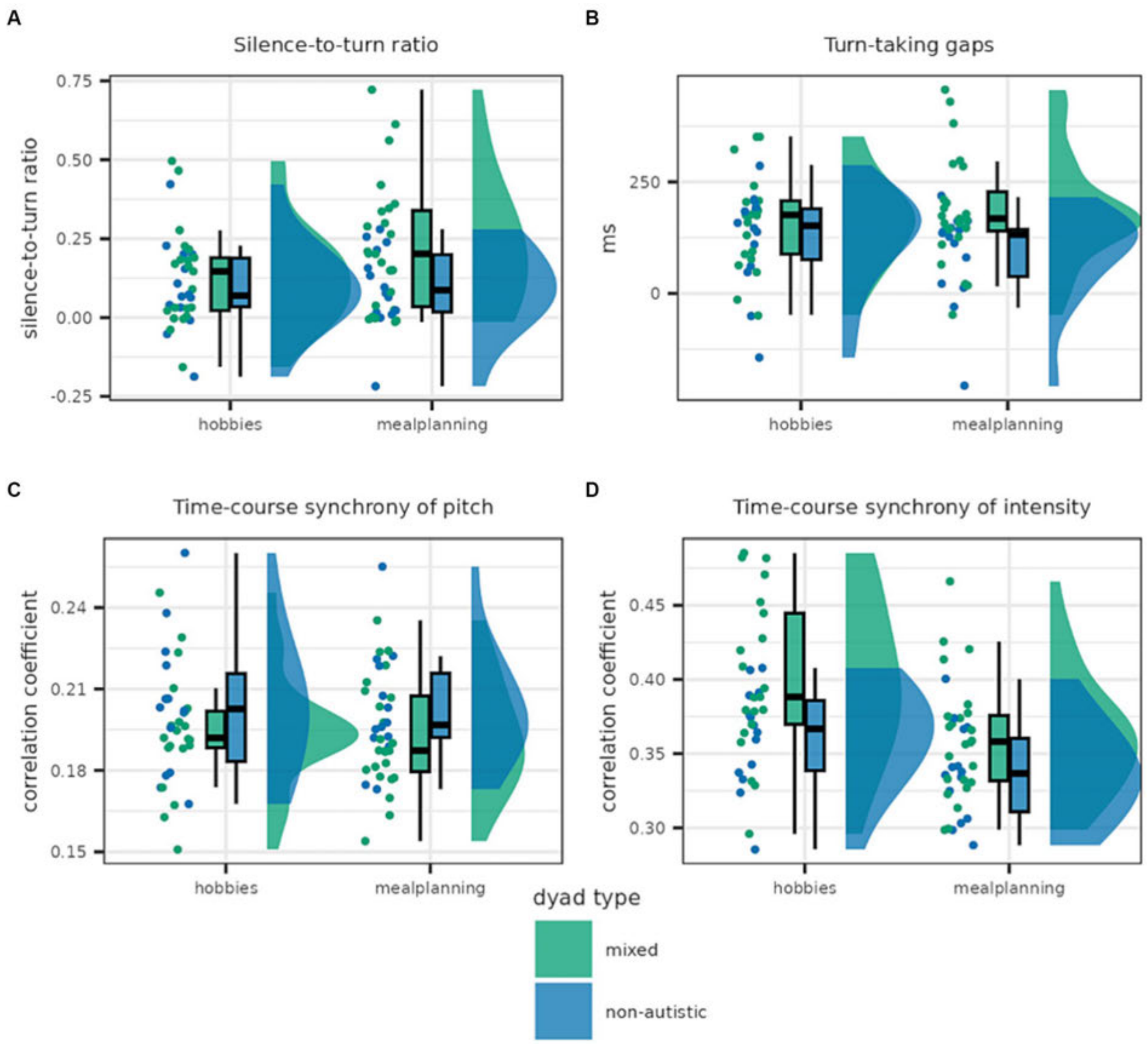

FIGURE 5
This graph shows the distribution of dyadic features for mixed and non-autistic dyads. Panel **(A)** shows the silence-to-turn ratio which was higher in mixed compared to non-autistic dyads. Panel **(B)** shows turn-taking gaps which were, on average, longer in mixed dyads. The lower panels show time-course synchrony of pitch **(C)** and intensity **(D)** with the latter being higher in mixed dyads.

autistic and one non-autistic adult. This study aimed at answering the following research question: (i) what is the potential of speech and interactional features of communication for objective, reliable and scalable classification of ASD? Additionally, we used the extracted features to answer the following questions: (ii) how do autistic and non-autistic people differ with regards to their speech, and (iii) how do interactions between an autistic and a non-autistic person differ from interactions between two non-autistic people?

Regarding our main research question (i), we are able to present a multivariable prediction model that is able to distinguish between mixed and non-autistic dyads with above 75% of balanced accuracy. Automated extraction of speech and interactional features of verbal conversations offer an exciting new avenue for investigating symptoms as well as assisting the diagnosis of ASD. First, automated extraction increases objectivity and replicability while also providing a more detailed and fine-grained perspective on actual speech differences. This fine-grained perspective could in turn inform intervention by focusing on the specific aspects that differ between autistic and non-autistic conversation partners. Additionally, the current diagnostic procedures are time consuming, and recommendations include a combination of semi-structured interviews and neuropsychological assessments (64). This increases psychological stress for the affected person and their families (64, 65). Recent studies have shown that machine learning algorithms based on automatically extracted features could assist in this process (18, 27, 28, 53). Koehler et al. (28) automatically extracted movement parameters from the video recordings of the dyadic interactions analysed here, although in a slightly larger sample. A support vector machine based on the synchrony of facial expressions led to a balanced accuracy of almost 80% and a stacked model of different modalities achieved a balanced accuracy of 77.9%, both outperforming the here-proposed model. However, the extraction from speech and interactional features based on audio recordings offers an especially low-tech and user-friendly data collection procedure that is scalable and economic. As long as the environment is quiet and each of the interaction partners has their own microphone, the proposed preprocessing pipeline is easily applicable. Additionally, this study shows the feasibility of recording free conversations with predefined topics without the need of a semi-structured interview or of a trained conversation partner.

The here-presented and other studies (18, 27, 28, 53) show the potential of developing a multivariable prediction model to assist diagnostics of ASD. However, sample sizes in all of these studies are limited, and while they serve as a proof-of-concept, it is paramount to develop and validate such a multivariable prediction model with significantly larger sample sizes. Such a large-scale study could also compare different machine learning algorithms to ensure optimal performance. Automated extraction of speech and conversation features from audio recordings of people performing the *meal planning* task may be especially fruitful for collecting a large data set, especially if the here-presented effects persist in virtual conversations. Additionally, it is important to note that the autistic adults in our sample are not representative for many autistic adults, for example those with an intellectual disability or those who are non-verbal. This also limits the applicability of any developed prediction model based on speech and interactional features to a subsample of the autistic population. Furthermore, although the non-autistic sample did not differ from the autistic sample in age and gender distribution, a questionnaire measuring autism-like traits indicated that the non-autistic sample was positioned at one end and the autistic sample at the other end of this spectrum. Although this is representative of a non-clinical population, higher autism-like traits are also observed in other clinical populations (66). Future research should include a representative sample of other psychiatric diagnoses than ASD. This can only be achieved by evaluating the performance of the here-presented multivariable prediction model in a large-scale study to ensure its adequate translation to the clinical reality of the diagnostic process.

Concerning research question (ii), our results regarding the speech differences between autistic and non-autistic adults differ from a recent meta-analysis (5). While we found *increased* pitch and intensity variance as well as articulation rate in non-autistic compared to autistic adults, the authors of the meta-analysis report *decreased* pitch variability for non-autistic compared to autistic people as well as no significant differences regarding intensity variability and speech rate. However, the meta-analysis included a wider sample ranging from infants to adults and several modes of speech production including conversations, narration, semi-standardised tests and crying. Focusing on an adult sample and a conversation paradigm, Ochi et al. (18) found a decrease in the standard deviation of intensity in autistic compared to non-autistic adults but no difference in speech rate. Additionally, Kaland et al. (67) also found a decrease in pitch range in autistic compared to non-autistic adults. However, autistic adults seem to show a larger pitch range or variability compared to non-autistic adults in less naturalistic contexts including the narrative subtext of an assessment scale (68), answering questions about pictures (69) and when asked to produce a phrase conveying specific emotions (70). Interestingly, Hubbard et al. (70) used produced emotional phrases to assess whether the emotion is recognisable. They found that while phrases produced by autistic adults were matched with the intended emotion more often, they were also perceived as sounding less natural. Therefore, it is possible that autistic adults exaggerate in artificial contexts more strongly than non-autistic adults, leading to less natural and, most importantly, less representative speech. This would explain the differences in pitch variability between more and less interactive speech paradigms and highlights that speech in monologues and interactive dialogues needs to be distinguished in order to contextualise decreased or increased pitch variability in ASD.

Despite several interventions aiming at improving verbal communication skills and turn-taking (71–73), there is little research on differences in interactional features of conversations including autistic people. In this study, we investigated the ratio of silence to turns as well as the duration of the gaps between turns to investigate research question (iii). We found that mixed dyads had a credibly higher ratio of silence to turns, especially when collaboratively planning a meal. This indicates that the amount they were silent was higher, and they were speaking less. This is in line with the findings by Ochi et al. (18). However, we only found anecdotal evidence in favour of a difference in turn-taking gaps between mixed and non-autistic dyads, while Ochi and colleagues reported a clear effect of credibly longer turn-taking gaps when the ADOS was conducted with autistic adults compared to non-autistic adults. This elongation of turn-taking gaps has also been reported for children taking the ADOS by Bone et al. (74). Additionally, they found that both less speaking time and longer turn-taking gaps correlated with ADOS severity, and that there was a significant difference in the length of turn-taking gaps between children who were diagnosed with ASD and those who were diagnosed with other developmental disorders. This discrepancy could be due to the conversation topics. In our studies, the difference in turn-taking gaps was smaller in the *hobbies* task which could suggest that differences are reduced when autistic adults are talking about their special interests.

It is important to note that studies have shown that some differences in interactions can be reduced or even diminished when autistic individuals are interacting with other autistic people [75–80; for a possible theory explaining this phenomenon see Milton, (81)]. Since this study did not include dyads consisting of two autistic people, it is unclear if the found differences would extend to such a scenario. Future research examining interactional features in verbal communication should investigate possible differences not only between mixed and non-autistic dyads, but should also include comparisons with dyads consisting of two autistic interaction partners.

In addition to interactional features of verbal conversations, we also assessed synchrony and turn-based adaptation of speech features between two interaction partners in a dyad. We found no difference in turn-based adaptation between autistic and non-autistic adults, meaning that the extent to which they adapted their pitch, intensity and articulation rate to the previous turn was comparable in both groups. Similarly, we also did not find any differences between time-course synchrony of pitch between mixed and non-autistic dyads. However, we found that time-course synchrony of intensity was higher in mixed dyads than in non-autistic dyads. This is in contrast to Ochi et al. (18) who found increased correlation of the blockwise mean of intensity in conversations with non-autistic compared to autistic adults in the context of the ADOS. They also found a trend towards increased correlation of the blockwise mean of pitch in the conversations with non-autistic adults. Similarly, Lahiri et al. (27) found decreased dissimilarity of prosodic features which suggests increased synchrony in non-autistic children when analysing conversations from the ADOS. This is also more in line with previous research on other modalities which consistently shows reduced interpersonal synchrony in mixed dyads including an autistic person compared to dyads consisting of two non-autistic people (21). More research is needed to assess in which context interpersonal synchrony of speech features differs between autistic and non-autistic adults or mixed and non-autistic dyads.

Despite the insights this study offers, it is still unclear how context influences speech production with respect to ASD. We aimed for a naturalistic conversation setting with one common (*hobbies*) and one uncommon (*meal planning*) conversation topic. Other studies have opted to focus on a more controlled speech production by pairing participants with a trained diagnostician (18, 27), asking participants to retell a story (68) or even to produce a specific phrase with the aim of conveying a predefined emotion (70). Some of these contexts may have led to the differences in the reported results. The influence of context could be investigated by combining a naturalistic conversation task with a more controlled speech production task. The first would allow to assess speech features in an interactive settings similar to everyday conversations, while the latter could provide a baseline for each participant. Additionally, the influence of the interaction partners themselves has not been investigated yet. In our study, all interaction partners were strangers before the experiment, and in other studies, the interaction partners were often part of the research team (18).

In this study, we investigated the potential of speech and interactional features of verbal communication for digitally assisted diagnostics. We used automatic feature extraction on two naturalistic 10-minute conversations between either two non-autistic strangers (non-autistic dyad) or one autistic and one non-autistic stranger (mixed dyad). We were able to classify between individuals from a non-autistic vs. from a mixed dyad based on these features with high accuracy which offers a low-tech, economic and scalable option for diagnostic classification. Additionally, we have shown differences in pitch and intensity variation as well as articulation rate between autistic and non-autistic adults and differences in silence-to-turn ratio, turn-taking gaps and time-course synchrony of intensity between non-autistic and mixed dyads. This study shows the potential of verbal markers for diagnostic classification of ASD and suggests multiple relevant features showing differences between autistic and non-autistic adults.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethikkommission der Medizinischen Fakultät der Ludwig-Maximilians-Universität München. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

IP: Conceptualization, Formal analysis, Investigation, Methodology, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. JK: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project–administration, Writing – review & editing. AN: Conceptualization, Data curation, Investigation, Methodology, Project–administration, Writing – review & editing. NK: Conceptualization, Methodology, Resources, Supervision, Writing – review & editing. CF-W: Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Stiftung Irene (PhD scholarship awarded to JK) and the German Research Council (grant numbers 876/3–1 and FA 876/5–1 awarded to CF-W).

Acknowledgments

We want to thank Elsa Sangaran for their part in writing the preregistration and inspecting the data. We also want to show appreciations to the interns in our group who aurally and visually inspected the raw and the preprocessed data (in alphabetical order): Alena Holy, Christian Aldenhoff and Sophie Herke. Additionally, we want to thank Stephanie Fischer and Johanna Späth for their contribution to the data collection and Mark Sen Dong for providing scripts to create cross validation structures. Last, we want to express our thanks to Francesco Cangemi for their input on the microphone setup and the preprocessing parameters.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1257569/full#supplementary-material

References

1. Lepschy G. F. de Saussure, Course in general linguistics, translated and annotated by Roy Harris. London: Duckworth, 1983. *J Linguist*. (1985) 21:250–4. doi: 10.1017/S0022226700010185

2. Pereira C, Watson C. Some acoustic characteristics of emotion. *5th Int Conf Spok Lang Process ICSLP 1998*. (1998). doi: 10.21437/icslp.1998-148

3. Whiteside SP. Acoustic characteristics of vocal emotions simulated by actors. *Percept Mot Skills*. (1999) 89:1195–208. doi: 10.2466/pms.1999.89.3f.1195

4. World Health Organization. (2019). International Classification of Diseases, Eleventh Revision (ICD-11). Available at: https://icd.who.int/browse11.

5. Asghari SZ, Farashi S, Bashirian S, Jenabi E. Distinctive prosodic features of people with autism spectrum disorder: a systematic review and meta-analysis study. *Sci Rep*. (2021) 11:23093–17. doi: 10.1038/s41598-021-02487-6

6. Pickett E, Pullara O, O'Grady J, Gordon B. Speech acquisition in older nonverbal individuals with autism: a review of features, methods, and prognosis. *Cogn Behav Neurol*. (2009) 22:1–21. doi: 10.1097/WNN.0b013e318190d185

7. Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, et al. Austism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord*. (1989) 19:185–212. doi: 10.1007/BF02211841

8. Fusaroli R, Lambrechts A, Bang D, Bowler DM, Gaigg SB. Is voice a marker for autism spectrum disorder? A systematic review and meta-analysis. *Autism Res*. (2017) 10:384–407. doi: 10.1002/aur.1678

9. Grossman RB, Bemis RH, Plesa Skwerer D, Tager-Flusberg H. Lexical and affective prosody in children with high functioning autism. *J Speech Lang Hear Res*. (2010) 53:778–93. doi: 10.1044/1092-4388(2009/08-0127)

10. Scharfstein LA, Beidel DC, Sims VK, Rendon Finnell L. Social skills deficits and vocal characteristics of children with social phobia or asperger's disorder: a comparative study. *J Abnorm Child Psychol*. (2011) 39:865–75. doi: 10.1007/s10802-011-9498-2

11. Bone D, Black MP, Ramakrishna A, Grossman R, Narayanan S. Acoustic-prosodic correlates of 'awkward' prosody in story retellings from adolescents with autism. *Proc Annu Conf Int Speech Commun Assoc INTERSPEECH*. (2015):1616–20. doi: 10.21437/interspeech.2015-374

12. Elias R, Muskett AE, White SW. Educator perspectives on the postsecondary transition difficulties of students with autism. *Autism*. (2019) 23:260–4. doi: 10.1177/1362361317726246

13. Garrels V. Getting good at small talk: student-directed learning of social conversation skills. *Eur J Spec Needs Educ*. (2019) 34:393–402. doi: 10.1080/08856257.2018.1458472

14. Grossman RB. Judgments of social awkwardness from brief exposure to children with and without high-functioning autism. *Autism*. (2015) 19:580–7. doi: 10.1177/1362361314536937

15. Tomprou M, Kim YJ, Chikersal P, Woolley AW, Dabbish LA. Speaking out of turn: how video conferencing reduces vocal synchrony and collective intelligence. *PloS One*. (2021) 16:e0247655–14. doi: 10.1371/journal.pone.0247655

16. Templeton EM, Chang LJ, Reynolds EA, LeBeaumont MDC, Wheatley TP, Cone Lebeaumont MD, et al. Fast response times signal social connection in conversation. *PNAS*. (2022) 119:1–8. doi: 10.1073/pnas.2116915119

17. McLaughlin ML, Cody MJ. Awkward silences: behavioral antecedents and consequences of the conversational lapse. *Hum Commun Res*. (1982) 8:299–316. doi: 10.1111/j.1468-2958.1982.tb00669.x

18. Ochi K, Ono N, Owada K, Kojima M, Kuroda M, Sagayama S, et al. Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PloS One*. (2019) 14:e0225377–22. doi: 10.1371/journal.pone.0225377

19. Nelson A. M. (2020). *Investigating the acoustic, prosodic, and interactional speech characteristics of autistic traits: An automated speech analysis in typically-developed adults*.

20. Georgescu AL, Koeroglu S, Hamilton AFC, Vogeley K, Falter-Wagner CM, Tschacher W. Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: a motion energy study. *Mol Autism*. (2020) 11:11–4. doi: 10.1186/s13229-019-0305-1

21. McNaughton KA, Redcay E. Interpersonal synchrony in autism. *Curr Psychiatry Rep*. (2020) 22:1–11. doi: 10.1007/s11920-020-1135-8

22. Gregory S, Webster S, Huang G. Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Lang Commun*. (1993) 13:195–217. doi: 10.1016/0271-5309(93)90026-J

23. Natale M. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *J Pers Soc Psychol*. (1975) 32:790–804. doi: 10.1037/0022-3514.32.5.790

24. Street RL. Speech convergence and speech evaluation in fact-finding interviews. *Hum Commun Res*. (1984) 11:139–69. doi: 10.1111/j.1468-2958.1984.tb00043.x

25. Ward A., Litman D. (2007). Measuring convergence and priming in tutorial dialog. Univ. Pittsburgh. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Measuring+Convergence+and+Priming+in+Tutorial+Dialog#0.

26. Wynn CJ, Borrie SA, Sellers TP. Speech rate entrainment in children and adults with and without autism spectrum disorder. *Am J Speech-Language Pathol*. (2018) 27:965–74. doi: 10.1044/2018_AJSLP-17-0134

27. Lahiri R, Nasir M, Kumar M, Kim SH, Bishop S, Lord C, et al. Interpersonal synchrony across vocal and lexical modalities in interactions involving children with autism spectrum disorder. *JASA Express Lett*. (2022) 2:095202. doi: 10.1121/10.0013421

28. Koehler JC, Dong MS, Nelson AM, Fischer S, Späth J, Plank IS, et al. Machine learning classification of autism Spectrum disorder based on reciprocity in naturalistic social interactions. *medRxiv*. (2022):22283571. doi: 10.1101/2022.12.20.22283571

29. Boersma P., Weenink D. (2022). Praat: doing phonetics by computer. Available at: http://www.praat.org/.

30. De Jong NH, Pacilly J, Heeren W. PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assess Educ Princ Policy Pract*. (2021) 28:456–76. doi: 10.1080/0969594X.2021.1951162

31. De Jong N. H., Pacilly J., Heeren W. (2021). Uhm-o-meter [computer software]. Available at: https://sites.google.com/view/uhm-o-meter/home.

32. R Core Team. (2021). R: A Language and Environment for Statistical Computing. Available at: https://www.r-project.org.

33. RStudio Team. (2020). RStudio: Integrated Development Environment for R. Available at: http://www.rstudio.com/.

34. JASP Team. (2022). JASP (Version 0.16.4)[Computer software]. Available at: https://jasp-stats.org/.

35. Koutsouleris N., Vetter C., Wiegand A. (2022). Neurominer [computer software]. Available at: https://github.com/neurominer-git/NeuroMiner_1.1.

36. The Mathworks Inc. (2022). MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com

37. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. (2015) 162:55–63. doi: 10.7326/M14-0697

38. Lehrl S. (1999). *Mehrfachwahl-Wortschatz-Intelligenztest: MWT-B*. Spitta.

39. Weiß R. (2006). *CFT 20-R: grundintelligenztest skala 2-revision*. Hogrefe.

40. World Health Organization. *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. Geneva: World Health Organization, (1993).

41. Kirby A, Edwards L, Sugden D, Rosenblum S. The development and standardization of the adult developmental co-ordination disorders/dyspraxia checklist (ADC). *Res Dev Disabil*. (2010) 31:131–9. doi: 10.1016/j.ridd.2009.08.010

42. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-Spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord*. (2001) 31:5–17. doi: 10.1023/A:1005653411471

43. Hautzinger M., Bailer M., Worall H., Keller F. (1994). *Beck-depressions-inventar (BDI)*. Huber.

44. Graf A. A German version of the self-monitoring scale. *Zeitschrift fur Arbeits- und Organ*. (2004) 48:109–21. doi: 10.1026/0932-4089.48.3.109

45. Paulus C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. Available at: http://hdl.handle.net/20.500.11780/3343.

46. Davis M. H. (1980). Interpersonal reactivity index. Available at: 10.1037/t01093-000

47. Popp K, Schäfer R, Schneider C, Brähler E, Decker O, Hardt J, et al. Faktorstruktur und Reliabilität der Toronto-Alexithymie-Skala (TAS-20) in der deutschen Bevölkerung. *Psychother Med Psychol*. (2008) 58:208–14. doi: 10.1055/s-2007-986196

48. Drimalla H., Baskow I., Behnia B, et al (2019). Imitation and recognition of facial emotions in autism: a computer vision approach. *Molecular Autism*. (2021) 12:27. doi: 10.1186/s13229-021-00430-0

49. Tschacher W, Rees GM, Ramseyer F. Nonverbal synchrony and affect in dyadic interactions. *Front Psychol*. (2014) 5:1–13. doi: 10.3389/fpsyg.2014.01323

50. Moulder RG, Boker SM, Ramseyer F, Tschacher W. Determining synchrony between behavioral time series: an application of surrogate data generation for establishing falsifiable null-hypotheses. *Psychol Methods*. (2018) 23:757–73. doi: 10.1037/met0000172

51. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. (2018) 14:91–118. doi: 10.1146/annurev-clinpsy-032816-045037

52. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev*. (2012) 36:1140–52. doi: 10.1016/j.neubiorev.2012.01.004

53. Georgescu AL, Koehler JC, Weiske J, Vogeley K, Koutsouleris N, Falter-Wagner CM. Machine learning to study social interaction difficulties in ASD. *Front Robot AI*. (2019) 6:1–7. doi: 10.3389/frobt.2019.00132

54. Koehler JC, Georgescu AL, Weiske J, Spangemacher M, Burghof L, Falkai P, et al. Brief report: specificity of interpersonal synchrony deficits to autism Spectrum disorder and its potential for digitally assisted diagnostics. *J Autism Dev Disord*. (2022) 52:3718–26. doi: 10.1007/s10803-021-05194-3

55. Fan R.-E., Chang K.-W., Hsieh C.-J., Wang X.-R., Lin C.-J. (2022). LIBLINEAR: a library for large linear classification. Available at: https://www.csie.ntu.edu.tw/~cjlin/papers/liblinear.pdf.

56. Pavlou M, Ambler G, Seaman S, De Iorio M, Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat Med*. (2016) 35:1159–77. doi: 10.1002/sim.6782

57. Cortes C, Vapnik VN. Support-vector networks. *Mach Learn*. (1995) 20:273–97. doi: 10.1007/BF00994018

58. Vapnik VN. An overview of statistical learning theory. *IEEE Trans Neural Netw*. (1999) 10:988–99. doi: 10.1109/72.788640

59. Vapnik VN. *The nature of statistical learning theory*. *2nd* ed. New York: Springer (2000).

60. Goss-Sampson M. Bayesian inference in JASP (2020):1–120. doi: 10.17605/OSF.IO/CKNXM,

61. Boorse J, Cola M, Plate S, Yankowitz L, Pandey J, Schultz RT, et al. Linguistic markers of autism in girls: evidence of a 'blended phenotype' during storytelling. *Mol Autism*. (2019) 10:1–12. doi: 10.1186/s13229-019-0268-2

62. Sturrock A, Adams C, Freed J. A subtle profile with a significant impact: language and communication difficulties for autistic females without intellectual disability. *Front Psychol*. (2021) 12:1–9. doi: 10.3389/fpsyg.2021.621742

63. Sturrock A, Chilton H, Foy K, Freed J, Adams C. In their own words: the impact of subtle language and communication difficulties as described by autistic girls and boys without intellectual disability. *Autism*. (2022) 26:332–45. doi: 10.1177/13623613211002047

64. Zwaigenbaum L, Penner M. Autism spectrum disorder: advances in diagnosis and evaluation. *BMJ*. (2018) 361:k1674–16. doi: 10.1136/bmj.k1674

65. Matson JL, Kozlowski AM. The increasing prevalence of autism spectrum disorders. *Res Autism Spectr Disord*. (2011) 5:418–25. doi: 10.1016/j.rasd.2010.06.004

66. Ruzich E, Allison C, Smith P, Watson P, Auyeung B, Ring H, et al. Measuring autistic traits in the general population: a systematic review of the autism-Spectrum quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Mol Autism*. (2015) 6:1–12. doi: 10.1186/2040-2392-6-2

67. Kaland C, Swerts M, Krahmer E. Accounting for the listener: comparing the production of contrastive intonation in typically-developing speakers and speakers with autism. *J Acoust Soc Am*. (2013) 134:2182–96. doi: 10.1121/1.4816544

68. Chan KKL, To CKS. Do individuals with high-functioning autism who speak a tone language show intonation deficits? *J Autism Dev Disord*. (2016) 46:1784–92. doi: 10.1007/s10803-016-2709-5

69. DePape AMR, Chen A, Hall GBC, Trainor LJ. Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Front Psychol*. (2012) 3:1–13. doi: 10.3389/fpsyg.2012.00072

70. Hubbard DJ, Faso DJ, Assmann PF, Sasson NJ. Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Res*. (2017) 10:1991–2001. doi: 10.1002/aur.1847

71. Bambara LM, Thomas A, Chovanes J, Cole CL. Peer-mediated intervention: enhancing the social conversational skills of adolescents with autism Spectrum disorder. *Teach Except Child*. (2018) 51:7–17. doi: 10.1177/0040059918775057

72. Rieth SR, Stahmer AC, Suhrheinrich J, Schreibman L, Kennedy J, Ross B. Identifying critical elements of treatment: examining the use of turn taking in autism intervention. *Focus Autism Other Dev Disabl*. (2014) 29:168–79. doi: 10.1177/1088357613513792

73. Thirumanickam A, Raghavendra P, McMillan JM, van Steenbrugge W. Effectiveness of video-based modelling to facilitate conversational turn taking of adolescents with autism spectrum disorder who use AAC. *AAC Augment Altern Commun*. (2018) 34:311–22. doi: 10.1080/07434618.2018.1523948

74. Bone D, Bishop S, Gupta R, Lee S, Narayanan SS. Acoustic-prosodic and turn-taking features in interactions with children with neurodevelopmental disorders. *Proc Annu Conf Int Speech Commun Assoc INTERSPEECH*. (2016):1185–9. doi: 10.21437/Interspeech.2016-1073

75. Heasman B, Gillespie A. Neurodivergent intersubjectivity: distinctive features of how autistic people create shared understanding. *Autism*. (2019) 23:910–21.

76. Crompton CJ, Ropar D, Evans-Williams CV, Flynn EG, Fletcher-Watson S. Autistic peer-to-peer information transfer is highly effective. *Autism*. (2020) 24:1704–12.

77. Crompton CJ, Hallett S, Ropar D, Flynn E, Fletcher-Watson S. 'I never realised everybody felt as happy as I do when I am around autistic people': A thematic analysis of autistic adults' relationships with autistic and neurotypical friends and family. *Autism*. (2020) 24:1438–48.

78. Crompton CJ, Sharp M, Axbey H, Fletcher-Watson S, Flynn EG, Ropar D. Neurotype-matching, but not being autistic, influences self and observer ratings of interpersonal rapport. *Front Psychol*. (2020) 11:2961.

79. Morrison KE, DeBrabander KM, Jones DR, Faso DJ, Ackerman RA, Sasson NJ. Outcomes of real-world social interaction for autistic adults paired with autistic compared to typically developing partners. *Autism*. (2020) 24:1067–80.

80. Alkire D, McNaughton KA, Yarger HA, Shariq D, Redcay E. Theory of mind in naturalistic conversations between autistic and typically developing children and adolescents. *Autism*. (2023) 27:472–88.

81. Milton DE. On the ontological status of autism: The 'double empathy problem'. *Disability and society*. (2012) 27:883–7.

4 Original Study III – Classifying autism in a clinical population based on motion synchrony: a proof-of-concept study using real-life diagnostic interviews

Contributions and reference

This project aimed to test the predictability of motion synchrony obtained through objective computer vision analyses of real-life diagnostic interviews for the diagnostic classification of autism in a clinical population. This study was an international collaboration with Seoul National University (SNU). I designed and supervised the implementation of the preprocessing pipeline at SNU Bundang Hospital, involving an extended on-site lab visit followed up by remote consultation. I was further involved in the screening of eligible videos. I analyzed the final data set using RStudio [69] and NeuroMiner [70] and interpreted the results in conjunction with the shared first author. I presented the results to the scientific community during talks and poster presentations at different conferences. In compliance with the endeavor of reproducibility and open science, I published the analysis scripts on GitHub (https://github.com/jckoe/SNU_ASDsync) and made a preprint available on MedRXiv. I wrote the manuscript and was primarily responsible to the revision process and finalizing the published article.

Koehler, J.C.*, Dong, M.S.*, Song, D.-Y., Bong, G., Koutsouleris, N., Yoo, H., & Falter-Wagner, C.M. (2024). **Classifying autism in a clinical population based on motion synchrony: a proof-of-concept study using real-life diagnostic interviews.** *Sci Rep,* 14, 5663. https://doi.org/10.1038/s41598-024-56098-y
*These authors contributed equally.

Note: For better readability, the supplementary material of this publication can be accessed online.

<u>Journal Citation Report:</u>
Impact Factor 2022: 4.6
5 Year Impact Factor 2022: 4.9
Ranked 21/134 Multidisciplinary Sciences (Journal Citation Indicator 2022)

scientific reports

OPEN

Classifying autism in a clinical population based on motion synchrony: a proof-of-concept study using real-life diagnostic interviews

Jana Christina Koehler[1,6]✉, Mark Sen Dong[1,4,6], Da-Yea Song[2,3], Guiyoung Bong[2], Nikolaos Koutsouleris[1,4,5], Heejeong Yoo[2,3,7] & Christine M. Falter-Wagner[1,7]✉

Predictive modeling strategies are increasingly studied as a means to overcome clinical bottlenecks in the diagnostic classification of autism spectrum disorder. However, while some findings are promising in the light of diagnostic marker research, many of these approaches lack the scalability for adequate and effective translation to everyday clinical practice. In this study, our aim was to explore the use of objective computer vision video analysis of real-world autism diagnostic interviews in a clinical sample of children and young individuals in the transition to adulthood to predict diagnosis. Specifically, we trained a support vector machine learning model on interpersonal synchrony data recorded in Autism Diagnostic Observation Schedule (ADOS-2) interviews of patient-clinician dyads. Our model was able to classify dyads involving an autistic patient (n = 56) with a balanced accuracy of 63.4% against dyads including a patient with other psychiatric diagnoses (n = 38). Further analyses revealed no significant associations between our classification metrics with clinical ratings. We argue that, given the above-chance performance of our classifier in a highly heterogeneous sample both in age and diagnosis, with few adjustments this highly scalable approach presents a viable route for future diagnostic marker research in autism.

Autism spectrum disorder is characterized by symptoms in social interaction and communication as well as repetitive behaviors. Typically diagnosed during childhood[1], autism is increasingly diagnosed in adulthood over the past years[2], with prevalence estimates around 1%[3]. Due to the lack of clear diagnostic markers, the current gold-standard diagnostic process requires multiple assessments with a trained interdisciplinary clinical team[4], including a diagnostic observation (e.g., Autism Diagnostic Observation Schedule, ADOS-2[5]), neuropsychological tests, and an interview with a caregiver about the early developmental history (e.g., Autism Diagnostic Interview, ADI-R[6]). While thorough assessments are vital for correct diagnosis, the process itself is lengthy and resource-heavy, causing long waiting times which, thus, comes at a great cost for all involved.

Due to the rising demand for diagnostics in recent years, attempts are increasingly made to advance the diagnostic process through personalized prediction approaches based on computational methods such as machine learning. One approach that naturally lends itself to further investigation is the data-driven investigation of existing diagnostic tools such as ADOS. Several studies have been conducted to improve the existing diagnostic algorithm by filtering out a subset of key items predictive for diagnosis. For example, using feature selection-based machine learning on a large data set of children's ADOS results, Kosmicki et al[7]. significantly reduced the number of relevant items for accurate diagnostic prediction by more than 55%. Küpper and colleagues[8] found that diagnostic prediction performance for adolescents and adults with only five ADOS items was comparable to

[1]Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Munich, Germany. [2]Department of Psychiatry, Seoul National University Bundang Hospital, Seongnam, South Korea. [3]Department of Psychiatry, Seoul National University College of Medicine, Seoul, Korea. [4]Max Planck Institute of Psychiatry, Munich, Germany. [5]Institute of Psychiatry, Psychology and Neuroscience, King's College, London, UK. [6]These authors contributed equally: Jana Christina Koehler and Mark Sen Dong. [7]These authors jointly supervised this work: Heejeong Yoo and Christine M. Falter-Wagner. ✉email: jana.koehler@med.uni-muenchen.de; Christine.falter@med.uni-muenchen.de

the originally proposed 11-item diagnostic algorithm. Nevertheless, this approach is prone to a certain circularity, given the outcome criterion, that is the clinical diagnosis of ASD, is heavily influenced by the features used for prediction[8]. Thus, using machine learning on objective and rater-independent datasets for the screening of potential markers is desirable. Hence, several studies investigated structural or functional brain abnormalities as predictive markers in ASD[9], with promising accuracies especially for younger children[10]. However, methods such as magnetic resonance imaging lack scalability and are impractical to implement in standardized clinical practice. Additionally, those approaches pose special challenges for a sensory-sensitive study population such as autistic individuals. Thus, a more translational approach uses machine learning for diagnostic classification in ASD through digitally assisted diagnostics or digital phenotyping[11], which directly taps the symptomatic behavior. This approach combines the advantages of moving away from the human coding of behaviors while using more scalable methods such as tablet-based movement data or video analysis via computer vision techniques. For instance, Anzulewicz and colleagues[12] reported that a machine learning model trained to identify children with ASD based on their tablet-recorded motion trajectories performed with an accuracy of 93%. In a recent study, Jin et al[13]. developed a pipeline to objectively extract movement features correlated with clinicians' ratings from children during ADOS interviews. Movement aberrances in autism are common, though its connection to autistic core symptomatology remains unclear[14].

Although autism is commonly referred to as a disorder of social interaction, thus, implying a certain degree of reciprocity, this aspect is challenging to assess objectively. The increasingly studied phenomenon of reduced interpersonal synchrony in ASD[15] provides such an opportunity. Interpersonal synchrony is commonly defined as the alignment of individuals within an interaction[16,17] and has found to be predictive for conversational features and outcomes, such as prosocial behavior[18] or empathy[19]. Given the frequent mismatch and often perceived awkwardness in autistic social interactions, a number of studies have investigated a potential link of interpersonal synchrony to the autistic phenotype, as well as potential interventions[15]. While interpersonal synchrony encompasses a range of aligned signals on multiple modalities, for this specific study we focused exclusively on interpersonal motor synchrony, i.e., the alignment of movement within a conversation. In a previous study[20], we found reduced interpersonal synchrony as derived from motion energy analysis (MEA[21]) in diagnostic interviews with autistic adults as compared to those who did not subsequently receive an autism diagnosis. Furthermore, we explored the predictiveness of interpersonal synchrony between autistic and non-autistic interactants on multiple modalities, finding high accuracy for the synchrony of facial and head movements[22]. However, these studies were conducted with adults, and while motor difficulties in autism tend to persist throughout adulthood[23], little is known about the predictive power of synchrony alterations in children.

In a study on video-based pose estimation, Kojovic et al[24]. investigated videos of ADOS interviews with small children. Their deep neural network analysis of multiple aspects of non-verbal interaction differentiated between autistic children and typically-developing (TD) children with an accuracy of 80.9% and additionally revealed associations between their model and the overall level of symptomatology. Thus, modeling based on direct extraction of predictive features from diagnostic videos opens a promising avenue for the clinical setting.

Our aim in this proof-of-concept study was to investigate automatic video analysis as a scalable approach to screen for synchrony alterations as an objective marker to classify autism in children and adolescents in transition to adulthood. To this end, we trained several support vector machine (SVM) classification models using synchrony features extracted from videos of real-life ADOS-2 interviews and investigated the associations of our classifiers' outputs with professional clinical ratings. Importantly, to explore model specificity in a realistic clinical scenario, we used a representative clinical sample that included participants who were subsequently diagnosed with ASD as well as patients with other psychiatric diagnoses.

Methods

In the following, we report the details of our prediction model following the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines[25].

Sample

The ADOS-2 videos and their related datasets were compiled from two different sources at Seoul National University Bundang Hospital: the patient pool of the psychiatric outpatient clinic for children and adolescents as well as from a study population of an unrelated study that included the ADOS-2. Therefore, the inclusion criteria and available data slightly differed. Patients referred to the outpatient clinic underwent extensive clinical examination to evaluate the presence of an ASD or differential diagnosis. Additional information on comorbidities and medication for this subsample is available in the supplementary material (see Supplementary Table S4.1) and was not included in the final analysis. For the patients from the unrelated study, ADOS-2 was performed as part of the study protocol, though the diagnosis had either already been suspected or given elsewhere. In contrast to the outpatient pool, exclusion criteria were applied in the unrelated study which comprised severe motor impairments restricting patients from engaging in the required ADOS-2 activities, as well as sensory-related issues or selective mutism. No age limit applied.

For all cases from both sources, the autism diagnosis was confirmed as a best clinical estimate consensus diagnosis by two psychiatrists, taking into account ADOS-2 and ADI-R results, as well as other neuropsychological assessments.

An overview of the current sample compilation procedure can be found in Fig. 1. All available ADOS-2 video materials were initially screened for the first occurrence of at least five minutes of consecutive and unobstructed footage for every participant based upon the following criteria: (a) steady camera position and constant lighting, (b) camera angle that includes the head and upper body of both participant and ADOS-2 administrator, (c) both participant and administrator being seated throughout all video frames (i.e., no freeplay, no running around),

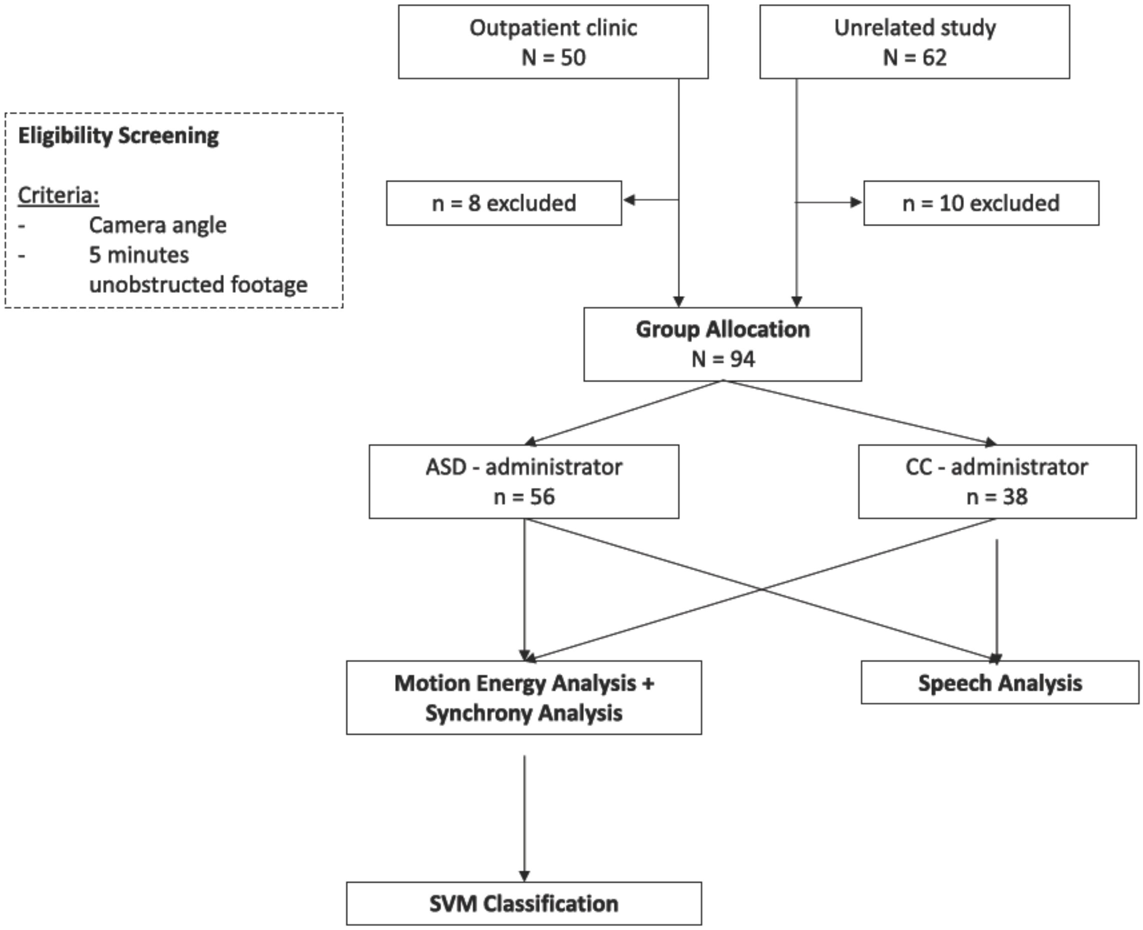

Figure 1. Consort chart of the current sample compilation.

(d) and no use of props. As only ADOS-2 modules three and four include longer instances of free-flowing conversation, the final sample was comprised of these modules. Excerpts were taken from the tasks *Emotion*, *Conversation and Reporting*, *Social Difficulties and Annoyance*, *Job/School Life*, *Friends, Relationships, and Marriage*, and *Loneliness*. Due to the semi-structured nature of ADOS-2, the final clips differed in length, ranging from 5:15 min to 14:37 min (mean length = 7:20 min). Interviews were conducted by six different administrators. All videos had a frame rate of 29.95 s.

The final dataset consisted of 56 participants with a diagnosis of ASD and 38 participants with other psychiatric conditions (i.e., n = 4 Intellectual Disability, n = 1 Developmental Delay, n = 10 ADHD, n = 1 Tourette Syndrome, n = 4 Depressive Disorder, n = 1 Social Phobia, n = 1 Anxiety Disorder, n = 2 Bipolar Disorder) or within the wider autism phenotype (n = 2), as well as n = 12 typically-developing (TD) children (including 8 unaffected siblings). This resulted in two diagnostic group allocations: ASD-administrator or clinical control (CC)-administrator.

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The study to use fully anonymized data collected retrospectively and prospectively were approved by the Institutional Review Board at Seoul National University Bundang Hospital (IRB no., B-1812-513-105; B-1912-580-304). Informed consent was obtained for the participant data collected prospectively from both participants and, in case the participant was a minor, their parent or legal guardian. A separate informed consent for the analysis of completely anonymized retrospective data was waived.

Video pre-processing and synchrony computation

Motion Energy Analysis (MEA)[21] was applied to all video clips, defining two regions of interest per participant and administrator (head and upper body). MEA extracts frame-to-frame gray-scale pixel changes. Keeping the camera position, lighting and background constant, all pixel changes above a manually set threshold represent movement within the regions of interest. After careful visual inspection of the resulting data quality, a threshold of eight was chosen for all videos.

Raw motion energy time series were subsequently forwarded to pre-processing using the RStudio package rMEA[26]. Videos were filmed in four different rooms. To account for potential biases of any distortions in the video signals, all MEA time series were scaled by standard deviation and smoothed with a moving average of

0.5 s, according to the default setting in rMEA. A comparison analysis of potential feature differences depending on the room can be found in the supplementary material (S2.2).

Interpersonal synchrony between participant and administrator in their head and body motion was computed with windowed cross-lagged correlations. In line with a previous analysis of diagnostic interviews with autistic adults[20], a window size of 60 s was chosen. To capture all instances of synchrony, time series were cross correlated with lags of 5 s and increments of 30 s. All values in the resulting cross-correlation matrices were converted to absolute Fisher Z values. Time series were subsequently shuffled and randomly paired into 500 pseudodyads. Cross correlations were conducted in the same manner, yielding a measure of pseudosynchrony per artificial dyad. They were subsequently compared to the interpersonal synchrony values to assess whether the interpersonal synchrony values were above-chance. Detailed results can be found in the supplementary material (S2.3).

Moreover, following procedures from Georgescu et al[27]., intrapersonal head and body coordination was computed for every patient, using window sizes of 30 s, lags of 5 s and a step size of 15 s.

Lastly, we derived the head and body movement quantity per participant from the respective MEA time series. Following previous procedures[20,28], they were defined as the number of frames with changes in motion energy divided by the total number of frames, resulting in four values per dyad (two for participant and administrator, respectively).

In addition to the processing of motion, we submitted our videos to an exploratory vocal output analysis. For this purpose, the audio tracks of the selected clips were processed with the software Praat[29] to semi-automatically extract annotations of intervals of vocalizations and silences. As there was no speaker distinction within the audio tracks, this analysis was considered exploratory and is not included in the main machine learning analysis. Details can be found in the supplementary material (S1.1).

Feature engineering

Because the videos in our sample varied in both length and conversational content (see Supplementary Material S2.1), as well as to account for the interview context, our aim was to gain the best estimate of the overall synchrony (i.e., instances in and out of synchrony), while simultaneously maintaining an adequate feature-to-sample ratio. For this reason, summary statistics of each cross-correlation matrix were computed (i.e., minimum, maximum, mean, median, standard deviation, skew, and kurtosis), resulting in seven features per participant-administrator dyad and region of interest (ROI). This procedure expands previous research investigating only the average of the entire matrix as a measure of synchrony (e.g[28,30]), therefore, providing additional insight and information on the richness of the data at hand. We additionally computed the same summary statistics for the intrapersonal head-body coordination of each participant. This approach slightly differed from a previous study[22], where we were interested in the trajectory of maximum synchrony instances during naturalistic and free-flowing conversations. To comply with previous procedures, we additionally computed a feature set using a peak-picking algorithm to obtain a measure of the trajectory of the highest synchrony instances during each interview. Details and results can be found in the supplementary material (S2.4).

The final feature set for each dyad consisted of 25 features per participant-administrator dyad (see Supplementary Table S4.2): seven interpersonal synchrony features per dyad and ROI (head and body), seven features for the intrapersonal head-body coordination of every participant, as well as four features for the individual amount of head and body movement of both interactants. IQ and sex of the participant were additionally included as features in a second model, as both are frequently associated with autism symptomatology and the likelihood of receiving a diagnosis[31,32].

Support vector machine (SVM) learning analyses

We trained two separate binary machine learning models to classify between dyad type: (1) a "behavioral" model containing only synchrony data objectively extracted from the videos (MEA), and (2) a model additionally containing sex and IQ as demographic features (MEA + DEMO). Taking into consideration the large age range in our sample, age was regressed out in both models. By constructing two separate models, we could explore whether demographic features frequently associated with ASD might improve the purely behavioral predictive performance. A L1-loss LIBSVM algorithm was chosen for both models, as it is frequently used in psychiatric research[33], known to perform robustly with reduced sample sizes[34]. In each model, the SVM algorithm independently modeled a linear relationship between features and classification labels by optimizing a linear hyperplane in a high-dimensional feature space to maximize separability between the dyads. Subsequently, the data was projected into the linear kernel space and their geometric distance to the decision boundary was measured. Thus, every dyad was assigned a predicted classification label and a decision score.

Machine learning analyses were conducted in NeuroMiner (Version 1.1; https://github.molgen.mpg.de/pages/LMU-Neurodiagnostic-Applications/NeuroMiner.io/)[35], an open-source mixed MATLAB[36]-Python-based machine learning library. To prevent any possibility of information leakage between training and testing data, our diagnostic classifiers were cross-validated in a repeated, nested, stratified cross-validation scheme. We used ten folds and ten permutations in the outer CV loop (CV2) and ten folds and one permutation in the inner loop (CV1). Specifically, at the CV2 level, we iteratively held back 9 or 10 participant-administrator dyads as test samples, while the rest of the data entered the CV1 cycle, where the data were again split into training and validation sets. This way training and testing data were strictly separated, with hyper-parameter tuning happening entirely within the inner loop while the outer loop was exclusively used to measure the classifier's generalizability to unseen data and generate decision scores for each dyad in this partition. This process was repeated for the remaining folds, after which the participants were reshuffled within their group and the process was repeated nine times, producing $10 \times 10 = 100$ decision scores for each held out participant. The final median decision score of each held out dyad was computed from the scores provided by the ensemble of models in which given dyad had not been used at the CV1 level for training or hyperparameter optimization. Additionally, the stratified design ensured that the proportion of the diagnostic groups in every fold would adequately reflect the proportion of the diagnostic group in the full sample and, thus, guarantee that each class is equally represented in each test fold to avoid bias during model training.

The preprocessing settings for the respective models can be found in Table 1.

Class imbalances were corrected for by hyperplane weighting. Balanced Accuracy (BAC = [sensitivity + specificity]/2) was used as the performance criterion for hyperparameter optimization. The C parameter was optimized in the CV1 cycle using 11 parameters within the following range: 0.0156, 0.0312, 0.0625, 0.1250, 0.2500, 0.5000, 1, 2, 4, 8, and 16, which represent the default settings in NeuroMiner[35]. Model significance was assessed through label permutation testing[37], with a significance level $\alpha = 0.05$ and 1000 permutations. The permutation testing procedure determines the statistical significance of a model's performances (i.e., BAC) by using the current data compared to models trained on the dataset but with the labels randomly permuted. Details regarding the permutation testing procedure can be found in the supplementary materials. The predictive pattern of the models was extracted using cross-validation ratio (CVR) and sign-based consistency. Firstly, CVR was computed as the mean and standard error of all normalized SVM weight vectors concatenated across the entire nested CV structure. CVR measures pattern element stability and was defined as the sum across CV2 folds of the CV1 median weights divided by their respective CV1 standard error, all of which was subsequently divided by the number of CV2 folds[38]. Secondly, we used the sign-based-consistency method[39] to test the stability of the predictive pattern by examining the consistency of positive and negative signs of the feature weight values across all models in the ensemble (see Supplementary Material S1.2 for additional information). Feature stability was assessed for statistical significance at $\alpha = 0.05$, using the Benjamini–Hochberg procedure of false discovery rate correction (FDR)[40].

Associations of SVM model and clinical variables

To investigate potential underlying clinical factors associated with our classification models, post-hoc correlation analyses with the SVM decision scores and ADOS-2, as well as ADI-R scores were performed in RStudio (version 2022–07.2)[41]. A dyad's predicted SVM decision score represents their distance from the hyperplane. ADOS-2 scores included domain scores for social affect (SA) and restricted and repetitive behaviors (RRB), as well as the total score (Total). Because our sample included data from both modules three and four, calibrated severity scores[42,43] were used for the correlation analyses for better comparison. For ADI-R, ratings on three subdomains based on caregiver report were used: reciprocal social interaction (A), social communication (B), and restricted and repetitive behaviors (C). Statistical significance was determined at $\alpha = 0.05$ and two-sided *p* values were corrected for multiple comparisons using FDR.

Exploratory SVM analysis

To further address the specificity of synchrony, given that phenotypic movement difficulties overlap in neurodevelopmental disorders (e.g., dyspraxia and autism, or hyperkinetic movement in ADHD), the MEA classifier was retrained within the same sample but using different class labels: (i) a neurodevelopmental disorders class, which grouped all 74 patients with a diagnosis of a neurodevelopmental disorder as defined by DSM-5[44] (n = 56 ASD, n = 10 ADHD, n = 1 Developmental Delay, n = 1 Tourette Syndrome, n = 4 Intellectual Disability, n = 2 Broad Spectrum/Pervasive Developmental Disorder–Not Otherwise Specified (PDD-NOS)), and (ii) a clinical control group consisting of the 20 patients with other psychiatric diagnoses or typically-developing participants (n = 12 TD including 8 unaffected siblings, n = 1 Anxiety Disorder, n = 2 Bipolar Disorder, n = 4 Depressive Disorder, n = 1 Social Phobia). The stratified CV structure was adapted accordingly.

Model	Features	Preprocessing pipeline
MEA	Interpersonal head synchrony (7)	1. Scaling between 0 and 1
	Interpersonal body synchrony (7)	2. Pruning of non-informative features (zero variance, infinite values)
	Intrapersonal head-body coordination of patient (7)	3. Age as covariate (partial correlation)
	Total head and body movement (4)	
MEA + DEMO	Interpersonal head synchrony (7)	1. Scaling between 0 and 1
	Interpersonal body synchrony (7)	2. Pruning of non-informative features (zero variance, infinite values)
	Intrapersonal head-body coordination of patient (7)	3. k-nearest neighbor imputation of missing values
	Total head and body movement (4)	4. Age as covariate (partial correlation)
	IQ (1)	
	Sex (1)	

Table 1. SVM Classification Model Descriptions. Number of features of respective modality in parentheses. Missing IQ values (16% of cases) were imputed using k-nearest neighbor imputation.

Results

Sample description

A description of the final sample grouped according to the ADOS-2 module can be found in Table 2. A chi-square test of independence revealed no significant association between the diagnostic group and sex ($\chi^2(1,94) = 0.045$, $p = 0.831$). Though naturally participants across both modules differed in age, there was no significant difference in age between diagnostic groups within each module. Because final diagnosis was partly based on ADOS-2 and ADI-R results, autistic patients across both modules had significantly higher ADOS-2 as well as ADI-R scores compared with the clinical control group. Best-estimate IQ values were significantly higher in the CC group for module 3. This effect was reversed in module 4, with autistic patients scoring significantly higher on their respective IQ assessment. SVM Classification Performance and Feature Importance.

Using only motion energy analysis data and regressing out age, our MEA model was able to classify interview dyads with an autistic participant as opposed to those with other psychiatric diagnoses with a BAC of 63.4% (Fig. 2). Detailed performance metrics, i.e., sensitivity, specificity, accuracy, positive and negative predictive values, and Area-Under-the-Receiver-Operating-Curve (AUC) can be found in Table 3. There was no significant residual association between age (M = 13.53, SD = 4.70) and the model's resulting decision scores (M = 0.19, SD = 0.89) after regressing out age during pre-processing ($r_{Pearson} = 0.06$, $p = 0.558$). The model that additionally included sex and IQ as features (MEA + DEMO) had a lower BAC of 59.4% (Sensitivity = 71.4%, Specificity = 47.4%, AUC = 0.58[CI = 0.46—0.70], also see S4.3Supplementary Table).

A closer investigation of the cross-validation ratio revealed that classification towards the autism-administrator dyads was driven by higher kurtosis and skewness of their body synchrony values (Fig. 3a). This means that a dyad with more pronounced outliers in their body synchrony, especially in the positive direction, was considered more autistic. In contrast, our model considered higher mean body synchrony values as non-autistic. Sign-based consistency revealed that this effect was relatively stable (Fig. 3b). Interestingly, the opposite effect was visible for head synchrony: higher kurtosis and skewness of head synchrony values were considered non-autistic, whereas higher mean head synchrony values were considered autistic. However, this was not consistent and of less feature importance than body synchrony.

A closer look at the movement parameters of both participant and administrator revealed that more movement by the administrator was taken into account when classifying an autistic dyad, whereas more movement by the participant was classified as a clinical control dyad.

A comprehensive list of cross-validation ratios and sign-based consistencies for all features of the MEA model can be found in the Supplementary Material (Supplementary Figs. 2 and 3).

Associations between SVM model scores and clinical variables

We conducted a range of correlation analyses of the resulting SVM scores of our models with ADOS-2 [5] and ADI-R[6] domain and total scores (Fig. 4). ADI-R data was incomplete for ten participants, who were discarded from the respective analysis.

Module		ASD (n = 56, 11 female)			CC (n = 38, 9 female)			*T test*			Effect size
		n	M	SD	n	M	SD	*t*	*df*	*p*	
3	Age	37	11.35	2.79	27	10.68	2.74	0.96	56.80	.339	.244
	Full Scale IQ	33	92.09	20.87	20	103.55	17.05	− 2.18	46.44	.035	− .601
	Verbal IQ	33	93.39	21.19	20	104.50	18.02	− 2.03	45.29	.048	− .56
	ADOS_SA	37	10.05	2.89	27	5.37	3.65	5.53	47.94	< .001	1.42
	ADOS_RRB	37	1.27	1.54	27	.26	.59	3.64	49.36	.001	.867
	ADI-R_A	36	17.58	6.38	24	6.25	5.53	7.30	53.99	< .001	1.90
	ADI-R_B	36	13.00	4.67	24	4.58	4.09	7.38	53.71	< .001	1.92
	ADI-R_C	36	4.81	2.62	24	1.12	1.39	7.07	55.75	< .001	1.76
4	Age	19	18.26	3.05	11	19.69	4.18	− 1.00	16.25	.334	− .392
	Full Scale IQ	18	95.56	17.52	8	77.25	12.34	3.05	18.91	.007	1.21
	Verbal IQ	17	102.65	19.20	8	90.25	11.12	2.03	21.72	.054	.79
	ADOS_SA	19	10.74	4.11	11	5.82	3.95	3.24	21.73	.004	1.22
	ADOS_RRB	19	2.05	1.75	11	.73	1.01	2.63	27.99	.014	.929
	ADI-R_A	16	14.31	7.91	8	6.50	4.87	2.98	20.77	.007	1.19
	ADI-R_B	16	10.44	4.70	8	4.88	3.80	3.12	17.16	.006	1.30
	ADI-R_C	16	4.81	2.61	8	1.75	1.49	3.65	21.44	.001	1.44

Table 2. Sample description and demographic group differences across subsamples. Full Scale IQ depicted as best estimate depending on age (WISC-III, WISC-IV, WPPSI-IV, WAIS → Korean versions); Verbal IQ as indicated by verbal comprehension index (VCI) of respective IQ test; group comparisons computes using Welch's t-test for unequal variances, *p* value adjusted for multiple testing with Bonferroni-Holm correction, Effect size cohen's d.

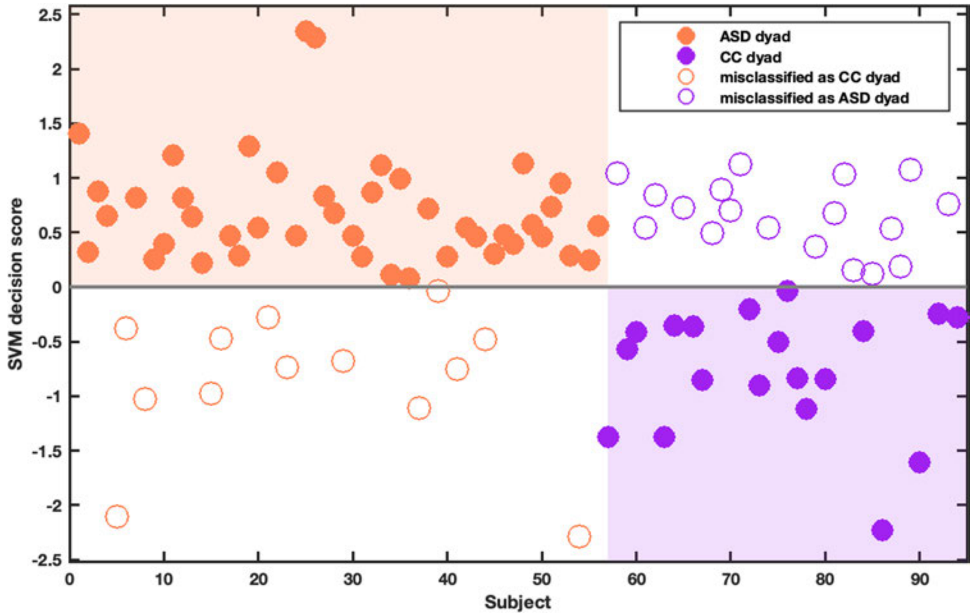

Figure 2. SVM classification results of ASD versus CC patient-administrator dyads. Figure depicts mean classifier scores of each dyad in the model containing only MEA data, resulting in a balanced classification accuracy of 63.4%. The further the score is from the decision boundary, the more likely this dyad was predicted as belonging to their respective class.

Classifier	BAC (%)	Sensitivity (%)	Specificity (%)	AUC [95% CI]	TN	TP	FN	FP	Accuracy (%)	Number needed to diagnose	Positive likelihood ratio	Diagnostic odds ratio	Permutation test, *p* value
ASD versus CC	63.4	76.8	50.0	.61 [.50—.72]	19	43	13	19	66.0	3.7	1.5	2.4	< .001
NDD versus CC	56.1	62.2	50.0	.52 [.38—.66]	10	46	28	10	59.6	8.2	1.2	1.5	.005

Table 3. Classification metrics for SVM classifiers based on Motion Energy Synchrony Analyses between Patient and Administrator. Both classifiers were trained on the same sample, regrouping patients under different labels. The NDD group contained, additionally to patients with ASD, patients with Intellectual Disability, Developmental Delay, ADHD, Tourette Syndrome, and Broad Spectrum/PDD-NOS. *ASD*—Autism Spectrum Disorder, *CC*—Clinical Control, *NDD*—Neurodevelopmental Disorder, *BAC*—Balanced Accuracy, *AUC*—Area Under The Receiver Operating Curve, *TN*—True Negatives, *TP*—True Positives, *FN*—False Negatives, *FP*—False Positives.

In general, classification towards the autistic group was loosely associated with higher ADI-R ratings on all three scales, although these findings were not statistically significant. No significant associations were found for the ADOS-2 ratings. Detailed correlation results can be found in S4.4 Supplementary Table.

Exploratory SVM analysis: NDD versus CC

When regrouping the present sample and classifying participants with neurodevelopmental disorders in general and clinical controls based on motion energy synchrony (analogous to the MEA model), the BAC decreased to 56.1% (Table 3).

Discussion

This proof-of-concept study aimed to explore the predictability of autism from non-verbal aspects of social interactions between participants and clinicians using videos of real-life diagnostic interviews. Our classification algorithm solely trained on objectively quantified synchrony values was able to predict autism in a representative clinical sample with a BAC of 63.4%. A separate model including demographic features frequently associated with the likelihood of an autism diagnosis (i.e., sex and IQ) yielded a lower balanced accuracy and, thus, did not improve predictive performance. Feature importance analyses revealed the impact of body synchrony and movement quantity for diagnostic classification. Slight but non-significant associations were found with ratings based on parent's reports (ADI-R), while we did not find any visible associations with ratings by clinicians. When

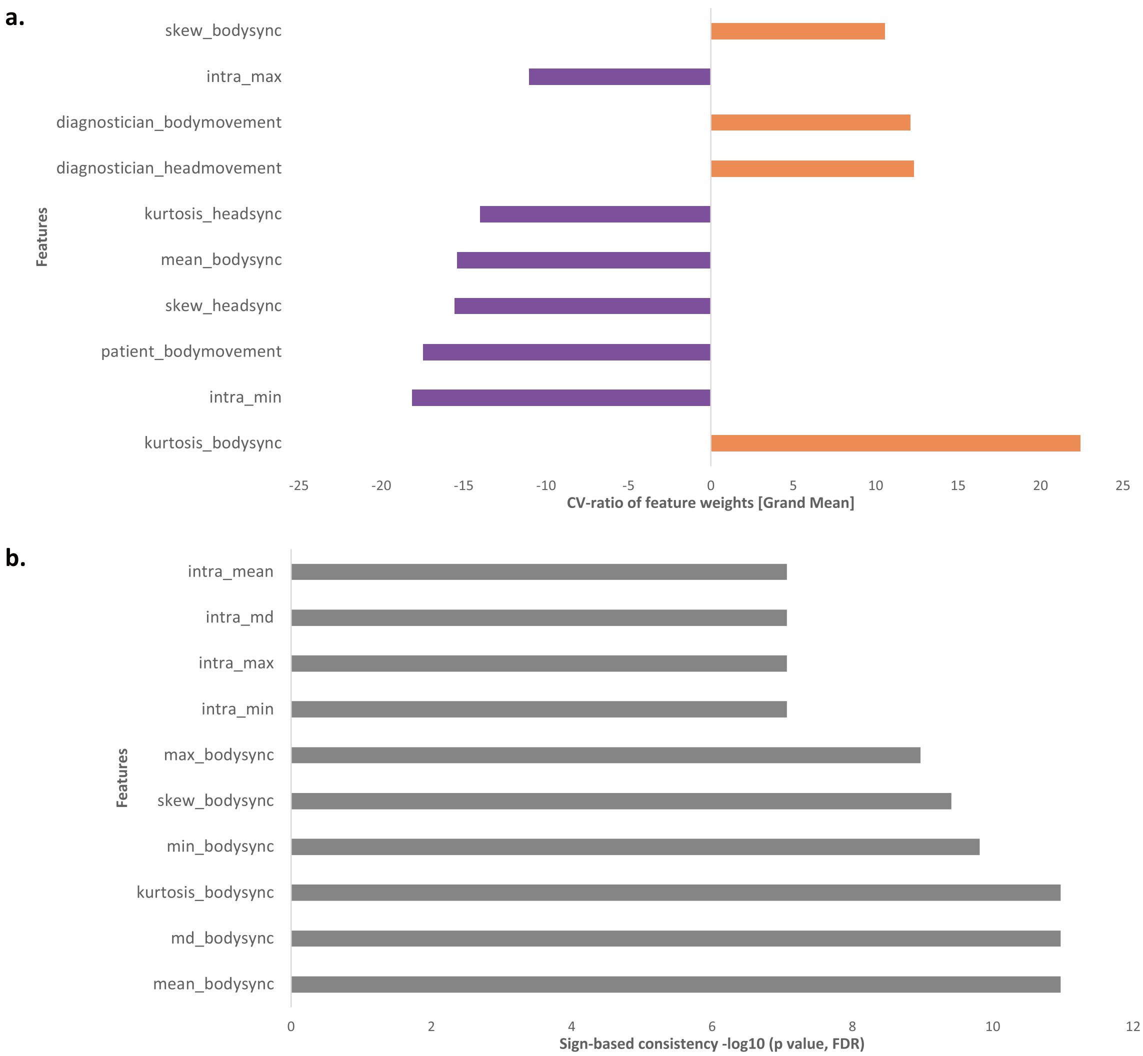

Figure 3. Feature importance of SVM model. Only the ten most important features are depicted. (**a**) Cross-validation ratio. Figure depicts the sum across CV2 folds of the selected CV1 median weights divided by the selected CV1 standard error, which is subsequently divided by the number of CV2 folds. Absolute values > = 2 correspond to p < = .05, absolute values > = 3 correspond to p < = .01. (**b**) Sign-based consistency. The importance of each feature was calculated as the number of times that the sign of the feature was consistent. The depicted scores represent the resulting negative logarithm of p values that were corrected using the Bonferroni-Holm false-discovery rate. Sign-based consistency -10log(p) > = 1.3 is equivalent to *p* < = .05.

classifying neurodevelopmental disorders in general against other psychiatric diagnoses, accuracy was lower than the base model, possibly suggesting a non-verbal social interaction signature specific to autism.

Compared to Kojovic et al[24]., the accuracy of our classifier based on motion energy synchrony data between participants and administrators was reduced. This might be due to several reasons: First, our sample was heterogeneous in terms of diagnosis and age. Instead of classifying ASD against TD children, our classifier was trained on a real-life clinical sample, including a range of diagnoses often co-occurring in autism. Reduced interpersonal synchrony has been reported for adults with other psychiatric diagnoses such as depression[45] and schizophrenia[46]; the former being a frequent co-occurring condition in ASD[47] and the latter sharing phenomenological overlaps with autism[48]. For the sake of completeness, we included information on comorbidities and medication in the supplementary material. However, due to the limited availability of this information for many participants, we did not run any analyses on these data. Future studies should investigate the influence of co-occurring and differential diagnoses by, e.g., running clustering analyses. We controlled for the large age range (5.5–28.7 years) present in our sample by including chronological age as a covariate, leaving no significant residual association of the model's decision scores with age. However, while reduced interpersonal synchrony has been found across the lifespan of individuals on the autism spectrum[15], they have yet to be investigated in direct comparison and the association to general motor skills remains unclear. In our sample, the continuing development of motor skills with age could have resulted in larger heterogeneity of the ability to synchronize and reduced classification performance.

Another approach to increase classification performance could incorporate multi-modal aspects of synchrony. In the present study, we focused on head and body motion synchrony. However, previous research has shown high predictability of, e.g., facial expression synchrony[49]. In fact, we previously found that facial expression

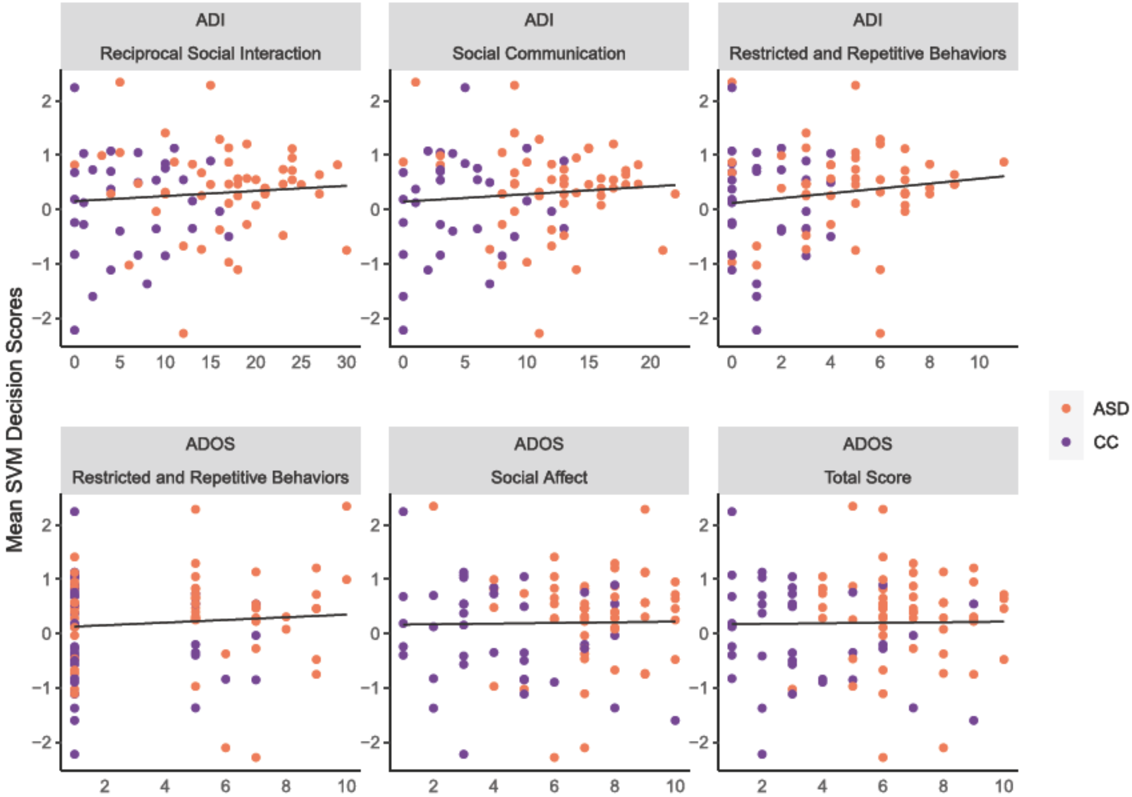

Figure 4. Association between SVM decision scores of MEA classifier and ADI and ADOS-2 domain scores. ADOS-2 scores were transformed to calibrated severity scores following procedures in[42,43]. It should be noted that while the initial class labelling was heavily influenced by both ADOS-2 and ADI-R results, nevertheless, they were not sufficient for diagnosis in this sample.

synchrony between two adults was superior to body movement synchrony in predicting autism[22]. As our videos were filmed from a side perspective, the automated analysis of facial expression with current algorithms requiring the presence of certain facial key points was not possible. However, slight changes in the setup, i.e., including frontal recording of distinct facial movements, could possibly improve predictive performance in the future. Additionally, the synchronization of speech and vocal output in interactions has been found to be reduced in autism[50,51]; although, the generalizability of vocal markers across studies is rather limited as suggested by a recent investigation[52]. Furthermore, closer investigation of more fine-grained non-verbal aspects of social interaction provides the distinct advantage such that markers across the entire spectrum could be explored, given that an estimated 25% of individuals on the autism spectrum are non-verbal[53]. Thus, the approach presented in this study is straightforward to adapt for this purpose.

In our sample, the model that additionally included sex and IQ as predictors, though frequently associated with the likelihood of autism diagnosis[31,32], performed worse than the model entirely based on synchrony and movement. A chi-square test revealed no significant associations between sex and group, suggesting that the distribution of sex between samples was homogeneous, resulting in less discriminative power. Additionally, our clinical control group included individuals with clinical conditions characterized by an imbalance in the male-to-female ratio and/or lower IQ levels (i.e., ADHD or intellectual disability), therefore, making them less distinct from individuals with ASD.

When closely assessing the feature weights, we found that the classification was driven by body synchrony and the clinician's total amount of body movement. More specifically, classification towards the autistic group was driven by greater movement by the administrator, while more participant movement was associated with classification towards the clinical control group. As MEA is a measure of motion energy rather than a measure of movement quality, this might possibly also reflect a unique feature of the diagnostic interview context, i.e., the clinician documenting on a clipboard and tending to document more meticulously if a patient exhibited more conspicuous behaviors. In contrast, our clinical control group included patients with attention deficit hyperactivity disorder (ADHD), a diagnosis commonly associated with elevated movement[44]. While this suggests a tendency of our model to classify movement, rather than synchrony, definite interpretations of the feature weights should be exhibited with caution before being validated on a larger sample.

Contrary to Kojovic and colleagues[24], we could not detect significant associations between our classifier based on synchrony data and ADOS-2 and ADI-R scores in our sample. This could be due to the differences of sample characteristics between both studies. Importantly, the former study classified children with autism against TD children. In clinical outpatient units the representative comparison group is heterogeneous concerning differential diagnoses. As such, our comparison group was more heterogenous with regard to diagnosis as it included children with other psychiatric diagnoses or social communication difficulties. Decreased specificity of ADOS in populations more representative of the real-world clinical setting has been reported in previous studies[54,55]. This was also visible in the overlap of ADOS-2 and ADI-R severity scores between both groups in our sample. On the other hand, the ADOS-2 and ADI-R scores, even though only one part of a clinical best estimate decision, made up the outcome criterion of our classifier (i.e., the diagnosis) to a large extent. Therefore, high associations between the decision scores of our classifier and the outcome criterion could imply a certain circularity (for a detailed discussion of this phenomenon see[56–58]). Though not available for this specific study, future research should employ different measures related to autism diagnosis to be able to further evaluate the underlying mechanisms involved in classification. Further, ADOS-2 is known to not comprehensively represent the entire autistic phenotype, with the scoring algorithm only encompassing a subset of behaviors. This, however, does not imply that other behaviors often manifested in autistic individuals are not associated with autism. One example are motor difficulties which are heavily prevalent in autism[59,60], though not part of the diagnostic algorithm of ADOS. Another example are first impression studies which show that a certain oddity is perceived implicitly at a first, non-verbal, glance, heavily driven by audio-visual, and not conversational content-related cues[61,62]. Moreover, eye-tracking studies reveal distinct eye gaze patterns predictive of autism[63,64], which are not entirely assessed in their quality within ADOS. Thus, automatic measurements provide the possibility to capture implicit, more nuanced behaviors and, therefore, could potentially augment the decision-making process in the future.

In an exploratory analysis to increase accuracy, we employed a SVM classification on a re-labelled sample, grouping ASD with other neurodevelopmental disorders as defined by the DSM-5[44]. However, this model performed slightly above chance, suggesting a synchrony signature specific to autism. Yet, we recognize that this finding needs external validation in order to be further interpreted.

Our study has several limitations that should be considered: First, the videos analyzed in this study were not initially recorded for the purpose of automated machine learning-based analysis procedures. For this reason, the setup varied regarding background and camera angles depending on the different rooms. This could also have contributed to the lack of significant differences in our comparison to pseudo-synchrony (see Supplementary materials S2.3). However, we consider this a feature, rather than a flaw of our approach. When comparing the synchrony values between the different rooms, we could not detect significant differences, underling the scalability of our setup. This is in line with Kojovic and colleagues[24] who investigated their computer vision algorithm with two validation samples, finding minimal influence of video conditions. However, for future reference, we have compiled recommendations for a more standardized recording protocol of ADOS-2 which can be found in the supplementary material (S3). Additionally, we recommend the use of separate microphones to allow for more elaborate analyses of verbal interaction, as well as the use of cameras for more fine-grained facial expression analyses.

Secondly, because our videos differed in length, the use of summary statistics as best estimate measures of interpersonal synchrony were deemed most suitable. However, this approach cannot capture the temporal dynamics of synchrony throughout a conversation. During free-flowing conversations, interactants tend to move in and out of synchrony over time[65], suggesting a certain flexibility in interpersonal alignment. However, no clear evidence exists regarding interview contexts. Thus, future research should investigate synchrony trajectories in more standardized experimental settings.

Moreover, the diagnostic label of the participants in our sample was partly influenced by the results of ADOS-2 and ADI-R. Thus, while the follow-up correlation analyses might shed light on underlying commonalities in autistic symptomatology between participants in our classification, they are not conclusive.

Further, regarding our aim to screen for synchrony as an objective marker to classify autism, we relied on one of the most widely used machine learning algorithms in psychiatric research[33]. Yet, different supervised and unsupervised machine learning algorithms tend to perform well with small data sets and could provide novel insights in the predictiveness of autistic social interaction. As an exploratory analysis, we retrained our winning models with both a random forest, as well as a GLM logistic regression algorithm, the results of which can be found under S2.5 in the supplementary material. However, future research could benefit from in-depth comparisons of detailed performance indices and selected feature spaces using other algorithms.

Finally, and importantly, even though we have implemented a careful and rather conservative cross-validation structure within our model, the sample size in this study is limited, and the results require external validation. K-fold, nested, external cross-validation is suggested as a gold-standard strategy to target the issue of overoptimistic model performances and overfitting, especially when dealing with small sample sizes[66]. As this study served as a proof-of-concept, the present videos were chosen based on a meticulous screening process, which consequently resulted in a high number of exclusions. For example, we only analyzed video excerpts of more than five minute in length and without the use of any external props; the latter of which is an important part of the ADOS-2 assessment. However, we are confident that the high scalability of the methodology used in this study will encourage future data collection and, hence, further external and cross-site validation. In this regard, it will be important to analyze any effects of relaxed inclusion criteria concerning, e.g., the minimum length of an analysis window for a feasible synchrony assessment. Conclusively, our results and the potential implications for their clinical usefulness should be interpreted with strict caution until further validation on larger cohorts. Therefore, further research is needed to assess the potential translation of our models into clinical practice.

In this research, we assessed the predictability of the interpersonal synchrony within excerpts of ADOS-2 as short as five minutes, finding a classification accuracy above chance. Importantly, we used objective motion extraction tools. While clinicians' judgments continue to outperform computational algorithms in their diagnostic precision[67], the notion of digital augmentation of the diagnostic process could prospectively loosen the current bottlenecks caused by resource-exhaustive clinical assessments. Experienced clinical judgement, as well as detailed accounts of the developmental history by caretakers, remain an invaluable element in the professional assessment. However, converging evidence in the field points towards the high potential of neuropsychological and behavioral markers for autism diagnosis (i.e., eye tracking[64,68], movement[12,13], synchrony[20,27]). Considering the aforementioned limitations, we present a viable route toward a digitally assisted diagnostic process in autism. Using a heterogeneous dataset, both in age and technical setup, our classification model could detect ASD in a clinical sample with an above-chance accuracy. With few adjustments regarding the standardization of the experimental setup, including possibilities to record nuanced facial expression and vocal output, the strength of our approach is the high scalability. Ultimately, it remains to assess which markers in combination will reach sufficient diagnostic power to be translated into clinical practice.

Data availability

The datasets generated or analyzed during the study are not publicly available as the IRB approved the data to be used within the research team but could be available from the corresponding author on reasonable request. The preprocessing scripts used during this study are available under https://github.com/jckoe/ SNU_ASDsync.git.

Received: 20 July 2023; Accepted: 1 March 2024
Published online: 07 March 2024

References

1. van't Hof, M. *et al.* Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. *Autism* **25**, 862–873. https://doi.org/10.1177/1362361320971107 (2021).
2. Huang, Y., Arnold, S. R. C., Foley, K. R. & Trollor, J. N. Diagnosis of autism in adulthood: A scoping review. *Autism* **24**, 1311–1327. https://doi.org/10.1177/1362361320903128 (2020).
3. Zeidan, J. *et al.* Global prevalence of autism: A systematic review update. *Autism Res.* **15**, 778–790. https://doi.org/10.1002/aur.2696 (2022).
4. Falkmer, T., Anderson, K., Falkmer, M. & Horlin, C. Diagnostic procedures in autism spectrum disorders: A systematic literature review. *Eur. Child Adolesc. Psychiatry* **22**, 329–340. https://doi.org/10.1007/s00787-013-0375-0 (2013).
5. Lord, C. *et al. Autism Diagnostic Observation Schedule* 2nd edn. (Western Psychological Services, 2012).
6. Lord, C., Rutter, M. & Le Couteur, A. Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* **24**, 659–685 (1994).
7. Kosmicki, J. A., Sochat, V., Duda, M. & Wall, D. P. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Transl. Psychiatry* https://doi.org/10.1038/tp.2015.7 (2015).
8. Küpper, C. *et al.* Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. *Sci. Rep.* **10**, 1–11. https://doi.org/10.1038/s41598-020-61607-w (2020).
9. Moon, S. J., Hwang, J., Kana, R., Torous, J. & Kim, J. W. Accuracy of machine learning algorithms for the diagnosis of autism spectrum disorder: Systematic review and meta-analysis of brain magnetic resonance imaging studies. *JMIR Ment. Health* https://doi.org/10.2196/14108 (2019).
10. Nogay, H. S. & Adeli, H. Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging. *Rev. Neurosci.* **31**, 825–841. https://doi.org/10.1515/revneuro-2020-0043 (2020).
11. Koehler, J. C. & Falter-Wagner, C. M. Digitally assisted diagnostics of autism spectrum disorder. *Front. Psychiatry* https://doi.org/10.3389/fpsyt.2023.1066284 (2023).
12. Anzulewicz, A., Sobota, K. & Delafield-Butt, J. T. Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism. *Sci. Rep.* **6**, 1–13. https://doi.org/10.1038/srep31107 (2016).
13. Jin, X., Zhu, H., Cao, W., Zou, X. & Chen, J. Identifying activity level related movement features of children with ASD based on ADOS videos. *Sci. Rep.* https://doi.org/10.1038/s41598-023-30628-6 (2023).
14. Fulceri, F. *et al.* Motor Skills as moderators of core symptoms in autism spectrum disorders: Preliminary data from an exploratory analysis with artificial neural networks. *Front. Psychol.* **9**, 2683 (2019).
15. McNaughton, K. A. & Redcay, E. Interpersonal synchrony in autism. *Curr. Psychiatry Rep.* https://doi.org/10.1007/s11920-020-1135-8 (2020).
16. Chartrand, T. L. & Bargh, J. A. The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* **76**, 893–910 (1999).
17. Condon, W. S. & Ogston, W. D. A segmentation of behavior. *J. Psychiatr. Res.* **5**, 221–235. https://doi.org/10.1016/0022-3956(67)90004-0 (1967).
18. Cirelli, L. K., Einarson, K. M. & Trainor, L. J. Interpersonal synchrony increases prosocial behavior in infants. *Dev. Sci.* **17**, 1003–1011. https://doi.org/10.1111/desc.12193 (2014).
19. Koehne, S., Hatri, A., Cacioppo, J. & Dziobek, I. Perceived interpersonal synchrony increases empathy: Insights from autism spectrum disorder. *Cognition* https://doi.org/10.1016/j.cognition.2015.09.007 (2015).
20. Koehler, J. C. *et al.* Brief report: Specificity of interpersonal synchrony deficits to autism spectrum disorder and its potential for digitally assisted diagnostics. *J. Autism Dev. Disord.* https://doi.org/10.1007/s10803-021-05194-3 (2021).
21. Ramseyer, F. & Ramseyer, F. Motion energy analysis (MEA). A primer on the assessment of motion from video. *J. Counsel. Psychol* **67**, 536 (2019).
22. Koehler, J.C., Dong, M.S., Nelson, A.M., Fischer, S., Späth, J., Plank, I.S., et al. Machine learning classification of autism spectrum disorder based on reciprocity in naturalistic social interactions. MedRxiv 2022:2022.12.20.22283571. https://doi.org/10.1101/2022.12.20.22283571.
23. Bhat, A. N., Landa, R. J. & Galloway, J. C. Current perspectives on motor functioning in infants, children, and adults with autism spectrum disorders. *Phys. Ther.* **91**, 1116–1129. https://doi.org/10.2522/ptj.20100294 (2011).
24. Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T. & Schaer, M. Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Sci. Rep.* https://doi.org/10.1038/s41598-021-94378-z (2021).
25. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMC Med.* https://doi.org/10.1186/s12916-014-0241-z (2015).
26. Kleinbub, J. R. & Ramseyer, F. T. rMEA: An R package to assess nonverbal synchronization in motion energy analysis time-series. *Psychother. Res.* https://doi.org/10.1080/10503307.2020.1844334 (2020).
27. Georgescu, A. L. *et al.* Machine learning to study social interaction difficulties in ASD. *Front. Robot. AI* **6**, 1–7. https://doi.org/10.3389/frobt.2019.00132 (2019).
28. Ramseyer, F. & Tschacher, W. Nonverbal synchrony of head- and body-movement in psychotherapy: different signals have different associations with outcome. *Front. Psychol.* **5**, 979. https://doi.org/10.3389/fpsyg.2014.00979 (2014).
29. Boersma, P., Weenink, D. No title. Praat: doing phonetics by computer 2019.

30. Georgescu, A. L. *et al.* Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: A motion energy study. *Mol. Autism* https://doi.org/10.1186/s13229-019-0305-1 (2020).
31. Loomes, R., Hull, L. & Mandy, W. P. L. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 466–474. https://doi.org/10.1016/j.jaac.2017.03.013 (2017).
32. Wolff, N., Stroth, S., Kamp-Becker, I., Roepke, S. & Roessner, V. Autism spectrum disorder and IQ–A complex interplay. *Front. Psychiatry* https://doi.org/10.3389/fpsyt.2022.856084 (2022).
33. Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* **14**, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037 (2018).
34. Cortes, C., Vapnik, V. & Saitta, L. *Support-Vector Networks Editor* Vol. 20 (Kluwer Academic Publishers, 1995).
35. Koutsouleris, N., Vetter, C., Wiegand, A. Neurominer (2022).
36. MATLAB 2022.
37. Golland P, Fischl B. Permutation tests for classification: towards statistical significance in image-based studies. **2732**. (2003).
38. Koutsouleris, N. *et al.* Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: A multimodal multisite machine learning analysis. *JAMA Psychiatry* **75**, 1156–1172. https://doi.org/10.1001/jamapsychiatry.2018.2165 (2018).
39. Gómez-Verdejo, V., Parrado-Hernández, E. & Tohka, J. Sign-consistency based variable importance for machine learning in brain imaging. *Neuroinformatics* **17**, 593–609. https://doi.org/10.1007/s12021-019-9415-3 (2019).
40. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**(1), 289–300 (1995).
41. RStudio_Team. RStudio: Integrated Development for R 2020. https://doi.org/10.1037/abn0000182.supp.
42. Hus, V., Gotham, K. & Lord, C. Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *J. Autism Dev. Disord.* **44**, 2400–2412. https://doi.org/10.1007/s10803-012-1719-1 (2014).
43. Hus, V. & Lord, C. The autism diagnostic observation schedule, module 4: Revised algorithm and standardized severity scores. *J. Autism Dev. Disord.* **44**, 1996–2012. https://doi.org/10.1007/s10803-014-2080-3 (2014).
44. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)* 5th edn. (American Psychiatric Pub, 2013).
45. Paulick, J. *et al.* Diagnostic features of nonverbal synchrony in psychotherapy: comparing depression and anxiety. *Cognit. Ther. Res.* https://doi.org/10.1007/s10608-018-9914-9 (2018).
46. Dean, D. J., Scott, J. & Park, S. Interpersonal coordination in schizophrenia: A scoping review of the literature. *Schizophr. Bull.* **47**, 1544–1556. https://doi.org/10.1093/schbul/sbab072 (2021).
47. Ghaziuddin, M., Ghaziuddin, N. & Greden, J. Depression in persons with autism: Implications for research and clinical care. *J. Autism Dev. Disord.* **32**, 299–306. https://doi.org/10.1023/A:1016330802348 (2002).
48. King, B. H. & Lord, C. Is schizophrenia on the autism spectrum?. *Brain Res.* **1380**, 34–41. https://doi.org/10.1016/j.brainres.2010.11.031 (2011).
49. Zampella, C. J., Bennetto, L. & Herrington, J. D. Computer vision analysis of reduced interpersonal affect coordination in youth with autism spectrum disorder. *Autism Res.* **13**, 2133–2142. https://doi.org/10.1002/aur.2334 (2020).
50. Ochi, K. *et al.* Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PLoS One* **14**, 1–22. https://doi.org/10.1371/journal.pone.0225377 (2019).
51. Fasano, R. M. *et al.* A granular perspective on inclusion: Objectively measured interactions of preschoolers with and without autism. *Autism Res.* **14**, 1658–1669. https://doi.org/10.1002/aur.2526 (2021).
52. Rybner, A. *et al.* Vocal markers of autism: Assessing the generalizability of machine learning models. *Autism Res.* **15**, 1018–1030. https://doi.org/10.1002/aur.2721 (2022).
53. Rose, V., Trembath, D., Keen, D. & Paynter, J. The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *J. Intellect. Disabil. Res.* **60**, 464–477. https://doi.org/10.1111/jir.12284 (2016).
54. Sappok, T. *et al.* Diagnosing autism in a clinical sample of adults with intellectual disabilities: How useful are the ADOS and the ADI-R?. *Res. Dev. Disabil.* **34**, 1642–1655. https://doi.org/10.1016/j.ridd.2013.01.028 (2013).
55. Maddox, B. B. *et al.* The accuracy of the ADOS-2 in identifying autism among adults with complex psychiatric conditions. *J. Autism Dev. Disord.* **47**, 2703–2709. https://doi.org/10.1007/s10803-017-3188-z (2017).
56. Kamp-Becker, I. *et al.* Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *Eur. Child. Adolesc. Psychiatry* **27**, 1193–1207. https://doi.org/10.1007/s00787-018-1143-y (2018).
57. Langmann, A., Becker, J., Poustka, L., Becker, K. & Kamp-Becker, I. Diagnostic utility of the autism diagnostic observation schedule in a clinical sample of adolescents and adults. *Res. Autism Spectr. Disord.* **34**, 34–43. https://doi.org/10.1016/j.rasd.2016.11.012 (2017).
58. Kamp-Becker, I. *et al.* Is the combination of ADOS and ADI-R necessary to classify ASD? Rethinking the "Gold Standard" in diagnosing ASD. *Front. Psychiatry* https://doi.org/10.3389/fpsyt.2021.727308 (2021).
59. Licari, M. K. *et al.* Prevalence of motor difficulties in autism spectrum disorder: Analysis of a population-based cohort. *Autism Res.* **13**, 298–306. https://doi.org/10.1002/aur.2230 (2020).
60. Bo, J., Lee, C.-M., Colbert, A. & Shen, B. Do children with autism spectrum disorders have motor learning difficulties?. *Res. Autism Spectr. Disord.* **23**, 50–62. https://doi.org/10.1016/j.rasd.2015.12.001 (2016).
61. Sasson, N. J. & Morrison, K. E. First impressions of adults with autism improve with diagnostic disclosure and increased autism knowledge of peers. *Autism* **23**, 50–59. https://doi.org/10.1177/1362361317729526 (2019).
62. Sasson, N. J. *et al.* Neurotypical peers are less willing to interact with those with autism based on thin slice judgments. *Sci. Rep.* **7**, 40700. https://doi.org/10.1038/srep40700 (2017).
63. Jones, W. & Klin, A. Attention to eyes is present but in decline in 2–6 month-olds later diagnosed with autism. *Nature* **504**, 427–431. https://doi.org/10.1038/nature12715.Attention (2013).
64. Liu, W., Li, M. & Yi, L. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Res* **9**, 888–898. https://doi.org/10.1002/aur.1615 (2016).
65. Mayo, O. & Gordon, I. In and out of synchrony—behavioral and physiological dynamics of dyadic interpersonal coordination. *Psychophysiology* https://doi.org/10.1111/psyp.13574 (2020).
66. Sanfelici, R., Dwyer, D. B., Antonucci, L. A. & Koutsouleris, N. Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: A meta-analytic view on the state of the art. *Biol Psychiatry* **88**, 349–360. https://doi.org/10.1016/j.biopsych.2020.02.009 (2020).
67. Kambeitz-Ilankovic, L., Koutsouleris, N. & Upthegrove, R. The potential of precision psychiatry: What is in reach?. *Br. J. Psychiatry* **220**, 175–178. https://doi.org/10.1192/bjp.2022.23 (2022).
68. Yaneva, V., Ha, L. A., Eraslan, S., Yesilada, Y. & Mitkov, R. Detecting high-functioning autism in adults using eye tracking and machine learning. *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**, 1254–1261. https://doi.org/10.1109/TNSRE.2020.2991675 (2020).

Acknowledgements

We thank Afton Nelson for proofreading this manuscript.

Author contributions
H.Y. and C.F.W. conceptualized the study. J.C.K. and C.F.W. designed the pre-processing procedure. D.S. and G.B. compiled and pre-processed the data. J.C.K. and M.S.D. analyzed and interpreted the data. N.K. supervised the machine learning analysis. J.C.K. wrote the manuscript. All authors read and approved the final manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL. The study design, analysis, interpretation of data and writing was supported by Stiftung Irene (PhD scholarship awarded to JCK) and the German Research Council (Grant numbers 876/3-1 and FA 876/5-1, awarded to CFW). This research was supported by the Institute for Information & Communications Technology Promotion (ITTP) grant funded by the Korean government (MSIT) (No.2019-0-00330, Development of AI Technology for Early Screening of Infant/Child Autism Spectrum Disorders based on Cognition of the Psychological Behavior and Response).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-56098-y.

Correspondence and requests for materials should be addressed to J.C.K. or C.M.F.-W.

5 References

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). 5th ed. Arlington, VA: American Psychiatric Pub; 2013.

[2] Huang Y, Arnold SRC, Foley KR, Trollor JN. Diagnosis of autism in adulthood: A scoping review. Autism 2020;24:1311–27. https://doi.org/10.1177/1362361320903128.

[3] Russell G, Stapley S, Newlove-Delgado T, Salmon A, White R, Warren F, et al. Time trends in autism diagnosis over 20 years: a UK population-based cohort study. J Child Psychol Psychiatry 2022;63:674–82. https://doi.org/10.1111/jcpp.13505.

[4] Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell 2020;180:568-584.e23. https://doi.org/10.1016/j.cell.2019.12.036.

[5] Frye RE, Vassall S, Kaur G, Lewis C, Karim M, Rossignol D. Emerging biomarkers in autism spectrum disorder: a systematic review. Ann Transl Med 2019;7:792–792. https://doi.org/10.21037/atm.2019.11.53.

[6] Bishop SL, Lord C. Commentary: Best practices and processes for assessment of autism spectrum disorder – the intended role of standardized diagnostic instruments. J Child Psychol Psychiatry 2023;64:834–8. https://doi.org/10.1111/jcpp.13802.

[7] Zwaigenbaum L, Penner M. Autism spectrum disorder: advances in diagnosis and evaluation. BMJ 2018;361:k1674. https://doi.org/10.1136/bmj.k1674.

[8] Penner M, Anagnostou E, Ungar WJ. Practice patterns and determinants of wait time for autism spectrum disorder diagnosis in Canada. Mol Autism 2018;9:16. https://doi.org/10.1186/s13229-018-0201-0.

[9] Helt M, Kelley E, Kinsbourne M, Pandey J, Boorstein H, Herbert M, et al. Can children with autism recover? If so, how? Neuropsychol Rev 2008;18:339–66. https://doi.org/10.1007/s11065-008-9075-9.

[10] Freitag CM. Autismus-Spektrum Störung nach DSM-5. Z Kinder Jugend-psychiatr Psychother 2014;42:185–92. https://doi.org/10.1024/1422-4917/a000288.

[11] Taylor LJ, Eapen V, Maybery M, Midford S, Paynter J, Quarmby L, et al. Brief report: An exploratory study of the diagnostic reliability for autism spectrum disorder. J Autism Dev Disord 2017;47:1551–8. https://doi.org/10.1007/s10803-017-3054-z.

[12] Sappok T, Diefenbacher A, Budczies J, Schade C, Grubich C, Bergmann T, et al. Diagnosing autism in a clinical sample of adults with intellectual disabilities: How useful are the ADOS and the ADI-R? Res Dev Disabil 2013;34:1642–55. https://doi.org/https://doi.org/10.1016/j.ridd.2013.01.028.

[13] Maddox BB, Brodkin ES, Calkins ME, Shea K, Mullan K, Hostager J, et al. The Accuracy of the ADOS-2 in Identifying Autism among Adults with Complex Psychiatric Conditions. J Autism Dev Disord 2017;47:2703–9. https://doi.org/10.1007/s10803-017-3188-z.

[14] Mayes SD, Calhoun SL, Mayes RD, Molitoris S. Autism and ADHD: Overlapping and discriminating symptoms. Res Autism Spectr Disord 2012;6:277–85. https://doi.org/10.1016/j.rasd.2011.05.009.

[15] Hommer RE, Swedo SE. Schizophrenia and Autism - Related Disorders. Schizophr Bull 2015;41:313–4. https://doi.org/10.1093/schbul/sbu188.

[16] Chakrabarti S, Fombonne E. Pervasive Developmental Disorders in Pre-school Children. JAMA 2001;285:3093–9. https://doi.org/10.1001/jama.285.24.3093.

[17] May T, Pilkington PD, Younan R, Williams K. Overlap of autism spectrum disorder and borderline personality disorder: A systematic review and

meta-analysis. Autism Research 2021;14:2688–710. https://doi.org/10.1002/aur.2619.

[18] Falter-Wagner CM. Erhöhtes Risiko für Komorbiditäten bei Autismus. InFo Neurologie + Psychiatrie 2020;22:18. https://doi.org/10.1007/s15005-020-1298-1.

[19] Lai MC, Kassee C, Besney R, Bonato S, Hull L, Mandy W, et al. Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. Lancet Psychiatry 2019;6:819–29. https://doi.org/10.1016/S2215-0366(19)30289-5.

[20] Dwyer DB, Falkai P, Koutsouleris N. Machine Learning Approaches for Clinical Psychology and Psychiatry. Annu Rev Clin Psychol 2018;14:91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037.

[21] Bzdok D, Meyer-Lindenberg A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018;3:223–30. https://doi.org/10.1016/j.bpsc.2017.11.007.

[22] Dwyer D, Koutsouleris N. Annual Research Review: Translational machine learning for child and adolescent psychiatry. J Child Psychol Psychiatry 2022. https://doi.org/10.1111/jcpp.13545.

[23] Lord C, Rutter M, DiLavore P, Risi S, Gotham K, Bishop CA: Western Psychological Services S %J T. Autism Diagnostic Observation Schedule Second Edition (ADOS-2) Manual (Part 1): Modules 1–4 2012.

[24] AWMF. Autismus-Spektrum-Störungen im Kindes-, Jugend- und Erwachsenenalter, Teil 1: Diagnostik: Interdisziplinäre S3-Leitlinie der DGKJP und der DGPPN sowie der beteiligten Fachgesellschaften, Berufsverbände und Patientenorganisationen. 2016.

[25] Küpper C, Stroth S, Wolff N, Hauck F, Kliewer N, Schad-Hansjosten T, et al. Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. Sci Rep 2020;10:1–11. https://doi.org/10.1038/s41598-020-61607-w.

[26] Wall DP, Dally R, Luyster R, Jung J-Y, DeLuca TF. Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism. PLoS One 2012;7:e43855. https://doi.org/10.1371/journal.pone.0043855.

[27] Wall DP, Kosmicki J, DeLuca TF, Harstad E, Fusaro VA. Use of machine learning to shorten observation-based screening and diagnosis of autism. Transl Psychiatry 2012;2:e100. https://doi.org/10.1038/tp.2012.10.

[28] Kosmicki JA, Sochat V, Duda M, Wall DP. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Transl Psychiatry 2015;5:e514–e514. https://doi.org/10.1038/tp.2015.7.

[29] Schulte-Rüther M, Kulvicius T, Stroth S, Wolff N, Roessner V, Marschik PB, et al. Using machine learning to improve diagnostic assessment of ASD in the light of specific differential and co-occurring diagnoses. J Child Psychol Psychiatry 2023;64:16–26. https://doi.org/10.1111/jcpp.13650.

[30] Bone D, Goodwin MS, Black MP, Lee CC, Audhkhasi K, Narayanan S. Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises. J Autism Dev Disord 2015;45:1121–36. https://doi.org/10.1007/s10803-014-2268-6.

[31] Zander E, Willfors C, Berggren S, Choque-Olsson N, Coco C, Elmund A, et al. The objectivity of the Autism Diagnostic Observation Schedule (ADOS) in naturalistic clinical settings. Eur Child Adolesc Psychiatry 2016;25:769–80. https://doi.org/10.1007/s00787-015-0793-2.

[32] Anagnostou E, Taylor MJ. Review of neuroimaging in autism spectrum disorders: What have we learned and where we go from here. Mol Autism 2011;2. https://doi.org/10.1186/2040-2392-2-4.

[33] de Marchena A, Eigsti IM. Conversational gestures in autism spectrum disorders: asynchrony but not decreased frequency. Autism Res 2010;3:311–22. https://doi.org/10.1002/aur.159.

[34] Özçalisikan S, Adamson LB, Dimitrova N. Early deictic but not other gestures predict later vocabulary in both typical development and autism. Autism 2016;20:754–63. https://doi.org/10.1177/1362361315605921.

[35] Stagg SD, Slavny R, Hand C, Cardoso A, Smith P. Does facial expressivity count? How typically developing children respond initially to children with autism. Autism 2014;18:704–11. https://doi.org/10.1177/1362361313492392.

[36] Ochi K, Ono N, Owada K, Kojima M, Kuroda M, Sagayama S, et al. Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. PLoS One 2019;14:1–22. https://doi.org/10.1371/journal.pone.0225377.

[37] Asghari SZ, Farashi S, Bashirian S, Jenabi E. Distinctive prosodic features of people with autism spectrum disorder: a systematic review and meta-analysis study. Sci Rep 2021;11. https://doi.org/10.1038/s41598-021-02487-6.

[38] Grossman RB. Judgments of social awkwardness from brief exposure to children with and without high-functioning autism. Autism 2015;19:580–7. https://doi.org/10.1177/1362361314536937.

[39] Sasson NJ, Faso DJ, Nugent J, Lovell S, Kennedy DP, Grossman RB. Neurotypical Peers are Less Willing to Interact with Those with Autism based on Thin Slice Judgments. Sci Rep 2017;7:40700. https://doi.org/10.1038/srep40700.

[40] Wedyan M, Al-Jumaily A. Early diagnosis autism based on upper limb motor coordination in high risk subjects for autism. IRIS 2016 - 2016 IEEE 4th International Symposium on Robotics and Intelligent Sensors: Empowering Robots with Smart Sensors 2017:13–8. https://doi.org/10.1109/IRIS.2016.8066059.

[41] Vabalas A, Gowen E, Poliakoff E, Casson AJ. Applying Machine Learning to Kinematic and Eye Movement Features of a Movement Imitation Task to

Predict Autism Diagnosis. Sci Rep 2020;10:1–13. https://doi.org/10.1038/s41598-020-65384-4.

[42] Ramseyer FT. Motion energy analysis (MEA): A primer on the assessment of motion from video. J Couns Psychol 2020;67:536–49. https://doi.org/10.1037/cou0000407.

[43] Colyer SL, Evans M, Cosker DP, Salo AIT. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. Sports Med Open 2018;4. https://doi.org/10.1186/s40798-018-0139-y.

[44] Zhao Z, Zhu Z, Zhang X, Tang H, Xing J, Hu X, et al. Atypical Head Movement during Face-to-Face Interaction in Children with Autism Spectrum Disorder. Autism Research 2021:1–12. https://doi.org/10.1002/aur.2478.

[45] Dawson G, Campbell K, Hashemi J, Lippmann SJ, Smith V, Carpenter K, et al. Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder. Sci Rep 2018;8:1–7. https://doi.org/10.1038/s41598-018-35215-8.

[46] Plank IS, Traiger LS, Nelson AM, Koehler JC, Lang SF, Tepest R, et al. The role of interpersonal synchrony in forming impressions of autistic and non-autistic adults. Sci Rep 2023;13. https://doi.org/10.1038/s41598-023-42006-3.

[47] McNaughton KA, Redcay E. Interpersonal Synchrony in Autism. Curr Psychiatry Rep 2020;22. https://doi.org/10.1007/s11920-020-1135-8.

[48] Glass D, Yuill N. Social motor synchrony in autism spectrum conditions: A systematic review. Autism 2023. https://doi.org/10.1177/13623613231213295.

[49] Bloch C, Tepest R, Jording M, Vogeley K, Falter-Wagner CM. Intrapersonal synchrony analysis reveals a weaker temporal coherence between gaze and gestures in adults with autism spectrum disorder. Sci Rep 2022;12. https://doi.org/10.1038/s41598-022-24605-8.

[50] Bloch C, Vogeley K, Georgescu AL, Falter-Wagner CM. INTRApersonal Synchrony as Constituent of INTERpersonal Synchrony and Its Relevance for Autism Spectrum Disorder. Front Robot AI 2019;6:1–8. https://doi.org/10.3389/frobt.2019.00073.

[51] Georgescu AL, Koehler JC, Weiske J, Vogeley K, Koutsouleris N, Falter-Wagner C. Machine Learning to Study Social Interaction Difficulties in ASD. Front Robot AI 2019;6:1–7. https://doi.org/10.3389/frobt.2019.00132.

[52] Lord C, Shulman C, DiLavore P. Regression and word loss in autistic spectrum disorders. J Child Psychol Psychiatry 2004;45:936–55. https://doi.org/10.1111/j.1469-7610.2004.t01-1-00287.x.

[53] Norrelgen F, Fernell E, Eriksson M, Hedvall Å, Persson C, Sjölin M, et al. Children with autism spectrum disorders who do not develop phrase speech in the preschool years. Autism 2015;19:934–43. https://doi.org/10.1177/1362361314556782.

[54] Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: A systematic review. Laryngoscope Investig Otolaryngol 2020;5:96–116. https://doi.org/10.1002/lio2.354.

[55] Drimalla H, Scheffer T, Landwehr N, Baskow I, Roepke S, Behnoush B, et al. Towards the automatic detection of social biomarkers in autism spectrum disorder: introducing the simulated interaction task (SIT). NPJ Digit Med 2020;3:1–10. https://doi.org/10.1038/s41746-020-0227-5.

[56] Bone D, Black MP, Ramakrishna A, Grossman R, Narayanan S. Acoustic-Prosodic Correlates of “Awkward” Prosody in Story Retellings from Adolescents with Autism. Interspeech, 2015, p. 1616–20.

[57] Briend F, David C, Silleresi S, Malvy J, Ferré S, Latinus M. Voice acoustics allow classifying autism spectrum disorder with high accuracy. Transl Psychiatry 2023;13. https://doi.org/10.1038/s41398-023-02554-8.

[58] Bone D, Bishop S, Gupta R, Lee S, Narayanan SS. Acoustic-Prosodic and Turn-Taking Features in Interactions with Children with Neurodevelopmental Disorders. In: file:///Users/janakoehler/Downloads/bone_IS161073.PDF, editor. Interspeech, 2016, p. 1185–9.

[59] Wolpert DH. Stacked generalization. Neural Networks 1992;5:241–59. https://doi.org/https://doi.org/10.1016/S0893-6080(05)80023-1.

[60] de Belen RAJ, Bednarz T, Sowmya A, Del Favero D. Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019. Transl Psychiatry 2020;10:333. https://doi.org/10.1038/s41398-020-01015-w.

[61] Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. New England Journal of Medicine 2018;378:981–3. https://doi.org/10.1056/NEJMp1714229.

[62] Cahan EM, Hernandez-Boussard T, Thadaney-Israni S, Rubin DL. Putting the data before the algorithm in big data addressing personalized healthcare. NPJ Digit Med 2019;2. https://doi.org/10.1038/s41746-019-0157-2.

[63] World Health Organization. International Statistical Classification of Diseases and Related Health Problems. 11th ed. 2019.

[64] Rødgaard EM, Jensen K, Vergnes JN, Soulières I, Mottron L. Temporal Changes in Effect Sizes of Studies Comparing Individuals with and Without Autism: A Meta-analysis. JAMA Psychiatry 2019;76:1124–32. https://doi.org/10.1001/jamapsychiatry.2019.1956.

[65] Parlett-Pelleriti CM, Stevens E, Dixon D, Linstead EJ. Applications of Unsupervised Machine Learning in Autism Spectrum Disorder Research: a Review. Rev J Autism Dev Disord 2023;10:406–21. https://doi.org/10.1007/s40489-021-00299-y.

[66] Molnar C. Interpretable Machine Learning A Guide for Making Black Box Models Explainable. 2020.

[67] Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: Experiments in behavior made easy. Behav Res Methods 2019;51:195–203. https://doi.org/10.3758/s13428-018-01193-y.

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

[69] RStudio_Team. RStudio: Integrated Development for R 2020. https://doi.org/10.1037/abn0000182.supp.

[70] Koutsouleris N, Vetter C, Wiegand A. Neurominer 2022.

Appendix: Perspective Article – Machine learning to study social interaction difficulties in ASD

Contributions and reference

This perspective article offered an outlook and proof-of-principle analysis regarding the use of social interaction data for the diagnosis of autism. I was involved in the data curation, as well as the statistical and machine learning analyses. I presented the results to the scientific community during talks and poster presentations at different conferences. I wrote the manuscript in close conjunction with the co-authors and was involved in the revisions and finalization of the published article.

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(November), 1–7. https://doi.org/10.3389/frobt.2019.00132 ***These authors contributed equally.**

<u>Journal Citation Reports:</u>

Impact Factor 2022: 3.4

5 Year Impact Factor 2022: 3.9

Ranked 23/42 in Robotics Journals (Journal Citation Indicator 2022)

frontiers in Robotics and AI

PERSPECTIVE
published: 29 November 2019
doi: 10.3389/frobt.2019.00132

Machine Learning to Study Social Interaction Difficulties in ASD

Alexandra Livia Georgescu[1,2*†], *Jana Christina Koehler*[3*†], *Johanna Weiske*[3], *Kai Vogeley*[2,4], *Nikolaos Koutsouleris*[3] and *Christine Falter-Wagner*[3,5]

[1] Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom, [2] Department of Psychiatry and Psychotherapy, University Hospital of Cologne, Cologne, Germany, [3] Department of Psychiatry and Psychotherapy, Medical Faculty, LMU Munich, Munich, Germany, [4] Institute of Neuroscience and Medicine, Cognitive Neuroscience (INM-3), Research Center Juelich, Jülich, Germany, [5] Institute of Medical Psychology, Medical Faculty, LMU Munich, Munich, Germany

OPEN ACCESS

Edited by:
Cigdem Beyan,
Italian Institute of Technology (IIT), Italy

Reviewed by:
Lori-Ann Rosalind Sacrey,
University of Alberta, Canada
Concetto Spampinato,
University of Catania, Italy

***Correspondence:**
Alexandra Livia Georgescu
alexandra.georgescu@kcl.ac.uk
Jana Christina Koehler
jana.koehler@med.uni-muenchen.de

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: 03 March 2019
Accepted: 13 November 2019
Published: 29 November 2019

Citation:
Georgescu AL, Koehler JC, Weiske J, Vogeley K, Koutsouleris N and Falter-Wagner C (2019) Machine Learning to Study Social Interaction Difficulties in ASD. Front. Robot. AI 6:132. doi: 10.3389/frobt.2019.00132

Autism Spectrum Disorder (ASD) is a spectrum of neurodevelopmental conditions characterized by difficulties in social communication and social interaction as well as repetitive behaviors and restricted interests. Prevalence rates have been rising, and existing diagnostic methods are both extremely time and labor consuming. There is an urgent need for more economic and objective automatized diagnostic tools that are independent of language and experience of the diagnostician and that can help deal with the complexity of the autistic phenotype. Technological advancements in machine learning are offering a potential solution, and several studies have employed computational approaches to classify ASD based on phenomenological, behavioral or neuroimaging data. Despite of being at the core of ASD diagnosis and having the potential to be used as a behavioral marker for machine learning algorithms, only recently have movement parameters been used as features in machine learning classification approaches. In a proof-of-principle analysis of data from a social interaction study we trained a classification algorithm on intrapersonal synchrony as an automatically and objectively measured phenotypic feature from 29 autistic and 29 typically developed individuals to differentiate those individuals with ASD from those without ASD. Parameters included nonverbal motion energy values from 116 videos of social interactions. As opposed to previous studies to date, our classification approach has been applied to non-verbal behavior objectively captured during naturalistic and complex interactions with a real human interaction partner assuring high external validity. A machine learning approach lends itself particularly for capturing heterogeneous and complex behavior in real social interactions and will be essential in developing automatized and objective classification methods in ASD.

Keywords: autism spectrum disorder, machine learning, nonverbal synchrony, support vector machine, motion energy analysis, classification, intrapersonal synchrony, nested cross-validation

INTRODUCTION

Autism spectrum disorder (ASD) is an umbrella term for neurodevelopmental conditions characterized by severe difficulties in social interaction and communication, as well as by repetitive behaviors and restricted interests (American Psychiatric Association, 2013). The prevalence rates of ASD are on the rise (Elsabbagh et al., 2012) and diagnostic services are experiencing an increased demand, in particular in adults seeking diagnostic advice (Murphy et al., 2011). Diagnostics according to medical guidelines are time-consuming, the clinical assessment is complicated by the phenotypic heterogeneity and the language-dependency of assessment with verbal skills being affected by the ASD.

Recently, computational methods of classification have been employed to increase diagnostic reliability and efficiency (Thabtah, 2018). In particular, machine learning (ML) employs algorithms to uncover patterns in complex datasets, which are utilized to improve decision making. ASD diagnostics come down to a decision-making problem that can be supported by automated models (classifiers) using ML to decide whether a newly assessed patient has ASD or not. This works by splitting available data into a training set, on which an algorithm is trained, which is then applied to a test set, resulting in a measure of accuracy of the resulting model. Without making assumptions ML finds classification solutions in a data-driven, bottom-up approach that can be applied to individual prediction making (Dwyer et al., 2018). The primary purposes of using ML are (1) to reduce assessment time to reach a diagnostic decision in order to provide quicker access to health care services, (2) to improve diagnostic reliability, and (3) diagnostic validity by reducing dimensionality of input data so as to identify those features that have the most diagnostic value in ASD (Thabtah, 2018). However, first applications of ML in studies on autism diagnostics have been inconsistent in terms of methodology and outcome, with inconsistent classification accuracy and specificity.

The aim of the present paper is twofold: First, we aim to give an overview of previous research that has attempted to apply ML methods to the classification of ASD, while suggesting guidelines for future research in terms of setup and algorithm design. Second, in a proof-of-principle analysis of data from a social interaction study we aim to establish the potential of using full-body non-verbal behavior data extracted from video recordings of naturalistic social interactions to classify autistic adults.

MACHINE LEARNING APPLICATIONS IN THE CLASSIFICATION OF ASD

First ML attempts in ASD have been used with the aim of shortening ADOS [Autism Diagnostic Observation Schedule, (Lord et al., 2000)] and ADI-R [Autism Diagnostic Interview, (Lord et al., 1994)] administration time by item-reduction yielding a classification accuracy of autism vs. typically-developing (TD) individuals of up to 99.9% (Wall et al., 2012a,b; Bone et al., 2016). In a similar attempt to predict case status words and expressions contained in 8 year old children's developmental evaluations across a network of multiple clinical sites were used for algorithm development (Maenner et al., 2016) with 86.5% prediction accuracy and high concordance with the respective clinician. Home videos of children have been rated by naïve and/or expert raters in terms of ASD-typical behavior and ratings fed into a predictive model along with other features of the diagnostic process (Glover et al., 2018; Tariq et al., 2018). However, while all these first studies using ML in ASD yield fairly high accuracies, the features utilized for classification are still highly subjective and not independent of the respective clinician who bases the diagnostic decision on just those features (circularity). Importantly, when using subjectively influenced data, resulting classification algorithms must be validated in an independent sample in order to prevent circularity.

An increasing number of studies are also using ML to separate individuals with ASD from TD individuals based on neuroimaging data. For example, Ecker et al. (2010) used regional gray and white matter volume measures from whole-brain structural MRI scans of individuals with ASD to investigate their diagnostic value. They used a common variant of ML, the support vector machine (SVM). This is an algorithm aiming at finding a boundary (the so-called "hyperplane") that can be used to optimally classify groups while being able to generalize to new cases (Dwyer et al., 2018). In their sample, the SVM correctly classified individuals with ASD and controls on the basis of their neuroanatomy with about 80% accuracy (Ecker et al., 2010). These original observations are supported by findings from several other neuroimaging studies with similar levels of classification accuracy in younger age groups (Wee et al., 2014), females with ASD (Calderoni et al., 2012) and with various anatomical and functional measurements (Coutanche et al., 2011). These results based on objective data are very promising, although not widely applicable due to high costs.

WHOLE-BODY MOVEMENTS AS A FEATURE IN ML ALGORITHMS IN ASD

Another source of objective data with high potential for diagnostics can be found in the motor domain. Approximately 80% of children with ASD are suspected to exhibit pronounced motor difficulties (Green et al., 2009). Difficulties with balance, gait, movement speed and timed movements have demonstrated to hold a high level of discrimination between children with ASD and TD children (Jansiewicz et al., 2006) and correlate strongly with measures of social and communicative functioning (Parma and de Marchena, 2016). Hence, movement parameters of social interactions in ASD should be investigated for their potential as a diagnostic marker.

Particularly relevant for ASD motor symptomology are gestures and non-verbal communicative behaviors (Georgescu et al., 2014). Accordingly, atypical non-verbal behavior has been included in the DSM-5 criteria for ASD. Yet, the assessment is not straightforward or standardized so far and is hampered by the fact that non-verbal behavior is not necessarily reduced in ASD, but abnormal in the *quality* of its temporal coordination with own verbal output (de Marchena and Eigsti, 2010) and that of an interaction partner. Literature provides evidence for aberrations in temporal processing (Allman and Falter, 2015) and time experience in ASD (Vogel et al., 2019), potentially affecting non-verbal communication. In fact, findings have shown that ASD can be characterized by increased temporal resolution associated with the severity of (non-verbal) communication impairments in ASD (Falter et al., 2012, 2013; Menassa et al., 2018; but see Isaksson et al., 2018).

Recently, movement in ASD has taken up increasing interest (for a review see Bo et al., 2016). In a proof-of-concept study to explore whether low-functioning children with ASD could be identified by means of a kinematic analysis of a simple motor task, 15 children with ASD and 15 TD children (2–4 years) were asked to pick up a ball and drop it into a hole while their movements were recorded using a motion tracker (Crippa et al., 2015). Seventeen kinematic parameters were extracted from the upper-limb movement and seven of these were found significant for discrimination. The classifier distinguished ASD from non-ASD with a classification accuracy of 96.7%, suggesting the validity of assuming a motor signature of ASD. Reach and throw movements of 10 ASD and 10 TD children were analyzed for "peculiar features" using ML and fed into a classification algorithm yielding an accuracy of 92.5% (Perego et al., 2009). Furthermore, Li et al. (2017) extracted 40 kinematic parameters of imitative movements and identified 9 of them that best describe variance of participant groups, resulting in a classification accuracy of 93%.

These studies demonstrate the potential of using kinematic biomarkers in diagnostics of ASD. However, the movements under investigation were staged, thus, highly unnatural. Yet, it has been established that individuals with ASD have particular difficulties with spontaneous "on-line" social interaction requiring intuitive decisions and behavior (Redcay et al., 2013) constituting an urgent need to move this type of research to more external validity and investigate movement in a more naturalistic context.

CLASSIFICATION USING INTRAPERSONAL SYNCHRONY: A PROOF-OF-CONCEPT STUDY

Whole-body movements in more naturalistic conversations were tested for their classification potential in 29 high functioning adults with ASD and 29 TD individuals. The data for this investigation came from a study on interpersonal coordination in dyadic interactions (Georgescu et al., under revision). The autistic participants were diagnosed and recruited at the Autism Outpatient Clinic of the Department of Psychiatry, University Hospital Cologne, Germany. The sample included only patients with the diagnoses high-functioning autism (ICD-10: F84.0) or Asperger syndrome (ICD-10: F84.5). Two medical specialists confirmed the diagnosis independently in clinical interviews, according to the criteria of the International Classification of Diseases (ICD-10) and supplemented by extensive neuropsychological examination. The TD sample was recruited online from the student and staff population at the University of Cologne and the University Hospital of Cologne, Germany. The study was conducted with the approval of the local ethics committee of the Medical Faculty of the University of Cologne. Participants were paired to conduct five 5 min social interaction tasks. Conversational dyads consisted of either two TD individuals, two individuals with ASD or a TD individual with an individual with ASD. An ice-breaker task, two debating tasks, a meal-planning task and a roleplay were included resulting in a total of 145 videos of social interactions (for more information, see Georgescu et al., under revision). All conversations were recorded in a room with standardized artificial lighting and using a high-definition video camera (Panasonic DV C Pro HD P2), mounted on a tripod 320 cm away from the chairs which were 60 cm apart from each other. Since one of the MIXED dyads did not understand instructions on the ice-breaker task, for the purpose of this analysis the whole task was abandoned, resulting in a total of 116 videos submitted for final analysis. Intrapersonal Synchrony between the head and upper body was quantified using Motion Energy Analysis, a widely used semi-automated frame-differencing method that continuously monitors the amount of movement occurring in manually pre-defined regions of interest and the method of lagged cross-correlations (Nagaoka and Komori, 2008; MEA; Altmann, 2011; Ramseyer and Tschacher, 2011). MEA offers the advantage of a constraint-free, objective analysis tool for non-verbal behavior (e.g., Ramseyer and Tschacher, 2011; Schmidt et al., 2012; Paxton and Dale, 2013). This method has been used to capture body movement in different contexts (e.g., Grammer et al., 1999; Ramseyer and Tschacher, 2011, 2014; Schmidt et al., 2012, 2014; Paxton and Dale, 2013). MEA and other frame-differencing methods have been successfully used in clinical research before (e.g., Kupper et al., 2015) and in particular in autism (Noel et al., 2017; Romero et al., 2017, 2018). We followed the MEA pipeline described in Ramseyer and Tschacher (2014). We manually selected two regions of interest (ROI) for each participant, covering (1) the head and (2) the rest of the body including the legs. Changes in grayscale values in these ROIs were detected and separately recorded as two continuous time series measuring the amount of movement in the head and the body region of each person. Data were submitted for quantification of Intrapersonal Synchrony (for more information on the MEA procedure in general, please see Ramseyer and Tschacher, 2014 and on this sample, Georgescu et al., under revision). Input time series were smoothed and scaled to account for different-sized ROIs using custom software in R (package rMEA, Kleinbub and Ramseyer, 2019) and cross-correlated in windows of 60 s with a time lag of ± 5 s (step size 0.04 s). Windows were not allowed to overlap. The resulting 1,004 lagged cross-correlations were then z-standardized and aggregated over the four conditions for every participant, yielding 4,016 features per participant which were implemented in the open-source machine learning tool NeuroMiner (https://www.pronia.eu/neurominer/). A support vector machine with linear kernel was chosen as a classification algorithm, a multivariate supervised learning technique widely

TABLE 1 | Performance metrics of the ASD vs. TD SVM classifier.

True positives/true negatives	28/16
False positives/false negatives	13/1
Accuracy [%]	75.9
Sensitivity [%]	96.6
Specificity [%]	55.2
Area under the curve	0.71

For detailed explanation of performance metrics please refer to Dwyer et al. (2018).

BOX 1 | Minimum requirements for reliable clinical application of ML in ASD research (adapted from Dwyer et al., 2018)

- Combination of objective variables and standard diagnostic measures as input features to classify ASD.
- Use of nested CV as a standard procedure.
- Prevent unstable model outcomes through *k*-fold CV.

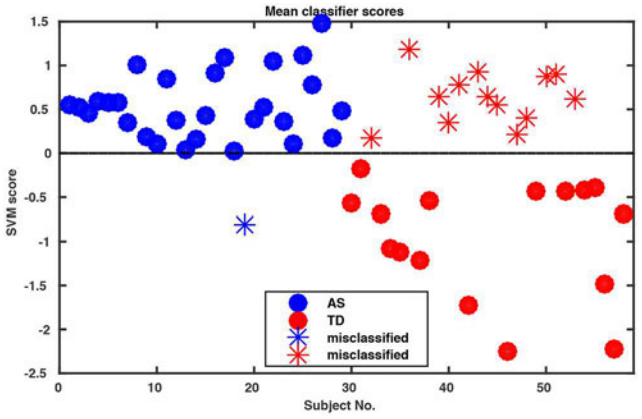

FIGURE 1 | Decision scores of SVM classification performance. The algorithm assigns a score to each participant indicating the probability of this participant as belonging to Group 1 or 2 (in our case ASD vs. TD) where the decision boundary between the two groups is zero. Notably, our algorithm misclassified only one of the ASD participants.

used in psychiatric research (Bone et al., 2016; Duda et al., 2016). Our repeated nested *k*-fold cross-validation (CV) structure consisted of 10-folds and five permutations for the outer cross-validation cycle (CV_2) and repeated 5-by-5-fold inner cross-validation cycle (CV_1), with participants being shuffled prior to each definition of folds. This way, the data available for training was maximized while ensuring enough heterogeneity within the inner test sample to avoid overfitting and create stable models. Parameter optimization was performed in CV_1, while model performance was evaluated in CV_2. Prior to analysis, data was preprocessed using principal component analysis (PCA) for dimensionality reduction, retaining the principal components that cumulatively explained 80% of the variance in each CV_1 fold, and subsequently, scaled feature-wise from 0 to 1. The slack parameter C was estimated in the inner CV cycle using eight parameters ranging from 0.015625 to 16. Overall classification performance resulted in 75.9% accuracy (**Table 1**). Remarkably, sensitivity was 96.6%, correctly classifying all but one individual with ASD (**Figure 1**).

Thus, with a portable and inexpensive video-setup in a naturalistic setting and a semi-automated analysis pipeline, we reached a good diagnostic classification of ASD within four 5 min interaction excerpts on the mere basis of objective motion data. Feeding further clinical and interaction variables into the algorithm promises a high potential for classification (see Future Perspectives section).

METHODOLOGICAL ISSUES IN MACHINE LEARNING APPROACHES TO CLASSIFYING ASD

Unlike e.g., Bone et al. (2016) or Li et al. (2017), most ML studies in ASD research have relied on simple cross-validation (CV) methods. This increases the likelihood of choosing an overly optimistic model (Cawley and Talbot, 2010). We therefore suggest the application of a second layer of CV to allow for parameter selection and model performance evaluation to not be performed on the same data and to prevent overfitting. The test fold is completely held out until parameter optimization within the inner CV cycle is achieved by splitting the training data once more into an (inner) test and (inner) training set. The optimized models can then be tested for generalizability on the outer test fold. This so-called nested CV maximizes generalizability and has now been established as a gold standard procedure in psychiatric research (Dwyer et al., 2018). In order to account for the small sample sizes in ASD research, often predictions are made in a leave-one-out approach whereby only one individual's data is held out in the test set while parameters are optimized on the others (Crippa et al., 2015; Li et al., 2017). Especially, for ASD with its highly heterogeneous phenotype, leave-one-out creates overly variable test sets, rendering model outcomes unstable (Varoquaux et al., 2017). This can be prevented through k-fold nested CV and simultaneous permutation of individual data sets within the inner cross-validation cycle (Dwyer et al., 2018). An overview of best-practice standards is outlined below.

FUTURE PERSPECTIVES

Impairments of non-verbal communication are seen across the entire spectrum of ASD warranting the use as a behavioral biomarker. Yet, its intricacy requires multivariate analysis methods to capture complex interdependencies across domains. Machine learning offers the potential to incorporate high-dimensional data for the detection of underlying mechanisms and classification if certain minimum practice requirements are fulfilled (see **Box 1**).

In our proof-of-principle study, we were able to classify high-functioning adults with ASD from TD adults on the mere basis of non-verbal intrapersonal motion synchrony in social interactions with an accuracy of 75.9%, which can be regarded a conservative estimate on the basis of a state-of-the art ML approach. Due to relatively small sample sizes available with high phenomenological heterogeneity in ASD, it is of utmost importance to choose adequate methods of cross-validation in order to maximize generalizability. The use of repeated nested cross-validation prevents overfitting and should be incorporated as a standard procedure in ML applications. However, given our rather limited sample size, the next steps for future research will be to apply the resulting algorithm to a completely new and larger data set and to investigate its transdiagnostic specificity across different psychiatric disturbances.

Future research should furthermore consider combining multiple non-verbal communication parameters and clinical data (e.g., questionnaires) in order to improve prediction and classification accuracy further and to possibly detect potential associations across domains. For instance, peculiarities in eye-gaze (Merin et al., 2007; Georgescu et al., 2013) and facial expression (McIntosh et al., 2006) in ASD demonstrate feasible approaches.

One future avenue would be to explore methods to quantify non-verbal behavior in a fully-automated fashion. In the present proof-of-principle study, a dataset was used that was analyzed using MEA, a classic frame-differencing approach. It has been shown that MEA is able to capture movements and even complex coordinative patterns to a similar extent as more expensive motion capture equipment such as the Polhemus (Romero et al., 2017). A main advantage for autism research of this method of extracting whole-body motor movement is that it does not involve any wearable technology. Given the hypersensitivity exhibited by many individuals with ASD, not having to add any attachable piece of equipment or body suit to their bodies is helpful. However, while MEA automatically detects pixel changes, corresponding regions of interest are drawn in manually. Although resulting values are standardized, there remains a subjective component. Computer vision tools that can estimate the coordinates of limb positions and even extract gaze location and body poses would offer similar benefits while balancing out subjective biases in the motion extraction process (Marín-Jiménez et al., 2014; Mehta et al., 2017; Tome et al., 2017; Cao et al., 2018). In addition, they offer even more flexibility, given it could be possible to include less strict and standardized experimental setups (no requirement for standardized camera or lighting conditions). However, the validity for movement extraction compared to other standard motion capture methods has not been demonstrated yet. Moreover, such tools vary greatly with respect to their susceptibility to tracking failures, or the type of videos they can support (single vs. multiple agent, indoor vs. outdoor etc.). Overall, with the current methodology that is available for motion extraction, the present semi-automated method offers a realistically applicable diagnostic value. Nevertheless, incredible advances are being made (Li et al., 2018; Tran et al., 2018) such that they are very promising tools for future non-verbal behavior in autism research and beyond.

Taken together, given the recent advances in predictive psychiatry, adequately applied ML offers the potential to fully capture the autistic phenotype in all its complexity with sufficient specificity across psychiatric disorders with a special focus on the spontaneous non-verbal behavior during social encounters with others and irrespective of clinician or site.

DATA AVAILABILITY STATEMENT

The video datasets generated and analysed during the current study are not publicly available due this being identifiable patient data from a sample that did not consent to their data being shared in any form.

ETHICS STATEMENT

Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki (1964). All participants received a monetary compensation for their participation of 50 Euros and were debriefed at the end. The study was conducted with the approval of the local ethics committee of the Medical Faculty of the University of Cologne.

AUTHOR CONTRIBUTIONS

AG, JK, and CF-W contributed equally to the drafting of this manuscript. AG provided the data. JW and JK performed the statistical analysis. All authors contributed to the manuscript revision, read, and approved the submitted version.

FUNDING

JK was supported by Stiftung Irene, LMU excellent seed funding (granted to CF-W; 867900-3) and a German Research Foundation grant (granted to CF-W; FA876/3-1). CF-W was supported by a Bavarian Gender Equality Grant. The data was collected under a postdoctoral grant awarded to AG under the Institutional Strategy of the University of Cologne within the German Excellence Initiative.

ACKNOWLEDGMENTS

Marius Kuschefski, Sevim Koeroglu, and Helen Fischer deserve much appreciation for their assistance with the data collection. We also thank the UK Media Team of the University Hospital of Cologne for their help and support with the video recordings. We are grateful to Wolfgang Tschacher and Fabian Ramseyer for help with Motion Energy Analysis.

REFERENCES

Allman, M., and Falter, C. (2015). "Abnormal timing and time perception in autism spectrum disorder?: a review of the evidence," in *Time Distortions in Mind: Temporal Processing in Clinical Populations,* eds A. Vatakis and M. Allman (Leiden; Boston, MA: Brill), 37–56.

Altmann, U. (2011). "Investigation of movement synchrony using windowed cross-lagged regression," in *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues.* Lecture Notes in Computer Science, Vol. 6800, eds A. Esposito, A. Vinciarelli, K. Vicsi, C. Pelachaud, and A. Nijholt (Berlin; Heidelberg: Springer).

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®).* American Psychiatric Publishing. doi: 10.1176/appi.books.9780890425596

Bo, J., Lee, C.-M., Colbert, A., and Shen, B. (2016). Do children with autism spectrum disorders have motor learning difficulties? *Res. Autism Spectr. Disord.* 23, 50–62. doi: 10.1016/j.rasd.2015.12.001

Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., and Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. *J. Child Psychol. Psychiatr.* 57, 927–937. doi: 10.1111/jcpp.12559

Calderoni, S., Retico, A., Biagi, L., Tancredi, R., Muratori, F., and Tosetti, M. (2012). Female children with autism spectrum disorder: an insight from mass-univariate and pattern classification analyses. *NeuroImage* 59, 1013–1022. doi: 10.1016/j.neuroimage.2011.08.070

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2018). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *ArXiv:1812.08008.* doi: 10.1109/CVPR.2017.143

Cawley, G. C., and Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Machine Learn. Res.* 11, 2079–2107.

Coutanche, M. N., Thompson-Schill, S. L., and Schultz, R. T. (2011). Multi-voxel pattern analysis of fMRI data predicts clinical symptom severity. *NeuroImage* 57, 113–123. doi: 10.1016/j.neuroimage.2011.04.016

Crippa, A., Salvatore, C., Perego, P., Forti, S., Nobile, M., Molteni, M., et al. (2015). Use of machine learning to identify children with autism and their motor abnormalities. *J. Autism Dev. Disord.* 45, 2146–2156. doi: 10.1007/s10803-015-2379-8

de Marchena, A., and Eigsti, I. M. (2010). Conversational gestures in autism spectrum disorders: asynchrony but not decreased frequency. *Autism Res.* 3, 311–322. doi: 10.1002/aur.159

Duda, M., Ma, R., Haber, N., and Wall, D. P. (2016). Use of machine learning for behavioral distinction of autism and ADHD. *Transl. Psychiatr.* 6:e732. doi: 10.1038/tp.2015.221

Dwyer, D. B., Falkai, P., and Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annu. Rev. Clin. Psychol.* 14, 91–118. doi: 10.1146/annurev-clinpsy-032816-045037

Ecker, C., Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Marquand, A., Daly, E. M., et al. (2010). Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. *NeuroImage* 49, 44–56. doi: 10.1016/j.neuroimage.2009.08.024

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., et al. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Res.* 5, 160–179. doi: 10.1002/aur.239

Falter, C. M., Braeutigam, S., Nathan, R., Carrington, S., and Bailey, A. J. (2013). Enhanced access to early visual processing of perceptual simultaneity in autism spectrum disorders. *J. Autism Dev. Disord.* 43, 1857–1866. doi: 10.1007/s10803-012-1735-1

Falter, C. M., Elliott, M. A., and Bailey, A. J. (2012). Enhanced visual temporal resolution in autism spectrum disorders. *PLoS ONE* 7:e32774. doi: 10.1371/journal.pone.0032774

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., and Vogeley, K. (2014). The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Front. Human Neurosci.* 8:807. doi: 10.3389/fnhum.2014.00807

Georgescu, A. L., Kuzmanovic, B., Schilbach, L., Tepest, R., Kulbida, R., Bente, G., et al. (2013). Neural correlates of "social gaze" processing in high-functioning autism under systematic variation of gaze duration. *NeuroImage* 3, 340–351. doi: 10.1016/j.nicl.2013.08.014

Glover, E., Garberson, F., Abbas, H., and Wall, D. P. (2018). Machine learning approach for early detection of autism by combining questionnaire and home video screening. *J. Am. Med. Inform. Assoc.* 25, 1000–1007. doi: 10.1093/jamia/ocy039

Grammer, K., Honda, M., Juette, A., and Schmitt, A. (1999). Fuzziness of nonverbal courtship communication unblurred by motion energy detection. *J. Pers. Soc. Psychol.* 77:487. doi: 10.1037//0022-3514.77.3.487

Green, D., Charman, T., Pickles, A., Chandler, S., Loucas, T., Simonoff, E., et al. (2009). Impairment in movement skills of children with autistic spectrum disorders. *Dev. Med. Child Neurol.* 51, 311–316. doi: 10.1111/j.1469-8749.2008.03242.x

Isaksson, S., Salomäki, S., Tuominen, J., Arstila, V., Falter-Wagner, C. M., and Noreika, V. (2018). Is there a generalized timing impairment in Autism Spectrum Disorders across time scales and paradigms? *J. Psychiatr. Res.* 99, 111–121.

Jansiewicz, E. M., Goldberg, M. C., Newschaffer, C. J., Denckla, M. B., Landa, R., and Mostofsky, S. H. (2006). Motor signs distinguish children with high functioning autism and Asperger's syndrome from controls. *J. Autism Dev. Disord.* 36, 613–621. doi: 10.1007/s10803-006-0109-y

Kleinbub, J. R., and Ramseyer, F. (2019). *rMEA Synchrony in Motion Energy Analysis (MEA) Time-Series.* R package version 1.1.0. Available online at: https://CRAN.R-project.org/package=rMEA

Kupper, Z., Ramseyer, F., Hoffmann, H., and Tschacher, W. (2015). Nonverbal synchrony in social interactions of patients with schizophrenia indicates socio-communicative deficits. *PLoS ONE* 10:e0145882. doi: 10.1371/journal.pone.0145882

Li, B., Sharma, A., Meng, J., Purushwalkam, S., and Gowen, E. (2017). Applying machine learning to identify autistic adults using imitation: an exploratory study. *PLoS ONE* 12:e0182652. doi: 10.1371/journal.pone.0182652

Li, Z., Gavrilyuk, K., Gavves, E., Jain, M., and Snoek, C. G. M. (2018). Videolstm convolves, attends and flows for action recognition. *Comp. Vis. Image Underst.* 166, 41–50. doi: 10.1016/j.cviu.2017.10.011

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule—generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* 30, 205–223. doi: 10.1023/A:1005592401947

Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Develop. Disord*. 24, 659–685. doi: 10.1007/BF02172145

Maenner, M. J., Yeargin-Allsopp, M., Van Naarden Braun, K., Christensen, D. L., and Schieve, L. A. (2016). Development of a machine learning algorithm for the surveillance of autism spectrum disorder. *PLoS ONE* 11:e0168224. doi: 10.1371/journal.pone.0168224

Marín-Jiménez, M. J., Zisserman, A., Eichner, M., and Ferrari, V. (2014). Detecting people looking at each other in videos. *Int. J. Comput. Vis.* 106, 282–296. doi: 10.1007/s11263-013-0655-7

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., and Wilbarger, J. L. (2006). When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Dev. Sci.* 9, 295–302. doi: 10.1111/j.1467-7687.2006.00492.x

Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, M., Seidel, H.-P., et al. (2017). Vnect: real-time 3d human pose estimation with a single rgb camera. *ACM Trans. Graph.* 36:44. doi: 10.1145/3072959.3073596

Menassa, D. A., Braeutigam, S., Bailey, A., and Falter-Wagner, C. M. (2018). Frontal evoked γ activity modulates behavioural performance in Autism Spectrum Disorders in a perceptual simultaneity task. *Neurosci. Lett.* 665, 86–91. doi: 10.1016/j.neulet.2017.11.045

Merin, N., Young, G. S., Ozonoff, S., and Rogers, S. J. (2007). Visual fixation patterns during reciprocal social interaction distinguish a subgroup of 6-month-old infants at-risk for autism from comparison infants. *J. Autism Dev. Disord.* 37, 108–121. doi: 10.1007/s10803-006-0342-4

Murphy, D. G. M., Beecham, J., Craig, M., and Ecker, C. (2011). Autism in adults. New biologicial findings and their translational implications to the cost of clinical services. *Brain Res.* 1380, 22–33. doi: 10.1016/j.brainres.2010.10.042

Nagaoka, C., and Komori, M. (2008). Body movement synchrony in psychotherapeutic counseling: A study using the video-based quantification method. *IEICE Trans. Inf. Syst.* 91, 1634–1640. doi: 10.1093/ietisy/e91-d.6.1634

Noel, J.-P., De Niear, M. A., Lazzara, N. S., and Wallace, M. T. (2017). Uncoupling between multisensory temporal function and nonverbal turn-taking in autism spectrum disorder. *IEEE Trans. Cognit. Develop. Syst.* 10, 973–982. doi: 10.1109/TCDS.2017.2778141

Parma, V., and de Marchena, A. B. (2016). Motor signatures in autism spectrum disorder: the importance of variability. *J. Neurophysiol.* 115, 1081–1084. doi: 10.1152/jn.00647.2015

Paxton, A., and Dale, R. (2013). Frame-differencing methods for measuring bodily synchrony in conversation. *Behav. Res. Methods* 45, 329–343. doi: 10.3758/s13428-012-0249-2

Perego, P., Forti, S., Crippa, A., Valli, A., and Reni, G. (2009). "Reach and throw movement analysis with support vector machines in early diagnosis of autism," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Minneapolis, MN), 2555–2558. doi: 10.1109/IEMBS.2009.5335096

Ramseyer, F., and Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *J. Consult. Clin. Psychol.* 79, 284–295. doi: 10.1037/a0023419

Ramseyer, F., and Tschacher, W. (2014). Nonverbal synchrony of head- and body-movement in psychotherapy: different signals have different associations with outcome. *Front. Psychol.* 5:979. doi: 10.3389/fpsyg.2014.00979

Redcay, E., Dodell-Feder, D., Mavros, P. L., Kleiner, M., Pearrow, M. J., Triantafyllou, C., et al. (2013). Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. *Hum. Brain Mapp.* 34, 2511–2523. doi: 10.1002/hbm.22086

Romero, V., Amaral, J., Fitzpatrick, P., Schmidt, R. C., Duncan, A. W., and Richardson, M. J. (2017). Can low-cost motion-tracking systems substitute a Polhemus system when researching social motor coordination in children? *Behav. Res. Methods* 49, 588–601. doi: 10.3758/s13428-016-0733-1

Romero, V., Fitzpatrick, P., Roulier, S., Duncan, A., Richardson, M. J., and Schmidt, R. C. (2018). Evidence of embodied social competence during conversation in high functioning children with autism spectrum disorder. *PLoS ONE* 13:e0193906. doi: 10.1371/journal.pone.0193906

Schmidt, R. C., Morr, S., Fitzpatrick, P., and Richardson, M. J. (2012). Measuring the dynamics of interactional synchrony. *J. Nonverbal. Behav.* 36, 263–279. doi: 10.1007/s10919-012-0138-5

Schmidt, R. C., Nie, L., Franco, A., and Richardson, M. J. (2014). Bodily synchronization underlying joke telling. *Front. Hum. Neurosci.* 8:633. doi: 10.3389/fnhum.2014.00633

Tariq, Q., Daniels, J., Schwartz, J. N., Washington, P., Kalantarian, H., and Wall, D. P. (2018). Mobile detection of autism through machine learning on home video: a development and prospective validation study. *PLoS Med.* 15:e1002705. doi: 10.1371/journal.pmed.1002705

Thabtah, F. (2018). Machine learning in autistic spectrum disorder behavioral research: a review and ways forward. *Inform. Health Soc. Care* 44, 278–297. doi: 10.1080/17538157.2017.1399132

Tome, D., Russell, C., and Agapito, L. (2017). "Lifting from the deep: convolutional 3d pose estimation from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 2500–2509. doi: 10.1109/CVPR.2017.603

Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., and Paluri, M. (2018). "A closer look at spatiotemporal convolutions for action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 6450–6459. doi: 10.1109/CVPR.2018.00675

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., and Thirion, B. (2017). Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *NeuroImage* 145, 166–179. doi: 10.1016/j.neuroimage.2016.10.038

Vogel, D., Falter-Wagner, C. M., Schoofs, T., Krämer, K., Kupke, C., and Vogeley, K. (2019). Interrupted time experience in autism spectrum disorder: empirical evidence from content analysis. *J. Autism Dev. Disord.* 49, 22–33. doi: 10.1007/s10803-018-3771-y

Wall, D. P., Dally, R., Luyster, R., Jung, J.-Y., and DeLuca, T. F. (2012a). Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PLoS ONE* 7:e43855. doi: 10.1371/journal.pone.0043855

Wall, D. P., Kosmicki, J., DeLuca, T. F., Harstad, E., and Fusaro, V. A. (2012b). Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl. Psychiatr.* 2:e100. doi: 10.1038/tp.2012.10

Wee, C.-Y., Wang, L., Shi, F., Yap, P.-T., and Shen, D. (2014). Diagnosis of autism spectrum disorders using regional and interregional morphological features. *Hum. Brain Mapp.* 35, 3414–3430. doi: 10.1002/hbm.22411

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

Acknowledgements

This thesis is the product of not only my hard work but also the support of many who I would like to express my gratitude to.

First, to my thesis advisory committee: My first supervisor, Prof. Christine Falter-Wagner, for introducing me to autism research, always providing valuable feedback, and teaching me about navigating my way in science. My second supervisor, Prof. Nikolaos Koutsouleris, for sharing his expert knowledge on machine learning and predictive modelling in psychiatry. My third supervisor, Prof. Thomas Geyer, for his knowledgeable advice on psychological research and providing helpful professional guidance.

To Stiftung Irene and Wissenschaftliche Gesellschaft Autismus Spektrum e.V. (WGAS) for providing me with a PhD scholarship.

To my co-authors and colleagues: Mark, whose expertise in guiding me through the analysis process has been invaluable for this work. Irene, whose professional and personal guidance has helped to keep me going throughout the rough patches. My office buddy Afton and the NEVIA lab colleagues, whose knowledge and kindness have enriched not only this work but also daily lab life.

To my family and friends: My deepest appreciation goes out towards my mom and dad, to whom I dedicate this thesis. Throughout life you have always ensured I had every opportunity to pursue my aspirations and taught me valuable lessons about balancing ambition and joy. My brother, my family and my friends, whose unwavering support and genuine interest in my endeavors have been a great source of comfort to me.

Lastly, to my husband: Words cannot express the endless amount of gratitude I feel to always have you by my side throughout this immense journey. It means the world to me, and I share this accomplishment with you.