

Manuscript

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

Koehler, Jana Christina <sup>1\*</sup>, Dong, Mark Sen <sup>1,4\*</sup>, Song, Da-Yea <sup>2,3</sup>, Bong, Guiyoung <sup>2</sup>, Koutsouleris, Nikolaos <sup>1,4,5</sup>, Yoo, Heejeong <sup>2,3+</sup>, Falter-Wagner, Christine M. <sup>1+</sup>

<sup>1</sup>Department of Psychiatry and Psychotherapy, LMU University Hospital, LMU Munich, Germany,

<sup>2</sup>Department of Psychiatry, Seoul National University Bundang Hospital, Seongnam, South Korea,

<sup>3</sup>Department of Psychiatry, Seoul National University College of Medicine, Seoul, Korea, <sup>4</sup>Max Planck

Institute of Psychiatry, Munich, Germany, <sup>5</sup>Institute of Psychiatry, Psychology and Neuroscience, King's College, London, United Kingdom

*\*equally contributing first authors*

*+equally contributing senior authors*

Corresponding author:

Jana C. Koehler, [jana.koehler@med.uni-muenchen.de](mailto:jana.koehler@med.uni-muenchen.de)

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

## Abstract

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 adolescents to predict diagnosis. Specifically, we trained a support vector machine learning model on interpersonal synchrony data recorded in Autism Diagnostic Observation Schedule (ADOS) 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.

## 1 Background

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 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

24 markers is desirable. Hence, several studies investigated structural or functional brain  
25 abnormalities as predictive markers in ASD [9], with promising accuracies especially for younger  
26 children [10]. However, methods such as magnetic resonance imaging lack scalability and are  
27 impractical to implement in standardized clinical practice. Additionally, those approaches pose  
28 special challenges for a sensory-sensitive study population such as autistic individuals. Thus, a more  
29 translational approach uses machine learning for diagnostic classification in ASD through digitally  
30 assisted diagnostics or digital phenotyping [11], which directly taps the symptomatic behavior. This  
31 approach combines the advantages of moving away from the human coding of behaviors while  
32 using more scalable methods such as tablet-based movement data or video analysis via computer  
33 vision techniques. For instance, Anzulewicz and colleagues [12] reported that a machine learning  
34 model trained to identify children with ASD based on their tablet-recorded motion trajectories  
35 performed with an accuracy of 93%. In a recent study, Jin et al. [13] developed a pipeline to  
36 objectively extract movement features correlated with clinicians' ratings from children during ADOS  
37 interviews.

38 Although autism is commonly referred to as a disorder of social interaction, thus, implying a certain  
39 degree of reciprocity, this aspect is challenging to assess objectively. The increasingly studied  
40 phenomenon of reduced interpersonal synchrony in ASD [14] provides such an opportunity. In a  
41 previous study [15], we found reduced interpersonal synchrony as derived from motion energy  
42 analysis (MEA [16]) in diagnostic interviews with autistic adults as compared to those who did not  
43 subsequently receive an autism diagnosis. Furthermore, we explored the predictiveness of  
44 interpersonal synchrony between autistic and non-autistic interactants on multiple modalities,  
45 finding high accuracy for the synchrony of facial and head movements [17]. However, these studies  
46 were conducted with adults, and while motor difficulties in autism tend to persist throughout  
47 adulthood [18], little is known about the predictive power of synchrony alterations in children.

48 In a study on video-based pose estimation, Kojovic et al. [19] investigated videos of ADOS  
49 interviews with small children. Their deep neural network analysis of multiple aspects of non-verbal  
50 interaction differentiated between autistic children and typically-developing (TD) children with an  
51 accuracy of 80.9% and additionally revealed associations between their model and the overall level  
52 of symptomatology. Thus, modeling based on direct extraction of predictive features from  
53 diagnostic videos opens a promising avenue for the clinical setting.

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

62

## 63 2 Methods

64 In the following, we report the details of our prediction model following the Transparent Reporting  
65 of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [20].

### 66 2.1 Sample

67 The ADOS videos and their related datasets were compiled from two different sources at Seoul  
68 National University Bundang Hospital: the patient pool of the psychiatric outpatient clinic for  
69 children and adolescents as well as from a study population of an unrelated study that included the  
70 ADOS. Therefore, the inclusion criteria and available data slightly differed. Patients referred to the

71 outpatient clinic underwent extensive clinical examination to evaluate the presence of an ASD or  
72 differential diagnosis. Additional information on comorbidities and medication for this subsample is  
73 available in the supplementary material (see Supplementary Table S4.1) and was not included in  
74 the final analysis. For the patients from the unrelated study, ADOS was performed as part of the  
75 study protocol, though the diagnosis had either already been suspected or given elsewhere. In  
76 contrast to the outpatient pool, exclusion criteria were applied in the unrelated study which  
77 comprised severe motor impairments restricting patients from engaging in the required ADOS  
78 activities, as well as sensory-related issues or selective mutism. No age limit applied.  
79 For all cases from both sources, the autism diagnosis was confirmed as a best clinical estimate  
80 consensus diagnosis by two psychiatrists, taking into account ADOS and ADI-R results, as well as  
81 other neuropsychological assessments.

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

88 As only ADOS modules three and four include longer instances of free-flowing conversation, the  
89 final sample was comprised of these modules. Excerpts were taken from the tasks *Emotion*,  
90 *Conversation and Reporting*, *Social Difficulties and Annoyance*, *Job/School Life*, *Friends*,  
91 *Relationships*, and *Marriage*, and *Loneliness*. Due to the semi-structured nature of ADOS, the final  
92 clips differed in length, ranging from 5:15 minutes to 14:37 minutes (mean length = 7:20 minutes).

93 Interviews were conducted by six different administrators. All videos had a frame rate of 29.95  
94 seconds.

95 The final dataset consisted of 56 participants with a diagnosis of ASD and 38 participants with other  
96 psychiatric conditions (i.e., n = 4 Intellectual Disability, n = 1 Developmental Delay, n = 10 ADHD, n =  
97 1 Tourette Syndrome, n = 4 Depressive Disorder, n = 1 Social Phobia, n = 1 Anxiety Disorder, n = 2  
98 Bipolar Disorder) or within the wider autism phenotype (n = 2), as well as n = 12 typically-  
99 developing (TD) children (including 8 unaffected siblings). This resulted in two diagnostic group  
100 allocations: ASD-administrator or clinical control (CC)-administrator.

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

109

110 [PLACEHOLDER FIGURE 1]

111

## 112 2.2 Video pre-processing and synchrony computation

113 Motion Energy Analysis (MEA) [16] was applied to all video clips, defining two regions of interest  
114 per participant and administrator (head and upper body). MEA extracts frame-to-frame gray-scale  
115 pixel changes. Keeping the camera position, lighting and background constant, all pixel changes  
116 above a manually set threshold represent movement within the regions of interest. After careful  
117 visual inspection of the resulting data quality, a threshold of eight was chosen for all videos.

118 Raw motion energy time series were subsequently forwarded to pre-processing using the RStudio  
119 package rMEA [21]. Videos were filmed in four different rooms. To account for potential biases of  
120 any distortions in the video signals, all MEA time series were scaled by standard deviation and  
121 smoothed with a moving average of 0.5 seconds, according to the default setting in rMEA. A  
122 comparison analysis of potential feature differences depending on the room can be found in the  
123 supplementary material (S2.2).

124 Interpersonal synchrony between participant and administrator in their head and body motion was  
125 computed with windowed cross-lagged correlations. In line with a previous analysis of diagnostic  
126 interviews with autistic adults [15], a window size of 60 seconds was chosen. To capture all  
127 instances of synchrony, time series were cross correlated with lags of 5 seconds and increments of  
128 30 seconds. All values in the resulting cross-correlation matrices were converted to absolute Fisher  
129 Z values. Time series were subsequently shuffled and randomly paired into 500 pseudodyads. Cross  
130 correlations were conducted in the same manner, yielding a measure of pseudosynchrony per  
131 artificial dyad. They were subsequently compared to the interpersonal synchrony values to assess  
132 whether the interpersonal synchrony values were above-chance. Detailed results can be found in  
133 the supplementary material (S2.3).

134 Moreover, following procedures from Georgescu et al. [22], intrapersonal head and body  
135 coordination was computed for every patient, using window sizes of 30 seconds, lags of 5 seconds  
136 and a step size of 15 seconds.

137 Lastly, we derived the head and body movement quantity per participant from the respective MEA  
138 time series. Following previous procedures [15,23], they were defined as the number of frames  
139 with changes in motion energy divided by the total number of frames, resulting in four values per  
140 dyad (two for participant and administrator, respectively).



141 In addition to the processing of motion, we submitted our videos to an exploratory vocal output  
142 analysis. For this purpose, the audio tracks of the selected clips were processed with the software  
143 Praat [24] to semi-automatically extract annotations of intervals of vocalizations and silences. As  
144 there was no speaker distinction within the audio tracks, this analysis was considered exploratory  
145 and is not included in the main machine learning analysis. Details can be found in the  
146 supplementary material (S1.1).

147

### 148 2.3 Feature engineering

149 Because the videos in our sample varied in both length and conversational content (see  
150 Supplementary Material S2.1), as well as to account for the interview context, our aim was to gain  
151 the best estimate of the overall synchrony (i.e., instances in and out of synchrony). For this reason,  
152 summary statistics of each cross-correlation matrix were computed (i.e., minimum, maximum,  
153 mean, median, standard deviation, skew, and kurtosis), resulting in seven features per participant-  
154 administrator dyad and region of interest (ROI). We additionally computed the same summary  
155 statistics for the intrapersonal head-body coordination of each participant. This approach slightly  
156 differed from a previous study [17], where we were interested in the trajectory of maximum  
157 synchrony instances during naturalistic and free-flowing conversations. To comply with previous  
158 procedures, we additionally computed a feature set using a peak-picking algorithm to obtain a  
159 measure of the trajectory of the highest synchrony instances during each interview. Details and  
160 results can be found in the supplementary material (S2.4).

161 The final feature set for each dyad consisted of 25 features per participant-administrator dyad (see  
162 Supplementary Table S4.2): seven interpersonal synchrony features per dyad and ROI (head and  
163 body), seven features for the intrapersonal head-body coordination of every participant, as well as  
164 four features for the individual amount of head and body movement of both interactants. IQ and

165 sex of the participant were additionally included as features in a second model, as both are  
166 frequently associated with autism symptomatology and the likelihood of receiving a diagnosis  
167 [25,26].

168

#### 169 2.4 Support Vector Machine (SVM) Learning Analyses

170 We trained two separate binary machine learning models to classify between dyad type: (1) a  
171 “behavioral” model containing only synchrony data objectively extracted from the videos (MEA),  
172 and (2) a model additionally containing sex and IQ as demographic features (MEA + DEMO). Age  
173 was regressed out in both models. By constructing two separate models, we could explore whether  
174 demographic features frequently associated with ASD might improve the purely behavioral  
175 predictive performance. A L1-loss LIBSVM algorithm was chosen for both models, as it is frequently  
176 used in psychiatric research [27], known to perform robustly with reduced sample sizes [28]. In  
177 each model, the SVM algorithm independently modeled a linear relationship between features and  
178 classification labels by optimizing a linear hyperplane in a high-dimensional feature space to  
179 maximize separability between the dyads. Subsequently, the data was projected into the linear  
180 kernel space and their geometric distance to the decision boundary was measured. Thus, every  
181 dyad was assigned a predicted classification label and a decision score.

182 Machine learning analyses were conducted in NeuroMiner (Version 1.1;  
183 <https://github.molgen.mpg.de/pages/LMU-Neurodiagnostic-Applications/NeuroMiner.io/>) [29], an  
184 open-source mixed MATLAB [30]-Python-based machine learning library. To prevent any possibility  
185 of information leakage between training and testing data, our diagnostic classifiers were cross-  
186 validated in a repeated, nested, stratified cross-validation scheme. We used ten folds and ten  
187 permutations in the outer CV loop (CV2) and ten folds and one permutation in the inner loop (CV1).  
188 Specifically, at the CV2 level, we iteratively held back 9 or 10 participant-administrator dyads as test

189 samples, while the rest of the data entered the CV1 cycle, where the data were again split into  
190 training and validation sets. This way training and testing data were strictly separated, with hyper-  
191 parameter tuning happening entirely within the inner loop while the outer loop was exclusively  
192 used to measure the classifier's generalizability to unseen data and generate decision scores for  
193 each dyad in this partition. This process was repeated for the remaining folds, after which the  
194 participants were reshuffled within their group and the process was repeated nine times, producing  
195 10x10=100 decision scores for each held out participant. The final median decision score of each  
196 held out dyad was computed from the scores provided by the ensemble of models in which given  
197 dyad had not been used at the CV1 level for training or hyperparameter optimization. Additionally,  
198 the stratified design ensured that the proportion of the diagnostic groups in every fold would  
199 adequately reflect the proportion of the diagnostic group in the full sample and, thus, guarantee  
200 that each class is equally represented in each test fold to avoid bias during model training.  
201 The preprocessing settings for the respective models can be found in Table 1.  
202 Class imbalances were corrected for by hyperplane weighting. Balanced Accuracy (BAC =  
203 [sensitivity+specificity]/2) was used as the performance criterion for hyperparameter optimization.  
204 The C parameter was optimized in the CV1 cycle using 11 parameters within the following range:  
205 0.0156, 0.0312, 0.0625, 0.1250, 0.2500, 0.5000, 1, 2, 4, 8, and 16, which represent the default  
206 settings in NeuroMiner [29]. Model significance was assessed through label permutation testing  
207 [31], with a significance level  $\alpha = .05$  and 1000 permutations. The permutation testing procedure  
208 determines the statistical significance of a model's performances (i.e., BAC) by using the current  
209 data compared to models trained on the dataset but with the labels randomly permuted. The  
210 predictive pattern of the models was extracted using cross-validation ratio (CVR) and sign-based  
211 consistency. Firstly, CVR was computed as the mean and standard error of all normalized SVM  
212 weight vectors concatenated across the entire nested CV structure. CVR measures pattern element

213 stability and was defined as the sum across CV2 folds of the CV1 median weights divided by their  
214 respective CV1 standard error, all of which was subsequently divided by the number of CV2 folds  
215 [32]. Secondly, we used the sign-based-consistency method [33] to test the stability of the  
216 predictive pattern by examining the consistency of positive and negative signs of the feature weight  
217 values across all models in the ensemble (see Supplementary Material S1.2 for additional  
218 information). Feature stability was assessed for statistical significance at  $\alpha = .05$ , using the  
219 Benjamini-Hochberg procedure of false discovery rate correction (FDR) [34].

220

221 [PLACEHOLDER TABLE 1]

222

## 223 2.5 Associations of SVM model and clinical variables

224

225 To investigate potential underlying clinical factors associated with our classification models, post-  
226 hoc correlation analyses with the SVM decision scores and ADOS, as well as ADI scores were  
227 performed in RStudio (version 2022-07.2) [35]. A dyad's predicted SVM decision score represents  
228 their distance from the hyperplane. ADOS scores included domain scores for social affect (SA) and  
229 restricted and repetitive behaviors (RRB), as well as the total score (Total). Because our sample  
230 included data from both modules three and four, calibrated severity scores [36,37] were used for  
231 the correlation analyses for better comparison. For ADI-R, ratings on three subdomains based on  
232 caregiver report were used: reciprocal social interaction (A), social communication (B), and  
233 restricted and repetitive behaviors (C). Statistical significance was determined at  $\alpha = .05$  and two-  
234 sided  $p$  values were corrected for multiple comparisons using FDR.

## 235 2.6 Exploratory SVM analysis

236 To further address the specificity of synchrony, given that phenotypic movement difficulties overlap  
237 in neurodevelopmental disorders (e.g., dyspraxia and autism, or hyperkinetic movement in ADHD),  
238 the MEA classifier was retrained within the same sample but using different class labels: i) a  
239 neurodevelopmental disorders class, which grouped all 74 patients with a diagnosis of a  
240 neurodevelopmental disorder as defined by DSM-5 [38] (n = 56 ASD, n = 10 ADHD, n = 1  
241 Developmental Delay, n = 1 Tourette Syndrome, n = 4 Intellectual Disability, n = 2 Broad  
242 Spectrum/Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS)), and ii) a  
243 clinical control group consisting of the 20 patients with other psychiatric diagnoses or typically-  
244 developing participants (n = 12 TD including 8 unaffected siblings, n = 1 Anxiety Disorder, n = 2  
245 Bipolar Disorder, n = 4 Depressive Disorder, n = 1 Social Phobia). The stratified CV structure was  
246 adapted accordingly.

247

## 248 3 Results

### 249 3.1 Sample Description

250

251 A description of the final sample grouped according to the ADOS module can be found in Table 2. A  
252 chi-square test of independence revealed no significant association between the diagnostic group  
253 and sex ( $\chi^2(1,94) = .045, p = .831$ ). Though naturally participants across both modules differed in  
254 age, there was no significant difference in age between diagnostic groups within each module.  
255 Because final diagnosis was partly based on ADOS and ADI-R results, autistic patients across both  
256 modules had significantly higher ADOS as well as ADI-R scores compared with the clinical control  
257 group. Best-estimate IQ values were significantly higher in the CC group for module 3. This effect

258 was reversed in module 4, with autistic patients scoring significantly higher on their respective IQ  
259 assessment.

[PLACEHOLDER TABLE 2]

### 260 3.2 SVM Classification Performance and Feature Importance

261 Using only motion energy analysis data and regressing out age, our MEA model was able to classify  
262 interview dyads with an autistic participant as opposed to those with other psychiatric diagnoses  
263 with a BAC of 63.4% (Figure 2). Detailed performance metrics, i.e., sensitivity, specificity, accuracy,  
264 positive and negative predictive values, and Area-Under-the-Receiver-Operating-Curve (AUC) can  
265 be found in Table 3. There was no significant residual association between age ( $M = 13.53$ ,  $SD =$   
266  $4.70$ ) and the model's resulting decision scores ( $M = .19$ ,  $SD = .89$ ) after regressing out age during  
267 pre-processing ( $r_{Pearson} = .06$ ,  $p = .558$ ). The model that additionally included sex and IQ as features  
268 (MEA + DEMO) had a lower BAC of 59.4% (Sensitivity = 71.4%, Specificity = 47.4%,  $AUC = .58$ [ $CI =$   
269  $.46 - .70$ ], also see S4.3 Supplementary Table).

[PLACEHOLDER FIGURE 2]

270 A closer investigation of the cross-validation ratio revealed that classification towards the autism-  
271 administrator dyads was driven by higher kurtosis and skewness of their body synchrony values  
272 (Figure 3a). This means that a dyad with more pronounced outliers in their body synchrony,  
273 especially in the positive direction, was considered more autistic. In contrast, our model considered  
274 higher mean body synchrony values as non-autistic. Sign-based consistency revealed that this effect  
275 was relatively stable (Figure 3b). Interestingly, the opposite effect was visible for head synchrony:  
276 higher kurtosis and skewness of head synchrony values were considered non-autistic, whereas

277 higher mean head synchrony values were considered autistic. However, this was not consistent and  
278 of less feature importance than body synchrony.

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

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

[PLACEHOLDER FIGURE 3]

### 284 3.3 Associations between SVM model scores and clinical variables

285  
286 We conducted a range of correlation analyses of the resulting SVM scores of our models with  
287 ADOS-2 [5] and ADI-R [6] domain and total scores (Figure 4). ADI-R data was incomplete for ten  
288 participants, who were discarded from the respective analysis.

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

293

[PLACEHOLDER FIGURE 4]

294

### 295 3.4 Exploratory SVM analysis: NDD vs. CC

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

299

300

[PLACEHOLDER TABLE 3]

301

## 302 4 Discussion

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

315 Compared to Kojovic et al. [19], the accuracy of our classifier based on motion energy synchrony  
316 data between participants and administrators was reduced. This might be due to several reasons:  
317 First, our sample was heterogeneous in terms of diagnosis and age. Instead of classifying ASD



318 against TD children, our classifier was trained on a real-life clinical sample, including a range of  
319 diagnoses often co-occurring in autism. Reduced interpersonal synchrony has been reported for  
320 adults with other psychiatric diagnoses such as depression [39] and schizophrenia [40]; the former  
321 being a frequent co-occurring condition in ASD [41] and the latter sharing phenomenological  
322 overlaps with autism [42]. For the sake of completeness, we included information on comorbidities  
323 and medication in the supplementary material. However, due to the limited availability of this  
324 information for many participants, we did not run any analyses on these data. Future studies  
325 should investigate the influence of co-occurring and differential diagnoses by, e.g., running  
326 clustering analyses. We controlled for the large age range (5.5 – 28.7 years) present in our sample  
327 by including chronological age as a covariate, leaving no significant residual association of the  
328 model's decision scores with age. However, while reduced interpersonal synchrony has been found  
329 across the lifespan of individuals on the autism spectrum [14], they have yet to be investigated in  
330 direct comparison and the association to general motor skills remains unclear. In our sample, the  
331 continuing development of motor skills with age could have resulted in larger heterogeneity of the  
332 ability to synchronize and reduced classification performance. Another approach to increase  
333 classification performance could incorporate multi-modal aspects of synchrony. In the present  
334 study, we focused on head and body motion synchrony. However, previous research has shown  
335 high predictability of, e.g., facial expression synchrony [43]. In fact, we previously found that facial  
336 expression synchrony between two adults was superior to body movement synchrony in predicting  
337 autism [17]. As our videos were filmed from a side perspective, the automated analysis of facial  
338 expression with current algorithms requiring the presence of certain facial key points was not  
339 possible. However, slight changes in the setup, i.e., including frontal recording of distinct facial  
340 movements, could possibly improve predictive performance in the future. Additionally, the  
341 synchronization of speech and vocal output in interactions has been found to be reduced in autism

342 [44,45]; although, the generalizability of vocal markers across studies is rather limited as suggested  
343 by a recent investigation [46]. Furthermore, closer investigation of more fine-grained non-verbal  
344 aspects of social interaction provides the distinct advantage such that markers across the entire  
345 spectrum could be explored, given that an estimated 25% of individuals on the autism spectrum are  
346 non-verbal [47]. Thus, the approach presented in this study is straightforward to adapt for this  
347 purpose.

348 When closely assessing the feature weights, we found that the classification was driven by body  
349 synchrony and the clinician's total amount of body movement. More specifically, classification  
350 towards the autistic group was driven by greater movement by the administrator, while more  
351 participant movement was associated with classification towards the clinical control group. As MEA  
352 is a measure of motion energy rather than a measure of movement quality, this might possibly also  
353 reflect a unique feature of the diagnostic interview context, i.e., the clinician documenting on a  
354 clipboard and tending to document more meticulously if a patient exhibited more conspicuous  
355 behaviors. In contrast, our clinical control group included patients with attention deficit  
356 hyperactivity disorder (ADHD), a diagnosis commonly associated with elevated movement [38].  
357 While this suggests a tendency of our model to classify movement, rather than synchrony, definite  
358 interpretations of the feature weights should be exhibited with caution before being validated on a  
359 larger sample.

360 Contrary to Kojovic and colleagues [19], we could not detect significant associations between our  
361 classifier based on synchrony data and clinical variables in our sample. This could be due to the  
362 differences of sample characteristics between both studies. Importantly, the former study classified  
363 children with autism against TD children. In clinical outpatient units the representative comparison  
364 group is heterogeneous concerning differential diagnoses. As such, our comparison group was  
365 more heterogenous with regard to diagnosis as it included children with other psychiatric diagnoses

366 or social communication difficulties. Decreased specificity of ADOS in populations more  
367 representative of the real-world clinical setting has been reported in previous studies [48,49]. This  
368 was also visible in the overlap of ADOS and ADI severity scores between both groups in our sample.  
369 In an exploratory analysis to increase accuracy, we employed a SVM classification on a re-labelled  
370 sample, grouping ASD with other neurodevelopmental disorders as defined by the DSM-5 [38].  
371 However, this model performed slightly above chance, suggesting a synchrony signature specific to  
372 autism. Yet, we recognize that this finding needs external validation in order to be further  
373 interpreted.

374 Our study has several limitations that should be considered: First, the videos analyzed in this study  
375 were not initially recorded for the purpose of automated machine learning-based analysis  
376 procedures. For this reason, the setup varied regarding background and camera angles depending  
377 on the different rooms. This could also have contributed to the lack of significant differences in our  
378 comparison to pseudo-synchrony (see Supplementary materials S2.3). However, we consider this a  
379 feature, rather than a flaw of our approach. When comparing the synchrony values between the  
380 different rooms, we could not detect significant differences, underling the scalability of our setup.  
381 This is in line with Kojovic and colleagues [19] who investigated their computer vision algorithm  
382 with two validation samples, finding minimal influence of video conditions. However, for future  
383 reference, we have compiled recommendations for a more standardized recording protocol of  
384 ADOS which can be found in the supplementary material (S3). Additionally, we recommend the use  
385 of separate microphones to allow for more elaborate analyses of verbal interaction, as well as the  
386 use of cameras for more fine-grained facial expression analyses.

387 Secondly, because our videos differed in length, the use of summary statistics as best estimate  
388 measures of interpersonal synchrony were deemed most suitable. However, this approach cannot  
389 capture the temporal dynamics of synchrony throughout a conversation. During free-flowing

390 conversations, interactants tend to move in and out of synchrony over time [50], suggesting a  
391 certain flexibility in interpersonal alignment. However, no clear evidence exists regarding interview  
392 contexts. Thus, future research should investigate synchrony trajectories in more standardized  
393 experimental settings.

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

398 Finally, and importantly, even though we have implemented a careful and rather conservative  
399 cross-validation structure within our model, the sample size in this study is limited, and the results  
400 require external validation. As this study served as a proof-of-concept, the present videos were  
401 chosen based on a meticulous screening process, which consequently resulted in a high number of  
402 exclusions. For example, we only analyzed video excerpts of more than five minute in length and  
403 without the use of any external props; the latter of which is an important part of the ADOS  
404 assessment. However, we are confident that the high scalability of the methodology used in this  
405 study will encourage future data collection and, hence, further external and cross-site validation. In  
406 this regard, it will be important to analyze any effects of relaxed inclusion criteria concerning, e.g.,  
407 the minimum length of an analysis window for a feasible synchrony assessment.

408 While clinicians' judgments continue to outperform computational algorithms in their diagnostic  
409 precision [51], the notion of digital augmentation of the diagnostic process could prospectively  
410 loosen the current bottlenecks caused by resource-exhaustive clinical assessments. Considering the  
411 aforementioned limitations, we present a viable route toward a digitally assisted diagnostic process  
412 in autism. Using a heterogeneous dataset, both in age and technical setup, our classification model  
413 could detect ASD in a clinical sample with an above-chance accuracy. With few adjustments

414 regarding the standardization of the experimental setup, including possibilities to record nuanced  
415 facial expression and vocal output, the strength of our approach has the potential for high  
416 scalability in everyday clinical practice.  
417

## 5 References

- [1] van 't Hof M, Tisseur C, van Berckeleer-Onnes I, van Nieuwenhuyzen A, Daniels AM, Deen M, et al. Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. *Autism* 2021;25:862–73. <https://doi.org/10.1177/1362361320971107>.
- [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] Zeidan J, Fombonne E, Scora J, Ibrahim A, Durkin MS, Saxena S, et al. Global prevalence of autism: A systematic review update. *Autism Research* 2022;15:778–90. <https://doi.org/10.1002/aur.2696>.
- [4] Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: A systematic literature review. *Eur Child Adolesc Psychiatry* 2013;22:329–40. <https://doi.org/10.1007/s00787-013-0375-0>.
- [5] Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop SL. *Autism Diagnostic Observation Schedule*. 2nd. Torrance, CA: 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* 1994;24:659–85.
- [7] 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. <https://doi.org/10.1038/tp.2015.7>.
- [8] 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>.
- [9] Moon SJ, Hwang J, Kana R, Torous J, Kim JW. 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* 2019;6. <https://doi.org/10.2196/14108>.
- [10] Nogay HS, Adeli H. Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging. *Rev Neurosci* 2020;31:825–41. <https://doi.org/10.1515/revneuro-2020-0043>.
- [11] Koehler JC, Falter-Wagner CM. Digitally assisted diagnostics of autism spectrum disorder. *Front Psychiatry* 2023;14. <https://doi.org/10.3389/fpsy.2023.1066284>.
- [12] Anzulewicz A, Sobota K, Delafield-Butt JT. Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism. *Sci Rep* 2016;6:1–13. <https://doi.org/10.1038/srep31107>.
- [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* 2023;13. <https://doi.org/10.1038/s41598-023-30628-6>.
- [14] McNaughton KA, Redcay E. Interpersonal Synchrony in Autism. *Curr Psychiatry Rep* 2020;22. <https://doi.org/10.1007/s11920-020-1135-8>.
- [15] 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>.

- [16] Ramseyer F. Ramseyer, F. (2019) Motion Energy Analysis (MEA). A primer on the assessment of motion from video. *Journal of Counseling Psychology* 2019;1:2019.
- [17] 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* 2022:2022.12.20.22283571. <https://doi.org/10.1101/2022.12.20.22283571>.
- [18] Bhat AN, Landa RJ, Galloway JC (Cole). Current Perspectives on Motor Functioning in Infants, Children, and Adults With Autism Spectrum Disorders. *Phys Ther* 2011;91:1116–29. <https://doi.org/10.2522/ptj.20100294>.
- [19] 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. <https://doi.org/10.1038/s41598-021-94378-z>.
- [20] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med* 2015;13. <https://doi.org/10.1186/s12916-014-0241-z>.
- [21] Kleinbub JR, Ramseyer FT. rMEA: An R package to assess nonverbal synchronization in motion energy analysis time-series. *Psychotherapy Research* 2020. <https://doi.org/10.1080/10503307.2020.1844334>.
- [22] 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>.
- [23] Ramseyer F, Tschacher W. Nonverbal synchrony of head- and body-movement in psychotherapy: different signals have different associations with outcome. *Front Psychol* 2014;5:979. <https://doi.org/10.3389/fpsyg.2014.00979>.
- [24] Boersma P, Weenink D. No Title. *Praat: Doing Phonetics by Computer* 2019.
- [25] Loomes R, Hull L, Mandy WPL. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J Am Acad Child Adolesc Psychiatry* 2017;56:466–74. <https://doi.org/https://doi.org/10.1016/j.jaac.2017.03.013>.
- [26] Wolff N, Stroth S, Kamp-Becker I, Roepke S, Roessner V. Autism Spectrum Disorder and IQ – A Complex Interplay. *Front Psychiatry* 2022;13. <https://doi.org/10.3389/fpsyg.2022.856084>.
- [27] 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>.
- [28] Cortes C, Vapnik V, Saitta L. *Support-Vector Networks* Editor. vol. 20. Kluwer Academic Publishers; 1995.
- [29] Koutsouleris N, Vetter C, Wiegand A. *Neurominer* 2022.
- [30] *MATLAB* 2022.
- [31] Golland P, Fischl B. *Permutation Tests for Classification: Towards Statistical Significance in Image-Based Studies*. vol. 2732. 2003.
- [32] Koutsouleris N, Kambeitz-Illankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, 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* 2018;75:1156–72. <https://doi.org/10.1001/jamapsychiatry.2018.2165>.
- [33] Gómez-Verdejo V, Parrado-Hernández E, Tohka J. Sign-Consistency Based Variable Importance for Machine Learning in Brain Imaging. *Neuroinformatics* 2019;17:593–609. <https://doi.org/10.1007/s12021-019-9415-3>.

- [34] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. vol. 57. 1995.
- [35] RStudio\_Team. RStudio: Integrated Development for R 2020. <https://doi.org/10.1037/abn0000182.supp>.
- [36] Hus V, Gotham K, Lord C. Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *J Autism Dev Disord* 2014;44:2400–12. <https://doi.org/10.1007/s10803-012-1719-1>.
- [37] Hus V, Lord C. The autism diagnostic observation schedule, module 4: Revised algorithm and standardized severity scores. *J Autism Dev Disord* 2014;44:1996–2012. <https://doi.org/10.1007/s10803-014-2080-3>.
- [38] American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). 5th ed. Arlington, VA: American Psychiatric Pub; 2013.
- [39] Paulick J, Rubel JA, Deisenhofer AK, Schwartz B, Thielemann D, Altmann U, et al. Diagnostic Features of Nonverbal Synchrony in Psychotherapy: Comparing Depression and Anxiety. *Cognit Ther Res* 2018;42. <https://doi.org/10.1007/s10608-018-9914-9>.
- [40] Dean DJ, Scott J, Park S. Interpersonal Coordination in Schizophrenia: A Scoping Review of the Literature. *Schizophr Bull* 2021;47:1544–56. <https://doi.org/10.1093/schbul/sbab072>.
- [41] Ghaziuddin M, Ghaziuddin N, Greden J. Depression in Persons with Autism: Implications for Research and Clinical Care. *J Autism Dev Disord* 2002;32:299–306. <https://doi.org/10.1023/A:1016330802348>.
- [42] King BH, Lord C. Is schizophrenia on the autism spectrum? *Brain Res* 2011;1380:34–41. <https://doi.org/10.1016/j.brainres.2010.11.031>.
- [43] Zampella CJ, Bennetto L, Herrington JD. Computer Vision Analysis of Reduced Interpersonal Affect Coordination in Youth With Autism Spectrum Disorder. *Autism Research* 2020;13:2133–42. <https://doi.org/10.1002/aur.2334>.
- [44] 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>.
- [45] Fasano RM, Perry LK, Zhang Y, Vitale L, Wang J, Song C, et al. A granular perspective on inclusion: Objectively measured interactions of preschoolers with and without autism. *Autism Research* 2021;14:1658–69. <https://doi.org/10.1002/aur.2526>.
- [46] Rybner A, Jessen ET, Mortensen MD, Larsen SN, Grossman R, Bilenberg N, et al. Vocal markers of autism: Assessing the generalizability of machine learning models. *Autism Research* 2022;15:1018–30. <https://doi.org/10.1002/aur.2721>.
- [47] 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. *Journal of Intellectual Disability Research* 2016;60:464–77. <https://doi.org/10.1111/jir.12284>.
- [48] 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>.
- [49] Maddox BB, Brodtkin 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>.
- [50] Mayo O, Gordon I. In and out of synchrony—Behavioral and physiological dynamics of dyadic interpersonal coordination. *Psychophysiology* 2020;57. <https://doi.org/10.1111/psyp.13574>.



- [51] Kambeitz-Illankovic L, Koutsouleris N, Upthegrove R. The potential of precision psychiatry: What is in reach? *British Journal of Psychiatry* 2022;220:175–8. <https://doi.org/10.1192/bjp.2022.23>.

## 6 Acknowledgements

We thank Afton Nelson for proofreading this manuscript.

## 7 Authors' contribution

HY and CFW conceptualized the study. JCK and CFW designed the pre-processing procedure. DS and GB compiled and pre-processed the data. JCK and MSD analyzed and interpreted the data. NK supervised the machine learning analysis. JCK wrote the manuscript. All authors read and approved the final manuscript.

## 8 Availability of data and materials

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](https://github.com/jckoe/SNU_ASDSync.git)

## 9 Additional Information

### 9.1 Competing Interests

The authors declare that they have no competing interests.

## 9.2 Funding

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).

## 10 Figure Legends

10.1 Figure 1. Consort chart of the current sample compilation.

10.2 Figure 2. SVM classification results of ASD vs. CC patient-administrator dyads

*Note. 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.*

10.3 Figure 3. Feature importance of SVM model

*Note. 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  $\geq 2$  correspond to  $p \leq .05$ , absolute values  $\geq 3$  correspond to  $p \leq .01$ . 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 -  $10\log(p) \geq 1.3$  is equivalent to  $p \leq .05$ .*

10.4 Figure 4. Association between SVM decision scores of MEA classifier and ADI and ADOS domain scores

*Note. ADOS scores were transformed to calibrated severity scores following procedures in [36,37]. 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.*

## 11 Tables

### 11.1 Table 1. SVM Classification Model Descriptions

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

*Note. Number of features of respective modality in parentheses. Missing IQ values (16% of cases) were imputed using k-nearest neighbor imputation.*

## 11.2 Table 2. Sample description and demographic group differences across subsamples

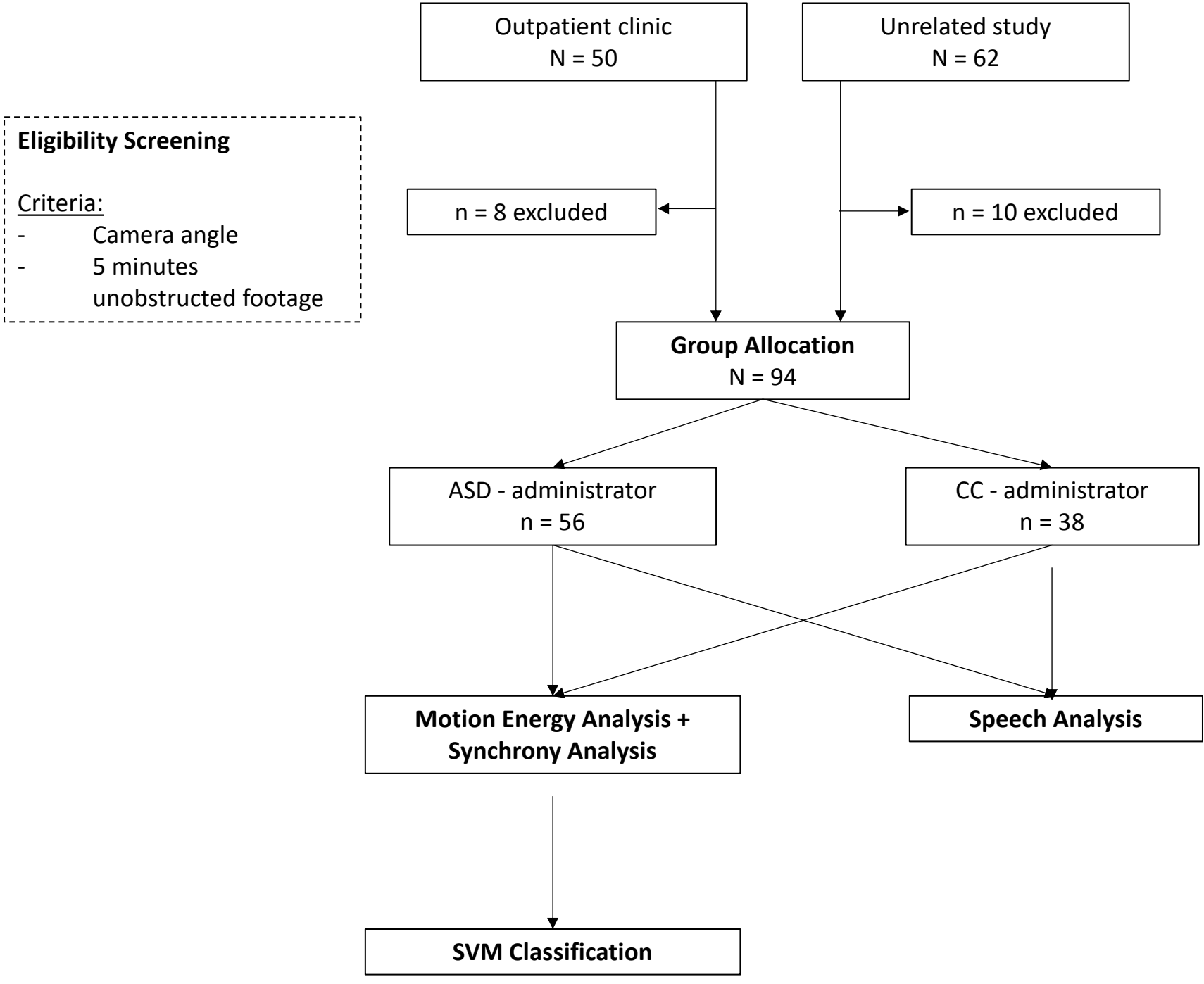
Module	ASD (n = 56, 11 female)			CC (n = 38, 9 female)			T test			Effect size	
	n	M	SD	n	M	SD	t	df	p		
3	Age	37	11.35	2.79	27	10.68	2.74	0.96	56.80	.339	.244
	IQ	33	92.09	20.87	20	103.55	17.05	-2.18	46.44	.035	-.601
	ADOS_SA	37	10.05	2.89	27	5.37	3.65	5.53	47.94	< .001	1.42
	ADOS_RRB	37	1.27	1.54	27	.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
	IQ	18	95.56	17.52	8	77.25	12.34	3.05	18.91	.007	1.21
	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

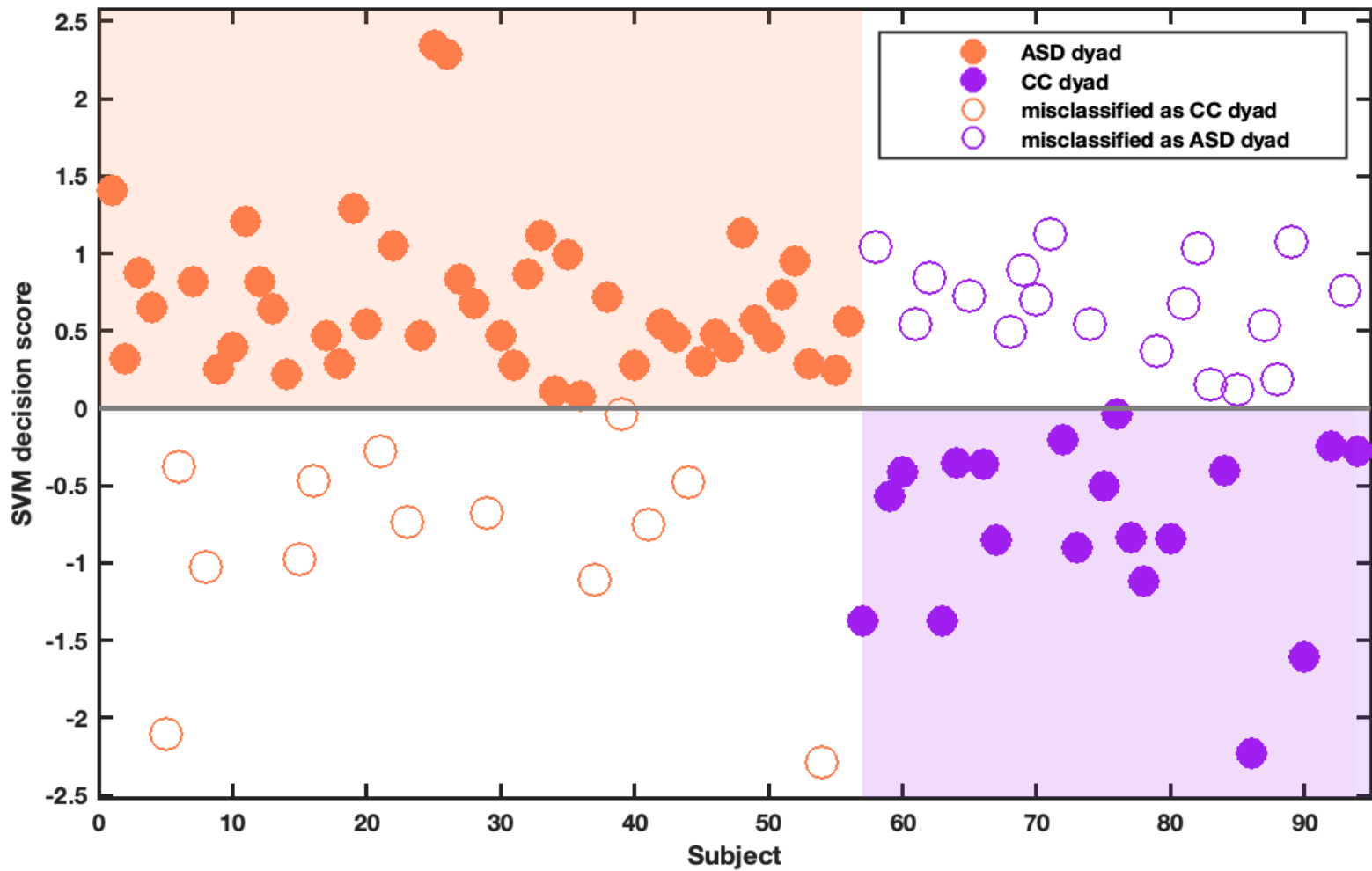
Note. Full Scale IQ depicted as best estimate depending on age (WISC-III, WISC-IV, WPPSI-IV, WAIS → Korean versions); group comparisons computes using Welch's t-test for unequal variances, p value adjusted for multiple testing with Bonferroni-Holm correction, Effect size cohen's d

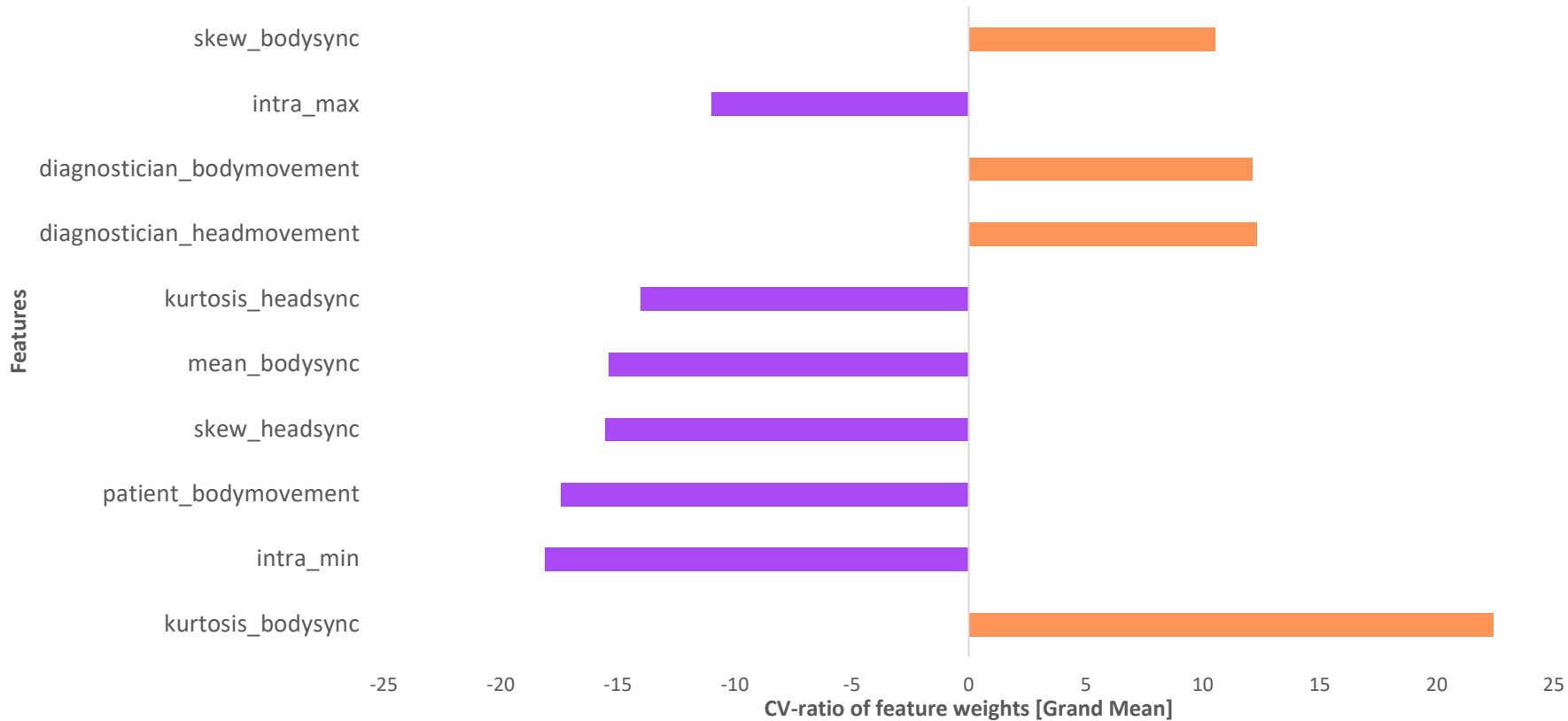
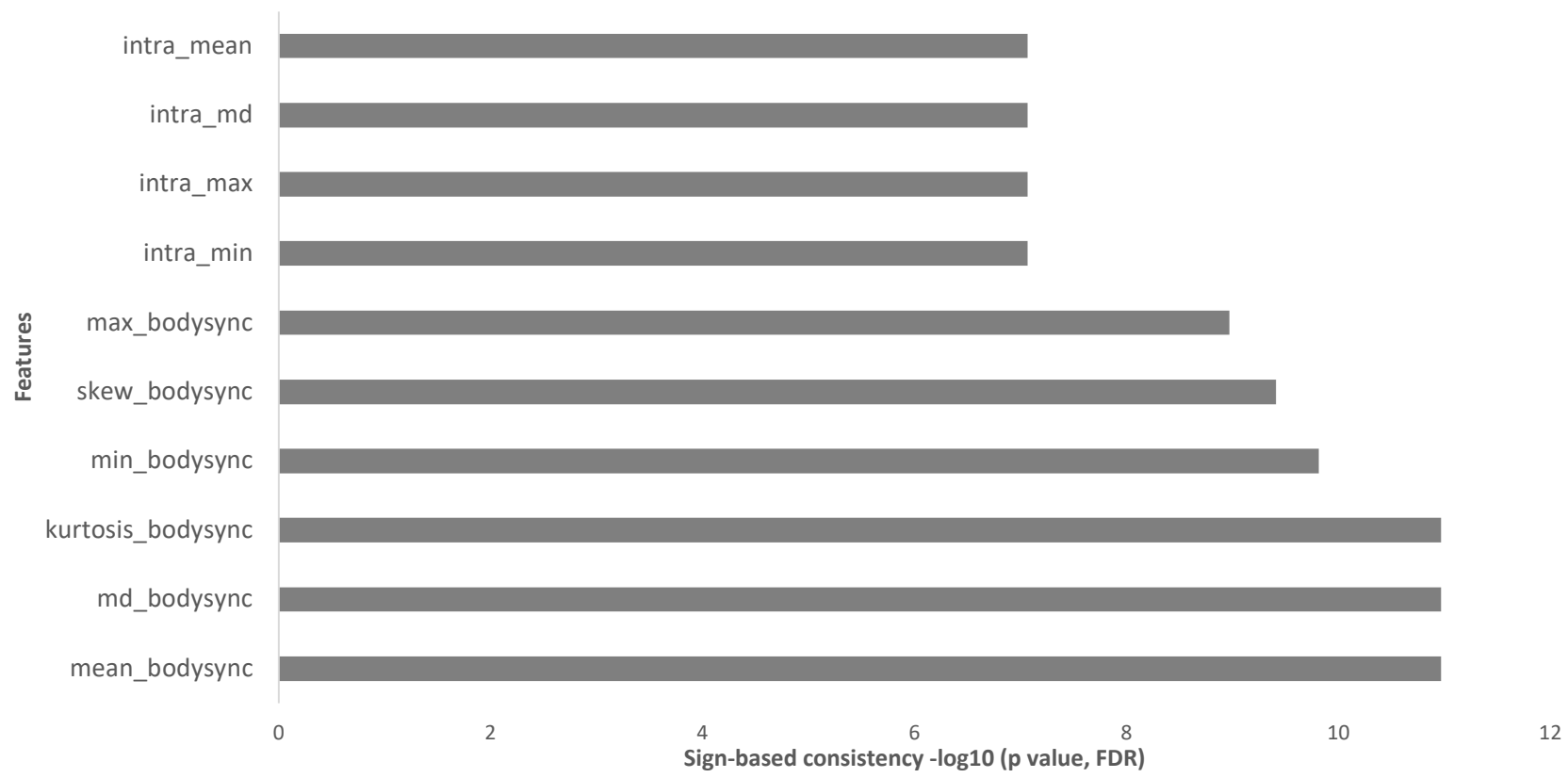
11.3 Table 3. Classification metrics for SVM classifiers based on Motion Energy Synchrony Analyses between Patient and Administrator.

Classifier	BAC (%)	Sensitivity (%)	Specificity (%)	AUC	TN	TP	FN	FP	Accuracy (%)	Number needed to diagnose	Positive likelihood ratio	Diagnostic odds ratio	Permutation test, <i>p</i> value
ASD vs. CC	63.4	76.8	50.0	.61	19	43	13	19	66.0	3.7	1.5	2.4	<.001
NDD vs. CC	56.1	62.2	50.0	.52	10	46	28	10	59.6	8.2	1.2	1.5	.005

*Note.* 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.





**a.****b.**



Mean SVM Decision Scores

