

*Annual Review of Biomedical Data Science*

# A Review of and Roadmap for Data Science and Machine Learning for the Neuropsychiatric Phenotype of Autism

Peter Washington<sup>1</sup> and Dennis P. Wall<sup>2</sup>

<sup>1</sup>Department of Information and Computer Sciences, University of Hawai'i at Mānoa, Honolulu, Hawai'i, USA

<sup>2</sup>Departments of Pediatrics (Systems Medicine), Biomedical Data Science, and Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California, USA; email: dpwall@stanford.edu

Annu. Rev. Biomed. Data Sci. 2023. 6:211–28

First published as a Review in Advance on  
May 3, 2023

The *Annual Review of Biomedical Data Science* is  
online at [biodatasci.annualreviews.org](http://biodatasci.annualreviews.org)

<https://doi.org/10.1146/annurev-biodatasci-020722-125454>

Copyright © 2023 by the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information.

ANNUAL  
REVIEWS **CONNECT**

[www.annualreviews.org](http://www.annualreviews.org)

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

## Keywords

autism, digital phenotyping, machine learning, crowdsourcing, digital health, digital psychiatry

## Abstract

Autism spectrum disorder (autism) is a neurodevelopmental delay that affects at least 1 in 44 children. Like many neurological disorder phenotypes, the diagnostic features are observable, can be tracked over time, and can be managed or even eliminated through proper therapy and treatments. However, there are major bottlenecks in the diagnostic, therapeutic, and longitudinal tracking pipelines for autism and related neurodevelopmental delays, creating an opportunity for novel data science solutions to augment and transform existing workflows and provide increased access to services for affected families. Several efforts previously conducted by a multitude of research labs have spawned great progress toward improved digital diagnostics and digital therapies for children with autism. We review the literature on digital health methods for autism behavior quantification and beneficial therapies using data science. We describe both case-control studies and classification systems for digital phenotyping. We then discuss digital diagnostics and therapeutics that integrate machine learning models of autism-related behaviors, including the factors that must be addressed for

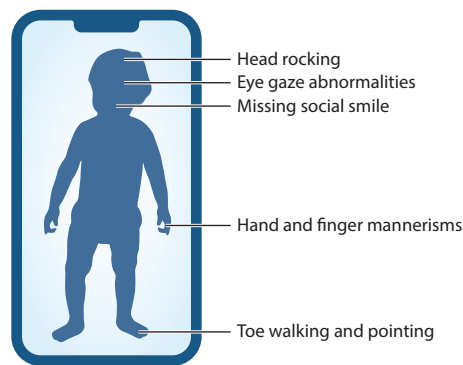
translational use. Finally, we describe ongoing challenges and potential opportunities for the field of autism data science. Given the heterogeneous nature of autism and the complexities of the relevant behaviors, this review contains insights that are relevant to neurological behavior analysis and digital psychiatry more broadly.

## INTRODUCTION

Autism spectrum disorder (autism) is a complex neuropsychiatric condition that manifests in a variety of phenotypic presentations. These include limited ranges of interest, social deficits, delays in communication, an inability to express and or recognize ranges of emotion, an avoidance of eye contact, and idiosyncratic motions of the hands, head, and body (**Figure 1**).

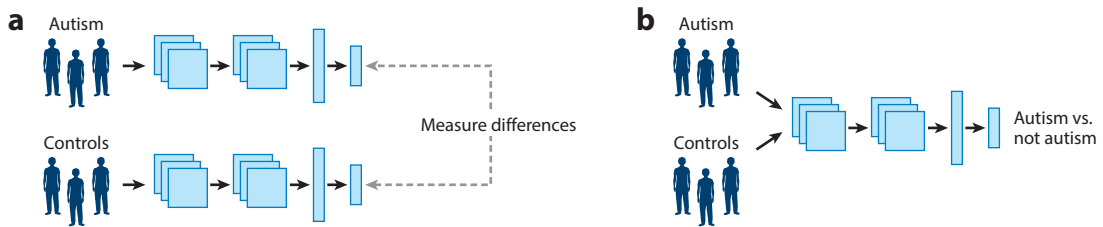
It is estimated that 1 in 44 children in the United States has autism (1), with the prevalence continuing to rise. Despite the high incidence rate, it is estimated that 83.86% of counties in the United States lack access to diagnostic and therapeutic resources (2). This dearth of resources leads to diagnoses later in life, often resulting in poorer psychiatric outcomes (3, 4).

Artificial intelligence (AI), and particularly machine learning (ML), has the potential to serve as the great equalizer for many behavioral healthcare concerns like autism. According to the Pew Research Center, 97% of adults in the United States own a cellular device and 85% own a smartphone, with the percentage increasing each year since 2011 (5). In emerging economies such as Mexico, Venezuela, Colombia, South Africa, Kenya, India, Vietnam, the Philippines, Tunisia, Jordan, and Lebanon, most adults have access to a mobile phone of some kind, with 53% having access to a phone that can connect to the Internet and run apps (a smartphone) (6). As these percentages continue to rise and Internet-powered devices become ubiquitous, access to digital services can become democratized on a global scale. While autism services are currently restricted to relatively privileged populations, digital solutions powered by emerging data science methodologies can make access to autism therapy universal.



**Figure 1**

Some of the observable features useful for identifying autism in a developing child at 2 years of age. These features can be captured through traditional clinical techniques involving structured electronic medical records, as well as alternative methods such as games on smartphones. The more such data can be passively collected (1) via ubiquitous devices, the higher the chances of capturing natural signals that have higher diagnostic precision. In contrast, in signals captured in the more artificial clinical settings, transiency and other factors can cause misleading readings or inaccurate measurements.



**Figure 2**

The two major classes of studies we discuss in this review. (a) Case-control studies (science) compare the outputs of a static model or computational analysis method across two or more groups, usually autism versus not autism. (b) Machine learning studies (engineering) have the goal of training a model to classify an individual or group as autism versus not autism.

This review focuses on digital data modalities that can help provide a quantitative understanding of the neuropsychiatric phenotype of autism. The field of autism behavioral data science is vast and broad, as autism is characterized by a variety of behavioral manifestations. The heterogeneous nature of autism presentation requires the quantification of autism's behavioral manifestations across several social and communicative dimensions, both to better understand the autism phenotype as it relates to differences in diagnostic cohorts (**Figure 2a**) and to engineer automated methods for autism classification (**Figure 2b**).

The field of autism data science is so vast that it is impossible for us to exhaustively cover the field within the scope of a single review. To maintain focus, we do not cover details of the models and ML techniques powering the studies we describe, although we briefly mention them. The state of the art of ML and data science is constantly evolving, and the details of the exact models used are not as relevant as the content of the data. We make an exception in cases where the model design is crucial to understanding the contribution of the work.

We first describe multimedia data modalities that are commonly used in digital phenotyping projects related to social human behavior, including common schemes for acquiring naturalistic data. We then discuss case-control studies for identifying statistically significant differences between cohorts with and without autism. We next review efforts to automatically classify autism using ML. Finally, we highlight ongoing challenges for the field, corresponding areas of opportunity for innovators to make a significant impact, and low-hanging fruit based on successful studies and methods applied to other behaviors and neuropsychiatric presentations.

## DATA MODALITIES AND ACQUISITION

There is a multitude of data types that encode behavior-rich data for digital phenotyping (**Table 1**). Data whose acquisition requires the involvement of humans, such as questionnaires about a child's behavior, are the least prone to computational errors and most comparable to the formal clinical diagnostic process. On the other hand, multimedia data streams obtained by fully automated procedures can capture subtle yet complex behavioral indications that would be missed by a human using predetermined prompts. Such multimedia data streams, however, are more difficult to successfully convert into a predictive embedding. Here, we discuss both categories of behavioral data modalities used in the field of autism ML.

One of the most common data sources for clinical phenotyping consists of multiple-choice questionnaires either containing or derived from clinical scoresheets for diagnosing or quantifying the magnitude of the child's presentation of autism. The most popular diagnostic assessments for autism include ADOS (Autism Diagnostic Observation Schedule) (7), ADI-R (Autism Diagnostic Interview-Revised) (8), CARS (Childhood Autism Rating Scale) (9), and GARS (Gilliam

**Table 1 Common data modalities used in both case-control and machine learning autism phenotyping studies<sup>a</sup>**

Modality	Data capture techniques	Feature extraction techniques (non-deep learning)
Questionnaire/instrument	Parent-filled questionnaire (on web or smartphone), clinician-filled questionnaire	Conversion to Likert scale
Eye gaze	Eye tracker, webcam, video camera	Gaze heat maps, gaze trajectories
Emotion evocation	Webcam, video camera	Histogram of oriented gradients, facial key points
Restrictive and repetitive motions	Webcam, video camera, motion sensor, depth sensor	Skeletal pose, optical flow, activity recognition
Idiosyncratic speech	Audio recorder	Many audio-specific features such as pitch, mel-frequency cepstral coefficients, temporal shape, temporal features, and spectral shape

<sup>a</sup>With the continually expanding popularity of deep learning models in behavioral data science (and the entirety of data science more broadly), many features are now automatically learned via neural networks rather than explicitly prespecified.

Autism Rating Scale) (10). Other scoresheets are used for screening or quantifying autism-related behaviors. Readily available datasets with such scoresheets filled by clinicians or parents include Autism Genetic Resource Exchange (11), Autism Consortium, National Database for Autism Research (12), Simons Simplex Collection (13), Simons Variation in Individuals Project (14), Autism Speaks, and the Autism Genome Project (15). Data-driven studies either directly analyze collections of these scoresheets or develop independent questionnaires of a similar nature to these studies. The primary benefits of this data modality are its constrained search space relative to the high-dimensional multimedia data and the high degree of reliability of the data due to its requirement of incorporating humans in the feature extraction process.

Rich multimedia streams are increasingly used in data science-based phenotyping efforts. Popular data streams enriched for social human behavior relevant to autism include single images, videos, eye tracking data, and movement records (16), as well as text (17) and audio recordings (18). The principal advantage of these data types is the rich and nuanced behavioral information that they encode. The challenge with such large and heterogeneous data, like in many other applications of data science, is finding the signal in the noise.

Most autism data science studies that collect phenotype-rich data utilize controlled lab settings for acquisition. This practice enables normalized data that can be easily compared across subjects. We cover such studies in this review. Recent innovations, however, center around collecting these data in naturalistic settings, and there is extensive room for novelty around the acquisition of homogeneous and structured data streams. Stanford University's *Guess What?* project, for example, provides a framework for capturing videos of children with autism enriched for social behaviors through a charades-style game (19). In this game, a parent or caregiver places a smartphone on their forehead while the child acts out charades-style prompts. The parent tilts the phone forward when the prompt is correctly guessed and backward otherwise (20). The game prompts, in conjunction with accelerometer and gyroscope logs of the phone's tilting motion, provide structured time-stamped metadata that can serve as automatically derived labels of emotions, body movements, and other behaviors (21). Videos collected from *Guess What?* sessions with emotion game prompts were demonstrated to be enriched with emotional facial expressions (22). The resulting frames enabled the construction of models that outperformed general-purpose facial emotion recognition classifiers (23–26). Beyond its value for digital phenotyping, *Guess What?* has been demonstrated to result in improvements in endpoints such as the SRS-2 (Social Responsiveness Score-2) and VABS-II (Vineland Adaptive Behavior Scales-II) among children with autism who have participated in the game at home with their families (27).

Phenotype-rich data can also be captured from digital therapies that were not originally developed for data acquisition. Socially assistive robots have been created to provide therapy through social play and to take the lead in social interactions with the child, thereby evoking social behavior by the child (28). Structured interactions with robots coupled with measurements from passive sensors can provide objective measurements that can be used for diagnoses (29). Wearable devices have also been explored as therapeutic tools. The Superpower Glass project is an educational tool deployed on a wearable headset to provide real-time facial emotion feedback to children with autism (30–32). This device provided pleasant experiences to children and their parents (33, 34), improved child performance on the SRS-2 (35, 36), and ultimately resulted in significant improvements on the VABS-II socialization subscale in a randomized controlled trial (37). While video and eye gaze data collected by the Superpower Glass system were evaluated for discriminative power in an autism-versus-neurotypical classification task, the classifier was unable to outperform classifiers that only used age and gender features (38), likely due to the nonstandardization of the interaction between users. Extra care must therefore be taken to ensure that data acquired from interactive systems are calibrated and standardized across users.

As we see in the following sections, dataset size and diversity are some of the greatest limitations of the current state of autism data science. Innovation in the structured data collection, cleaning, and labeling processes is likely to enable well-powered studies and robust diagnostic models.

## CASE-CONTROL STUDIES

In this section, we describe scientific studies that use data science methods to quantify differences between cohorts of individuals with autism and matched neurotypical control cohorts. Such experiments aim to gather equivalent data from multiple groups and determine whether there are statistically significant differences.

Gaze features are frequently used for case-control studies since avoidance of eye contact is a well-documented autism behavior. Several studies use an eye tracker to measure a child's attention to structured socially enriched video scenes. Jones & Klin (39) compared fixation patterns over several months for 59 children at high risk for autism (defined as full siblings of a child with autism) against 51 typically developing children, finding that children later diagnosed with autism exhibit a decline in fixation time from 2 to 6 months of age. Riby & Hancock (40) compared 26 children with autism and 18 with Williams syndrome, a rare developmental disorder characterized by intellectual disabilities, distinctive facial features, and cardiovascular issues. The children in the study watched cartoon scenes while their gaze was measured. The authors found that children with autism watched faces for less time than is typical, whereas children with Williams syndrome attended to the face for a longer duration than is typical. Chawarska et al. (41) compared the visual attention to a social scene of 67 children at high risk for autism (full sibling of a child with autism) against that of 50 children at low risk using eye tracking, finding that infants later diagnosed for autism viewed the scene for a shorter duration and spent less time looking at the actress in the scene and her face. Campbell et al. (42) compared the visual attention to a scene of 22 toddlers with autism against that of 82 neurotypical controls, finding that 8% of participants with autism oriented in response to name calling compared with 63% of neurotypical participants. Sadria et al. (43) applied network analysis to eye tracking data by considering each area of interest as a node and each saccadic transition between two areas defining an edge. They found that there were statistically significant differences between children with autism and neurotypical controls in the degree centrality for areas pertaining to the mouth and right eye. Breaking from the more standard practice of using highly structured interactions and in-lab settings for data collection, Varma et al. (44) measured eye gaze differences between cohorts using network

analysis from crowdsourced data collected during use of a mobile autism therapeutic, finding a statistically significant difference between the groups for a single area of interest. Alvari et al. (45) analyzed eye contact during unconstrained therapist–child interactions by applying unsupervised clustering on data from 62 children with autism, identifying three distinct subgroups defined by eye contact dynamics. In one of the largest autism-related eye tracking studies containing 563 subjects with autism and 1,300 other subjects, children with autism exhibited a higher percent fixation to dynamic geometric images compared with other children (46). In an eye tracking study of 21 children with autism and 31 neurotypical children playing the Go/No-Go game, Putra et al. (47) found that children with autism had more unstable gaze modulation.

Difficulty understanding and evoking emotion is another core difficulty in children with autism. Affective data can therefore be useful in distinguishing autistic children from neurotypical controls. Guha et al. (48) compared the facial dynamics of 20 participants with high-functioning autism against 19 participants with neurotypical development, finding reduced complexity in the dynamics of the eye region in the autism group. In an uncharacteristically large-scale study for the field, Egger et al. (49) compared emotional features in videos collected from a ResearchKit-based iPhone app from 1,756 families, finding that children at high risk of autism evoked statistically significantly more neutral emotions and less positive emotions compared with children at low risk of autism.

Restrictive and repetitive head and body movements are another core trait defining autism, motivating work in analyzing quantitative differences in head movement, body movement, and other motor features between cases and controls. Martin et al. (50) compared the yaw, pitch, and roll of the head of children with and without autism (21 children in each group), finding that children with autism exhibited faster head turning and inclination compared to neurotypical controls. Dawson et al. (51) conducted computer vision analysis of head postural control on 104 toddlers, 22 of whom had an autism diagnosis, finding that the rate of head movement for children with autism was 2.22 times greater than for neurotypical toddlers.

Another diagnostic criterion for autism is the use of idiosyncratic speech, giving rise to projects that analyze audio and text features. In a study of 15 children with autism and 30 typically developing children, Hudenko et al. (52) found that children with autism produced no unvoiced laughter, which largely consists of atonal laughs without evident periodicity that are typically used to provide social affirmation rather than convey positive affect. In contrast, typically developing children produced laughter that was 37–48% unvoiced. In a study of 12 controls and 2 infants at high risk for autism (siblings of children already diagnosed with autism), Orlandi et al. (53) found differences in the mean fundamental frequency in the crying audio of high-risk infants and controls at 10 days, 6 weeks, and 12 weeks.

Advances in human behavior analysis techniques are sure to bring about analyses that will elucidate new differences between autistic individuals and neurotypical controls. As these methods continue to advance, incorporation with genomic and neural data will allow for an increasingly precise understanding of how small biological changes can affect behavior, helping to reveal which aspects of autism are preset biologically and which are affected by social and environmental factors. One of the largest opportunities for future case–control studies is a dramatic increase in scale. Given the diversity of humans, any study with tens of subjects in each cohort will contain a multitude of biases.

## **CLASSIFICATION OF AUTISM AND RELATED BEHAVIORS**

Supervised ML is one of the most widely applied data science techniques in most data-driven fields (54). Naturally, the framework of training a model to make predictions from data can enable

the creation of automated diagnostics for autism, which can easily be formulated as a supervised learning problem where data from one of the modalities of autism we discussed above are used as the input to the model and a diagnosis is emitted as the model's output. In this section, we discuss research efforts following this general framework for autism and the behaviors it comprises.

While automatic extraction of behaviors related to autism is feasible with current technologies (e.g., eye tracking for attention and facial emotion classification for affect), there are many behaviors that diagnosticians use to classify autism that are beyond the scope of current technologies. For example, “indicates pleasure to others,” “shares excitement,” and “social participation” are all diagnostic criteria that are top predictors for autism, according to feature selection algorithms applied to clinical scoresheets (55). For such behaviors, ML and other automated approaches are not yet feasible. Recently, ML-powered human-in-the-loop approaches to autism diagnosis have emerged. Such approaches use either untrained humans such as crowd workers (56–58) as human feature extractors of audiovisual data or direct in-person child observations such as parent-filled questionnaires in conjunction with clinical experts and trained video analysts (59–63), as is the case with the pediatric digital health company Cognoa. After the human-extracted features are obtained, they are provided as input into an ML model, as in a study by Tariq et al. (64), who achieved an 92% AUC (area under the curve) score using 116 videos of children with autism and 46 videos of typically developing children. This pipeline was replicated in 50 neurotypical children and 50 children with autism from Bangladesh, achieving an AUC of 85% (65) and demonstrating that this method has the potential to generalize to diverse global populations.

An important design consideration when constructing diagnostic models is the identification of small feature subsets to minimize the time to diagnosis. Wall et al. (55) pioneered this application of feature selection methods to diagnostic scoresheets of autism. Multiple studies have subsequently validated the use of ML feature reduction methods to algorithmically identify small subsets of 4–10 questions of diagnostic and screening instruments that alone can be used to predict autism (66–68).

Another important consideration for translation of these methods into real-world diagnostics is child privacy. Washington et al. (69) applied privacy-preserving transformations to videos of children with autism, including obfuscating the child's face with a box using frame-by-frame face detection, pitch-shifting the audio, and applying global image transformations such as Gaussian blurring, pixelation, and optical flow (70). Each of these mechanisms reduced performance of the human-in-the-loop models by less than 7% on metrics such as AUROC (area under the receiver operating curve) and AUPRC (area under the precision recall curve). There is ample room for future work to explore higher-precision privacy mechanisms.

These human-in-the-loop diagnostic techniques have yet to be applied to behavioral conditions beyond autism, although we suspect that these methods could aid diagnostic practices for heterogeneous behavioral conditions more broadly, including attention-deficit/hyperactivity disorder (ADHD), anxiety, depression, and other developmental delays. We next discuss classification using such rich multimedia data streams. Although fully automated ML methods (as of 2023) cannot yet support the complexity of several social behaviors that encompass autism (e.g., “Does the child enjoy participating in social games and interactions?”), applying ML approaches to raw data streams for autism diagnostics has the benefit of picking up subtle cues and representations with speed and objectivity. Below we discuss these approaches, which leverage many of the same modalities as the case-control studies.

Just as for case-control studies, gaze patterns are a common feature for diagnosing autism. Liu et al. (71) used face scanning patterns to predict autism with an accuracy of 88.5% on a dataset of 29 children with autism, 29 neurotypical children matched by age, and 29 neurotypical children matched by IQ. Duan et al. (72) created a dataset of fixation maps and scanpaths of 14 children with

autism and 14 controls; such datasets can help develop predictive models for autism. Chang et al. (73) used a decision tree classifier to predict autism from a sample of 40 children with autism and 936 typically developing controls watching highly structured and strategically designed movies displayed on an iPhone or iPad with an AUC of 0.9. Oliveira et al. (74) used gaze fixation maps from an eye tracker to predict autism on a dataset of 76 subjects with autism and 30 typically developing subjects, reaching 90% precision, 69% recall, and 93% specificity.

Affective computing is not as popular in automated diagnostic approaches as it is for case-control studies, although preliminary studies exist. Drimalla et al. (75) developed a tool to facilitate a simulated social interaction with a child, recording the subject's eye gaze, voice, and facial expressions during the interaction. Using both facial expressions and vocal analysis collected from the simulated social interaction in a study with 37 adults with autism and 43 neurotypical controls, Drimalla et al. (76) detected autism with an accuracy of 73%, specificity of 79%, and sensitivity of 67%.

Pose and movement features have been quite successful in distinguishing autism cases from controls. Khosla et al. (77) used the arrangement of a child's eyes, nose, and lips in a front-facing image of the face to classify autism with 87% accuracy. Hashemi et al. (78) tracked the head motion of children with autism using bounding boxes over the child's eyes, ears, and nose. Lidstone et al. (79) used 3D depth cameras from Microsoft's motion-sensing input device Kinect to distinguish children with autism from neurotypical controls, reaching an AUROC of 0.94 on a sample of 23 children with autism and 17 neurotypical controls. Kojovic et al. (80) achieved an  $F_1$  score of 0.89 in a binary autism diagnosis task on 68 children with autism and 68 neurotypical controls by using a convolutional neural network (CNN) to extract visual features from pose images and feeding these into a long short-term memory network. Cook et al. (81) presented a new dataset of video clips consisting of examples of 35 atypical motor movements associated with autism and 33 typical movements. They used this dataset to train a decision tree with skeletal key points extracted from OpenPose and temporal velocities encoding body movements over time, reaching an average  $F_1$  score across folds of 0.71. Li et al. (82) classified autism using kinematic features extracted from hand movement imitations by 16 children with autism and matched controls, reaching an overall accuracy of 70.5%. Anzulewicz et al. (83) measured tablet touch and gesture kinematics on 37 children with autism and 45 typically developing children, using these features to measure autism with 93% accuracy. Cavallo et al. (84) classified using hand movement data of 20 children with autism and 20 neurotypical children performing a structured grasping task, reaching 83% accuracy, 80% sensitivity, and 85% specificity.

Audio and text features are increasingly being used for ML diagnostics of autism. Li et al. (85) built a classifier to predict atypical prosody and stereotyped idiosyncratic speech associated with autism with accuracies of 88.1% and 77.8%, respectively, on a dataset of 118 children who were administered the ADOS Module 2 and 71 children who were administered the ADOS Module 3. Chi et al. (86) used crowdsourced videos from a mobile autism therapeutic game given to 20 children with autism and 38 neurotypical children to classify autism with 79% balanced accuracy using a CNN predicting from spectrogram images. Lau et al. (87) used acoustic features pertaining to rhythmic and tonal aspects of prosody to classify autism, which achieved an AUC of 0.90 using rhythm-relevant features and an AUC of 0.695 using intonation-relevant features on a dataset of 55 English-speaking people with autism and 39 English-speaking controls, as well as 28 Cantonese-speaking people with autism and 24 Cantonese-speaking controls. Maenner et al. (88) constructed a random forest classifier of autism using words and phrases from developmental evaluations, achieving a sensitivity of 84%, a positive predictive value (PPV) of 89.4%, and an AUROC of 0.932 on a dataset of 1,450 children, of whom 754 met the criteria for an autism diagnosis.



Data modalities have recently been combined to predict autism. Vabalas et al. (89) combined both eye and motion data to predict a diagnosis from 22 children with autism and 22 neurotypical controls, reaching 78% accuracy with both modalities, 73% for only motion features, and 70% for only eye features. Javed & Park (90) used both facial expressions and upper body movements recorded during child–robot interactions to predict autism among five children with autism and seven typically developing children, with accuracy, precision, and recall in the high eighties.

Automated diagnostics for autism are sure to improve as ML methods continue to rapidly advance and as data collection systems become more ubiquitous. The classifiers will likely need continuous recalibration, as the symptom presentation varies and as the definition of autism continues to evolve with subsequent iterations of the DSM (*Diagnostic and Statistical Manual of Mental Disorders*). We anticipate a future in which AI-enabled tools enable global reach and inclusion while maintaining the highest levels of clinical rigor and reliability, not just for diagnostics but also for precise and dynamic tracking of personalized treatments that are guided by the quantified predictions such tools make.

## TRANSLATION OF DIAGNOSTIC MODELS

The vibrant research using AI for automatic evaluation and measurement of observable phenotypes of autism (and other neurodevelopmental phenotypes) has promise to translate into practical clinical and educational solutions that facilitate autism’s diagnosis, tracking, and treatment. This requires careful considerations of design, feasibility testing, and tests for clinical effectiveness using useful endpoints. In this section, we focus on only the diagnostic and therapeutic tools that have passed through regulatory review and have received market clearance by the US Food and Drug Administration (FDA).

Diagnostic tools for autism have started to enter the market. Cognoa (59–63) has the first FDA-approved digital diagnostic, called Canvas Dx, that uses AI for clinical decision support in the identification of autism in children of 18–72 months of age. This diagnostic device is indicated for remote use on smartphones as well as for use in both primary settings and more specialty settings. Another diagnostic tool, Earlitec (39, 91–97), incorporates the single modality of eye tracking for autism diagnosis support in children of 16–30 months old and is specifically approved for use in specialty clinical settings.

Several researchers have noted the importance of managing uncertainty with AI in medicine (98–100). Due to the importance of precision measurements when making AI-guided diagnoses on developing children, where ephemeral signals are common and expected, managing uncertainty is particularly valuable. Cognoa’s Canvas Dx (59–63) is one such tool that addresses this issue. Canvas Dx emits a probability score of autism but makes a positive call of autism or a negative call of no autism only if the emitted probability exceeds an upper or a lower threshold, respectively. Otherwise, the classifier abstains from classification, indicating that the model is not confident enough to make a prediction at that time. The ability to abstain allows the model to avoid making diagnostic calls when presented with insufficient information to make a call in either direction. This is an important safety measure, but one that can be tuned and optimized with new information, as under, for example, the FDA’s recent guidance for the incorporation of a predetermined change control plan, which has been granted with the authorization of de novo and derivative devices including Canvas Dx, Medtronic, Caption Guidance, and Apple’s atrial fibrillation detection system. We anticipate that all medical devices using AI for diagnostics will be required by the FDA to have an abstention feature and that many will include algorithm modification plans that enable careful tuning with good ML practices given real-world use data.

Evaluating these diagnostic models that can abstain from making a prediction requires careful consideration of evaluation metrics. We argue that PPV and negative predictive value (NPV)

are the optimal evaluation metrics for models that make abstentions, as they consider only the positive and negative calls that are made, rather than metrics like sensitivity and specificity, which consider all positives and negatives regardless of whether the classifier made a call on those cases. We therefore recommend using PPV and NPV as the primary clinical endpoints in diagnostic classifiers.

Another crucial endpoint for diagnostic systems is model fairness. A common issue with ML systems for healthcare is that they have differing performance levels across demographic groups (101, 102). There are several quantitative fairness metrics that have been proposed in recent ML literature, including demographic parity, equalized odds, equal opportunity, test fairness, and counterfactual fairness, among many others (102). While no fairness metrics have been standardized yet in the field, we believe that in the future it will be important to report metrics like demographic parity along with standard performance metrics. Large, crowdsourced data collection efforts such as SPARK (Simons Foundation Powering Autism Research for Knowledge) (103) and remotely and freely accessible smartphone applications (19–22) are important steps toward achieving fair and unbiased data collection.

## **TRANSLATION OF MODELS FOR DIGITAL THERAPEUTICS**

There is tremendous promise for digital solutions that use ML to improve the delivery and quality of treatments for autism and other aspects of behavioral health. Because AI-powered digital therapy is a blossoming field of study, there are uncertainties about the reactions of patient populations to digital therapies' trustworthiness, usability, and understandability, among several other dimensions.

We describe two AI-enabled therapies for pediatric autism that have followed a path to translation outside the research lab. The Superpower Glass system is one such therapy, which is delivered on a pair of Google Glass smart glasses worn by a child with autism. In therapy using Superpower Glass, the parent interacts with the child wearing the Glass system. The Glass provides real-time emotion cues to the child via a computer vision facial emotion detection system. Superpower Glass was developed and refined through an iterative design process. In early prototyping sessions, possible user interface designs were explored through storyboarding and low-fidelity prototyping through codesign sessions involving children with autism and their parents (31–34). These changes were made prior to initial feasibility testing (35, 36), which led to further refinements until the system was formally tested in a randomized controlled clinical trial (37). This process is crucial for novel therapeutic paradigms in digital health given the uncertainty and novelty of AI-powered solutions.

Another therapy, EndeavorRx by Akili, is the first game-based digital therapy and one of the first digital therapies to pass through regulatory review (104). This system has a sophisticated design with underlying software models for independent game play by a child with ADHD, a condition that has numerous symptomatic overlaps with and is often comorbid with autism. Both EndeavorRx and Superpower Glass have translated to market and have received Breakthrough Therapy status by the FDA.

Digital applications for psychiatry and behavioral sciences are traditionally developed as one-size-fits-all solutions. However, the integration of the ML models discussed in this review can enable adaptive digital tools that analyze the user's behavior and recalibrate accordingly. Measures of autism symptoms can be embedded within gamified therapeutics so that the AI can understand the child's response to a particular "dose" of the therapy in real time. With such an understanding of the child's dose–response curve, digital solutions can then increase or decrease the dose of the therapy.

## CHALLENGES AND OPPORTUNITIES

The field of autism behavioral data science is a special case of computational behavioral phenotyping. As such, the field is likely to continue improving as computational behavior analysis tools become more sophisticated, granular, and robust. Similarly, autism data science can benefit from advancements in behavioral data science for other psychiatric conditions. In this section, we briefly discuss challenges and opportunities in the field of digital behavioral phenotyping for neuropsychiatry. Many of the opportunities we discuss involve reapplying data science innovations used in other conditions to autism, and many of the challenges for autism data science also arise in data science used for complex human behavior analysis more broadly.

One of the most important difficulties in the field of autism data science is the lack of large, standardized datasets reported in studies. Classification performances reported by ML diagnostic systems vary drastically across papers, but it is unclear whether these performance differences are the result of datasets with intrinsically varying differences. Therefore, there is a need for standardized datasets. Similarly, the sample sizes in both case-control and ML analyses are usually small, as there tend to be only tens of children represented in each group. The reported numbers from these studies limit the ability to compare across studies even if the same study conditions are used. There is a strong need for larger and more diverse studies, as these will allow findings to be generalized. Crowdsourced domain-specific data collection platforms (19–22) hold great promise for enabling these sample sizes. However, unlike controlled lab settings, such systems face issues of data consistency, with potential for overfitting. An alternative approach that bypasses this issue is model personalization: The model is optimized for each subject's unique features and the environment in which the model will be deployed. There are some preliminary efforts published in the autism data science field (105) that demonstrate improved performance when personalizing ML classifiers to the subject. These approaches require increased interaction between the child and the subject, creating opportunity for novel human-computer interactions.

Self-supervised learning (SSL) is the automatic learning of feature representations from unlabeled data. SSL can help accelerate model personalization when applied to a single user's data streams. Yu & Sano (106) applied SSL to wearable sensor data for stress detection, improving performance compared with supervised baselines by between 7.7% and 13.8% on three independent datasets.

Distinguishing autism from related conditions is an emerging challenge for the field. An increasing body of literature from autism genetics, neuroscience, and psychiatry suggests that autism consists of not a single spectrum but instead independent subtypes. There have been some efforts to quantitatively subtype autism from behavioral data modalities, and there is an opportunity for such efforts to validate and complement efforts in genetics, neuroscience, and psychiatry. For example, Gardner-Hoag et al. (107) analyzed 854 children with autism and identified seven distinct clusters of autism behaviors, with each cluster representing a single dominant prototypical behavioral characteristic.

Most of the analyses reported here distinguish between having autism and not having autism or being neurotypical, as few efforts to date have attempted to distinguish between a variety of psychiatric conditions with overlapping symptoms. Binary classifiers are useful if a condition is already suspected, but the ideal scenario will be to distinguish autism from conditions such as ADHD, schizophrenia, anxiety, depression, and speech delays. There has been some preliminary work in this vein. Duda et al. (108) built a classifier distinguishing ADHD ( $N = 174$ ) from autism ( $N = 248$ ) using a subset of 15 questions from the SRS, reaching an AUROC of 0.89. Wawer et al. (109) used text utterances to classify both autism ( $N = 37$ ) and schizophrenia ( $N = 37$ ) separately, finding that pretraining a model for schizophrenia results in improved performance when using transfer learning to predict autism. Demetriou et al. (110) used a battery of tests created for various

psychiatric conditions to classify autism ( $N = 62$ ), early psychosis ( $N = 48$ ), and social anxiety disorder ( $N = 83$ ), suggesting that combining inputs for different overlapping conditions can be a useful approach. Iakovidou et al. (111) used accelerometer, blood volume pulse, and electrodermal activity data from wearable sensors to differentiate autism from Rett syndrome with 95% accuracy in a sample containing 10 children with autism and 10 with Rett syndrome. Since neuropsychiatric conditions are often comorbid, future efforts will likely need to be formulated as a multiclass classification problem or a similar variant.

In addition to case-cohort analyses and ML diagnostic systems, data science methods have the potential to provide automated longitudinal outcome tracking and treatment response measurements. Some preliminary work in this area has been conducted, but there are ample opportunities for more work in this field, which would benefit digital interventions and treatments for autism symptoms. McKernan et al. (112) measured the amount of time for a child to respond to a social partner using automated speech processing methods, finding that children receiving interventions demonstrated a larger reduction in latency compared with those not receiving the intervention (105 children in each group). Kołakowska et al. (113) measured tablet sensors during gameplay, finding that these sensors can be used to predict therapy progress in 40 children with autism, with accuracy surpassing 80%.

The study of autism spans several academic disciplines, and there are myriad data science studies for autism using genomics (114–117), epigenomics (118), proteomics (119–122), metagenomics (123–125), and brain imaging (126–129). While we admire and hope to see more of these studies, they are beyond the scope of this review. We note, however, that there is an opportunity for efforts in biology-based fields to eventually converge with the types of analyses described here, as increasingly granular behavioral phenotypes using digital phenotyping methods will likely enable more robust and precise biological association studies. We note that many of the data science methods that we have reviewed can apply more broadly to neurological, developmental, and behavioral conditions.

We invite all readers to join us in this promising, burgeoning field. We look forward to the increases in access to care and scientific understanding of complex social human behavior that will inevitably flourish over the next two decades and beyond.

## DISCLOSURE STATEMENT

D.P.W. is the founder of Cognoa, a company focused on building scalable digital and AI-medicine solutions for pediatric healthcare. P.W. is not aware of any memberships, affiliations, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

The writing of this review was supported in part by funds to D.P.W. from the National Institutes of Health (1R01EB025025-01, 1R01LM013364-01, 1R21HD091500-01, 1R01LM013083); National Science Foundation (award 2014232); Hartwell Foundation; Bill & Melinda Gates Foundation; Coulter Foundation; Lucile Packard Foundation for Children's Health; Auxiliaries Endowment; IsDB Transform Fund; Weston Havens Foundation; program grants from Stanford's Human-Centered Artificial Intelligence Program, Precision Health and Integrated Diagnostics Center, Beckman Center, Bio-X Center, Predictives and Diagnostics Accelerator, Spectrum, Spark Program in Translational Research, and MediaX; and from the Wu Tsai Neurosciences Institute's Neuroscience:Translate Program. We also acknowledge generous support from David Orr, Imma Calvo, Bobby Dekesyer, and Peter Sullivan.

## LITERATURE CITED

1. Maenner MJ, Shaw KA, Bakian AV, Bilder DA, Durkin MS, et al. 2021. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2018. *MMWR Surveill. Summ.* 70(11):1–16
2. Ning M, Daniels J, Schwartz J, Dunlap K, Washington P, et al. 2019. Identification and quantification of gaps in access to autism resources in the United States: an infodemiological study. *J. Med. Internet Res.* 21(7):e13094
3. Bargiela S, Steward R, Mandy W. 2016. The experiences of late-diagnosed women with autism spectrum conditions: an investigation of the female autism phenotype. *J. Autism Dev. Disord.* 46(10):3281–94
4. Fenske EC, Zalenski S, Krantz PJ, McClannahan LE. 1985. Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Anal. Interv. Dev. Disabil.* 5(1/2):49–58
5. Pew Res. Cent. 2021. *Mobile fact sheet*. Fact Sheet, Pew Res. Cent., Washington, DC. <https://www.pewresearch.org/internet/fact-sheet/mobile>
6. Silver L, Smith A, Johnson C, Taylor K, Jiang J, et al. 2019. *Mobile connectivity in emerging economies*. Tech. Rep., Pew Res. Cent., Washington, DC
7. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, et al. 2000. The Autism Diagnostic Observation Schedule—Generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* 30(3):205–23
8. Rutter M, Le Couteur A, Lord C. 2003. *Autism Diagnostic Interview—Revised*. Los Angeles: West. Psychol. Serv.
9. Van Bourgondien ME, Marcus LM, Schopler E. 1992. Comparison of DSM-III-R and childhood autism rating scale diagnoses of autism. *J. Autism Dev. Disord.* 22(4):493–506
10. South M, Williams BJ, McMahon WM, Owley T, Filipek PA, et al. 2002. Utility of the Gilliam Autism Rating Scale in research and clinical populations. *J. Autism Dev. Disord.* 32:593–99
11. Geschwind DH, Sowsinski J, Lord C, Iversen P, Shestack J, et al. 2001. The Autism Genetic Resource Exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.* 69(2):463–66
12. Hall D, Huerta MF, McAuliffe MJ, Farber GK. 2012. Sharing heterogeneous data: the National Database for Autism Research. *Neuroinformatics* 10(4):331–39
13. Fischbach GD, Lord C. 2010. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* 68(2):192–95
14. Simons VIP (Var. Individ.) Consort. 2012. Simons Variation in Individuals Project (Simons VIP): a genetics-first approach to studying autism spectrum and related neurodevelopmental disorders. *Neuron* 73(6):1063–67
15. Hu-Lince D, Craig DW, Huentelman MJ, Stephan DA. 2005. The Autism Genome Project. *Am. J. Pharmacogenom.* 5(4):233–46
16. de Belen RA, Bednarz T, Sowmya A, Del Favero D. 2020. Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019. *Transl. Psychiatry* 10:333
17. Abbe A, Grouin C, Zweigenbaum P, Falissard B. 2016. Text mining applications in psychiatry: a systematic literature review. *Int. J. Methods Psychiatr. Res.* 25(2):86–100
18. Fusaroli R, Lambrechts A, Bang D, Bowler DM, Gaigg SB. 2017. Is voice a marker for autism spectrum disorder? A systematic review and meta-analysis. *Autism Res.* 10(3):384–407
19. Kalantarian H, Washington P, Schwartz J, Daniels J, Haber N, et al. 2019. Guess what? *J. Healthc. Inform. Res.* 3(1):43–66
20. Kalantarian H, Jedoui K, Washington P, Wall DP. 2018. A mobile game for automatic emotion-labeling of images. *IEEE Trans. Games* 12(2):213–18
21. Kalantarian H, Jedoui K, Washington P, Tariq Q, Dunlap K, et al. 2019. Labeling images with facial emotion and the potential for pediatric healthcare. *Artif. Intell. Med.* 98:77–86
22. Kalantarian H, Washington P, Schwartz J, Daniels J, Haber N, et al. 2018. A gamified mobile system for crowdsourcing video for autism research. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 350–52. Washington, DC: IEEE Comput. Soc.

23. Hou C, Kalantarian H, Washington P, Dunlap K, Wall DP. 2021. Leveraging video data from a digital smartphone autism therapy to train an emotion detection classifier. medRxiv 2021.07.28.21260646. <https://doi.org/10.1101/2021.07.28.21260646>
24. Kalantarian H, Jedoui K, Dunlap K, Schwartz J, Washington P, et al. 2020. The performance of emotion classifiers for children with parent-reported autism: quantitative feasibility study. *JMIR Ment. Health* 7(4):e13174
25. Washington P, Kalantarian H, Kent J, Husic A, Kline A, et al. 2022. Improved digital therapy for developmental pediatrics using domain-specific artificial intelligence: machine learning study. *JMIR Pediatr. Parent.* 5(2):e26760
26. Washington P, Kalantarian H, Kent J, Husic A, Kline A, et al. 2020. Training an emotion detection classifier using frames from a mobile therapeutic game for children with developmental disorders. arXiv:2012.08678 [cs.CV]
27. Penev Y, Dunlap K, Husic A, Hou C, Washington P, et al. 2021. A mobile game platform for improving social communication in children with autism: a feasibility study. *Appl. Clin. Inform.* 12(5):1030–40
28. Scassellati B, Admoni H, Matarić M. 2012. Robots for use in autism research. *Annu. Rev. Biomed. Eng.* 14:275–94
29. Scassellati B. 2007. How social robots will help us to diagnose, treat, and understand autism. In *Robotics Research*, ed. S Thrun, R Brooks, H Durrant-Whyte, pp. 552–63. Berlin: Springer
30. Haber N, Voss C, Wall D. 2020. Making emotions transparent: Google Glass helps autistic kids understand facial expressions through augmented-reality therapy. *IEEE Spectr.* 57(4):46–52
31. Kline A, Voss C, Washington P, Haber N, Schwartz H, et al. 2019. Superpower glass. *GetMobile Mobile Comput. Commun.* 23(2):35–38
32. Voss C, Washington P, Haber N, Kline A, Daniels J, et al. 2016. Superpower Glass: delivering unobtrusive real-time social cues in wearable systems. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1218–26. New York: Assoc. Comput. Mach.
33. Washington P, Voss C, Haber N, Tanaka S, Daniels J, et al. 2016. A wearable social interaction aid for children with autism. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2348–54. New York: Assoc. Comput. Mach.
34. Washington P, Voss C, Kline A, Haber N, Daniels J, et al. 2017. SuperpowerGlass: a wearable aid for the at-home therapy of children with autism. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1(3):112
35. Daniels J, Haber N, Voss C, Schwartz J, Tamura S, et al. 2018. Feasibility testing of a wearable behavioral aid for social learning in children with autism. *Appl. Clin. Inform.* 9(1):129–40
36. Daniels J, Schwartz J, Haber N, Voss C, Kline A, et al. 2017. Design and efficacy of a wearable device for social affective learning in children with autism. *J. Am. Acad. Child Adolesc. Psychiatry* 56(10):S257 (Abstr.)
37. Voss C, Schwartz J, Daniels J, Kline A, Haber N, et al. 2019. Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: a randomized clinical trial. *JAMA Pediatr.* 173(5):446–54
38. Nag A, Haber N, Voss C, Tamura S, Daniels J, et al. 2020. Toward continuous social phenotyping: analyzing gaze patterns in an emotion recognition task for children with autism through wearable smart glasses. *J. Med. Internet Res.* 22(4):e13810
39. Jones W, Klin A. 2013. Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* 504(7480):427–31
40. Riby D, Hancock PJ. 2009. Looking at movies and cartoons: eye-tracking evidence from Williams syndrome and autism. *J. Intellect. Disabil. Res.* 53(2):169–81
41. Chawarska K, Macari S, Shic F. 2013. Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biol. Psychiatry* 74(3):195–203
42. Campbell K, Carpenter KL, Hashemi J, Espinosa S, Marsan S, et al. 2019. Computer vision analysis captures atypical attention in toddlers with autism. *Autism* 23(3):619–28
43. Sadria M, Karimi S, Layton AT. 2019. Network centrality analysis of eye-gaze data in autism spectrum disorder. *Comput. Biol. Med.* 111:103332
44. Varma M, Washington P, Chrisman B, Kline A, Leblanc E, et al. 2022. Identification of social engagement indicators associated with autism spectrum disorder using a game-based mobile app: comparative study of gaze fixation and visual scanning methods. *J. Med. Internet Res.* 24(2):e31830

45. Alvari G, Coviello L, Furlanello C. 2021. EYE-C: eye-contact robust detection and analysis during unconstrained child-therapist interactions in the clinical setting of autism spectrum disorders. *Brain Sci.* 11(12):1555
46. Wen TH, Cheng A, Andreason C, Zahiri J, Xiao Y, et al. 2022. Large scale validation of an early-age eye-tracking biomarker of an autism spectrum disorder subtype. *Sci. Rep.* 12:4253
47. Putra PU, Shima K, Alvarez SA, Shimatani K. 2021. Identifying autism spectrum disorder symptoms using response and gaze behavior during the Go/NoGo game CatChicken. *Sci. Rep.* 11:22012
48. Guha T, Yang Z, Grossman RB, Narayanan SS. 2016. A computational study of expressive facial dynamics in children with autism. *IEEE Trans. Affect. Comput.* 9(1):14–20
49. Egger HL, Dawson G, Hashemi J, Carpenter KL, Espinosa S, et al. 2018. Automatic emotion and attention analysis of young children at home: a ResearchKit autism feasibility study. *npj Digital Med.* 1:20
50. Martin KB, Hammal Z, Ren G, Cohn JF, Cassell J, et al. 2018. Objective measurement of head movement differences in children with and without autism spectrum disorder. *Mol. Autism* 9:14
51. Dawson G, Campbell K, Hashemi J, Lippmann SJ, Smith V, et al. 2018. Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder. *Sci. Rep.* 8:17008
52. Hudenko WJ, Stone W, Bachorowski JA. 2009. Laughter differs in children with autism: an acoustic analysis of laughs produced by children with and without the disorder. *J. Autism Dev. Disord.* 39(10):1392–400
53. Orlandi S, Manfredi C, Bocchi L, Scattoni ML. 2012. Automatic newborn cry analysis: a non-invasive tool to help autism early diagnosis. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2953–56. Piscataway, NJ: IEEE
54. Jordan MI, Mitchell TM. 2015. Machine learning: trends, perspectives, and prospects. *Science* 349(6245):255–60
55. Wall DP, Kosmicki J, Deluca TF, Harstad E, Fusaro VA. 2012. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl. Psychiatry* 2(4):e100
56. Washington P, Leblanc E, Dunlap K, Penev Y, Kline A, et al. 2020. Precision telemedicine through crowdsourced machine learning: testing variability of crowd workers for video-based autism feature recognition. *J. Pers. Med.* 10(3):86
57. Washington P, Leblanc E, Dunlap K, Penev Y, Varma M, et al. 2020. Selection of trustworthy crowd workers for telemedical diagnosis of pediatric autism spectrum disorder. *Pac. Symp. Biocomput.* 26:14–25
58. Washington P, Kalantarian H, Tariq Q, Schwartz J, Dunlap K, et al. 2019. Validity of online screening for autism: crowdsourcing study comparing paid and unpaid diagnostic tasks. *J. Med. Internet Res.* 21(5):e13668
59. Abbas H, Garberson F, Glover E, Wall DP. 2018. Machine learning approach for early detection of autism by combining questionnaire and home video screening. *J. Am. Med. Inform. Assoc.* 25(8):1000–7
60. Abbas H, Garberson F, Glover E, Wall DP. 2017. Machine learning for early detection of autism (and other conditions) using a parental questionnaire and home video screening. In *2017 IEEE International Conference on Big Data*, pp. 3558–61. Piscataway, NJ: IEEE
61. Abbas H, Garberson F, Liu-Mayo S, Glover E, Wall DP. 2020. Multi-modular AI approach to streamline autism diagnosis in young children. *Sci. Rep.* 10:5014
62. Fusaro VA, Daniels J, Duda M, DeLuca TF, D’Angelo O, et al. 2014. The potential of accelerating early detection of autism through content analysis of YouTube videos. *PLOS ONE* 9(4):e93533
63. Kanne SM, Carpenter LA, Warren Z. 2018. Screening in toddlers and preschoolers at risk for autism spectrum disorder: evaluating a novel mobile-health screening tool. *Autism Res.* 11(7):1038–49
64. Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, et al. 2018. Mobile detection of autism through machine learning on home video: a development and prospective validation study. *PLOS Med.* 15(11):e1002705
65. Tariq Q, Fleming SL, Schwartz JN, Dunlap K, Corbin C, et al. 2019. Detecting developmental delay and autism through machine learning models using home videos of Bangladeshi children: development and validation study. *J. Med. Internet Res.* 21(4):e13822
66. Duda M, Daniels J, Wall DP. 2016. Clinical evaluation of a novel and mobile autism risk assessment. *J. Autism Dev. Disord.* 46(6):1953–61

67. Levy S, Duda M, Haber N, Wall DP. 2017. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Mol. Autism* 8:65
68. Washington P, Paskov KM, Kalantarian H, Stockham N, Voss C, et al. 2020. Feature selection and dimension reduction of social autism data. *Pac. Symp. Biocomput.* 25:707–18
69. Washington P, Tariq Q, Leblanc E, Chrisman B, Dunlap K, et al. 2021. Crowdsourced privacy-preserved feature tagging of short home videos for machine learning ASD detection. *Sci. Rep.* 11:7620
70. Washington P, Chrisman B, Leblanc E, Dunlap K, Kline A, et al. 2022. Crowd annotations can approximate clinical autism impressions from short home videos with privacy protections. *Intell.-Based Med.* 6:100056
71. Liu W, Li M, Yi L. 2016. Identifying children with autism spectrum disorder based on their face processing abnormality: a machine learning framework. *Autism Res.* 9(8):888–98
72. Duan H, Zhai G, Min X, Che Z, Fang Y, et al. 2019. A dataset of eye movements for the children with autism spectrum disorder. In *Proceedings of the 10th ACM Multimedia Systems Conference*, pp. 255–60. New York: Assoc. Comput. Mach.
73. Chang Z, Di Martino JM, Aiello R, Baker J, Carpenter K, et al. 2021. Computational methods to measure patterns of gaze in toddlers with autism spectrum disorder. *JAMA Pediatr.* 175(8):827–36
74. Oliveira JS, Franco FO, Revers MC, Silva AF, Portolese J, et al. 2021. Computer-aided autism diagnosis based on visual attention models using eye tracking. *Sci. Rep.* 11:10131
75. Drimalla H, Landwehr N, Baskow I, Behnia B, Roepke S, et al. 2019. Detecting autism by analyzing a simulated social interaction. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 193–208. Cham, Switz.: Springer
76. Drimalla H, Scheffer T, Landwehr N, Baskow I, Roepke S, et al. 2020. Towards the automatic detection of social biomarkers in autism spectrum disorder: introducing the simulated interaction task (SIT). *npj Digital Med.* 3:25
77. Khosla Y, Ramachandra P, Chaitra N. 2021. Detection of autistic individuals using facial images and deep learning. In *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions*, pp. 1–5. Piscataway, NJ: IEEE
78. Hashemi J, Spina TV, Tepper M, Esler A, Morellas V, et al. 2012. A computer vision approach for the assessment of autism-related behavioral markers. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics*, pp. 1–7. Piscataway, NJ: IEEE
79. Lidstone DE, Rochowiak R, Pacheco C, Tunçgenç B, Vidal R, et al. 2021. Automated and scalable Computerized Assessment of Motor Imitation (CAMI) in children with Autism Spectrum Disorder using a single 2D camera: a pilot study. *Res. Autism Spectr. Disord.* 87:101840
80. Kojovic N, Natraj S, Mohanty SP, Maillart T, Schaer M. 2021. Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Sci. Rep.* 11:15069
81. Cook A, Mandal B, Berry D, Johnson M. 2019. Towards automatic screening of typical and atypical behaviors in children with autism. In *2019 IEEE International Conference on Data Science and Advanced Analytics*, pp. 504–10. Piscataway, NJ: IEEE
82. 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
83. Anzulewicz A, Sobota K, Delafield-Butt JT. 2016. Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism. *Sci. Rep.* 6:31107
84. Cavallo A, Romeo L, Ansuini C, Battaglia F, Nobili L, et al. 2021. Identifying the signature of prospective motor control in children with autism. *Sci. Rep.* 11:3165
85. Li M, Tang D, Zeng J, Zhou T, Zhu H, et al. 2019. An automated assessment framework for atypical prosody and stereotyped idiosyncratic phrases related to autism spectrum disorder. *Comput. Speech Lang.* 56:80–94
86. Chi NA, Washington P, Kline A, Husic A, Hou C, et al. 2022. Classifying autism from crowdsourced semistructured speech recordings: machine learning model comparison study. *JMIR Pediatr. Parent.* 5(2):e35406
87. Lau JC, Patel S, Kang X, Nayar K, Martin GE, et al. 2022. Cross-linguistic patterns of speech prosodic differences in autism: a machine learning study. *PLOS ONE* 17(6):e0269637



88. Maenner MJ, Yeargin-Allsopp M, Van Naarden Braun K, Christensen DL, Schieve LA. 2016. Development of a machine learning algorithm for the surveillance of autism spectrum disorder. *PLOS ONE* 11(12):e0168224
89. Vabalas A, Gowen E, Poliakoff E, Casson AJ. 2020. Applying machine learning to kinematic and eye movement features of a movement imitation task to predict autism diagnosis. *Sci. Rep.* 10:8346
90. Javed H, Park CH. 2020. Behavior-based risk detection of autism spectrum disorder through child-robot interaction. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 275–77. New York: Assoc. Comput. Mach.
91. Constantino JN, Kennon-McGill S, Weichselbaum C, Marrus N, Haider A, et al. 2017. Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* 547(7663):340–44
92. Jones W, Carr K, Klin A. 2008. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch. Gen. Psychiatry* 65(8):946–54
93. Klin A, Jones W, Schultz R, Volkmar F, Cohen D. 2002. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* 59(9):809–16
94. Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W. 2009. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* 459(7244):257–61
95. Rice K, Moriuchi JM, Jones W, Klin A. 2012. Parsing heterogeneity in autism spectrum disorders: visual scanning of dynamic social scenes in school-aged children. *J. Am. Acad. Child Adolesc. Psychiatry* 51(3):238–48
96. Shultz S, Klin A, Jones W. 2011. Inhibition of eye blinking reveals subjective perceptions of stimulus salience. *PNAS* 108(52):21270–75
97. Shultz S, Klin A, Jones W. 2018. Neonatal transitions in social behavior and their implications for autism. *Trends Cogn. Sci.* 22(5):452–69
98. Grote T. 2021. Trustworthy medical AI systems need to know when they don't know. *J. Med. Ethics* 47(5):337–38
99. Hendrickx K, Perini L, Van der Plas D, Meert W, Davis J. 2021. Machine learning with a reject option: a survey. arXiv:2107.11277 [cs.LG]
100. Kompa B, Snoek J, Beam AL. 2021. Second opinion needed: communicating uncertainty in medical machine learning. *npj Digital Med.* 4:4
101. Char DS, Shah NH, Magnus D. 2018. Implementing machine learning in health care—addressing ethical challenges. *N. Engl. J. Med.* 378(11):981–83
102. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. 2021. A survey on bias and fairness in machine learning. *ACM Comput. Surv.* 54(6):115
103. Feliciano P, Daniels AM, Snyder LG, Beaumont A, Camba A, et al. 2018. SPARK: a US cohort of 50,000 families to accelerate autism research. *Neuron* 97(3):488–93
104. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, et al. 2020. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. *Lancet Digit. Health* 2(4):e168–78
105. Rudovic O, Lee J, Dai M, Schuller B, Picard RW. 2018. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci. Robot.* 3(19):eaao6760
106. Yu H, Sano A. 2022. Semi-supervised learning and data augmentation in wearable-based momentary stress detection in the wild. arXiv:2202.12935 [eess.SP]
107. Gardner-Hoag J, Novack M, Parlett-Pelleriti C, Stevens E, Dixon D, et al. 2021. Unsupervised machine learning for identifying challenging behavior profiles to explore cluster-based treatment efficacy in children with autism spectrum disorder: retrospective data analysis study. *JMIR Med. Inform.* 9(6):e27793
108. Duda M, Haber N, Daniels J, Wall DP. 2017. Crowdsourced validation of a machine-learning classification system for autism and ADHD. *Transl. Psychiatry* 7(5):e1133
109. Wawer A, Chojnicka I, Okruszek L, Sarzynska-Wawer J. 2022. Single and cross-disorder detection for autism and schizophrenia. *Cogn. Comput.* 14(1):461–73

110. Demetriou EA, Park SH, Ho N, Pepper KL, Song YJ, et al. 2020. Machine learning for differential diagnosis between clinical conditions with social difficulty: autism spectrum disorder, early psychosis, and social anxiety disorder. *Front. Psychiatry* 11:545
111. Iakovidou N, Lanzarini E, Singh J, Fiori F, Santosh P. 2020. Differentiating females with Rett syndrome and those with multi-comorbid autism spectrum disorder using physiological biomarkers: a novel approach. *J. Clin. Med.* 9(9):2842
112. McKernan EP, Kumar M, Di Martino A, Shulman L, Kolevzon A, et al. 2022. Intra-topic latency as an automated behavioral marker of treatment response in autism spectrum disorder. *Sci. Rep.* 12:3255
113. Kołakowska A, Landowska A, Anzulewicz A, Sobota K. 2017. Automatic recognition of therapy progress among children with autism. *Sci. Rep.* 7:13863
114. De Rubeis S, Buxbaum JD. 2015. Genetics and genomics of autism spectrum disorder: embracing complexity. *Hum. Mol. Genet.* 24(R1):R24–31
115. Mefford HC, Batshaw ML, Hoffman EP. 2012. Genomics, intellectual disability, and autism. *N. Engl. J. Med.* 366(8):733–43
116. Vorstman JA, Parr JR, Moreno-De-Luca D, Anney RJ, Nurnberger JI Jr., et al. 2017. Autism genetics: opportunities and challenges for clinical translation. *Nat. Rev. Genet.* 18(6):362–76
117. Willsey HR, Willsey AJ, Wang B, State MW. 2022. Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. *Nat. Rev. Neurosci.* 23(6):323–41
118. Waye MM, Cheng HY. 2018. Genetics and epigenetics of autism: a review. *Psychiatry Clin. Neurosci.* 72(4):228–44
119. Abraham J, Szoko N, Natowicz MR. 2019. Proteomic investigations of autism spectrum disorder: past findings, current challenges, and future prospects. In *Reviews on Biomarker Studies in Psychiatric and Neurodegenerative Disorders*, ed. PC Guest, pp. 235–52. Cham, Switz.: Springer
120. Anderson GM. 2015. Autism biomarkers: challenges, pitfalls and possibilities. *J. Autism Dev. Disord.* 45(4):1103–13
121. Al-Ayadhi L, Halepoto DM. 2013. Role of proteomics in the discovery of autism biomarkers. *J. Coll. Physicians Surg. Pak.* 23:137–43
122. Ristori MV, Mortera SL, Marzano V, Guerrera S, Vernocchi P, et al. 2020. Proteomics and metabolomics approaches towards a functional insight onto autism spectrum disorders: phenotype stratification and biomarker discovery. *Int. J. Mol. Sci.* 21(17):6274
123. Kotsiliti E. 2022. Gut microbiome and autism spectrum disorder. *Nat. Rev. Gastroenterol. Hepatol.* 19:6
124. Mulle JG, Sharp WG, Cubells JF. 2013. The gut microbiome: a new frontier in autism research. *Curr. Psychiatry Rep.* 15(2):337
125. Pulikkan J, Mazumder A, Grace T. 2019. Role of the gut microbiome in autism spectrum disorders. In *Review on Biomarker Studies in Psychiatric and Neurodegenerative Disorders*, ed. PC Guest, pp. 253–69. Cham, Switz.: Springer
126. Khodatars M, Shoeibi A, Sadeghi D, Ghaasemi N, Jafari M, et al. 2021. Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: a review. *Comput. Biol. Med.* 139:104949
127. Moon SJ, Hwang J, Kana R, Torous J, Kim JW. 2019. Accuracy of machine learning algorithms for the diagnosis of autism spectrum disorder: systematic review and meta-analysis of brain magnetic resonance imaging studies. *JMIR Ment. Health* 6(12):e14108
128. Nogay HS, Adeli H. 2020. Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging. *Rev. Neurosci.* 31(8):825–41
129. Pagnozzi AM, Conti E, Calderoni S, Fripp J, Rose SE. 2018. A systematic review of structural MRI biomarkers in autism spectrum disorder: a machine learning perspective. *Int. J. Dev. Neurosci.* 71:68–82