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),
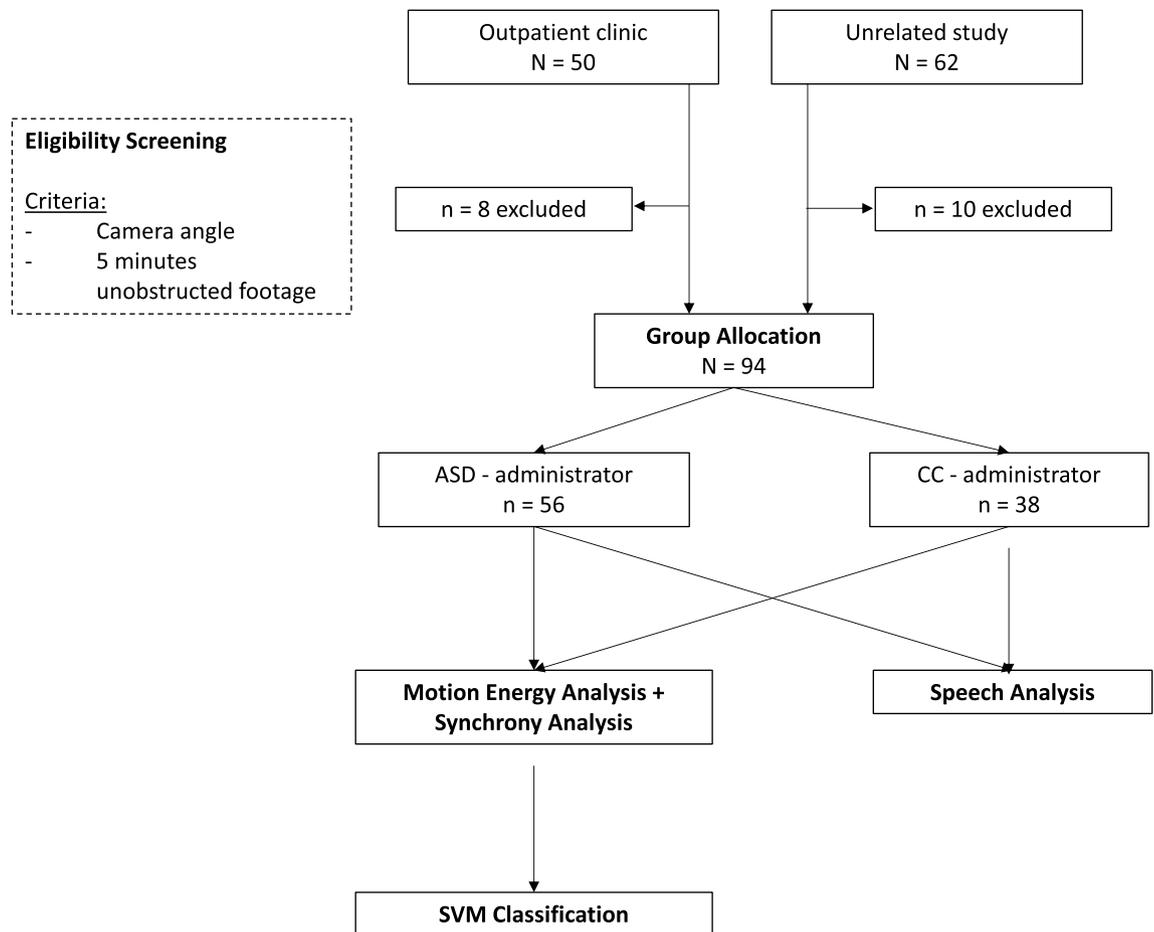
Figure 1. Consort chart of the current sample compilation.

(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

3

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 α = 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 α = 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 α = 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.
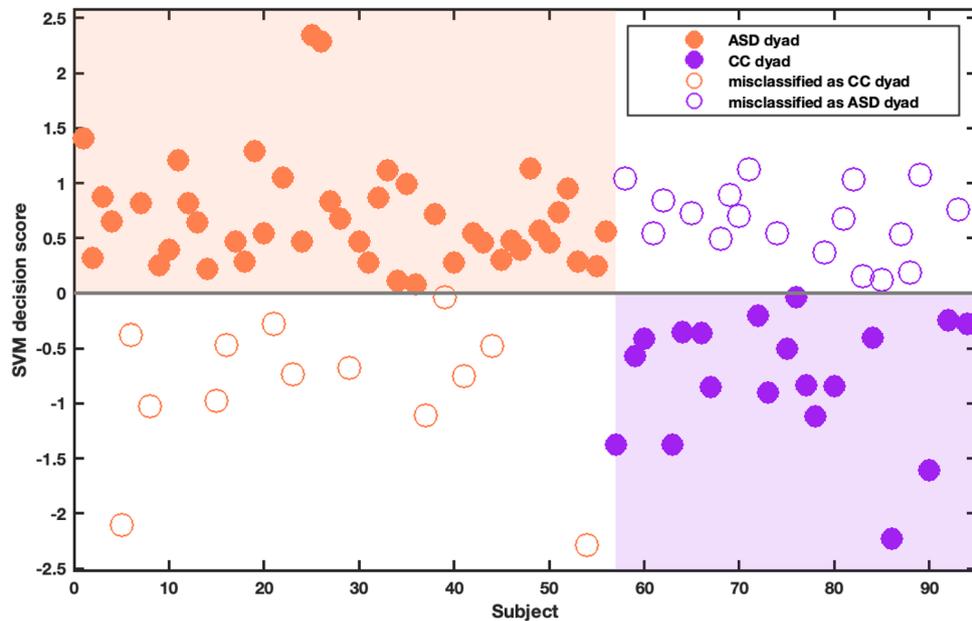
Figure 2. SVM classification results of ASD versus CC patient-administrator dyads. Figure depicts mean classifier scores of each dyad in the model containing only MEA data, resulting in a balanced classification accuracy of 63.4%. The further the score is from the decision boundary, the more likely this 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
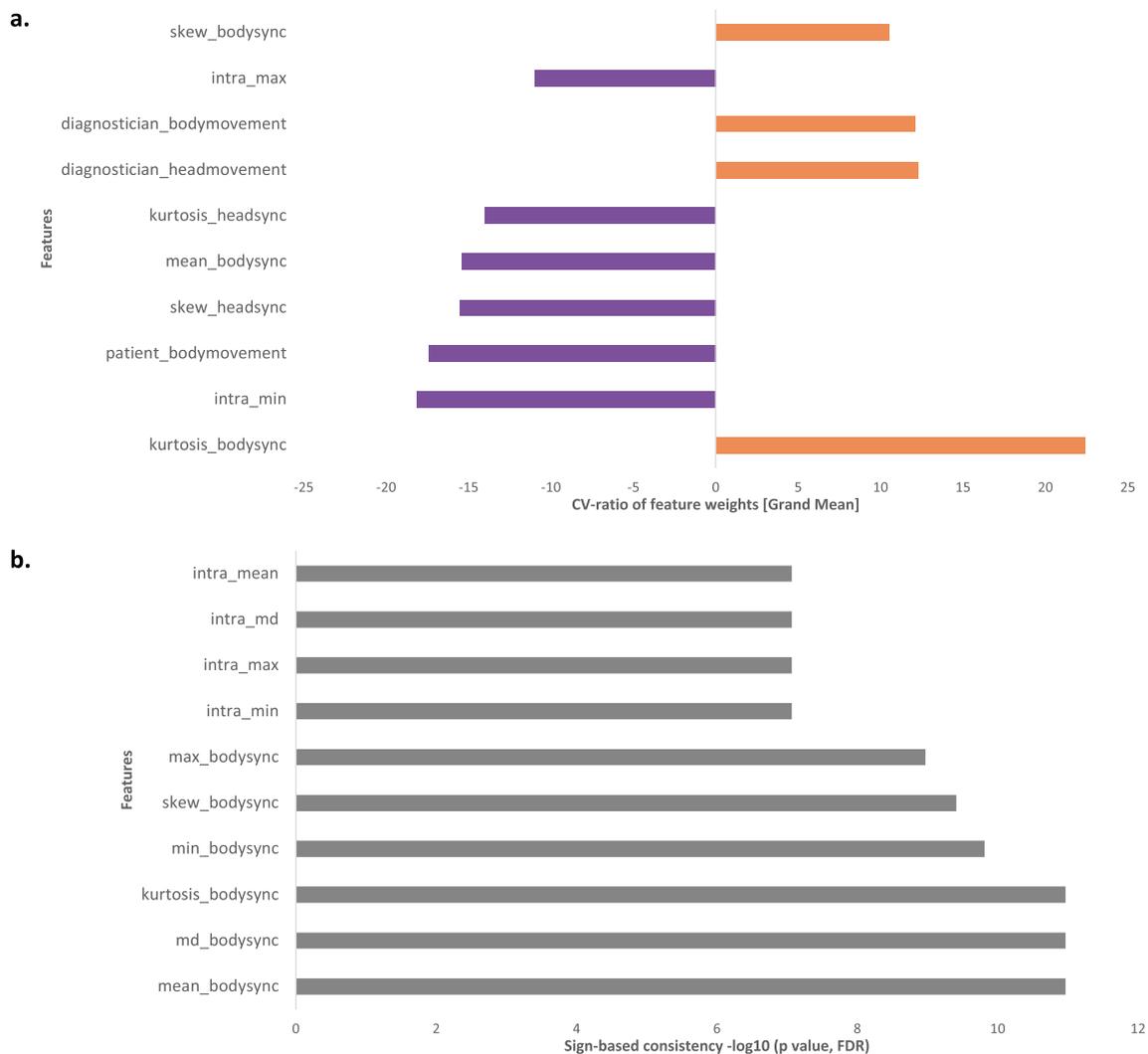
Figure 3. Feature importance of SVM model. Only the ten most important features are depicted. (**a**) Cross-validation ratio. Figure depicts the sum across CV2 folds of the selected CV1 median weights divided by the selected CV1 standard error, which is subsequently divided by the number of CV2 folds. 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
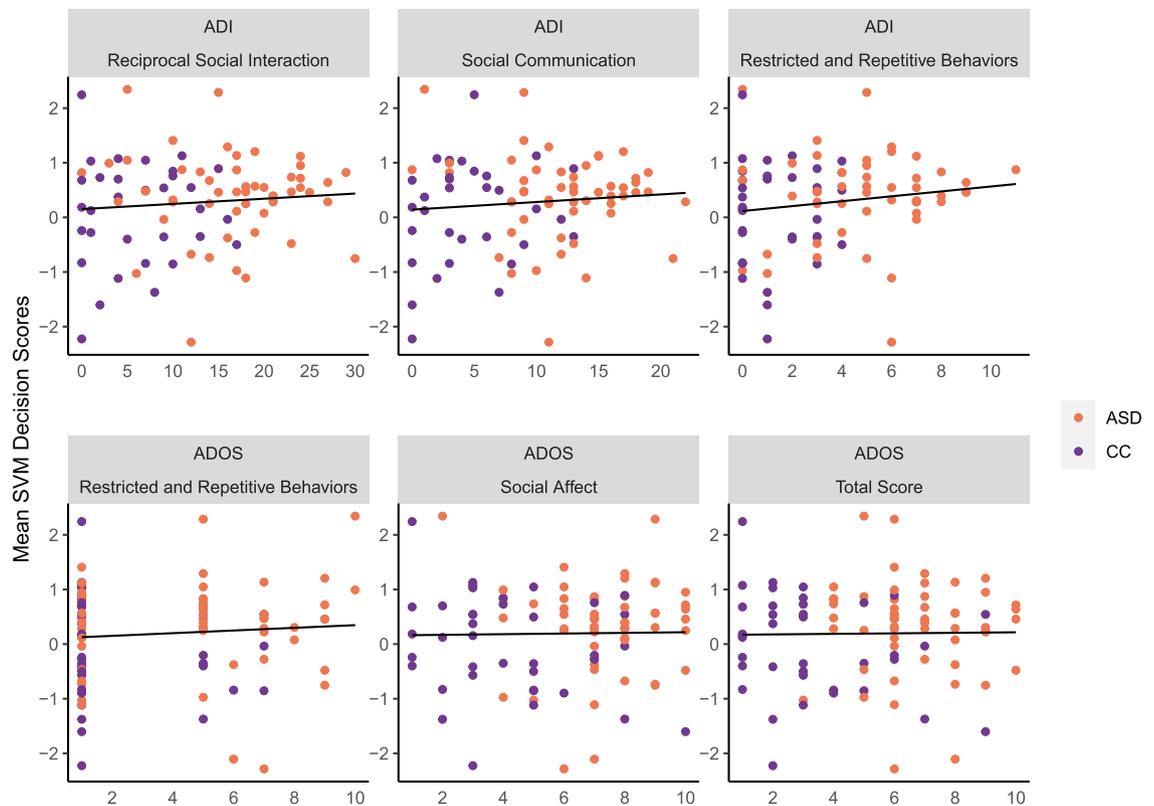
Figure 4. Association between SVM decision scores of MEA classifier and ADI and ADOS-2 domain scores. ADOS-2 scores were transformed to calibrated severity scores following procedures in[42,43]. It should be noted that while the initial class labelling was heavily influenced by both ADOS-2 and ADI-R 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.a., 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

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

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.

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.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.