REVIEW PAPER

Using Machine Learning for Motion Analysis to Early Detect Autism Spectrum Disorder: A Systematic Review

Roberta Simeoli[1,2] · Angelo Rega[1,2] · Mariangela Cerasuolo[3,4] · Raffaele Nappo[2] · Davide Marocco[1]

Abstract

Diagnosis of autism spectrum disorder (ASD) is typically performed using traditional tools based on behavioral observations. However, these diagnosis methods are time-consuming and can be misleading. Integrating machine learning algorithms with technological screening tools within the typical behavioral observations can possibly enhance the traditional assessment and diagnostic process. In the last two decades, to improve the accuracy and reliability of autism detection, many clinicians and researchers began to develop new screening methods by means of advanced technology like machine learning (ML). These methods include artificial neural networks (ANN), support vector machines (SVM), a priori algorithms, and decision trees (DT). Mostly, these methods have been applied to pre-existing datasets, derived from the standard diagnostic and assessment tools, to implement and test predictive models. On the other hand, the detection of new objective behavioral measures such as biomarkers could lead to a significant strengthening of existing screening tools. In the present study, we carried out a critical review of the literature about the latest findings in this field. The aim was to shed light about the effectiveness of using ML systems for motion analysis to enhance both clinical assessment and diagnostic processes. Specifically, we discussed the contribution of ML systems in promoting early diagnosis of ASD. The literature review showed that motion patterns ML analysis predicts ASD classification as accurately as that of classical gold standard tools. However, the application of these methods is still challenging, as discussed in this review.

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder which includes deficits in social communication and social interaction with the presence of restricted and repetitive behaviors. These behaviors are referred to as stereotyped or repetitive motor movements, insistence on sameness, difficulties with changes in routine and rigid patterns of behavior, restricted interest, and unusual response to sensory aspects. Although motor impairments have been generally observed in ASD, their importance is not taken into account. Importantly, motor impairments in ASD significantly impact the quality of life and social development of ASD (Lai et al., 2014). Motor abnormalities in ASD can occur very early in development (Brian et al., 2008; Teitelbaum et al., 1998), they become more evident over the years (Fournier et al., 2010) and they may evolve into a pervasive feature of the disorder. Several manifestations of these motor abnormalities have been detected and they can include abnormalities in walking patterns (e.g., Gong et al., 2020; Nobile et al., 2011; Rinehart & McGinley, 2010), hand movements such as reaching and grasping (e.g., Bäckström et al., 2021; Forti et al., 2011), and eye-hand coordination (e.g., Crippa et al., 2013). Considering the need of a more reliable and earlier diagnostic process for ASD, motion features could represent an effective precursor.

In the field of neurodevelopmental disorders, early detection can lead to early intervention, which is generally crucial for both children and families. Recently,

✉ Roberta Simeoli
  roberta.simeoli@unina.it

[1] Department of Humanistic Studies, University of Naples Federico II, 80138 Naples, Italy

[2] Neapolisanit S.R.L. Rehabilitation Center, 80044 Ottaviano, Italy

[3] Associazione Italiana Per L'Assistenza Spastici Onlus Sez Di Cicciano, 80033 Cicciano, Italy

[4] Faculty of Medicine, University of Ostrava, 70103 Ostrava, Czech Republic

researchers highlighted the importance of early identification to improve the impact and effectiveness of interventions (Koegel et al., 2014). However, currently, the diagnostic process involves a series of tests that may take hours of clinical examination, and their validity is provided after the child is 3 years old (e.g., Autism Diagnostic Observation Schedule, ADOS; Lord et al., 1999). Moreover, the variability of symptoms makes ASD identification even more complicated. In that regard, computer technologies have been embraced to provide useful guidance through the assessment and diagnostic process. The gold standard for ASD diagnosis has been formalized with clinical symptom judgment through semi-structured game-based behavioral observations (Lord et al., 2000) and standardized interviews or questionnaires. Recent studies have begun to explore the predictive role of some behavioral measures such as eye-gaze or motion patterns as potential objective measures of the disorder, aiming at identifying a well-defined phenotype and enabling a computer-aided diagnosis perspective. These studies typically implement classification methods based on machine learning (ML) algorithms to predict or classify individuals of different groups by maximizing the distance between groups of data sets. Specifically, they seek to identify the patterns or distinctive traits that clearly separate the different groups of individuals within the dataset, to provide an accurate classification.

Recent literature suggests that motor abnormalities could be a consistent marker of ASD (Dowd et al., 2012). In fact, some studies have shown that the severity of motor deficits is related to the degree of social withdrawal and with the severity of symptoms in general (Freitag et al., 2007). Furthermore, it appeared that motor disorders in children with ASD differ in specific characteristics from those of children with other neurodevelopmental disorders (e.g., attention deficit/ hyperactivity-ADHD; Ament et al., 2015; Izawa et al., 2012).

Other evidence in that regard has been provided using wearable sensors or video analysis systems for motion detection integrated with artificial intelligence (AI) systems (e.g., Perego et al., 2009; Vabalas et al., 2019). They reported that it could be possible to observe and quantitatively define the motor abnormalities and the underlying problems of interaction with the surrounding world. Recent advances in computer-assisted analysis such as ML have been applied to the clinical field for ASD assessment and detection with the aim of enhancing the accuracy, time, and quality of the diagnostic process. ML methods include artificial neural network (ANN), supporting vector machine (SVM), a priori algorithms, and decision trees (DT), most of which have been applied to autism-related datasets to build predictive models.

Thus, the assessment of motor functions in ASD and their analysis through ML may clarify the neurobiological basis of the disease and improve the diagnostic process, especially for the early diagnosis perspective.

This literature review aims at collecting and critically discussing recent findings on ML application for ASD detection through motion analysis. The studies analyzed the use of different technological devices and different assessment settings and tasks for a new perspective diagnostic model.

Specifically, in this literature review, the focus has been on studies that applied ML models to measures derived from both body movement and eye movement. As these studies are relatively recent, there is not yet an established gold standard for the standardized application of these methods. Researchers are still struggling to find the most appropriate features, technological tool, and ML methods for optimal efficacy.

Therefore, the present work aims at critically describe the potentials and limitations of these new approaches, trying to summarize the different features, methods, and tools considered so far, while also critically considering the necessity for future research to evaluate the interaction of different and more specific motion features. This is crucial for achieving a more comprehensive and nuanced understanding of the underlying processes of ASD and its associated diagnostic frameworks.

Machine Learning Application in Clinical Dataset

Supervised ML commonly refers to all procedures that train a computer algorithm to identify a complex pattern of data (i.e., "features") that can then be used to predict group membership of new subjects (e.g., patients vs. controls). The procedure to diagnose ASD mainly involves identifying the right class, namely ASD or typically developing (TD), depending on input features. The classification procedure is a predictive work that uses intelligent, automated methods to diagnose whether a child fits with ASD criteria or not. The automated tool uses the input data to perform learning, and its efficiency is measured by the extent to which it can predict the diagnostic class. ML techniques require a well-defined dataset for the training phase to extract the classification algorithm that best separates the groups, whereas, in the testing phase, the classification algorithm can be used to predict the class membership of a participant not involved in the previous training phase. It is important to specify that we can distinguish at least two different approaches in the field of ML, the one described so far refers to supervised ML approach. On the other hand, unsupervised ML is a type of learning where the model is trained on input data without corresponding labels. In this case, the main goal is to discover patterns or hidden structures within the data. The model attempts to group the data based on similarities or find recurring patterns. Unsupervised learning is useful when there are no output labels available to train the model

or when you want to explore the data to identify interesting patterns or clusters of similar data. Some examples of unsupervised learning algorithms include principal component analysis (PCA), clustering (such as k-means), and dimensionality reduction. We could therefore summarize that supervised learning requires labeled input data to train the model to make accurate predictions, while unsupervised learning aims to discover hidden structures or groupings in unlabeled input data (Milano et al, 2023).

In this scenario, the use of pattern recognition methods to predict group membership should not be conceived merely in a potentially ''diagnostic'' perspective but also as a useful tool to identify new objective measures for each individual, leading to a phenotypic redefinition of the disorder.

In contrast to traditional programming, ML attempts to extrapolate algorithms from data exclusively. The development of ML can be broadly organized into supervised or unsupervised models. The main goal of supervised classification is to learn rules from examples in different groups and use these rules to predict unseen cases into prospective classes as accurately as possible. Predictive models are developed based on observed characteristics from examples in each class. The model is first trained using labeled samples. A set of input feature variables is used as input and a desired class value is used as output. The obtained classifier model is then used to recognize new samples and categorize their class. On the other hand, unsupervised ML uses training data that do not include any output information (i.e., labels). Unsupervised models (e.g., clustering and association rules) can provide descriptive knowledge to help understand the inherent structure or properties of the data.

The reliability of the classification typically depends on the learning algorithm used by the classifier, the size and quality of the training data (Sarker, 2021). Since high-dimensional datasets retain irrelevant, noisy, and redundant traits, the more the data dimensionality is, the more the optimization of features will get slower. For this reason, feature selection methods are used to decrease data dimensionality, choosing the most relevant features to enhance classification accuracy and minimize computational cost.

Evaluation metrics are used to measure the performances of ML predictive models. The most popular evaluation metrics for predicting ASD are (i) accuracy: it finds the number of test cases that have been appropriately classified from the absolute number of test cases; (ii) sensitivity: it detects the proportion of actual positive test cases that have been correctly identified; and (iii) specificity: it indicates the percentage of actual negatives test cases correctly identified.

One of the major risks in using ML is overfitting. Specifically, overfitting is a phenomenon where a predictive model becomes excessively complex and starts to memorize the noise or random fluctuations present in the training data, rather than learning the underlying patterns that generalize

well to unseen data. This leads to poor performance on new data because the model is overly specialized to the idiosyncrasies of the training set.

In order to control these risks cross-validation methods are used to assess the predictive ability of a model and generate its efficiency. Specifically, cross-validation uses different portions of the data to test and train a model on different iterations.

The most used cross-validation methods are (i) the k-fold cross-validation method, according to this strategy, the dataset is divided into k subsets of the same size. One of them is selected as a test set, and the other k − 1 composes the training set. Each set becomes a test set for one time, and thus, this process is repeated k times; (ii) the leave-one-out (LOO) cross-validation consisting in applying the algorithm once for each instance, using all other instances as a training set and using the selected instance as a single-item test set; and (iii) the leave-one-subject-out (LOSO) cross-validation method is a technique used in the field of cross-validation to assess the performance of a predictive model. Specifically, it involves excluding a complete subject from the training set and using it as the test set for each iteration. This process is repeated for all available subjects, allowing for a reliable estimation of the model's generalization abilities.

These methods help in constraining the model complexity, generating synthetic data, and evaluating the model's performance on multiple subsets of the limited data, respectively.

Method

Extensive research was carried out by the properly trained reviewer (R.S.) on Pubmed. The search strategy consisted of the following keywords algorithm: (((autism[Title/Abstract]) AND ("machine learning"[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta-analysis[Title/Abstract]).

Articles were included if:

a. The sample included individuals with a clinical diagnosis of autism, infantile autism, ASD, PDD-NOS, or Asperger's Syndrome (AS) with and without an intellectual disability.
b. ML was used for data analysis.
c. Features selected for the analysis included eye or body patterns of movements.

Articles were excluded if:

a. The dataset includes qualitative behavioral data, traditional tools results or parent-report measures, medical or genetic data, patterns of vocal production.

b. The study was a prospective validation, preliminary pilot study, or single-subject study.
c. ML was used to analyze rehabilitation effects.
d. ML was used to detect specific symptoms or behavior (e.g., self-injury behavior).
e. Analysis was applied on biomarkers data recorded during tasks or activities that required previous acquisition (e.g., eye-movement during a text reading task).

The titles and abstracts of the search results were then screened, and the relevant papers were identified. As shown in Fig. 1, database searches produced 447 records. After the screening of titles and abstracts, 64 full-text articles were obtained for further analysis, and 25 articles met our inclusion criteria. Selected articles were primarily divided based on the type of features used for classification (e.g., eye motion-related features, body motion-related features), then organized according to the task (e.g., imitation task, and reach and grasp tasks) and to the specific body part analyzed (e.g., finger, and skeletal keypoints).

Results

Results of our review have been organized into two separate sections:

i. Body motion-related features extraction, which includes studies that analyzed measures of body movements collected using different tools (e.g., smart tablet device; motion sensor) and during specific tasks (e.g., reach and grasp tasks; imitation tasks). All these studies are collected in Table 1;

ii. Eye motion-related feature extraction, which includes all the studies that used eye-gaze measures to classify ASD. These studies are collected in Table 2.

Body Motion-Related Features Extraction

The symptomatology related to ASD syndrome essentially derives from an impairment of learning processes that normally occur early in childhood. Several studies showed that the processes of active perception are compromised in ASD, which could lead to abnormalities in planning processes, serial and prospective coordination (von Hofsten & Rosander 2012). An ineffective information processing from the outside world may result in cognitive and language developmental delay and therefore in social interaction. Thus, by properly tracing the way of moving of ASD people within the environment, it is possible to detect the specific motor patterns that characterize their interaction with the world. ML systems can identify typical motor patterns of the disorder and effectively support the diagnostic process. In particular, the potential of identifying such invisible and objective features enables early diagnosis of the disorder.

Fig. 1 Adapted from the PRISMA flow diagram (Moher et. al., 2009), showing the steps of the selection process undergone by the studies in this review
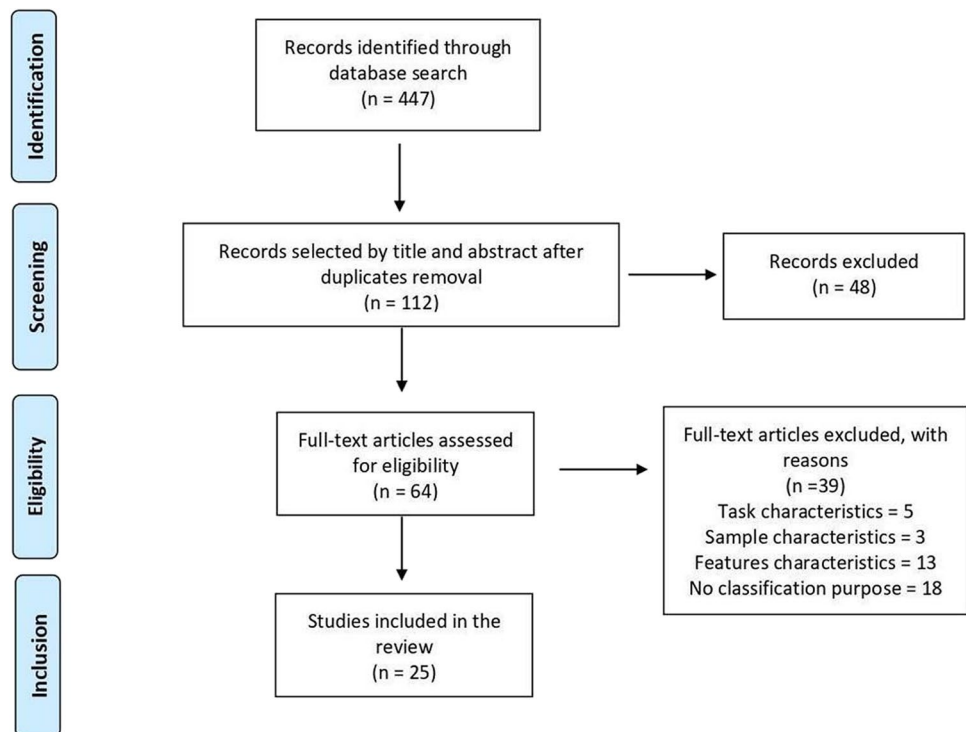
Table 1 The table summarizes the key components of the studies that analyzed body motion data, including participant characteristics (e.g., number of participants, diagnosis, age); the type of tool used for measurements; the task participants had to perform during the assessment; the type of features recorded; the adopted machine learning method; and the obtained results

Study	Participants	Apparatus	Task	Features	ML methods	Results
Crippa et al. (2015)	15 ASD 15 TD Age: preschool-aged	Optoelectronic system with 8 infrared-motion analysis cameras and passive markers attached to participants' wrists and to the hand dorsum Two markers placed on the ball and four on the box	Reach to-drop task consisting in grasping a rubber ball and dropping it in a hole	17 kinematic measures based on: Duration; velocity; acceleration; Wrist angle at the time of the ball drop was included	SVM was used for classification analysis and a LOO cross-validation strategy was applied	The classification accuracy reached a maximum accuracy of 96.7% (specificity 93.8% and sensitivity 100%) by using seven features selected by the Fisher discriminant ratio-based technique. Overall mean accuracy, specificity, and sensitivity rates were also calculated over a number of selected features ranging from 1 to 17. The overall mean classification accuracy (specificity/sensitivity) was 84.9% (mean specificity 89.1% and mean sensitivity 82.2%)
Cavallo et al. (2021)	20 ASD 20 TD Mean age: TD = 9.5 ± 1.5 yrs ASD = 9.8 ± 1.5 yrs	A near-infrared camera motion capture system 8 retro-reflective hemispheric markers positioned on the right hand	Reaching towards and grasping a bottle with one of three prospective intentions: place the bottle into a box; pour some water; pass the bottle to another person	Kinematic measures: wrist velocity; wrist acceleration; wrist jerk; grip aperture; wrist height	Analysis was based on a SVM-G model Record-wise cross-validation and subject-wise cross-validation were applied	Classifiers trained and evaluated using record-wise splits achieved perfect accuracy in discriminating ASD versus TD accuracy 100% Classification accuracy of classifiers trained with subject-wise data splits reached 75%
Vabalas et al. (2019)	24 ASD 22 TD Age range: 31.5–30.8 yrs	Polhemus Fastrak motion sensor attached to the distal phalange of the index finger	Pointing task performed with both dominant and non-dominant hand	Movement measures: X, Y, Z coordinates; Velocity; acceleration; jerk; amplitude	A SVM algorithm was used Four feature selection methods were used: SVM-RFE; Student's T test; ReliefF; mRMR A nested cross-validation method was used	Statistically significant classification accuracy of 71%, with a sensitivity of 75%, specificity 66% was achieved with t-test feature selection 70% accuracy, 72% sensitivity, 68% specificity was obtained with ReliefF feature selection When the SVM-RFE and mRMR were used, classification results were not significant

Table 1 (continued)

Study	Participants	Apparatus	Task	Features	ML methods	Results
Vabalas et al. (2020)	22 ASD 22 TD Mean age: TD = 32.6 yrs ASD = 28 yrs	A Polhemus Fastrak motion sensor attached to the distal phalange of the index finger An EyeLink 1000 Plus eye-tracker was used to collect eye movement	Movement imitation task performed with both dominant and non-dominant hand	Kinematic features based on: velocity; acceleration; jerk; amplitude Eye movement features based on: saccade measures and visual attention	SVM algorithm and a nested cross-validation strategy were used Feature selection method applied: Student's t-test; ReliefF; K-nearest neighbor; mRMR	Analysis of kinematic data reached a classification accuracy of 73%, with a sensitivity of 88%, specificity of 59% Eye data analysis showed a 70% accuracy, 43% sensitivity, 97% specificity Analysis of combined data achieved a 78% accuracy, 57% sensitivity, 99% specificity The t-test feature selection consistently produced the most stable feature sets
Anzulewicz et al. (2016)	37 ASD 45 TD Age range: 3–6 yrs	Touch screen and inertial sensors inside a smart tablet device	Children were engaged in drawing and matching tasks during gameplay displayed on a tablet	262 kinematic features based on: velocity, acceleration; pressure; etc	Three ML algorithms were employed: ET; RF; RGF 10 repetitions of a tenfold cross-validation were applied	Each algorithm differentiated individuals within the autism group from the control group using the 262 features derived from the touch screen and inertial sensors Creativity gameplay produced greater predictive accuracy than those from Sharing gameplay The most effective algorithm was the RGF with age and gender features excluded, reaching an accuracy up to 93%
Li et al. (2017)	14 ASD 14 TD Mean age: ASD = 32 ± 7.69 TD = 29.31 ± 8.38	Polhemus Liberty motion sensor attached to the index finger of the dominant hand	Hand movement imitation task	Kinematic features were extracted from each trial and were based on position, velocity and acceleration	Four classification models were used: Naive Bayes; SVM; DT; RF LOO cross-validation was applied	The Naive Bayes classification models achieved the highest average accuracies of 64.5% and 56.7%, in specific hand conditions. The SVM achieved a relatively better result than Naive Bayes when the kinematic features were reduced from 40 to 9 with an accuracy of 86.7%

Table 1 (continued)

Study	Participants	Apparatus	Task	Features	ML methods	Results
Emanuele et al. (2021)	13 ASD 13 TD Age range: ASD=44–82 mts TD=43–71 mts	Optoelectronic motion capture system with 3 infrared cameras 10 passive markers were attached to the upper limbs and one to the manubrium of the sternum	Children were instructed to reach for and grasp a felt-tip pen to complete a figure printed on a sheet of paper	17 Kinematic features were extracted from each trial and were based on position, velocity, and acceleration	A SVM was used LOSO cross-validation was applied	Classification accuracy reached up to 94.7% in distinguishing ASD from TD
Raya et al. (2020a)	24 ASD 25 TD Age range: 4–7 yrs	An RGB-D camera was used to estimate the pose in real-time of the body	Imitation task during a virtual experience enriched by visual, auditory and olfactive stimuli delivered as different conditions	Kinematic features of different body area: head (nose, eyes, and ears), trunk (neck and mid hip), arms (shoulders, elbows, and wrists), legs (hips, knees, ankles), and feet (big toes, small toes and heels)	SVM with a LOSO cross-validation procedure has been applied	The best classification accuracy reached 82.98% for the head and trunk parameter, independently by the specific stimuli condition
Robles et al. (2022)	6 ASD 13 TD Mean age: ASD=28 yrs TD=23 yrs	HTC VIVE Pro Eye VR System and the Tobii XR SDK	Participants involved in a virtual reality experience were instructed to purchase items shown on a list in the virtual environment representing a supermarket by pointing to and selecting the product in the shop	27 kinematic and eye-gaze features, including hand, head, and gaze behavior	LR, SVM, and a neural network were used. Each algorithm was evaluated applying five-fold cross-validation	LR: average accuracy of 80%, sensitivity of 80% and specificity of 80%; SVM achieves an average accuracy of 63.3%, sensitivity of 80% and specificity of 60% for the matched data set and an accuracy of 80%, sensitivity of 80% and specificity of 80% for the random data set; A neural network consisting of one hidden layer achieved an accuracy of 76.7%, sensitivity of 80% and specificity of 70% for the matched data set. An accuracy of 93.30%, sensitivity of 80% and specificity of 100% for the random data set is achieved

Table 1 (continued)

Study	Participants	Apparatus	Task	Features	ML methods	Results
Kojovic et al. (2021)	34 ASD 34 TD Age range: 1.2–5.1 yrs	Multi-person 2D pose estimation OpenPose technology	Interaction with the experimenter during ADOS	skeletal keypoints (e.g., nose, shoulder, elbow, and wrist)	CNN	The trained model achieves an accuracy of 80.9% in discriminating the two groups
Perego et al. (2009)	10 ASD 10 TD Mean age: 41 mts	Optoelectronic system, with 8 infrared cameras	The experimental task consisted of grasping a rubber ball placed over a support and throwing it in a see-through squared basket	Kinematic variables of reach-and-throw movement. Measures were based on position, velocity and acceleration	SVM has been applied. A 10 k-fold cross-validation method was used	The best accuracy (100%) was obtained with a soft margin classification, where high accuracy is achieved by means of a narrow separation margin. Very good accuracy (92.5%) was obtained even in a more conservative case with a harder margin classification, assuring the power of the classification system in view of new collecting data
Simeoli et al. (2021)	30 ASD 30 TD Age range: 6–7 yrs	Smart tablet device	Cognitive tasks that required finger movement on the screen of the tablet	Finger movement coordinates	ANN and tenfold cross-validation method has been applied	ANN correctly differentiated individuals within ASD and TD groups with an accuracy of 93% (sensitivity 87%, specificity 98%)
Li et al. (2020)	25 ASD 25 TD Age range: 5–12 yrs	COP data were collected using a force plate	Participants were instructed to stand barefoot on two feet and maintain a stationary stance for 20 s during two conditions: (1) eyes open and (2) eyes closed	12 features (six COP variables by two conditions) of postural sway. The COP variables were computed, including linear displacement, total distance, sway area, and complexity	The ML classifiers included: discriminant analysis, K-nearest neighbor, naïve Bayes; DT; SVM; RF fivefold cross-validation method was applied	All ML classifiers used in the present study successfully identified ASD postural control patterns based on the COP features with high accuracy rates (>80%). Among the ML classifiers, the naïve Bayes method was the optimal means to identify ASD postural control with the highest accuracy rate (90%). The second-best algorithm was the K-nearest neighbor method that had an accuracy rate of 86%

Table 1 (continued)

Study	Participants	Apparatus	Task	Features	ML methods	Results
Zhao et al. (2022)	20 ASD 23 TD Age range: 7–13 yrs	The OpenFace 2.0 was utilized to extract the time series of the participant's head movement	Participants were engaged in a conversation with a 33-year-old female interviewer	The head rotation range (RR) and the amount of rotation per minute (ARPM) in the pitch (head nodding direction), yaw (head shaking direction) and roll (lateral head inclination)	Five classifiers were used including: SVM; LDA; DT; RF, and ensembles learning with boosting (ENS) LOO cross-validation was applied	The maximum classification accuracy of 92.11% was achieved with the DT classifier with two features (i.e., RR_Pitch and ARPM_Yaw)

ML machine learning, *SVM* support vector machine, *SVM-RFE* support vector machine—recursive feature elimination, *SVM-G* support vector machine Gaussian kernel, *LOO* leave-one-out, *LOSO* leave-one-subject-out, *ET* ExtraTrees, *RF* random forest, *RGF* regularized greedy forest, *DT* decision tree, *LR* logistic regression, *CNN* convolutional neural network, *ANN* artificial neural network, *COP* center of pressure, *mRMR* minimum redundancy maximum relevance, *ADOS* Autism Diagnostic Observation Schedule

The studies listed in Table 1 analyzed such motion features. The following sections describe these studies grouped according to the task and the type of motion features examined.

Reach and Grasp Tasks

Four studies collected and analyzed kinematic measures during a reach and grasp task. Authors used different tools to record the data and different classification models. Crippa et al., (2015), identified 17 kinematic measures of a reach and grasp movements and recorded them through an optoelectronic system and passive markers. They applied a SVM classifier method with leave-one-out (LOO) cross-validation strategy and a Fisher discriminant ratio (FDR)-based technique (Padilla et al., 2012) was used for feature selection. Results revealed a maximum classification accuracy of 96.7% using 7 of the 17 features extracted. Authors pointed out that all these 7 kinematic features are related to the second part of the movement in which the child transported the ball from a support to the target hole where the ball was to be dropped, suggesting that goal-oriented movements may be critical in separating children with ASD from typical developing (TD) children.

Cavallo et al. (2021) also collected kinematic measures of a reach and grasp movement through a camera motion capture and hand markers. Analyses were based on a SVM model that used a Gaussian kernel (SVM-G) to compute the hyperplane that best separated the different trials in each diagnosis group. Models were trained, validated, and tested using a record-wise cross-validation data split. Classification accuracy was computed as the fraction of correctly classified trials for each cross-validation iteration reaching an accuracy of 75%.

Emanuele et al. (2021) used an optoelectronic motion capture system and ten passive markers attached symmetrically to the upper limbs and one to the manubrium of the sternum to record reach and grasp movement during a reach and grasp task. They applied a SVM algorithm reaching up to 95% of accuracy. In Perego et al.'s study (2009), kinematic variables of reach-and-throw movement were recorded by eight infrared cameras, during an experimental task consisting in grasping a rubber ball placed over a support and throwing it in a see-through squared basket. A SVM has been applied with a 10 k-fold cross-validation method and 92.5% of accuracy was achieved.

Imitation Task

Vabalas et al., (2019, Vabalas et al. 2020) and Li et al. (2017) analyzed finger movement while participants were engaged in a pointing imitation task. The researchers collected data using a motion sensor attached to the distal phalange of the index

finger and eye movement data with a remote eye tracker. Kinematic features of finger movement were based on velocity, acceleration, jerk, and amplitude of movement. The authors used a SVM algorithm with nested cross-validation for both datasets. Results revealed that the classification accuracy increased to 78% when the two datasets were combined. An accuracy of 73% was achieved for movement data, and 70% for eye-gaze data (Vabalas et al., 2019, Vabalas et al. 2020).

Li et al. (2017) extracted 40 features of finger movement based on position, velocity, and acceleration, and applied different classification methods (Naive Bayes; SVM; DT; Random Forest (RF)); They used three features selection methods (SVM weights; leave-one-parameter-out; PCA weights) the highest accuracy was reached when the SVM was applied on features reduced from 40 to 9 (accuracy of 86.7%). Authors highlighted that the most suitable model must be sought for each specific dataset and it can affect the accuracy of the classification (Li et al., 2017).

Smart Tablet-Based Feature Extraction

Some studies in this review shed light on the potential of using ecological tools, widely used in everyday life, such as smart-table device, to obtain significant classification results. Anzulewicz et al. (2016) and Simeoli et al. (2021) reported significant classification results assessing children during game-play activities. The remarkable advantage is that it is possible to do classification without using sophisticated tools that require a certain degree of expertise from researchers and clinicians. In particular, Anzulewicz et al. (2016) used touch screen and inertial sensors inside a smart tablet device to record finger kinematics during a game-play. Authors collected 262 kinematic data and analyzed the dataset with different algorithms, reaching a classification accuracy of 93%. Simeoli et al. (2021) also collected finger motion coordinates during a simple cognitive task administered through a smart tablet device. Authors used 12 motion features based on acceleration, velocity, straightness, directional changes, etc., and applied a feedforward multi-layer perceptron ANN, obtaining a classification accuracy of 93% (Simeoli et al., 2021).

Pose Estimation Features Extraction

Several studies in this review reported results of classification using features of movement of different parts of the body (e.g., pose in real-time, skeletal keypoints, and center of pressure (COP)). For example, Raya et al., (2020a, 2020b) used an RGB-D camera to estimate the pose in real-time of the body. The authors analyzed 25 joints divided in different parts of the body, during a virtual experience enriched by visual, auditory, and olfactive stimuli delivered as different conditions. A SVM with leave-one-subject-out (LOSO)

cross-validation procedure was applied to the data and results revealed an accuracy of 82.98% using the head and trunk parameters, independently by the stimuli conditions.

Robles et al. (2022) also extracted a total of 27 features, including hand, head, and gaze behavior, recorded using the HTC VIVE Pro Eye VR System and the Tobii XR SDK, respectively. Logistic regression (LR), SVM, and an ANN were used to analyze the data and each algorithm was evaluated applying fivefold cross-validation. Authors stressed the difference in classification accuracy obtained on random and matched dataset, where matched dataset means grouped for comparable age and IQ scores. In fact, the highest accuracy was reached when the random dataset was used. In particular, the ANN on the random dataset achieved an accuracy of 93.3%, whereas an accuracy of 86.70% was reached on the matched dataset (Robles et al., 2022).

Kojovic et al. (2021) used a multi-person 2D pose estimation OpenPose technology to extract several skeletal keypoints during an interaction with the experimenter providing an ADOS. A Convolutional Neural Network (CNN) was applied on the dataset achieving an accuracy of 80.9%.

Zhao et al. (2022) also used an OpenFace 2.0 and extracted the time series of the participant's head movement from the video clip while they were engaged in a conversation. Head rotation range, the amount of rotation per minute in the pitch (head nodding direction), yaw (head shaking direction), and roll (lateral head inclination) were recorded, and five classifiers were used (SVM; Linear Discriminant Analysis (LDA); DT; RF). The maximum classification accuracy of 92.11% was achieved with the DT classifier and two features. Differently, Li et al. (2020) used a force plate to extract 12 features of COP. The authors tested different ML classifiers (discriminant analysis, K-nearest neighbor, naïve Bayes, DT, SVM, RF), and all these methods successfully identified ASD postural control patterns with an average accuracy of 80%. Among them, the naïve Bayes method was the optimal means with the highest accuracy of 90%. The second-best algorithm was the K-nearest neighbor method with an accuracy of 86% (Li et al., 2020).

Eye Motion-Related Feature Extraction

Eye tracking technology has received particular attention in the context of ASD, as ocular gaze abnormalities have been recognized as one of the hallmarks of the disorder (Frazier et al., 2017). It can be described as the process of capturing, tracking, and measuring eye movements or the absolute point of gaze (POG), which refers to the point where the eye gaze is focused in the visual scene. This technology allows an objective and quantitative method for recording the characteristics of eye movement. Eye trackers capture three basic categories of eye movements: fixation, saccade, and blink. A fixation is the moment that occurs while pausing the gaze on

Table 2 The table summarizes the key components of the studies that analyzed eye motion data, including participant characteristics (e.g., number of participants, diagnosis, age); the type of tool used for measurements; the task participants had to perform during the assessment; the type of features recorded; the adopted machine learning method; and the obtained results

Study	Participants	Apparatus	Task	Features	ML methods	Results
Cilia et al. (2021)	30 ASD 29 TD Mean age: 8 yrs	A screen-based eye tracker	The participants were invited to watch a set of photographs and videos, which included scenarios tailored specifically to stimulate the eye movement across the screen area	Coordinates of eye movements and participant's POG during the experiment runtime	CNN composed of 4 convolutional layers, 4 pooling layers, and 2 fully connected layers. In addition, dropout layers were used, which help reduce the possibility of overfitting threefold cross-validation was used	The model could provide a notable prediction accuracy of 90%, sensitivity; 83%, and precision 80%. The data set was split into training and test sets based on a threefold cross-validation using 3 stepwise procedures. The accuracy declined to 71% after splitting the dataset in training and test set
Wan et al. (2019)	37 ASD 37 TD Age range: 4–6 yrs	Remote eye-tracker	Participants were invited to watch a 10-s video of a woman speaking	Fixation time on specific AOI	A SVM was applied on fixation time using five-fold cross-validation	Classification accuracy was 85.1%, sensitivity was 86.5%, and specificity was 83.8%
Tsuchiya et al. (2021)	39 ASD 102 TD 24 CONTR Age range: 5–17 yrs	Gazefinder to capture eye gaze patterns and stimulus movie clips equipped in a personal computer with a monitor	Participants were invited to watch 12 short movie clips automatically presented in a fixed order	On 100 AOIs two types of eye gaze indices were collected: the AOI rate score and the AOI count score	four diagnostic algorithms based on the age and rate/count measures were applied The general classification performance was evaluated using the LOO method	The best-fit algorithm derived from merging for the younger individuals the final AOI rate score algorithm, and for the older individuals the final AOI count score algorithm. The classification accuracy reached was 78%, sensitivity 74%, specificity 80%
Elbattah et al. (2019)	29 ASD 30 TD Mean age: 7.8 yrs	Eye tracker	N/S	Eye movement scanpath as a sequence of consecutive fixations and saccades traced through time and space	Autoencoders, ANN K-means	The clustering experiments empirically confirmed that eye-tracking scanpaths could be grouped into coherent clusters, which largely resembled the original grouping of samples
Alcañiz et al. (2022)	35 ASD 20 TD Age range: 4–7 yrs	Eye tacker	Children were engaged in a virtual experience	Specific eye tracking variables related to the fixation on specific AOIs	naïve Bayes, XGBoost, kNN, RF, and SVM fivefold cross-validation method was used	The main result demonstrated 86% accuracy in the recognition of autistic children when using all eye-tracking variables and the SVM methods

Table 2 (continued)

Study	Participants	Apparatus	Task	Features	ML methods	Results
Jiang et al. (2019)	25 ASD 35 TD Age: ≥8 yrs	Eye-tracking data were collected with two Tobii Pro eye-tracker	Dynamic Affect Recognition Evaluation (DARE) task that requires an individual to recognize one of six emotions while observing a slowly transitioning face video	Task Features: The first category of features described the behavioral performance in the facial emotion recognition task. Gaze features: Fixation time, fixation frequency, fixation duration, and saccade amplitude. Face Features: face features extracted using OpenFace	An RF classifier was used for the three categories of features independently, and then using all features combined. A LOO cross-validation strategy was used in the experiments	The combination of task, gaze, and face features achieved 72.5% classification accuracy for individual fixations. Task features had very low sensitivity, but in combination with gaze and face features, the sensitivity increased to 91.3%, which suggests the important role of eye-tracking data for distinguishing subjects with ASD
Kanhirakadavath and Chandran (2022)	30 ASD 29 TD Age range: 3–13 yrs	SMI RED eye tracker	Eye-tracking scanpath data from existing dataset	Velocity, acceleration, and jerk of eye movement	Boosted Decision Tree (BDT); Deep Support Vector Machine (DSVM); Decision Jungle (DJ); DNN. LOO cross-validation was applied	The BDT model could discriminate ASD from TD with a mean AUC of 67.6%, detecting 64.40% of true positive cases and 57.40% of true negative cases. In comparison, the DNN model achieved a mean AUC of 78.6%, successfully detected 78.57% true positive cases and 75.47% of true negative cases. After data augmentation, the resulting scores show that the AUC of the classical and DNN models improved significantly, with the BDT showing a 10% improvement (AUC 74.6%), and the DNN's improvement was 23.4% (AUC 97.00%)

Table 2 (continued)

Study	Participants	Apparatus	Task	Features	ML methods	Results
Liaqat et al. (2021)	14 ASD 14 TD Age range: 5–12 yrs	Tobii T120 eye-tracker	Free-viewing tasks of natural images	The scan-path data consist of coordinates and duration of eye fixation	Classical multilayer perceptron (MLP) networks and convolutional networks Data augmentation methods was applied to alleviate the overfitting problem	Results revealed that the ASD prediction accuracy reaches 67.23% accuracy on the validation dataset and 62.13% accuracy on the test dataset
Liu et al. (2016)	29 ASD 29 TD/IQ matched 29 TD/AGE matched Age range: 4–11 yrs	Eye tracker	Face recognition task	Face scanning pattern as fixation area	SVM with a LOO cross-validation strategy	The model achieved: accuracy of 88.51%; specificity of 86.21%; sensitivity of 93.10%
Zhao et al. (2021)	20 ASD 23 TD Age range: 6–13 yrs	Head-mounted eye tracker (Tobii Pro Glasses 2)	Eye tracking was performed on children while they were engaged in face-to-face conversations with an interviewer	Four AOIs fixation data were analyzed, including the eyes, mouth, whole face, and whole body. We computed the percentage of visual fixation time on each AOI as features 16 AOI-based features were acquired (4 sessions×4 AOIs)	SVM, LDA, DT, RF LOO cross-validation was used	A maximum classification accuracy of 92.31% was achieved with the SVM classifier by combining features on both visual fixation and session length. The classification accuracy of combined features was higher than that obtained using visual fixation features (maximum classification accuracy 84.62%) or session length (maximum classification accuracy 84.62%) alone
Lin et al. (2022)	55 HAT 52 LAT Mean age: HAT=24.7 LAT=25	Embedded module aGlass is used to track eye movement data in VR	Children engaged in a virtual reality experience	Feature based on fixation on specific region of the scene	LR; Naive Bayes Classifier; SVM; DT; RF; Gradient Boosting Decision Tree (GBDT) threefold cross-validation was applied	The highest accuracy of 70% belonged to the RF model, DT and GBDT. But the RF model showed the highest value even in the test experiment

ML machine learning, *CNN* convolutional neural network, *SVM* support vector machine, *AOI* area-of-interest, *LOO* leave-one-out, *ANN* artificial neural network, *RF* random forest, *DT* decision tree, *DNN* deep neural network, *AUC* area under the ROC curve, *LR* logistic regression, *HAT* high autistic traits, *LAT* low autistic traits, *N/S* not specified

an object so that the brain can perform the perception process. Accurate perception requires constant scanning of the object with rapid eye movements, which are called saccades. On the other hand, a blink is often a sign that the system has lost track of the eye gaze. Eye-tracking scan-paths have been commonly used as a practical means of representing gaze behavior. A scan-path represents a sequence of consecutive fixations and saccades over time and across space.

This literature review section describes results from recent studies that tried to disentangle the question about how ML analytics can help identify and classify ASD based on eye movement (Table 2). Several authors analyzed the visual fixation time on different areas-of-interest (AOI). Some of these authors restricted the field of observation to specific face-areas (Jiang et al., 2019; Liu et al., 2016; Wan et al., 2019; Zhao et al., 2021). In particular, Zhao et al. (2021) analyzed the percentage of visual fixation time on different AOIs using a head-mounted eye tracker during a face-to-face conversation. The authors applied different classification methods (SVM, LDA, DT, and RF) and results revealed that the maximum classification accuracy of 92.31% was achieved by the SVM classifier by combining features on both visual fixation and session length (Zhao et al., 2021). The classification accuracy of combined features was higher than that obtained using visual fixation features (84.62%) or session length (84.62%) separately. Liu et al. (2016) observed face-scanning eye movement patterns of fixation during a face recognition task. They adopted a data-driven feature extraction method and a SVM for the classification. The model achieved an accuracy of 88.51% and the most discriminative facial areas able to explain the 79.31% of accuracy were identified. Particularly, the TD group looked longer than the ASD group in the right eye, and the area slightly above the month, while the ASD group looked longer than the TD group at the area slightly below the left eye (Liu et al., 2016).

Jiang et al. (2019) combined task, face, and gaze features to classify individual fixations. A RF classification method was applied, and results revealed that task features only had very low sensitivity, but in combination with gaze and face features, the sensitivity increased to 91.3%, which suggested the important role of eye-tracking data in distinguishing subjects with ASD (Jiang et al., 2019). Wan et al. (2019) also observed fixation time on specific face areas through a remote eye-tracker tool, while the children were engaged in watching a video of a woman speaking. A SVM method was applied and classification accuracy of 85.1% was achieved.

Based on the assumption that the social saliency within the image content can lead to different eye responses between TD and ASD individuals, some other authors collected eye-fixation data considering wider scenarios (Alcañiz et al., 2022; Cilia et al., 2021; Elbattah et al., 2019; Liaqat et al., 2021; Tsuchiya et al., 2021).

Liaqat et al. (2021) analyzed coordinates and duration of eye fixation while participants were engaged in free-viewing tasks of natural images. Authors found that the ASD prediction accuracy reaches 67.23% using a classical multilayer perceptron (MLP) network. Alcañiz et al. (2022) observed fixation on specific AOI, during a virtual experience. They identified four eye-tracking variables: number of frames in which the participant did not see anything defined as an AOI; number of frames in which the participant did see something defined as an AOI; number of frames in which the participant saw any defined character; and number of frames in which the participant saw any defined item. SVM methods, reached 86% of classification accuracy in the recognition of autistic children when using all eye-tracking variables.

Elbattah et al. (2019) analyzed eye movement scan-path as a sequence of consecutive fixations and saccades traced through time and space using an autoencoder to extract features from the dataset and K-means models to classify their data. The clustering experiments empirically confirmed that eye-tracking scan-paths could be grouped into coherent clusters, which largely resembled the original grouping of samples. Tsuchiya et al., (2021), demonstrated that, using a Gazefinder to capture eye gaze patterns and a specific best-fit diagnostic algorithm, it was possible to discriminate between school-aged children and adolescents with ASD from typically developing individuals of the same age range. Two types of eye gaze indices were collected: the AOI rate score (the percentage of gaze fixation time allocated to each AOI divided by the duration of each movie clip) and the AOI count score (the representation of the presence or absence of a fixed gaze over each AOI, regardless of the duration of the eye gaze). The proposed algorithm was based on age and rate/count score measures. The best-fit algorithm is derived from merging, for the younger individuals, the final AOI rate score algorithm, and for the older individuals, the final AOI count score algorithm. The classification accuracy reached was 78%, while sensitivity and specificity were respectively 74% and 80% (Tsuchiya et al., 2021).

Cilia et al. (2021) used a CNN to analyze POG while children were engaged in watching photographs and videos. Their results showed a classification accuracy of 71%.

On the other hand, Kanhirakadavath and Chandran (2022) did not focus on fixation data but analyzed velocity, acceleration, and jerk of eye movement. They applied a deep neural network (DNN) that achieved an accuracy of 78.6%. Authors underlined that, after a data augmentation process, used to populate the original images dataset, the resulting scores showed a significant improvement of 23.4% (Kanhirakadavath & Chandran 2022).

Unlike previous studies focused on ASD prediction, Lin et al. (2022) proposed a novel technique that combined eye-movement data and ML algorithms for predicting high levels of autistic traits. They engaged children in a virtual reality

experience and analyzed features based on fixation on specific regions of the scene. Authors applied several classification models, but results revealed that the highest accuracy of 70% belonged to the RF model, DT, and gradient boosting decision tree (GBDT), during the preliminary experiment. However, the RF model showed the highest value even in the test experiment (Lin et al., 2022).

Discussion

Traditionally, ASD diagnosis and assessment include a series of explicit qualitative and quantitative measures characterized by semi-structured behavioral tasks' observations in which the examiner rates and scores an individual's response to prompted situations (e.g., the Autism Diagnostic Observation Schedule, ADOS; Lord et al., 1999). These instruments principally focus on social behavior and communication analysis. Assessment of sensory processing, motor abnormalities, and general interaction with the environment remains restricted and unworked.

Although classical gold-standard tools have been widely adopted in ASD clinical practice, several limitations remain (Volkmar et al., 2009) and among them are (i) the absence of explicit sensory functioning assessment; (ii) the subjective evaluation and the examiner's expertise; (iii) the ecological validity of the assessment setting.

The studies reported in this review shed light on the advantages of using ML models and non-invasive technological tools to measure ASD symptoms in ecological situations and assessment settings. Indeed, these studies showed that it is possible to classify ASD and TD individuals within more ecological settings than the classical assessment laboratories and administer tasks more generalizable to different ASD severity degree and developmental age (Crippa et al., 2015).

Contemporaneously summarizing the innovations in the field, this literature review focused on studies that have analyzed both eye and body movement data collected through different technological tools. Tables 3 and 4 summarize the different ML models and the different types of data and technological tools used for analysis. Since, it is not possible to establish a direct comparison between the various methods proposed, considering the extreme variability of conditions, settings, types of data and tasks, and analysis models, we provide an overview of the trend in research within the field.

While examining these studies, it becomes apparent that certain ones illustrate how it is possible to leverage the enjoyment of using serious games and smart table devices to gather information useful for diagnosis, obtaining a good level of classification accuracy (Anzulewicz et al., 2016; Simeoli et al., 2021). Such studies shed light on the potential of identifying objective and invisible measures that do not require behavioral or social skills or previous communications competencies, to increase the chances of working in the direction of early diagnosis.

It is worth noting that the research progressing in this direction, in addition to enhancing diagnostic procedures, has the potential to enhance our understanding of the disorder. The way these individuals move in space can provide further insights into the etiopathogenesis of the disorder, and this literature is moving in this direction, focusing on identifying specific motion features to be associated with symptoms' severity of ASD.

For example, Vabalas and Freeth (2016) observed that people with elevated autistic symptoms exhibited shorter and less frequent saccades. On the other hand, Liberati et al. (2017) showed greater saccade amplitude and higher frequency in children with ASD. However, the authors of these two studies used two distinct eye-tracking tools with a different sampling rate. Therefore, this issue is far to be disentangled, largely depending on the equipment and data used. Furthermore, another variable is the type of condition and stimulation provided during data collection. In fact, other studies showed that eye-tracking was useful to identify children diagnosed with ASD based on the duration of fixations and the number of saccades depending on the type of images displayed (Pierce et al., 2011). These evidences suggest that methods (i.e., materials and procedure) should be more consistent across the studies. Our review highlighted that when the AOI for the analysis was restricted to face area, classification reached higher level of accuracy (Jiang et al., 2019; Liu et al., 2016; Wan et al., 2019; Zhao et al., 2021).

Results from our literature review showed that ML algorithm applied on biomarker dataset can achieve a classification accuracy comparable to classic gold standard tools. However, the use of these methods leads to several controversial issues that future research will have to untangle.

Firstly, the sample size is one of crucial limitation claimed by most author. In fact, a consistent part of the studies collected data from a small sample to test the classification with an independent untrained dataset (Cavallo et al., 2021; Crippa et al., 2015; Vabalas et al., 2020). Small sample size can lead to different problems, among them, training overfittings is one of the most critical in ML. Specifically, a small sample size can exacerbate the risk of overfitting due to limited representation of the underlying population. When the available data is scarce, the model has fewer diverse examples to learn from, making it easier for it to inadvertently fit noise or outliers in the training set. With insufficient data, the model may erroneously attribute importance to random variations or outliers, resulting in an overly complex representation that fails to generalize.

Moreover, a small sample size can lead to high variance in the estimated model parameters. The model may excessively respond to individual training examples, causing fluctuations in its predictions. This sensitivity to small changes

in the training data can lead to overfitting, as the model tries to capture every detail, including random fluctuations specific to the limited training sample.

To mitigate overfitting with a small sample size, techniques such as regularization (e.g., Anzulewicz et al., 2016), data augmentation (e.g., Kanhirakadavath and Chandran 2022), and cross-validation are commonly employed.

Most of the studies in this review report a lack in sample size (e.g., Tsuchiya et al., 2021; Cilia F, et al., 2021; Alcañiz M, et al., 2022). In order to address this issue, authors proposed different solutions such as data augmentation method (e.g., Kanhirakadavath and Chandran 2022), involving an independent dataset (the second control group) to be tested, and applying cross-validation strategy.

Cross-validation is required not only for checking the predictive validity, but also for achieving optimal diagnostic performance. In this review, one of the most used cross-validation strategies was the LOO that is assumed to perform better than other cross-validation methods because it allows a less biased estimation of generalization error. Specifically, since the LOO method trains the model on all samples except one and evaluates on that excluded sample, the estimation of generalization error is less biased compared to techniques like k-fold cross-validation, where training is done on smaller subsets and evaluation is performed on a limited number of samples. This makes the LOO method more accurate for estimating the model's generalization ability (Pontil, 2002).

On the other hand, we can address to LOO also a minimal data loss, since that, for each iteration, a single sample is excluded as the test set, and all other samples are used as the training set. This maximizes the utilization of the available data, avoiding significant information loss compared to other cross-validation methods that may divide the dataset into larger portions.

The studies that evaluated reach and grasp movements showed a high accuracy (average ± 90%) using SVM method (Emanuele et al. 2021; Perego et al. 2009; Crippa et al., 2015; Cavallo et al., 2021). Cavallo et al. (2021) highlighted the different accuracy reached depending on the different methods used to arrange the dataset for classification.

However, being able to choose the appropriate classification method and the corresponding cross-validation model based on the type of data at hand becomes essential to advance research in this field, and reliable feature selection methods are also valid strategies to overcome the high dimensionality issue. (e.g., Alcañiz Raya et al., 2020a, 2020b).

Furthermore, models developed with small samples and high dimensional data are prone to noise. For example, Vabalas et al., (2019, 2020) and Li et al. (2017) examined finger movement during an imitation task and using a SVM algorithm, resulting in a classification accuracy of 76% and 87%, respectively. Importantly, the 87% of accuracy was reached when features were reduced from 40 to 9. These results pointed out the importance of finding the best fit between model, data dimensionality, and sample size.

In general, a larger sample should allow the model to be tested, and recruiting participants with various presentations (e.g., different degrees of severity, different cultural and biological settings) might be considered to ensure the generalizability of the ML model and therefore, improve the generalization of the results.

Table 3 Eye motion data results. The table presents the overall results of the studies related to eye motion analysis. Results are arranged based on the ML method used (SVM; ANN; others) and the type of features analyzed (AOI; scanpath)

Eye motion data	SVM		ANN		OTHERS	
Features	AOI	Scanpath	AOI	Scanpath	AOI	Scanpath
Num. of studies	3	1	0	3	1	3
Accuracy range	85%–92%	88%	-	67%–90%	78	70%–74%

N.B. For articles that reported results from various ML models, we have considered for this table only the model that achieved the highest accuracy level

Table 4 Body motion data results. The table presents the overall results of the studies related to body motion analysis. Results are arranged based on the ML method used (SVM; ANN; others) and the type of recording tool method (camera; motion sensor)

Body motion data	SVM		ANN		OTHERS	
Tools	Camera	Sensor	Camera	Sensor	Camera	Sensor
Num. of studies	5	2	2	1	0	4
Accuracy range	75%–100%	70%–73%	80%–93%	93%	-	67%–93%

N.B. For articles that reported results from various ML models, we have considered for this table only the model that achieved the highest accuracy level

Another important remark is that the high specificity of the samples could limit the generalizability of the classification. Most of the studies analyzed in this review chose to measure very specific and homogeneous samples, carefully selecting specific sets of inclusion and exclusion criteria to rule out potential confounders. However, models based on homogeneous samples are less likely to generalize to real-life clinical settings. As ASD is a heterogeneous condition, the possibility of training the models with larger groups, would be useful to improve the generalization to a wide condition range of the syndrome. Considering that autistic traits exist on a spectrum, moving the research in this field could lead to find measures that can identify autism regardless of its position on the spectrum. Specifically, although there may be variability within the spectrum itself, this variability should be a specificity compared with typically developing population, given that we expect there to be specific motor pattern typical for all the autistic spectrum. This hypothesis is based on specific theories and scientific research of recent years (Friston and Kiebel, 2009; De Jaegher, 2013; Torres et al., 2013) that focused on sensory processing models and general body-world interaction as the foundation for the entire autistic symptomatology and as an underlying explanation to support the idea that there may be a distinctiveness in autistic motor patterns that characterizes all individuals on the spectrum, differentiating them from the rest of the population.

Delving deeper into this topic is necessary to demonstrate this specificity analyzing these manifestations around the entire spectrum and also exploring the difference with other neurodevelopmental conditions and disorders. Also, including other neurodevelopmental disorders with movement impairments (e.g., the attention deficit hyperactivity disorder (ADHD), could be valuable for ML to determine if these groups present similar body movements that are distinct for the different neurodevelopment conditions. Nevertheless, in many cases the experimental groups were not matched on measures related to the intelligence quotient performance, cognitive abilities, and motor comorbidities (e.g., dyspraxia), limiting the confidence that the experimental differences observed in the experiment were due to a diagnosis of ASD, and not to differences in other cognitive or non-cognitive factors (e.g., Anzulewicz et al., 2016; Cavallo et al., 2021; Simeoli et al., 2021). Results from Robles et al. (2022) stressed this aspect. The authors compared the accuracy reached with matched vs. random dataset, revealing that matching the dataset according to IQ measure leads to a decrease in accuracy. This result confirms the need to control these variables in order to verify the classification validity.

Furthermore, for almost all studies the classification model did not give any indication about the severity of the symptoms except for the study conducted by Lin et al. (2022). In general, these studies show a lack in indicating whether the indices collected were associated with clinical correlates, severity, or prognosis (e.g., Tsuchiya et al., 2021). Future studies should include other neurodevelopmental disorders involving social and communication impairments, such as ADHD, in order to improve discrimination between groups compared with TD children and identify with greater precision the degrees of severity in each disorder, highlighting patterns, similarities, and differences.

Another aspects that seem to be related to the ASD abnormalities are sensory dysfunctions. Multiple studies investigated the relationship between ASD and electrodermal activity (EDA). The evidence from these studies is controversial: some research found no differences in EDA levels in response to sensory stimuli (e.g., McCormick et al., 2014), whereas others confirm these differences (e.g., Schoen et al., 2009). Raya et al., (2020a, 2020b) analyzed electrodermal activity (EDA) during different virtual experiences characterized by the presentation of different sensory stimuli (auditory, visual, olfactive) and analyzed these data using ML methods. Authors used SVM with a LOSO cross-validation procedure. Results revealed the potential of classifying ASD according to these parameters highlighting the need for future research in this field (Raya, et al., 2020).

Future ML studies are encouraged to generate as many features as possible to allow for specification of the globally optimal set of features for ASD identification (Zhao et al., 2021). These methods could be improved with more sophisticated models to determine which and how many (at least) features could be identified to be used in the clinical practice to identify children with ASD. For example, measure of the scanning eye movement patterns should be combined with other types of psychological and physiological measures (e.g., brain activities, skin conductance, speech, motions, body gestures, and facial expressions) to obtain a more comprehensive multimodal measure of the risk of ASD and aid the process of diagnosis and early detection.

Conclusion

Although, in recent years, research on ASD biomarkers has grown exponentially, and the integration of technological aids in assessment procedures is almost certain, it is worth noting that such biomarkers should still be deemed as supplementary data. However, the results of this review highlight that ML methods and technological tools can improve and refine the diagnostic process, providing additional information about such "invisible" traits of the disorder. Research in the field is still needed to investigate the correlations between certain biomarkers and specific symptoms of the disorder. Anyway, one of the greatest advantages that research in this field can provide is the potential of identifying additional biomarkers as potential predictors of ASD and

ensuring new prospective for early diagnosis. One revolutionary aspect of using these methods lies in the ability to move beyond mere categorical diagnosis and instead utilize them to gain further insights into the disorder. While current research using these methods may be focused on the accuracy of classifications based on an existing diagnosis, future studies could shift towards a more exploratory approach to the disorder. This approach could refine the assessment towards a phenotypic description of the disorder.

References

Alcañiz, M., Chicchi-Giglioli, I. A., Carrasco-Ribelles, L. A., Marín-Morales, J., Minissi, M. E., Teruel-García, G., Sirera, M., & Abad, L. (2022). Eye gaze as a biomarker in the recognition of autism spectrum disorder using virtual reality and machine learning: A proof of concept for diagnosis. *Autism Research: Official Journal of the International Society for Autism Research, 15*(1), 131–145. https://doi.org/10.1002/aur.2636

Ament, K., Mejia, A., Buhlman, R., Erklin, S., Caffo, B., Mostofsky, S., & Wodka, E. (2015). Evidence for specificity of motor impairments in catching and balance in children with autism. *Journal of Autism and Developmental Disorders, 45*(3), 742–751. https://doi.org/10.1007/s10803-014-2229-0

Anzulewicz, A., Sobota, K., & Delafield-Butt, J. T. (2016). Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism. *Science and Reports, 6*, 1–13. https://doi.org/10.1038/srep31107

Bäckström, A., Johansson, A. M., Rudolfsson, T., Rönnqvist, L., von Hofsten, C., Rosander, K., & Domellöf, E. (2021). Motor planning and movement execution during goal-directed sequential manual movements in 6-year-old children with autism spectrum disorder: A kinematic analysis. *Research in Developmental Disabilities, 115*, 104014. https://doi.org/10.1016/j.ridd.2021.104014

Brian, J., Bryson, S. E., Garon, N., Roberts, W., Smith, I. M., Szatmari, P., & Zwaigenbaum, L. (2008). Clinical assessment of autism in high-risk 18-month-olds. *Autism, 12*(5), 433–456. https://doi.org/10.1177/1362361308094500

Cavallo, A., Romeo, L., Ansuini, C., Battaglia, F., Nobili, L., Pontil, M., Panzeri, S., & Becchio, C. (2021). Identifying the signature of prospective motor control in children with autism. *Scientific Reports, 11*(1), 3165. https://doi.org/10.1038/s41598-021-82374-2

Cilia, F., Carette, R., Elbattah, M., Dequen, G., Guérin, J. L., Bosche, J., Vandromme, L., & Le Driant, B. (2021). Computer-aided screening of autism spectrum disorder: Eye-tracking study using data visualization and deep learning. *JMIR Human Factors, 8*(4), e27706. https://doi.org/10.2196/27706

Crippa, A., Forti, S., Perego, P., & Molteni, M. (2013). Eye-hand coordination in children with high functioning autism and Asperger's disorder using a gap-overlap paradigm. *Journal of Autism and Developmental Disorders, 43*(4), 841–850. https://doi.org/10.1007/s10803-012-1623-8

Crippa, A., Salvatore, C., Perego, P., Forti, S., Nobile, M., Molteni, M., & Castiglioni, I. (2015). Use of machine learning to identify children with autism and their motor abnormalities. *Journal of Autism and Developmental Disorders, 45*(7), 2146–2156. https://doi.org/10.1007/s10803-015-2379-8

Dowd, A. M., McGinley, J. L., Taffe, J. R., & Rinehart, N. J. (2012). Do planning and visual integration difficulties underpin motor dysfunction in autism? A kinematic study of young children with autism. *Journal of Autism and Developmental Disorders, 42*(8), 1539–1548. https://doi.org/10.1007/s10803-011-1385-8

Elbattah M, Carette R, Dequen G, Guérin L. & Cilia, F. (2019). Learning clusters in autism spectrum disorder: Image-based clustering of eye-tracking scanpaths with deep autoencoder. Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS 1417–1420. https://doi.org/10.1109/EMBC.2019.8856904

Emanuele, M., Nazzaro, G., Marini, M., Veronesi, C., Boni, S., Polletta, G., D'Ausilio, A., & Fadiga, L. (2021). Motor synergies: Evidence for a novel motor signature in autism spectrum disorder. *Cognition, 213*, 104652. https://doi.org/10.1016/j.cognition.2021.104652

Forti, S., Valli, A., Perego, P., Nobile, M., Crippa, A., & Molteni, M. (2011). Motor planning and control in autism A kinematic analysis of preschool children. *Research in Autism Spectrum Disorders, 5*(2), 834–842. https://doi.org/10.1016/j.rasd.2010.09.013

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: A synthesis and meta-analysis. *Journal of Autism and Developmental Disorders, 40*(10), 1227–1240. https://doi.org/10.1007/s10803-010-0981-3

Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., & Youngstrom, E. A. (2017). A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. *Journal of the American Academy of Child and Adolescent Psychiatry, 56*(7), 546–555. https://doi.org/10.1016/j.jaac.2017.05.005

Freitag, C. M., Kleser, C., Schneider, M., & von Gontard, A. (2007). Quantitative assessment of neuromotor function in adolescents with high functioning autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 37*(5), 948–959. https://doi.org/10.1007/s10803-006-0235-6

Gong, L., Liu, Y., Yi, L., Fang, J., Yang, U., & Wei, K. (2020). Abnormal gait patterns in autism spectrum disorder and their correlations with social impairments. *Autism Research, 13*, 1215–1226. https://doi.org/10.1002/aur.2302

Izawa, J., Pekny, S. E., Marko, M. K., Haswell, C. C., Shadmehr, R., & Mostofsky, S. H. (2012). Motor learning relies on integrated sensory inputs in ADHD, but over-selectively on proprioception in autism spectrum conditions. *Autism Research : Official Journal of the International Society for Autism Research, 5*(2), 124–136. https://doi.org/10.1002/aur.1222

Jiang, M., Francis, S.M., Srishyla, D., Conelea, C.A., Zhao, Q., & Jacob, S. (2019). Classifying individuals with ASD through facial emotion recognition and eye-tracking. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6063–6068

Kanhirakadavath, M. R., & Chandran, M. S. M. (2022). Investigation of eye-tracking scan path as a biomarker for autism screening using

machine learning algorithms. *Diagnostics (Basel, Switzerland), 12*(2), 518. https://doi.org/10.3390/diagnostics12020518

Koegel, L. K., Koegel, R. L., Ashbaugh, K., & Bradshaw, J. (2014). The importance of early identification and intervention for children with or at risk for autism spectrum disorders. *International Journal of Speech-Language Pathology, 16*(1), 50–56. https://doi.org/10.3109/17549507.2013.861511

Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M. (2021). Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Scientific Reports, 11*(1), 15069. https://doi.org/10.1038/s41598-021-94378-z

Lai, M. C., Lombardo, M. V., & Baron-Cohen, S. (2014). Autism. *Lancet (London, England), 383*(9920), 896–910. https://doi.org/10.1016/S0140-6736(13)61539-1

Li, B., Sharma, A., Meng, J., Purushwalkam, S., & Gowen, E. (2017). Applying machine learning to identify autistic adults using imitation: An exploratory study. *PLoS ONE, 12*(8), e0182652. https://doi.org/10.1371/journal.pone.0182652

Li, Y., Mache, M. A., & Todd, T. A. (2020). Automated identification of postural control for children with autism spectrum disorder using a machine learning approach. *Journal of Biomechanics, 113*, 110073. https://doi.org/10.1016/j.jbiomech.2020.110073

Liaqat, S., Wu, C., Duggirala, P. R., Cheung, S. S., Chuah, C. N., Ozonoff, S., & Young, G. (2021). Predicting ASD diagnosis in children with synthetic and image-based eye gaze data. *Signal Processing. Image Communication, 94*, 116198. https://doi.org/10.1016/j.image.2021.116198

Liberati, A., Fadda, R., Doneddu, G., Congiu, S., Javarone, M. A., Striano, T., & Chessa, A. (2017). A statistical physics perspective to understand social visual attention in autism spectrum disorder. *Perception, 46*(8), 889–913. https://doi.org/10.1177/0301006616685976

Lin, Y., Gu, Y., Xu, Y., Hou, S., Ding, R., & Ni, S. (2022). Autistic spectrum traits detection and early screening A machine learning based eye movement study. *Journal of child and adolescent psychiatric nursing official publication of the Association of Child and Adolescent Psychiatric Nurses Inc, 35*(1), 83–92. https://doi.org/10.1111/jcap.12346

Liu, W., Li, M., & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research : Official Journal of the International Society for Autism Research, 9*(8), 888–898. https://doi.org/10.1002/aur.1615

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). 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.

McCormick, C., Hessl, D., Macari, S. L., Ozonoff, S., Green, C., & Rogers, S. J. (2014). Electrodermal and behavioral responses of children with autism spectrum disorders to sensory and repetitive stimuli. *Autism Research : Official Journal of the International Society for Autism Research, 7*(4), 468–480. https://doi.org/10.1002/aur.1382

Milano, N., Simeoli, R., Rega, A., & Marocco, D. (2023). A deep learning latent variable model to identify children with autism through motor abnormalities. *Frontiers in Psychology, 14*, 1194760. https://doi.org/10.3389/fpsyg.2023.1194760

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Nobile, M., Perego, P., Piccinini, L., Mani, E., Rossi, A., Bellina, M., & Molteni, M. (2011). Further evidence of complex motor dysfunction in drug naïve children with autism using automatic motion

analysis of gait. *Autism, 15*(3), 263–283. https://doi.org/10.1177/1362361309356929

Padilla, P., Lopez, M., Gorriz, J. M., Ramirez, J., Salas-Gonzalez, D., & Alvarez, I. (2012). NMF–SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease. *IEEE Transactions on Medical Imaging, 31*(2), 207–216.

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

Pierce, K., Carter, C., Weinfeld, M., Desmond, J., Hazin, R., Bjork, R., & Gallagher, N. (2011). Detecting, studying, and treating autism early: The one-year well-baby check-up approach. *The Journal of Pediatrics, 159*(3), 458-465.e4656. https://doi.org/10.1016/j.jpeds.2011.02.036

Pontil, M. (2002). Leave-one-out error and stability of learning algorithms with applications. *International Journal of Systems Science*

Raya, M. A., Giglioli, I. A., Marín-Morales, J., Higuera-Trujillo, J. L., Olmos, E., Minissi, M. E., Teruel Garcia, G., Sirera, M., & Abad, L. (2020b). Application of supervised machine learning for behavioral biomarkers of autism spectrum disorder based on electrodermal activity and virtual reality. *Frontiers in Human Neuroscience, 14*, 90. https://doi.org/10.3389/fnhum.2020.00090

Raya, M. A., Marín-Morales, J., Minissi, M. E., Garcia, G. T., Abad, L., & Chicchi Giglioli, I. A. (2020a). Machine learning and virtual reality on body movements' behaviors to classify children with autism spectrum disorder. *Journal of Clinical Medicine, 9*(5) https://doi.org/10.3390/jcm9051260

Rinehart, N., & McGinley, J. (2010). Is motor dysfunction core to autism spectrum disorder? *Developmental Medicine and Child Neurology, 52*(8), 697. https://doi.org/10.1111/j.1469-8749.2010.03631.x

Robles, M., Namdarian, N., Otto, J., Wassiljew, E., Navab, N., Falter-Wagner, C., & Roth, D. (2022). A virtual reality based system for the screening and classification of autism. *IEEE Transactions on Visualization and Computer Graphics, 28*(5), 2168–2178. https://doi.org/10.1109/TVCG.2022.3150489

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science, 2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Schoen, S. A., Miller, L. J., Brett-Green, B. A., & Nielsen, D. M. (2009). Physiological and behavioral differences in sensory processing: A comparison of children with autism spectrum disorder and sensory modulation disorder. *Frontiers in Integrative Neuroscience, 3*, 29. https://doi.org/10.3389/neuro.07.029.2009

Simeoli, R., Milano, N., Rega, A., & Marocco, D. (2021). Using technology to identify children with autism through motor abnormalities. *Frontiers in Psychology, 12*, 635696. https://doi.org/10.3389/fpsyg.2021.635696

Teitelbaum, P., Teitelbaum, O., Nye, J., Fryman, J., & Maurer, R. G. (1998). Movement analysis in infancy may be useful for early diagnosis of autism. *Proceedings of the National Academy of Sciences of the United States of America, 95*(23), 13982–13987. https://doi.org/10.1073/pnas.95.23.13982

Torres, E. B., Brincker, M., Isenhower, R. W., III, Yanovich, P., Stigler, K. A., Nurnberger, J. I., Jr., et al. (2013). Autism: The micro-movement perspective. *Frontiers in Integrative Neuroscience, 7*, 32. https://doi.org/10.3389/fnint.2013.00032

Tsuchiya, K. J., Hakoshima, S., Hara, T., Ninomiya, M., Saito, M., Fujioka, T., Kosaka, H., Hirano, Y., Matsuo, M., Kikuchi, M., Maegaki, Y., Harada, T., Nishimura, T., & Katayama, T. (2021). Diagnosing autism spectrum disorder without expertise: A pilot

study of 5- to 17-Year-Old Individuals Using Gazefinder. *Frontiers in Neurology, 11*, 603085. https://doi.org/10.3389/fneur.2020.603085

Vabalas, A., & Freeth, M. (2016). Brief report: Patterns of eye movements in face to face conversation are associated with autistic traits: Evidence from a student sample. *Journal of Autism and Developmental Disorders, 46*(1), 305–314. https://doi.org/10.1007/s10803-015-2546-y

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Kinematic features of a simple and short movement task to predict autism diagnosis. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2019*, 1421–1424. https://doi.org/10.1109/EMBC.2019.8857307

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2020). Applying machine learning to kinematic and eye movement features of a movement imitation task to predict autism diagnosis. *Science and Reports, 10*, 1–13. https://doi.org/10.1038/s41598-020-65384-4

Volkmar, F. R., State, M., & Klin, A. (2009). Autism and autism spectrum disorders: Diagnostic issues for the coming decade. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 50*(1–2), 108–115. https://doi.org/10.1111/j.1469-7610.2008.02010.x

von Hofsten, C., & Rosander, K. (2012). Perception-action in children with ASD. *Frontiers in Integrative Neuroscience, 6*, 115. https://doi.org/10.3389/fnint.2012.00115

Wan, G., Kong, X., Sun, B., Yu, S., Tu, Y., Park, J., Lang, C., Koh, M., Wei, Z., Feng, Z., Lin, Y., & Kong, J. (2019). Applying eye tracking to identify autism spectrum disorder in children. *Journal of Autism and Developmental Disorders, 49*(1), 209–215. https://doi.org/10.1007/s10803-018-3690-y

Zhao, Z., Tang, H., Zhang, X., Qu, X., Hu, X., & Lu, J. (2021). Classification of children with autism and typical development using eye-tracking data from face-to-face conversations: Machine learning model development and performance evaluation. *Journal of Medical Internet Research, 23*(8), e29328. https://doi.org/10.2196/29328

Zhao, Z., Zhu, Z., Zhang, X., Tang, H., Xing, J., Hu, X., Lu, J., & Qu, X. (2022). Identifying autism with head movement features by implementing machine learning algorithms. *Journal of Autism and Developmental Disorders, 52*(7), 3038–3049. https://doi.org/10.1007/s10803-021-05179-2