Detecting Autism by Analyzing a Simulated Social Interaction

Hanna Drimalla[1,2,3], Niels Landwehr[1,5], Irina Baskow[2], Behnoush Behnia[4],
Stefan Roepke[4], Isabel Dziobek[2,3], and Tobias Scheffer[1]

[1] Department of Computer Science, University of Potsdam, Potsdam, Germany
tobias.scheffer@uni-potsdam.de
[2] Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany
{hanna.drimalla, irina.baskow, isabel.dziobek}@hu-berlin.de
[3] Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany
[4] Department of Psychiatry and Psychotherapy, Campus Benjamin Franklin,
Charité-Universitätsmedizin Berlin*, Berlin, Germany
{behnoush.behnia, stefan.roepke}@charite.de
[5] Leibniz Institute for Agricultural Engineering and Bioeconomy, Potsdam, Germany
NLandwehr@atb-potsdam.de

Abstract. Diagnosing autism spectrum conditions takes several hours by well-trained practitioners; therefore, standardized questionnaires are widely used for first-level screening. Questionnaires as a diagnostic tool, however, rely on self-reflection—which is typically impaired in individuals with autism spectrum condition. We develop an alternative screening mechanism in which subjects engage in a simulated social interaction. During this interaction, the subjects' voice, eye gaze, and facial expression are tracked, and features are extracted that serve as input to a predictive model. We find that a random-forest classifier on these features can detect autism spectrum condition accurately and functionally independently of diagnostic questionnaires. We also find that a regression model estimates the severity of the condition more accurately than the reference screening method.

1 Introduction

Autism spectrum conditions (ASC) encompass a range of neurodevelopmental conditions that affect how an individual perceives the world and interacts with others. Around 1 in every 100 individuals has some form of autism [30] and shows the characteristic impairments in social communication and interaction as well as restricted interests and repetitive behaviors [4].

Autism is nowadays seen as a spectrum of conditions: its severity and impact on the individual's life vary [40]. Mild autism conditions with normal intelligence levels have been described as *high-functioning autism* and *Asperger syndrome*. Although symptoms of autism are already occurring early in life, individuals

* Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health

with high-functioning autism are often diagnosed later [8], due to compensation strategies [19] or the subtlety of the autistic symptoms. A study with college students points to a substantial proportion of individuals with normal intelligence and autism that are undiagnosed [50]. Despite the higher social-functional level, their lifetime rate of psychiatric consultations is high, reflecting a need of earlier support and diagnosis [24].

Existing diagnostic tools for adults with autism examine the altered social communication and interaction patterns in semi-structured activities with the individual (*Autism Diagnostic Observation Schedule, ADOS* [26]) as well as in diagnostic interviews with the parents (*Autism Diagnostic Interview-Revised, ADI-R* [27]). These diagnostic processes are considered to establish a "gold-standard" diagnosis and concentrate on diagnostic criteria defined by DSM-V [4] and ICD-10 [41]. But they take several hours of time and have to be carried out by well-trained practitioners. Thus, they cannot be administered to any significant share of the population to screen for high-functioning autism.

Therefore, standardized questionnaires are often used for screening. One of the most widely applied ones is the *Autism-Spectrum Quotient (AQ)* [9], a brief self-administered questionnaire that measures traits associated with the autistic spectrum in adults with normal intelligence. In a clinical study in which both an AQ screening and a diagnosis by a medical practitioner have been observed for adults who sought out a diagnostic clinic, the AQ screening has shown an AUC of 0.78 [54]. One general concern about the AQ screening and other self-reports is the universal bias towards giving socially desirable answers [37]. In autism diagnosis, another aspect may affect the results of self-reports even more: individuals with autism spectrum condition often have an impaired introspection [18] and problems in abstract reasoning [35]. Therefore, an easy but still precise screening mechanism for autism that does not rely on self-reflection would be beneficial.

Deficits in social interaction that are part of the diagnostic criteria [4, 41] include a lack of social-emotional reciprocity, lack of facial expressions, and abnormalities in eye contact and voice modulation. In this paper, we develop and evaluate a screening approach for high-functioning adults of both genders that automatically analyzes these criteria in a simulated social interaction. Section 2 presents related work to autism detection via machine learning. Section 3 describes the Simulated Interaction screening method. Section 4 presents empirical results and Section 5 concludes.

2 Related Work

To date, there have been only a few studies using automatic behavioral analysis to detect autism. Crippa et al. [12] monitor upper-limb movements during a specifically designed manual task with an optoelectronic system, and use features extracted from these kinematics data to detect children with autism spectrum condition. They observe a maximum accuracy of 96.7% (with a maximum precision of 93.8% and a maximum recall of 100%) on a small sample of 15 children with ASC and 15 neurotypical children. However, as the method is developed for this very specific task and is applicable only for optoelectronic systems, it provides no scalable screening mechanism. Furthermore, it focuses on a behavioral feature that is not very tightly linked to ASC.

Hashemi et al. [20] use a computer vision approach to analyze activities that are assessed by the *Autism Observation Scale for Infants* [10] in video recordings. In a small sample of three infants with and three without indication of ASC risk they find differences between the groups in head motion and gait. However, the paper makes no attempt to classify the children based on the video analysis.

Liu et al. [25] observe eye movements during a face-recognition task; such face scanning patterns have been reported to differ in autism [42]. Based on the gaze patterns of all participants, they are able to discriminate between autistic and neurotypical children with accuracy of 88.5%. Since this process involves a high-acuity eye tracker, it does not suggest itself as a scalable screening process. Another study points at the potential of eye gaze as a feature for detection of autism: Gliga et al. [16] found that eye movements of nine-month-old infants significantly predict a higher level of autism symptoms at two years of age.

A recent study [38] used machine-learning-based voice analysis to classify word utterances of children with autism spectrum condition and children with typical development. The study focuses on the classification at the level of word utterances and not individuals. Therefore, the value of this approach for diagnostic purposes remains unclear.

Beyond autism, there has been remarkable progress in using machine learning technologies to infer underlying medical and psychological conditions from behavior or appearance. Some promising results could be achieved in learning to detect depression [39], predict suicidal ideation [23] or recognize schizophrenia [48]. Using audio and video recordings of the participants, some studies focused on speech and vocal cues [36]. Other studies investigated the predictive value of facial expressions [11], gaze direction or head pose [1].

3 Simulated Interaction

In this section, we develop the *Simulated Interaction* screening method. The aim of this procedure is to detect autism via a simulated social interaction, using only a screen, a webcam, and a microphone. This problem can be divided into two subproblems: The first is to predict whether a practitioner will diagnose the individual as on the autism spectrum condition (binary classification). The second is an assessment of the degree of autism, conceptualized as the individual's value of the ADI-R, a diagnostic clinical interview with their parents (regression). Specifically, we focus on the score for the reciprocal social interaction subdomain of the ADI-R (social subscale), as we expect this score to be most sensitive to high-functioning individuals and closely corresponding to the naturalistic setting of a simulated social interaction.

The core symptoms of autism are deficits in social communication and interactions [2]. These deficits manifest themselves in a number of nuances that guide the design of the screening method. First, in a social interaction, individuals tend to involuntarily mimic facial expressions of their conversation partner [43]. Individuals with autism spectrum condition are less likely to mimic the facial expressions of others spontaneously [33]. Similarly, a reduced intensity [46] or at least qualitative differences [17] of facial expressions in natural conversations have been observed for autism conditions.

Secondly, autism spectrum condition typically manifests itself in altered gaze patterns in complex environments [58] or in emotion recognition paradigms [51]. Madipakkam et al. [28] observe that patients have a tendency to avoid eye contact and instead let their gaze stray over a wider range of angles than neurotypical subjects. Thirdly, effects of autism spectrum condition on individual's voice have been reported; *e.g.,* for prosody [45] or pitch [44]—not in all studies, but especially in naturalistic settings [13].

Motivated by these findings, we design Simulated Interaction as a "dialog" between the recording of an actress and the participant about positive and negative food experiences. The actress addresses the participant directly and asks simple questions. The participant listens to the actress and answers her question while the actress nods and appears to listen. The first two parts are emotionally neutral. The third part is about the participant's favorite food and addresses a joyful topic. The final part about the participant's least favorite food raises an emotional response of disgust. The exact schedule is described in Table 1.

Table 1. Simulated Interaction schedule

Speaker	Topic	Time (s)
Actress	introduces herself, asks "what is your name?"	4
Participant	answers.	2
Actress	describes her way to the institute, asks "how did you get to the institute today?"	90
Participant	answers.	20
Actress	thanks the participant, switches the topic towards dining and describes how she sets the table for dinner; asks "How do you prepare a table for dinner?"	40
Participant	answers.	25
Actress	describes her favorite food, asks "what is your favorite food?"	25
Participant	answers.	25
Actress	describes her least favorite food, asks "which food do you dislike?"	25
Participant	answers.	25
Actress	thanks the participant and concludes the conversation.	8

3.1 Feature Extraction

Based on the described phenomenology of autism condition, we concentrate our feature on gaze, voice, and facial expressions. Facial expressions can be broken down into *facial action units* according to the facial action coding system developed by [14]—each action unit is comprised of visually detectable muscle movements in the face. A major advantage of this sign-based over message-based approaches is its objectivity as it does not require any interpretation [15]. Moreover, it does not reduce the facial expression to a small set of prototypical emotional expressions [47].

We employ the OpenFace library 1.0.0 [7] to extract the occurrence and intensity of 18 facial action units as well as gaze angles for both eyes from each frame. OpenFace is an open-source tool that is capable of facial-landmark detection, head-pose estimation, facial-action-unit recognition and eye-gaze estimation. OpenFace detects and tracks facial landmarks with conditional local neural fields [6], and aligns the face.

In order to detect emotions, OpenFace extracts HOG features, and reduces the HOG features space by principal component analysis. To correct for person-specific neutral expressions, OpenFace subtracts the median value of each of the remaining features over the entire observation sequence. Finally, the activation of each action unit is determined by a support-vector classifier and its intensity by support-vector regression. OpenFace has been tested on the SEMAINE [34], DISFA [31] and BP4D-Spontaneous [56] datasets and demonstrated state-of-the-art results [5] as well as outperformed the baselines of the FERA 2015 challenge [49].

OpenFace performs appearance-based gaze estimation. On the MPIIGaze data set [57], it achieves a mean absolute error of under 10 degrees which exceeds the performance of other tools (*e.g.,* EyeTab [53]). OpenFace detects eye-region landmarks including the eyelids, iris, and pupil with conditional local neural fields, trained on the SynthesEyes dataset [52]. Based on the location of the eye and pupil, it estimates the center of the eyeball and infers the gaze vector from this center through the pupil.

To extract features of the audio recording of the participant's voice, we use the librosa library [32]. For each frame, we extract prosodic (root-mean-square energy) as well as spectral features (forty mel-frequency cepstral coefficients, MMCC). Both are standard features in speech recognition [22] and have shown to be altered under autism condition [29].

From these primary features, we extract secondary features that aggregate the values for each feature. For action units' intensity and gaze angle, we calculate arithmetic mean, standard deviation, skewness, kurtosis, the maximum and the time point of the maximum. For action unit's occurrence and the voice features, we calculate the mean values. All aggregated values are calculated for seven parts of the conversation.

3.2 Machine Learning Methods

We use SVMs and random forests as base machine-learning methods. For the SVM we use a radial basial function kernel and tune the regularization parameters in a nested cross-validation with grid search. For the random forest we use an ensemble of 1,000 different trees on different subsets of data and input variables. We tune the maximal depth of the trees and the minimum number of samples per leaf with a nested grid-search.

Additionally, we explore the use of two different convolutional neural networks all employing the 1-dimensional convolution operation. The convolution uses a stride of one and zero padding. We tune the number of filters, size of convolution and pooling, dropout rate and number of units of the dense layer via a nested grid search; Table 2 shows the search space for all hyperparameters. Both CNNs use the rectified linear activation function for the hidden units, and have one output unit with a logistic activation function. We use the cross-entropy loss function and train the networks with the gradient descent algorithm Adam. To avoid over-fitting, training is stopped when the validation-loss does not improve for three epochs.

The *StackedCNN* contains four learned layers: Two 1D-convolutional layers, one dense layer and one fully-connected output layer. Each convolutional layer is followed by a max-pooling layer. One additional drop-out layer is followed by a dense layer. The structure of the network can be seen in Figure 1. We tune the hyperparameters via a full grid search, leading to a minimal number of 135 and a maximal number of 938,769 parameters.

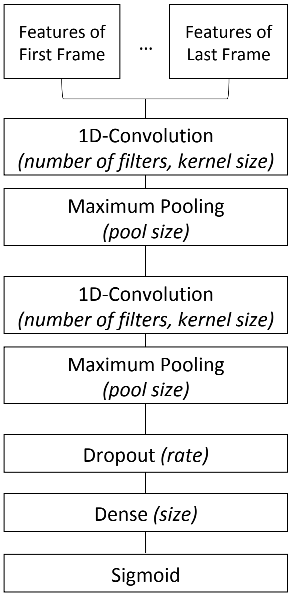
Features of First Frame
...
Features of Last Frame
1D-Convolution (number of filters, kernel size)
Maximum Pooling (pool size)
1D-Convolution (number of filters, kernel size)
Maximum Pooling (pool size)
Dropout (rate)
Dense (size)
Sigmoid

Fig. 1. Architecture of the StackedCNN (tuned hyperparameters are printed in italics).

PooledCNN is a CNN with multiple inputs that incorporates some domain knowledge into its structure. It contains seven learned convolutional layers—one for each conversation part. The input is split into the conversation parts and distributed accordingly into the seven convolutional layer. Thus, each convolutional layer receives only a part of the original input. Every convolutional layer is followed by a pooling layer. The size of the pooling equals the number of units of the previous convolutional layer leading to seven units as input for the following dense layer. The next and last layer is the fully-connected output layer. The structure of the network can be seen in figure 2. The hyperparameters are tuned via a random search with 20 iterations, leading to a minimal number of 618 parameters and a maximal number of 110,343 parameters.

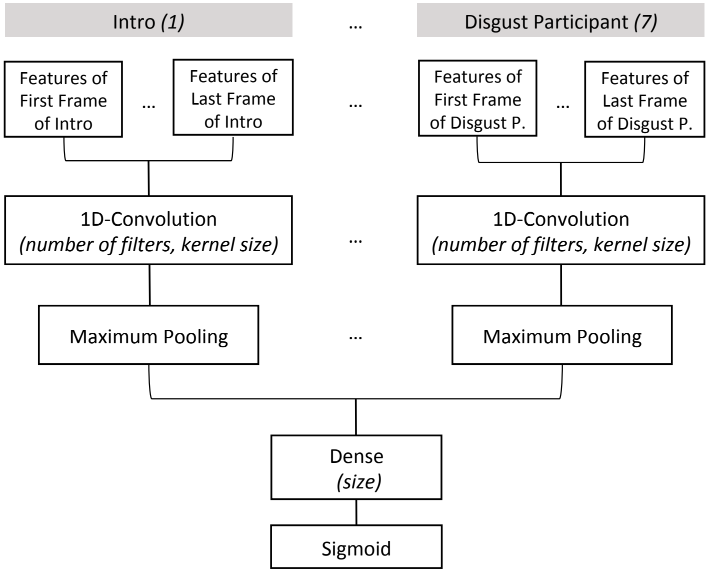
Intro *(1)*
...
Disgust Participant *(7)*
Features of First Frame of Intro
...
Features of Last Frame of Intro
...
Features of First Frame of Disgust P.
...
Features of Last Frame of Disgust P.
1D-Convolution *(number of filters, kernel size)*
...
1D-Convolution *(number of filters, kernel size)*
Maximum Pooling
...
Maximum Pooling
Dense *(size)*
Sigmoid

Fig. 2. Architecture of the PooledCNN (tuned hyperparameters are printed in italics).

Table 2. Hyperparameter Space

Hyperparameter	Stacked	Pooled
number of filters	{1, 2, 4, 8}	{1, 2, 4, 8}
size of kernel	{2, 4, 8, 16, 32}	{2, 4, 8, 16, 32}
size of pooling	{5, 25, 50}	-
rate of dropout	{0.25, 0.5}	-
size of dense layer	{4, 8, 16, 32, 64, 128, 256}	{8, 16, 32, 64, 128}

4 Empirical Study

This section explores the effectiveness of the Simulated Interaction screening method on a sample of patients that have been diagnosed with autism spectrum condition and a neurotypical control group. The AQ questionnaire will serve as reference screening method.

4.1 Data Collection

We record the audio and webcam stream of 44 neurotypical participants and 37 participants with autism spectrum condition. The sample is balanced regarding gender. The neurotypical participants have been selected based on a questionnaire that asks for a history of a wide range of neurological and psychological conditions; participants with any such history have been excluded from the study. Each participant with autism spectrum condition been diagnosed by a practitioner and scored according to ADOS and ADI-R. We measure autistic traits of all participants via a German version of the AQ questionnaire [9].

We film the faces of the participants in experiment rooms with constant lighting conditions and no disturbing background. The participants are recorded with the internal microphone and a webcam with a rate of 30 frames per second and a resolution of 640 x 480 pixels, leading to a total number of 11,340 frames. These video and audio recordings of each participant during the conversation are the raw input to detect participant's autism. OpenFace provides a success value (0/1) for the face tracking of each frame. Participants with a success rate of less than 90% of the frames are excluded from the experiment.

4.2 Evaluation Protocol

To validate the results, we use a *nested cross-validation* strategy with an outer leave-one-out cross-validation loop and an inner 3-fold cross-validation loop in which we tune all hyperparameters. The hyperparameters of the neural networks are listed in Table 1.

4.3 Prediction of the Clinical Diagnosis

We will first study the ability of Simulated Interaction to predict the clinical diagnosis of an autism condition.

Comparison of base machine-learning methods. We first compare the different machine learning methods under investigation. Figure 3 compares ROC curves for the base learning methods using the full set of features. The random forest and SVM achieve the best detection using all features. The random forest achieves an area under the curve of 0.84 and the SVM an AUC of 0.81. The SVM ($p < 0.01$ according to a sign test) and the random forest ($p < 0.01$ according to a sign test) perform significantly better than the majority baseline. Their predictions correlate strongly with the autism diagnosis of the participants: the class probabilities predicted by the random forest reach a point-biserial correlation with the diagnosis of $r = 0.53$ ($p < 0.0001$).

Both neural network models perform worse than the random forest and SVM. The CNN without knowledge about the interaction parts, performs close to chance level with an AUC of 0.53 (stacked). The PooledCNN produces a better result with an AUC of 0.64. However, with a set threshold of 0.5, none of the CNNs achieves a higher accuracy than the naive baseline, which always predicts the majority class. The class probabilities predicted by the PooledCNN correlate positively with the autism diagnosis at trend-level ($r = 0.21$, $p < 0.1$).

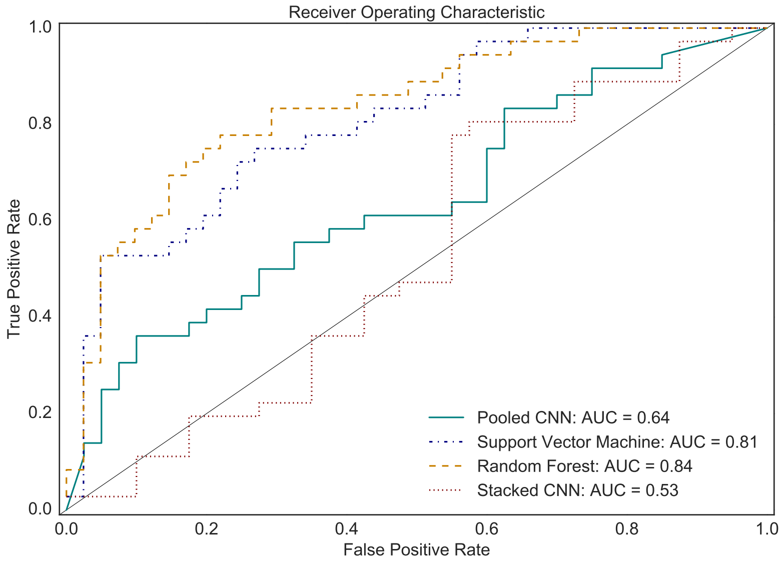

Fig. 3. ROC Curves for the Different Classifiers

Comparison of feature subsets. We now explore the relative contributions of the different types of features; Figure 4 compares the ROC curves of various feature subsets. As on the complete feature set the random forest performs best, we concentrate on this classifier for the next steps.

We compare the performance of different groups of features: the occurrence of action units, the intensity of action units, the gaze angles, all video features and the vocal features. The best prediction of the autism diagnosis is achieved with the combination of all features, as Figure 4 shows. Statistical testing reveals that a significantly better detection is possible with the vocal features, the intensity of action units, all video features, or all features together than with the baseline ($p < 0.05$ for all). There was no evidence that the performance differs significantly between these three feature groups ($p > 0.05$ for all). The predic-

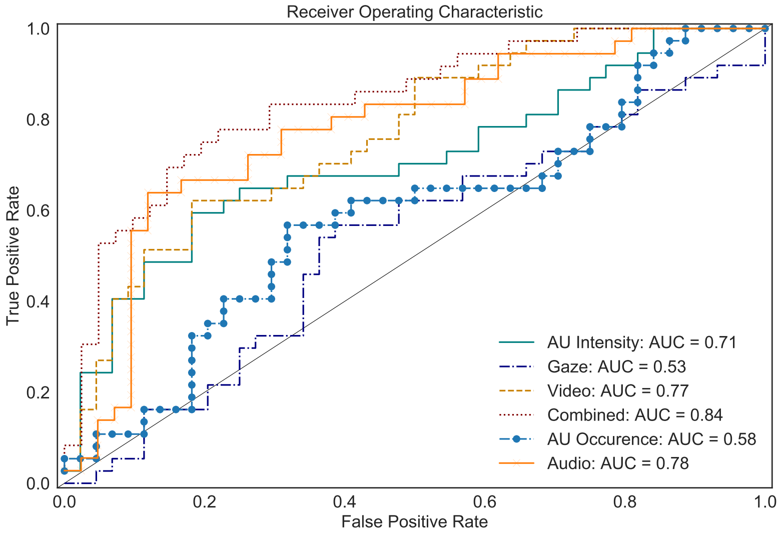

Fig. 4. ROC Curves for the Different Features

tion based on gaze was significantly worse ($p < 0.05$ for all of them). Comparing the facial features revealed that on a trend-level the prediction was significantly worse for the occurrence than for the intensity ($p < 0.05$).

Comparison to AQ. We compare the Simulated Interaction screening method to the AQ questionnaire. Using the AQ score as decision-function value we obtain an AUC of 0.99 whereas Simulated Interaction using random forests and the full feature set attains an AUC of 0.84. This difference is significant ($p < 0.001$) according to a sign test.

In previous studies using a clinical sample of patients, the AQ has shown an AUC of 0.78 [54] which starkly contrasts our observation of 0.99. This difference can be explained by the selection criteria for our neurotypical sample. Subjects with any history of psychological, psychiatric, and neurological treatments have been excluded from the neurotypical sample. This results in a much easier classification problem than can be expected in any clinical or broad screening setting.

Patients in a clinical environment usually seek out medical treatment because of an impaired quality of life. The AQ questionnaire has a known propensity to higher scores for individuals with obsessive-compulsive disorder, social-anxiety [21] or schizophrenia [55]. Furthermore, individuals with suspected ASD typically enter the diagnostic process at a specialized center as in the study mentioned above [54]. Despite signs of autism and self-diagnosis, only around 73%

[3, 54] receive an official diagnosis. In such a difficult setting, the AQ's low AUC of 0.78 and high false-positive rate of 64% [3] are unsatisfactory.

Thus, it is natural to ask whether a combination of the AQ questionnaire and the Simulated Interaction screening method is more accurate than just the AQ. This would be the case if the error cases of AQ and Simulated Interaction were independent of one another. Unfortunately, our limited sample does not allow us to answer this question. The AQ questionnaire misclassifies only one single patient in our sample, which makes it impossible to draw any conclusions about the independence of error cases or the accuracy of a combined detection model. However, since the AQ and Simulated Interaction are based on fundamentally different functional principles, our results motivate a follow-up study in with a larger clinical sample from a clinical distribution of patients.

4.4 Estimation of the Autism Degree

We will now evaluate Simulated Interactions' ability to estimate the severity of the autism condition, measured by the score of the clinical interview of the parents, ADI-R (social subscale). We use a random forest regression model; in order to further reduce the dimensionality of the feature space, we aggregate each of the secondary features over the seven parts of the interview into a single value. Figure 5 plots the gold-standard ADI-R social subscale score over the score of the regression model. For the individuals with ASC, it is possible to estimate the ADI-R social subscale with a random forest better than a mean baseline, according to a paired t-test ($p < 0.05$).

Comparison to AQ. The root mean squared error of the tree is 5.40, while the baseline produces an error of 6.42. The ADI-R social subscale values predicted by the random forest correlate positively with the true values at trend-level ($p < 0.1$). By contrast, the values of the AQ and the ADI-R social subscale are not significantly positively correlated (trend towards a negative correlation: $r = -0.42$, $p < 0.1$) and a prediction with a random forest regressor on the autism questionnaire score lead to a mean absolute error of 8.02, which is significantly worse than the prediction with the Simulated Interaction ($p < 0.05$).

5 Conclusion

Diagnosing autism spectrum condition and quantifying the severity of the condition require time and well-trained practitioners. Quantifying the severity accurately also requires access to the individual's parents. In this paper, we presented and evaluated the Simulated Interaction screening method for autism spectrum condition of high-functioning individuals. We find that it is possible to predict the binary diagnosis of autism spectrum condition with high accuracy from the facial expressions, vocal features, and gaze patterns of the individuals.

The intensity of the facial expressions and the vocal features turned out to be more informative than the occurrence of facial expressions and the gaze patterns.

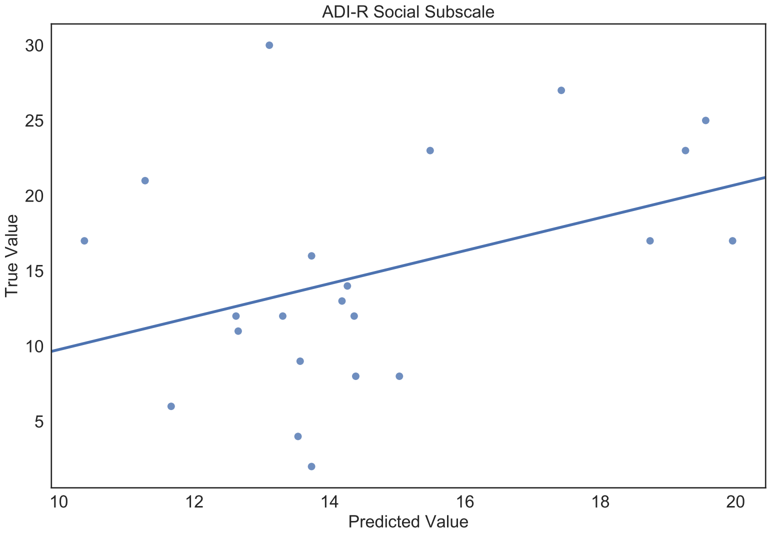

Fig. 5. Regression for the ADI-R Social

The webcam might not be capable to track the gaze behavior sensitively enough to detect individual differences. The results concerning the facial expressions fit the literature about qualitative differences in facial expressions.

For the prediction of the binary diagnosis, the Simulated Interaction screening method did not outperform the AQ questionnaire. However, Simulated Interaction is functionally independent of a self-assessment because it is based on fundamentally different features. While our limited sample does not allow us to draw any conclusions on the accuracy of a combination of AQ and Simulated Interaction, our findings motivate a follow-up study on a larger and more difficult clinical sample consisting of subjects with a suspected ASD.

Regarding the estimation of the severity, we can conclude that Simulated Interaction estimates the ADI-R values significantly better than a mean baseline and the AQ questionnaire.

The Simulated Interaction screening method has two principal advantages over questionnaires: first, it is not as easily biased by social desirability tendencies as a questionnaire. Secondly, it is independent of introspection—which is typically impaired in subjects with autism spectrum condition. Simulated Interaction could potentially be used as an online self-test, possibly in combination with a questionnaire.

Acknowledgment

This work was partially funded by the German Science Foundation under grant LA3270/1-1.

References

1. Alghowinem, S., Goecke, R., Cohn, J.F., Wagner, M., Parker, G., Breakspear, M.: Cross-cultural detection of depression from nonverbal behaviour. IEEE International Conference on Automatic Face & Gesture Recognition and Workshops **1** (2015)
2. American Psychiatric Association: Diagnostic and statistical manual of mental disorders: DSM-5. American Psychiatric Association, Arlington, VA, fifth edition edn. (2013)
3. Ashwood, K., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F., Findon, J., Eklund, H., Spain, D., Wilson, C., et al.: Predicting the diagnosis of autism in adults using the autism-spectrum quotient (aq) questionnaire. Psychological medicine **46**(12), 2595–2604 (2016)
4. Association, A.P.: Diagnostic and Statistical Manual of Mental Disorders: Dsm-5. Diagnostic and Statistical Manual of Mental Disorders, Amer Psychiatric Pub Incorporated (2013), https://books.google.de/books?id=EIbMlwEACAAJ
5. Baltrusaitis, T., Mahmoud, M., Robinson, P.: Cross-dataset learning and person-specific normalisation for automatic action unit detection. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). pp. 1–6. IEEE (2015)
6. Baltrusaitis, T., Robinson, P., Morency, L.P.: Constrained local neural fields for robust facial landmark detection in the wild. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 354–361 (2013)
7. Baltrusaitis, T., Robinson, P., Morency, L.P.: Openface: An open source facial behavior analysis toolkit. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1–10. IEEE (2016)
8. Barnard, J., Harvey, V., Potter, D.: Ignored or ineligible? The reality for adults with autism spectrum disorders. National Autistic Society (2001)
9. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E.: The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. Journal of autism and developmental disorders **31**(1), 5–17 (2001)
10. Bryson, S.E., Zwaigenbaum, L.: Autism observation scale for infants. In: Comprehensive guide to autism, pp. 299–310. Springer (2014)
11. Cohn, J.F., Kruez, T.S., Matthews, I., Yang, Y., Nguyen, M.H., Padilla, M.T., Zhou, F., de La Torre, F.: Detecting depression from facial actions and vocal prosody. In: Staff, I. (ed.) 2009 3rd International Conference on Affective Computing and Intelligent Interaction. pp. 1–7. IEEE (2009)
12. Crippa, A., Salvatore, C., Perego, P., Forti, S., Nobile, M., Molteni, M., Castiglioni, I.: Use of machine learning to identify children with autism and their motor abnormalities. Journal of autism and developmental disorders **45**(7), 2146–2156 (2015)
13. Diehl, J.J., Watson, D., Bennetto, L., McDonough, J., Gunlogson, C.: An acoustic analysis of prosody in high-functioning autism. Applied Psycholinguistics **30**(3), 385–404 (2009)
14. Ekman, P., Friesen, W.V.: Facial action coding system. Consulting Psychologists Press, Palo Alto, CA (1978)
15. Fasel, B., Luettin, J.: Automatic facial expression analysis: A survey. Pattern Recognition **36**(1), 259–275 (2003), http://www.sciencedirect.com/science/article/pii/S0031320302000523

16. Gliga, T., Bedford, R., Charman, T., Johnson, M.H.: Enhanced visual search in infancy predicts emerging autism symptoms. Current biology : CB **25**(13), 1727–1730 (2015)
17. Grossman, R.B., Edelson, L.R., Tager-Flusberg, H.: Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism. Journal of Speech, Language, and Hearing Research **56**(3), 1035–1044 (2013)
18. Happé, F.: Theory of mind and the self. Annals of the New York Academy of Sciences **1001**(1), 134–144 (2003)
19. Harms, M.B., Martin, A., Wallace, G.L.: Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. Neuropsychology review **20**(3), 290–322 (2010)
20. Hashemi, J., Spina, T.V., Tepper, M., Esler, A., Morellas, V., Papanikolopoulos, N., Sapiro, G.: A computer vision approach for the assessment of autism-related behavioral markers. In: Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on. pp. 1–7. IEEE (2012)
21. Hoekstra, R.A., Bartels, M., Cath, D.C., Boomsma, D.I.: Factor structure, reliability and criterion validity of the autism-spectrum quotient (aq): a study in dutch population and patient groups. Journal of autism and developmental disorders **38**(8), 1555–1566 (2008)
22. Ittichaichareon, C., Suksri, S., Yingthawornsuk, T.: Speech recognition using mfcc. In: International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July. pp. 28–29 (2012)
23. Laksana, E., Baltrušaitis, T., Morency, L.P., Pestian, J.P.: Investigating facial behavior indicators of suicidal ideation. In: Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on. pp. 770–777. IEEE (2017)
24. Lehnhardt, F.G., Gawronski, A., Volpert, K., Schilbach, L., Tepest, R., Vogeley, K.: Das psychosoziale funktionsniveau spätdiagnostizierter patienten mit autismus-spektrum-störungen–eine retrospektive untersuchung im erwachsenenalter. Fortschritte der Neurologie-Psychiatrie **80**(2), 88–97 (2012)
25. Liu, W., Li, M., Yi, L.: Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. Autism research : official journal of the International Society for Autism Research **9**(8), 888–898 (2016)
26. Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., Pickles, A., Rutter, M.: The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. Journal of autism and developmental disorders **30**(3), 205–223 (2000)
27. 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. Journal of Autism and Developmental Disorders **24**(5), 659–685 (1994)
28. Madipakkam, A.R., Rothkirch, M., Dziobek, I., Sterzer, P.: Unconscious avoidance of eye contact in autism spectrum disorder. Scientific reports **7**(1), 13378 (2017)
29. Marchi, E., Schuller, B., Batliner, A., Fridenzon, S., Tal, S., Golan, O.: Emotion in the speech of children with autism spectrum conditions: Prosody and everything else. In: Proceedings 3rd Workshop on Child, Computer and Interaction (WOCCI 2012), Satellite Event of INTERSPEECH 2012 (2012)
30. Mattila, M.L., Kielinen, M., Linna, S.L., Jussila, K., Ebeling, H., Bloigu, R., Joseph, R.M., Moilanen, I.: Autism spectrum disorders according to dsm-iv-tr and

comparison with dsm-5 draft criteria: An epidemiological study. Journal of the American Academy of Child and Adolescent Psychiatry **50**(6), 583–592.e11 (2011)

31. Mavadati, S.M., Mahoor, M.H., Bartlett, K., Trinh, P., Cohn, J.F.: Disfa: A spontaneous facial action intensity database. IEEE Transactions on Affective Computing **4**(2), 151–160 (2013)
32. McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E., Nieto, O.: librosa: Audio and music signal analysis in python. In: Proceedings of the 14th python in science conference. pp. 18–25 (2015)
33. McIntosh, D.N., Reichmann-Decker, A., Winkielman, P., Wilbarger, J.L.: When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. Developmental science **9**(3), 295–302 (2006)
34. McKeown, G., Valstar, M.F., Cowie, R., Pantic, M.: The semaine corpus of emotionally coloured character interactions. In: 2010 IEEE International Conference on Multimedia and Expo. pp. 1079–1084. IEEE (2010)
35. Minshew, N.J., Meyer, J., Goldstein, G.: Abstract reasoning in autism: A disassociation between concept formation and concept identification. Neuropsychology **16**(3), 327 (2002)
36. Moore, E., Clements, M.A., Peifer, J.W., Weisser, L.: Critical analysis of the impact of glottal features in the classification of clinical depression in speech. IEEE transactions on bio-medical engineering **55**(1), 96–107 (2008)
37. Van de Mortel, T.F., et al.: Faking it: social desirability response bias in self-report research. Australian Journal of Advanced Nursing, The **25**(4), 40 (2008)
38. Nakai, Y., Takiguchi, T., Matsui, G., Yamaoka, N., Takada, S.: Detecting abnormal word utterances in children with autism spectrum disorders: Machine-learning-based voice analysis versus speech therapists. Perceptual and motor skills **124**(5), 961–973 (2017)
39. Nasir, M., Jati, A., Shivakumar, P.G., Nallan Chakravarthula, S., Georgiou, P.: Multimodal and multiresolution depression detection from speech and facial landmark features. In: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. pp. 43–50. ACM (2016)
40. Newschaffer, C.J., Croen, L.A., Daniels, J., Giarelli, E., Grether, J.K., Levy, S.E., Mandell, D.S., Miller, L.A., Pinto-Martin, J., Reaven, J., Reynolds, A.M., Rice, C.E., Schendel, D., Windham, G.C.: The epidemiology of autism spectrum disorders. Annual review of public health **28**, 235–258 (2007)
41. Organization, W.H.: The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. ICD-10 classification of mental and behavioural disorders / World Health Organization, World Health Organization (1992), https://books.google.de/books?id=DFM0DgAAQBAJ
42. Pelphrey, K.A., Sasson, N.J., Reznick, J.S., Paul, G., Goldman, B.D., Piven, J.: Visual scanning of faces in autism. Journal of autism and developmental disorders **32**(4), 249–261 (2002)
43. Seibt, B., Mühlberger, A., Likowski, K., Weyers, P.: Facial mimicry in its social setting. Frontiers in psychology **6**, 1122 (2015)
44. Sharda, M., Subhadra, T.P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., Sen, A., Singhal, N., Erickson, D., Singh, N.C.: Sounds of melodypitch patterns of speech in autism. Neuroscience letters **478**(1), 42–45 (2010)
45. Shriberg, L.D., Paul, R., McSweeny, J.L., Klin, A., Cohen, D.J., Volkmar, F.R.: Speech and prosody characteristics of adolescents and adults with high-functioning autism and asperger syndrome. Journal of Speech, Language, and Hearing Research **44**(5), 1097–1115 (2001)

46. Stagg, S.D., Slavny, R., Hand, C., Cardoso, A., Smith, P.: Does facial expressivity count? how typically developing children respond initially to children with autism. Autism **18**(6), 704–711 (2014)
47. Tian, Y.L., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE transactions on pattern analysis and machine intelligence **23**(2), 97–115 (2001)
48. Tron, T., Peled, A., Grinsphoon, A., Weinshall, D.: Automated facial expressions analysis in schizophrenia: A continuous dynamic approach. In: Serino, S., Matic, A., Giakoumis, D., Lopez, G., Cipresso, P. (eds.) Pervasive Computing Paradigms for Mental Health, Communications in Computer and Information Science, vol. 604, pp. 72–81. Springer International Publishing and Imprint and Springer, Cham (2016)
49. Valstar, M.F., Almaev, T., Girard, J.M., McKeown, G., Mehu, M., Yin, L., Pantic, M., Cohn, J.F.: Fera 2015 - second facial expression recognition and analysis challenge. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). pp. 1–8. IEEE (2015)
50. White, S.W., Ollendick, T.H., Bray, B.C.: College students on the autism spectrum: prevalence and associated problems. Autism : the international journal of research and practice **15**(6), 683–701 (2011)
51. Wieckowski, A.T., White, S.W.: Eye-gaze analysis of facial emotion recognition and expression in adolescents with asd. Journal of clinical child and adolescent psychology : the official journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53 **46**(1), 110–124 (2017)
52. Wood, E., Baltrusaitis, T., Zhang, X., Sugano, Y., Robinson, P., Bulling, A.: Rendering of eyes for eye-shape registration and gaze estimation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3756–3764 (2015)
53. Wood, E., Bulling, A.: Eyetab: Model-based gaze estimation on unmodified tablet computers. In: Proceedings of the Symposium on Eye Tracking Research and Applications. pp. 207–210. ACM (2014)
54. Woodbury-Smith, M.R., Robinson, J., Wheelwright, S., Baron-Cohen, S.: Screening adults for asperger syndrome using the aq: A preliminary study of its diagnostic validity in clinical practice. Journal of autism and developmental disorders **35**(3), 331–335 (2005)
55. Zhang, L., Sun, Y., Chen, F., Wu, D., Tang, J., Han, X., Ye, J., Wang, K.: Psychometric properties of the autism-spectrum quotient in both clinical and non-clinical samples: Chinese version for mainland china. BMC psychiatry **16**(1), 213 (2016)
56. Zhang, X., Yin, L., Cohn, J.F., Canavan, S., Reale, M., Horowitz, A., Liu, P., Girard, J.M.: Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. Image and Vision Computing **32**(10), 692–706 (2014)
57. Zhang, X., Sugano, Y., Fritz, M., Bulling, A.: Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)
58. Zhao, S., Uono, S., Yoshimura, S., Kubota, Y., Toichi, M.: Atypical gaze cueing pattern in a complex environment in individuals with asd. Journal of autism and developmental disorders **47**(7), 1978–1986 (2017)