

Received 29 August 2022, accepted 11 September 2022, date of publication 22 September 2022, date of current version 7 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3208587



The Role of Intelligent Technologies in Early Detection of Autism Spectrum Disorder (ASD): A Scoping Review

MANU KOHLI^{ID}, (Member, IEEE), ARPAN KUMAR KAR, AND SHUCHI SINHA

Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

Corresponding author: Manu Kohli (manu.kohli@dms.iitd.ac.in)

ABSTRACT **Background:** Two-year delay is reported between the first developmental concern raised by the parents and the diagnosis of ASD (Autism Spectrum Disorder), delaying the start of early intervention programs most beneficial within the first three years. **Aim:** Evaluate the role of technology in ASD detection by answering four research questions analyzing 1) evolution of technology, 2) use of various bio-behavioral data sources, 3) demographic categories, databases, controls, comparators, and assessment instruments, and 4) data collection, processing, and outcomes of the technology-based methods in ASD detection. **Methods:** Scoping review included behavioral-based ASD screening and diagnostic studies, published between 1st January 2011 to 31st December 2021 in PUBMED, SCOPUS, and IEEE Xplore databases for children under six years. The studies were evaluated using the Critical Appraisal Skills Programm (CASP) and the PRISMA scoping review checklist (PRISMA-ScR). **Results:** The shortlisted 35 studies were categorized into seven bio-behavioral categories. The review highlighted the extensive use of machine learning (ML) and Deep Learning (DL) to detect infants (as young as 9 to 12 months) at risk of ASD and Other developmental delays (ODD) using multimodal structured and unstructured data. However, the review reported various internal and external validity threats. **Conclusion:** Technology can significantly improve the current ASD detection process. The validation and adoption of technology can be fast-tracked by 1) designing robust study protocols, 2) executing multi-cultural field trials, 3) standardizing datasets, data quality, and feature engineering methods, 4) recruiting statistically significant participants from ASD, typically developing (TD) and other developmental disorders (ODD) groups to ensure technological generalization, validation, and adoption outside laboratory settings.

INDEX TERMS Autism, screening, diagnosis, technology, machine learning, mobile technology, artificial intelligence.

I. INTRODUCTION

Autism Spectrum Disorder (ASD) roughly affects 1.3 million children annually, on a conservative one in hundred diagnosis rate [1], and has increased 700% since 1996 [2].

Generally, the best-recommended practices to detect ASD and other development disorders (ODD) are developmental monitoring, screening, and diagnosis [3]. The World Health Organization [4] recommends developmental monitoring in

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar^{ID}.

Low and Medium Income countries (LMICs) for the early identification of developmental challenges. Screening is a more formal, standardized method that includes routine pediatric evaluations [5], usually practiced in high-income countries (HIC). Practitioners such as doctors, nurses, and school teachers request families to respond to level 1 screeners [6] such as M-CHAT-R/F [7], evaluating children's social communication, peer interaction, eye contact, motor skills, and fixated behaviors if any. Further, High-Risk (HR) children, for example, siblings of children with ASD or with birth complications, low birth weight, or admitted to newborn intensive

care unit (NICU), are recommended to undergo additional developmental risk assessments [8].

An exhaustive developmental examination can confirm diagnosis and referral to intervention if the screening instrument indicates a developmental concern. Clinicians usually implement gold-standard tools such as Autism Diagnostic Observation Schedule (ADOS-2) [9] and Autism Diagnostic Interview-Revised (ADI-R) [10] to confirm ASD diagnosis.

Though early ASD indicators are evident at 12 months, and diagnosis is possible at earlier than 18 months [11], most children are diagnosed between 48-60 months [12], [13], highlighting a delay of two years. Delayed diagnosis slows the initiation of early intervention services by 12-14 months [14], which can improve children's IQ by 10-15 points if started under the age of three [15] due to the brain's high neuroplasticity. Therefore, early ASD identification and intervention can ensure a better quality of life for ASD children.

There are various reasons for the delayed or misdiagnosis of ASD among children. Firstly, children with ASD exhibit high variability in typical ASD features such as stereotypical interests, repetitive behaviors, and limited communication and social skills [1]. The high behavioral variance makes it challenging for the clinician to establish an early diagnosis for borderline and high-functioning ASD children, for example, with Asperger syndrome [16]. Moreover, with 80% of ASD cases diagnosed in males [1], women with ASD [17] are susceptible to diagnostic delays and misdiagnosis attributed to stereotypical gender biases [18]. Secondly, the symptomatic similarity of ASD with Attention Deficit Hyperactive Disorder (ADHD) and speech delays [19] often leads to delayed or misdiagnoses [20]. An accurate diagnosis is critical to identifying the child's area of strength and developing a personalized need-based intervention plan per the child's need [21]. Thirdly, the gold standard ASD diagnostic and screening tools such as ADOS [9], ADI-R [10], M-CHAT-R/F [7], and CARS2 [22] are designed for the western world. Therefore, these tests are sensitive to evaluation biases and subjective decision-making of clinicians from Low and Medium income countries (LMICs), resulting in incorrect results, primarily influenced due to lack of training and cultural disparities [23]. Fourthly, the availability of clinicians and infrastructure globally to assist ASD detection and management is limited [24], especially in LMICs, a challenge further constrained by the poor awareness of the disorder [25]. Also, families have limited access to clinicians and infrastructure and usually travel considerable distances or relocate to access services [26]. These limitations lead to lengthy wait times, delayed diagnosis, and causing stress to individuals and families [26], [27].

In addition, the current ASD detection process has limitations. The clinicians require significant training and time to implement diagnostic instruments [28]. A 93-point ADI-R questionnaire, for example, can take 2.5 hours to complete [29] across multiple visits. Further, interview responses are based on the caregiver's subjective comprehension of

assessment questions and their reliance on memory recall of the child's developmental history, contributing to evaluation and assessment biases [30]. Moreover, developmental evaluations are seldom conducted in children's natural contexts, such as in their homes. An encounter with a new clinician in a new environment with social performance pressure may trigger discomfort for the child resulting in assessment and diagnostic biases.

Artificial Intelligence (A.I.) based innovations have fast-tracked ASD diagnostics [31], [32], increased clinician capacity, and improved access to early intervention programs [26]. The adoption of these technologies has surged during the COVID-19 pandemic [33]. These solutions have the following benefits over traditional face-to-face methods: 1) enhancing ASD management solution access to rural and underserved persons and families, 2) reducing doctors' and patients' expenditures (such as travel duration and cost), and 3) expanding providers' coverage areas. The preliminary findings provide evidence of technological innovation's feasibility and efficacy in improving current ASD detection and behavioral intervention methods, enhancing access, quality, and affordability. However, more in-depth analysis and information can confirm the impact and outcomes of these innovations.

Scoping reviews are a descriptive method that aids in analyzing complicated or varied research projects by identifying the critical concepts, theories, and evidence sources to guide and evaluate the adoption of new methods into practice [34]. The results of scoping reviews can identify gaps in the existing literature and indicate areas with limited evidence to merit additional studies or a systematic review. We, therefore, performed a scoping review to evaluate the use of innovative technologies for ASD detection. We investigated a body of literature to examine the extent, nature, and scope of current research activities and answer the following four research questions based on the PICO framework [35], [36] aligned toward diagnostic innovations. In the framework definition, "P" signifies the population in focus, "I" for intervention or researched condition, "C" for the comparators, and "O" for psychometric outcomes.

- 1) RQ1 How has the literature on technology-based ASD detection methods evolved?
- 2) RQ2 How do researchers use the various bio-behavioral markers to detect ASD?
- 3) RQ3 What demographic categories, databases, controls, comparators, and assessment instruments are a part of the technology-facilitated ASD detection process?
- 4) RQ4 How have researchers gathered and processed multimodal data? How do technological innovation's results compare to conventional ASD detection methods?

The review is based on PRISMA scoping review guidelines and includes the following sections. Section II details eligibility criteria for study selection, keyword definition and justification, study search process, data extraction, and

analysis. The results are listed in section III, where we synthesized the review finding and answered four research questions. We present the result under seven multimodal data categories, technological subdivisions, analyzing data sources, data extraction, synthesis, and outcomes. Discussion section IV highlights internal and external validity threats, advantages, disadvantages, ethical, legal, and cultural constraints, high-level limitations, and mitigation measures and recommends future directions. Section V lists the study's limitations and section VI lists future directions and additional focus areas for research. Finally, in section VII, we conclude our findings.

II. MATERIALS AND METHODS

This section describes the study's selection criteria, search strategy, justification, data extraction, and analysis. The review is conducted using the PRISMA Extension for Scoping Reviews (PRISMA-ScR) checklist [37]. The 22-point checklist is attached in the appendix section (See Appendix C).

A. ELIGIBILITY CRITERIA

The inclusion criteria for this study are as follows: (1) Studies that leveraged technology and included behavioral-based ASD screening or diagnostic methods; (2) included children under the age of six; (3) published between January 1, 2011, and December 31, 2021; (4) included quantitative ASD detection methods including cross-sectional experiments, longitudinal data analysis, and dataset investigations; and (5) were part of one of the three electronic databases: PUBMED, IEEE Xplore, and SCOPUS. The following are the search criteria justifications:

- 1) Most evidence-based ASD detection methods [38], and tools [9], [10], [22], [30] track social communication, eye contact, challenging behavioral, and notable play-based landmarks to identify children with ASD. We, therefore, shortlisted studies that used these behavioral landmarks and excluded studies focusing on medicine, biology, genetics, EEG (electroencephalogram), MRI (Magnetic resonance imaging) usage, and non-technology-based ASD screening or diagnosis methods.
- 2) We excluded conference papers to ensure we included only high-quality peer-review journal publications selected from PUBMED, IEEE Xplore, and SCOPUS.
- 3) We excluded literature reviews as we focussed on studies that conducted experiments, trials, datasets, or longitudinal multimodal data analysis.
- 4) Since 2011, the growth in mobile and edge-based A.I. innovations can be attributed to the emergence of low-cost, scalable cloud computing infrastructure and sensors [39], [40], [41], [42], [43]. Therefore, we selected studies published between January 1, 2011, and December 31, 2021, to evaluate the role of technology in ASD evaluation.

- 5) Given the importance and effectiveness of early ASD detection and intervention due to the brain's strong neuroplasticity, the emphasis of the review was limited to studies that included children under the age of six.

B. SEARCH STRINGS

We searched the following search strings in the title, abstract, and keywords fields: ("Autism spectrum disorder" OR "ASD" OR "Autism" OR "AUTISTIC") AND ("Detect*" OR "Predict*" OR "Diagnos*" OR "Screening" OR "Identif*" OR "Suspect" OR "Classif*" OR "Distinguish*" OR "Differentiate" OR "Risk") AND ("Technology" OR "A.I." OR "Artificial Intelligence" OR "Machine Learning" OR "Mobile").

The search string justification is as follows:

- 1) The keywords "Autism spectrum disorder" OR "ASD" OR "Autism" OR "AUTISTIC" shortlisted studies focused on Autism Spectrum Disorder.
- 2) To ensure shortlisted studies focused on screening, diagnosis, detection, and identification of ASD, we included the following keywords: "Detect*" OR "Predict*" OR "Diagnos*" OR "Screening" OR "Identif*" OR "Suspect" OR "Risk". The keywords "Classif*" OR "Distinguish*" OR "Differentiate" were included to shortlist studies that differentiate between ASD, T.D. (Typical development), and ODD (other developmental disorder) groups.
- 3) The keywords "Technology" OR "A.I." OR "Artificial Intelligence" OR "Machine Learning" OR "Mobile" helped shortlist studies with technology usage.
- 4) The search outcomes of the above three criteria were combined with an AND operator.

C. SEARCH RESULTS PROCESSING

Two authors MK and AKK (Manu Kohli & Arpan Kumar Kar), completed each phase of the PRISMA scoping review depicted in Figure 1. Any contradictory results were resolved with the consultation and mediation of a third author SS (Shuchi Sinha). The search results were downloaded, compiled, and imported into Zotero[©] for the presence of duplicates and subsequent removal. Zotero[©] assists reference management by syncing citations with bibliographies, DOI (Digital object identifiers), and metadata. Each unique article's title and abstract were screened for relevancy, followed by a full-text analysis per the inclusion-exclusion criteria listed in subsection II-A. Thirty-two studies were shortlisted post-full-text analysis, and additional three publications [44], [45], [46] were uncovered by analyzing the shortlisted study's references, making the total shortlisted study count to 35.

D. DATA EXTRACTION AND ANALYSIS

The review included extracting the below-listed data from thirty-five shortlisted studies listed in two tables. Table 1 includes multimodal input data, feature reduction steps, environment setting, data processing algorithms, and

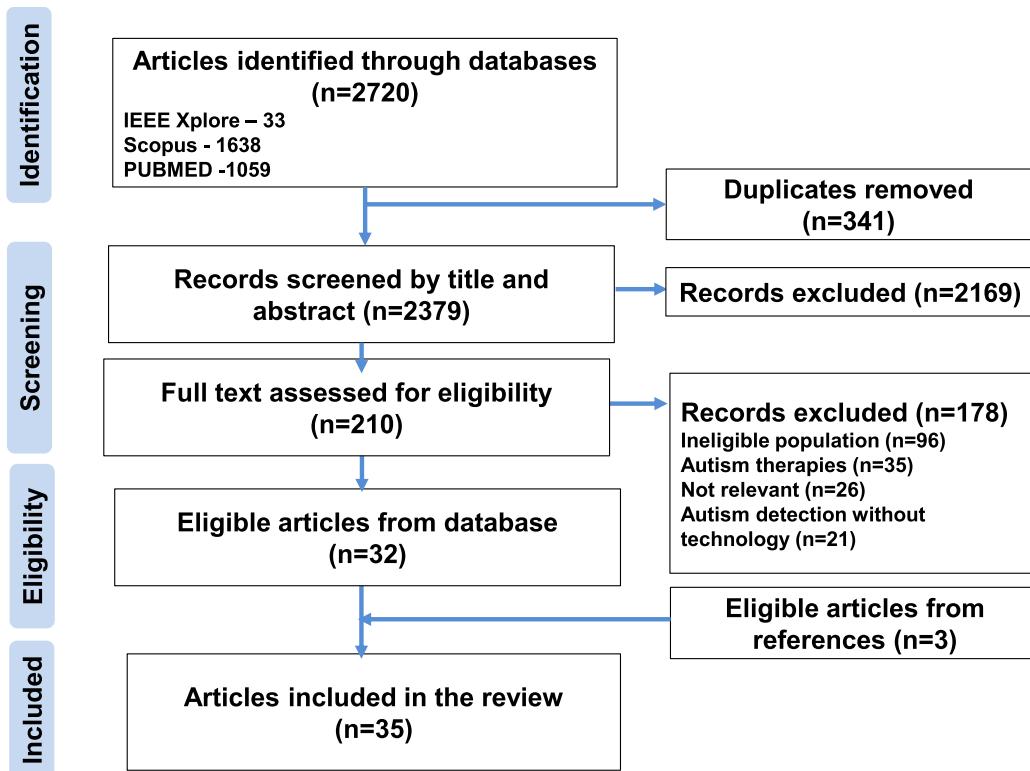


FIGURE 1. PRISMA flow diagram- Literature screening and shortlisting studies.

psychometric outcomes, i.e., sensitivity, specificity, and accuracy. Table 2 summarizes the enrolment counts, software or hardware devices used, assessment tools, assessment duration, limitations, and future direction of each study. In addition, a quality evaluation using the Critical Appraisal Skills Programme (CASP) was performed for each shortlisted study.

The technical terms used in the review are explained in Appendix B in Table 5.

- 1) Study objective, methods, and experiment locations
- 2) Participant's group size and diagnosis status
- 3) Datasets used in the study
- 4) Bio-behavioral markers for data extraction
- 5) Assessment duration, tools, and methods
- 6) List of software, material, or devices used
- 7) Multimodal data collection steps
- 8) Data processing steps
- 9) Technology used in the study, and
- 10) Outcomes, limitations, and future direction

III. RESULTS

This section answers four research questions and presents quality assessment results.

A. QUALITY EVALUATION

Two authors (MK and AKK) undertook the quality evaluation of shortlisted studies using the Critical Appraisal Skills

Programme (CASP) tool [80]. The studies were scored with three possible responses: a) criterion met, b) partially met, or c) not applicable, not met, or not mentioned, with scores of 2,1 and 0, respectively. Table 3 shows implemented rating scales, referring to previous clinical studies [81], [82], to rank studies into high, medium, and moderate categories. The quality evaluation sheet for shortlisted studies is attached in Appendix C section.

The quality evaluation suggested ten studies with moderate quality [46], [48], [49], [50], [52], [53], [56], [57], [63], [76]. While twenty-five studies were classified as high-quality [44], [45], [47], [51], [54], [55], [59], [60], [61], [62], [64], [65], [66], [67], [68], [69], [70], [71], [72], [74], [75], [77], [78], [83], [84]. In general, the quality analysis highlighted the following limitations for most studies: 1) small sample sizes, 2)Unclear study questions, objectives, inclusion and exclusion criteria, 3) Insufficient information on participant sampling and recruitment,4) Imprecise data analysis and outcomes reporting.

In the following sections, four research questions are answered.

B. RQ1 HOW HAS THE LITERATURE ON TECHNOLOGY-BASED ASD DETECTION METHODS EVOLVED?

We respond to the research question by assessing the selected study's 1) temporal publishing, 2) co-authorships, and

TABLE 1. Data input, processing, and outcomes summary from (N=35) shortlisted studies.

Ref	Input Data	Reduced Features	Setting	Data Analysis	Sensitivity	Specificity	Accuracy
[47]	ADOS, ADI-R records	For PQ 17-21 / 157	Home	LR	>0.95	>0.95	-
[48]	Questionnaire responses from three modules	For PQ 13/157	Home, Clinic	GBDT	0.90	0.83	-
[49]	Thirty features from video annotations	8/30	Home	LR	0.94	0.78	0.89
[50]	Features from video annotations	(2-8)/31	Home	LR	0.76	0.58-0.77	0.70-0.76
[51]	Features from video annotations	10	Home	SVM	-	-	FF 0.82, SF 0.83
[52]	Case-similar video selection	-	Home	VIRSA scores (SA)	0.78	0.53	0.59
[53]	Features from video annotations		Home	SA	-	-	0.91
[54]	HM and audio features from videos	4-10	Home	DS; CVA	0.83	0.81	0.765-0.81
[55]	Video recordings from front facing camera	-	Home, Clinic	IntraFace algorithm; Facial Action Coding system	0.70-0.95	-	0.85-0.94
[56]	Fixation data, AOI	-	Clinic	SA	-	-	-
[57]	Fixation data, AOI	-	Clinic	DA	0.86	0.84	0.85
[58]	Fixation data, AOI	-	Clinic	SA	-	-	-
[59]	Fixation data, AOI	-	Clinic	Centrality scores	-	-	-
[60]	Fixation data, AOI	-	Clinic	AdaBoost	-	-	0.65-0.886
[61]	Camera recorded joint attention behavior videos, AOI	1726/273	Clinic	Cohen's D	-	-	-
[62]	Videos of Participants watching movies		Home;Clinic	CVA;SVM	-	-	-
[63]	Video recorded expressions during imitation	7/7	Clinic	CVA;VGG19; Resnet18	-	-	-
[64]	Features from video annotations	-	Home, Clinic	OpenFace;CNN; CE-CLM;HOG;PCA	-	-	-
[65]	Videos of Participants watching movies		Clinic	CVA	-	-	-

TABLE 1. (Continued.) Data input, processing, and outcomes summary from (N=35) shortlisted studies.

[66]	Child Robot interaction videos	Clinic	CVA	-	-	-	
[67]	Video recorded skeletal body movements	-	Clinic	CNN LSTM network	with 0.854	0.765	0.809
[68]	Hand movements recorded via sensors	10/262	Home;Clinic; Classroom	RGF2	0.76-0.83	0.67-0.88	-
[69]	Camera recorded upper Limb movements	7/17	Clinic	SVM	0.967	1	0.938
[70]	Video recorded body movements	1/20 to 2/3	VR room	Open Pose;SVM	-	-	0.69-0.893
[71]	Infrared video recorded with motor movement kinematics	-	Clinic	SVM;PCA	0.826-0.923	0.846-0.86	0.73
[72]	Scanned referral records	150-300 /18962	Clinic	Doc2Vec;SVM	-	-	0.65- 0.83
[73]	Free text followed by a question	-	Home	Regression Tree	-	-	0.65- 0.83
[74]	EMR records	20/89	Clinic	RF	0.243	0.993	0.965
[75]	UC Irvine dataset			MGOA	1	1	1
[76]	Assessment records at 8, 14, 24 and 36 months	-	Clinic	SVM	0.796	0.522	0.575
[77]	Q-CHAT records	-	Clinic	SVM-RFE	0.39-1.0	0.78-1.0	0.66-0.95
[78]	-	10/25	Clinic	PCA;Z score;	0.68-1.0	0.739-1.0	0.752-1.0
[44]	Preprocessed audio clips	54-88	Clinic	AE BLSTM	-	-	0.6818
[45]	Videos annotated for social-situations and speech patterns		Home	LR	0.826	0.643	0.75
[46]	Preprocessed audio crying data	2	Home, Clinic	SVM	0.7142-0.8571	0.8-1.0	0.85-0.928

Misc

AE – Auto encoder; BLSTM – Bidirectional Long Short Term Memory Networks; CE-CLM – Convolutional Experts Constrained Local Model; CNN – Convolutional Neural Network; CVA – Computer Vision Analysis ; DA – Discriminant Analysis; DS – Decision Stump; FF – Face to Face; GBDT – Gradient Boosted Decision Tree; HM – Hand movement; LR – Logistics Regression; HOG – Histogram of Oriented Gradients; LSTM – Long short-term memory; MGOA – Modified Grasshopper Optimization Algorithm; PCA – Principal Component Analysis; PQ – Parental Questionnaire; SA – Statistical Analysis; SF – Still-face; SVM – Support Vector Machine; SVM-RFE – Support Vector Machine-Recursive Feature Elimination; RF – Random forest; RGF2 – Regularized Greedy Forest

3) keywords trends. In addition, we highlight prominent journals where shortlisted articles were published.

1) PUBLICATION TRENDS

The temporal publication patterns suggested that around 80% of the shortlisted studies were published between 2018 and

2021 (Figure 2). Even though the use of technology in ASD management and in general has shown growth since 2011 [85], [86], [87], the review highlight 2018 to 2021 as dominant years in the adoption of Machine Learning (ML) and Deep Learning (DL) technologies. This aberration can be attributed to the following inclusion criteria for shortlisting

TABLE 2. Shortlisted studies (N=35) participants, evaluation duration, hardware, software, limitations and future Directions summary.

Ref	Evaluation Duration	Algorithms	Devices & Psychometric Tools	Participant Count	Limitations	Future Recommendations
[47]	Ten mins	DF;RF;LR	ADOS;ADI-R;M-CHAT-R;SRS-2;CBCL	ASD (N=121);TD (N=12);ODD (N=29)	Individual results not available for questionnaire and video screener	Large sample; Include elderly with physical and psychological conditions
[48]	Twenty five mins	CVA;SVM; DL; GBDT	ADOS;ADI-R;M-CHAT-R;SRS-2;CBCL	ASD (N=272);NOT ASD (N=103)	Compounded results using ML for three modules as probability	Include younger population, calculate ASD severity
[49]	Eight-ten minutes	DT;SVM;LR	MSEL;VABS;AOSI	ASD (N=116);TD (N=46)	Validation performed on children with existing condition	Shortlist relevant ASD diagnostic features using video tagging
[50]	Two-five mins video	LR;DT	ADOS;M-CHAT	ASD (N=50);TD (N=50);SLC (N=50)	Moderate ML model performance on cross-cultural population with SLC	Include ODD condition, create video data-sets for DL and CVA
[51]	FF 2 mins; SF- 1 mins	RF;SVM;NB; Chi-square;T-test	CSBS-DP;CARS; ADI-R; ADOS; GDSQ; ABC; DSM-V	HR (N=45);TD (N=43)	Small sample sizes; The HR and TD groups not age matched	Controlled age-matched groups with large sample size.
[52]	Seven-Eight minutes	Chi-square;T -test	MSEL;ADOS-2	HR (N=76);LR (N=37)	Large number of HR participants may bias community based evaluations	ViRSA as a part of two-stage screener may improve accuracy.
[53]	Sixty- Seventy minutes	Video tagging	-	ASD (N=4);TD (N=1)	Data privacy;Quality challenges when video shot by parents; time-intensive video tagging	Larger sample; Video tagging to be context-specific rather than generic
[54]	Three 3-mins videos	CVA;DS;J48; SVM	CARS;ADI-R	WS (N=32); TD(N=19)	WS rarity limits sample size; blocked camera view restrict video quality	Not-specified ;
[55]	Five minutes	Intra-class correlation coefficient; IntraFace algorithm; Facial Action Coding; LBP	M-CHAT-R/F; ADOS T	ASD (N=15); TD (N=18)	Several obstacles prevented actions coding, for example, missing head turns to a name-call, and poor image quality.	Incorporate hand detection, gaze analysis, eye-tracking, RGB-Depth data, and new emotion classification methods.
[56]	Each image 5-seconds	T-Test; Two-Way ANOVA	Tobii eye-tracker;DSM-IV; ABC;CCMD-III	ASD (N=35);TD (N=17)	Small sample size; Same CI images used for all participants	Include personalized CI images, preference assessment for complex; social and nonsocial stimuli
[57]	Each image 10-seconds	DA;T-Test;Chi-square	LCD monitor;BeGaze software;CARS; DSM-V; GDS	ASD (N=37);TD (N =37)	Small sample size and missing IQ assessment.	Larger sample; ADOS and ADI-R to establish ground truth ASD diagnosis
[58]	Each image 5-seconds	Chi-square test	Mirametrix S2 Eye Tracker; M-CHAT;	Preterm infants (N =31)	Small convenience sample; missing control group; only M-CHAT was used	Longitudinal studies with bigger sample size and control group
[59]	Each image 3-seconds	Centrality scores	SensoMotoric with infrared tech; DSM-IV;ABC;CCMD-III	ASD (N=17);TD (N=23)	Fixation duration of 2-3s to observe the image might be insufficient.	Longer experiment test times
[60]	Five minutes	AdaBoost; T-test; Mann-Whitney U; ANOVA; Cohen's D test	DQ	ASD (N=21); TD (N=31)	Small sample size; Computation based on response time and game performance; Difference between group performance at response's preparation stage is unclear.	Bigger sample; evaluate variations of gaze differences on sex and age group.

TABLE 2. (Continued.) Shortlisted studies (N=35) participants, evaluation duration, hardware, software, limitations and future Directions summary.

[61]	Ten minutes	LMM with restricted maximum likelihood; ANOVA; MSEL	ADI-R; ADOS-2; Tobii TX300	ASD (N=81); TD (N=31)	Small sample; reliance on only community diagnosis; unwarranted accuracy and precision measures of the eye tracker.	Include auditory social attention and bigger sample size
[62]	Five movies; each 4-5 minutes	CVA; Chi-square; SVM; LR	M-CHAT-R/F; ADOS-T; MSEL	ASD (N=22); TD (N=74); ODD (N=8)	Small sample; CVA model trained on adult faces; CG included both typical and ODD	Capture facial expressions for engaged scenarios; CVA testing on larger datasets with ASD, ODD, ADHD, and TD
[63]	-	SVM; CNN; VGG19; Resnet18	-	ASD (N=10); TD (N=10)	Dataset - CK; FER2013, culturally and age-wise different to experiment; Video quality affected by light	Culturally relevant samples can improve accuracy
[64]	Around 2 mins	OpenFace frame-work; PCA	ADOS-2; DSM-IV/V; WSI; GMDS	ASD (N=18); TD (N=15)	Small sample size and scripted social interaction rules with limited responsiveness and engagement intensity in home videos.	Include ODD samples and infants at risk of ASD, a bigger sample size
[65]	Five mins	CVA; GLRM	ADOS-T; MSEL	ASD (N=22); ODD (N=8); TD (N=74)	Small sample size	Larger sample size; Evaluate postural control variations on age, sex, ODD and/or ADHD
[66]	Single session	CVA; RAT; CNN; CLNF; ML Supervised & reinforcement	RGBD sensors; workstations; camera; Robot Ono with 17 DOF;	ASD risk (N=3); TD (N=3)	Small sample size; CVA failed to detect child's responses blocked by hand or hairs	Larger sample; Use CV to analyze phrases and stereotypical movements.
[67]	Around 42 mins	ANN; Spearman rank; SVD	ADOS-2; VABS-2; BEIQ	ASD (N=68); TD (N=68)	Limited features derived from social interaction videos; The sliding window approach is ineffective for diagnosis.	Use of Spatio-temporal attention will enhance the manifestation of other disorders.
[68]	Two- Five minutes	KS test; RGF2	iPad; inertial sensors (tri-axial accelerometer; gyroscope; magnetometer)	ASD (N=35); TD (N=45)	No subgroup analysis with IQ measure; Single center recruitment may induce bias.	Include large sample size and broader evaluation parameters for algorithm generalization.
[69]	-	ANCOVA; Chi-square; FDR; SVM	ADOS; DSM-IV; Optoelectronic system	ASD(N=15); TD (N=15)	Small sample; Not all participants completed ADOS;	Larger sample; Include HF ASD, ODD, females
[70]	Fourteen mins	OpenPose; SVM with LOSO	NVidia GTX1060; ADOS2; ADI-R	ASD (N=24); TD (N=25);	Small sample size, Stimuli order (V-VA-VAO) not randomized	Larger ODD sample with matched CG on age, gender and IQ. Randomization of stimuli sequence
[71]	10 trials; Five minutes each	SVM; PCA; ANOVA	ADOS	ASD (N=13); TD (N=13)	ASD heterogeneity may not have been captured by small sample size.	Include ODD samples to identify atypical development with motor challenges.
[72]	-	SVM; LDA; Word2Vec	OCR software Omni-page	SD (N=56); TD (N=143)	Lack of labelled data; Privacy issues limit open-source data labelling.	Balanced CG and EG dataset; Generate ASD severity index and assess interventions outcome.
[73]	-	Regression tree	M-CHAT-R/F; M-CHAT-R ; ASQ;	HR (N=115);	Unusual implementation sequences of M-CHAT-R/F, ASQ and M-CHAT-R	Randomise M-CHAT-R and ASQ; Include Low-risk sample to validate results.

TABLE 2. (Continued.) Shortlisted studies (N=35) participants, evaluation duration, hardware, software, limitations and future Directions summary.

[74]	LR;ANN;RF	EMR dataset	ASD (N=1397); TD (N=94741)	Data quality; Lack of genetic focus; Lost ASD diagnosis post therapies.	Studies to focus on genetics.
[75]	-	GOA	ASD dataset [79]	-	MGOA falls short on convergence speed.
[76]	-	ML;SVM	MSEL;VABS;AOSI	HR (N=161); LR (N=71)	Limited statistical accuracy to predict ASD at 14 months.
[77]	-	SVM;RF;NB; LR;KNN	ASQ;Q-CHAT; Q-CHAT-10	ASD (N=139); TD (N=126)	Large-scale validation studies needed in multiple settings.
[78]	-	DNN	Q-CHAT; Q-CHAT-10	ASD (N=139); TD (N=126)	High rate of false negatives (FN) and false positives (FP); Small dataset.
[44]	-	AE;SVM; LSTM	ADOS-2;ADI-R;BeDevel-I;BeDevel-P; K-CARS;SCQ;SRS-2;DSM-V	ASD (N=10);TD (N=29); eGEMAPS speech data	Low sample size and performance scores
[45]	Two 5-min videos	Chi-Square; LR; T-score; Log OR; Mixed ANOVA	DSM-IV; CARS; ADOS; MSEL;VABS	ASD (N=23); TD(N=14)	Limited count and quality of video-recordings; lack of non-ASD CG; Vocalization lack variety as not elicited
[46]	300ms-3s	Sound Forge Pro;SVM with RBF	DSM-V;GARS-2	ASD (N=31); TD (N=31); Belalcazar-Bolaños dataset	Limited sample size and ODD participants; time-intensive data preprocessing.

Algorithm AE – Autoencoder; ANN – Artificial Neural Network;BLSTM – Bidirectional Long Short-Term Memory Networks; CLNF – Constrained Local Neural Field; CNN – Convolutional Neural Network; CVA – Computer Vision Analysis;DA – Discriminant Analysis; DF – Decision Forest; DL – Deep Learning; DNN – Deep Neural Network; DS – Decision Stump; DT – Decision Tree; FDR – Fisher Discriminant Ratio ; GBDT – Gradient Boosted Decision Trees; GLRM – Generalized Linear Mixed Regression Model; GLRM – Generalized Linear Mixed Regression Model; KNN – K-Nearest Neighbor; RF – Random Forest; MGOA – Modified Grasshopper Optimization Algorithm; J48 – Tree-based Algorithm; KS – Kolmogorov-Smirnov; LBP – Local binary pattern; LMM – Linear Mixed Model; Log OR – Log Odd Ratio; LOSO (Leave one subject out); LR – Logistics Regression; LSTM – Long Short-Term Memory; ML – Machine Learning; NB – Naïve Bays; NN: Neural Network; OCR – Optical Character Reader; PCA – Principal Component Analysis; RAT – Robot Assisted Therapy; RGF2 – Regularized Greedy Forest; SVD – Singular Value Decomposition; SVM – Support Vector Machine

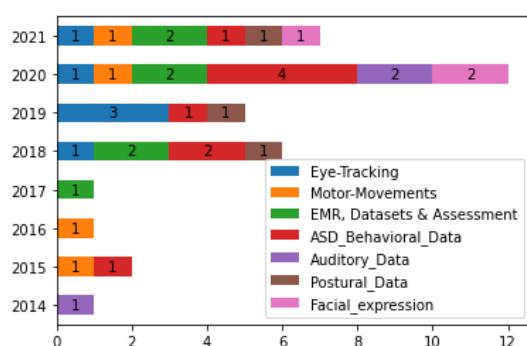
Psychometric properties ABC – Autism Behavioral Checklist; ADI R – Autism Diagnostic Interview-Revised; ADOS T – Autism Diagnostic Observation Schedule for Toddlers; ADOS – Autism Diagnostic Observation Schedule; AOSI – Autism Observation Scale for Infants; ASQ – Autism Spectrum Quotient; BEIQ – Best Estimate Intellectual Quotient; BeDevel-I – Behavior Development Screening for Toddlers Interview; BeDevel-P – Behavior Development Screening for Toddlers Play; CARS – Childhood Autism rating scale; CBCL – Child Behavior

TABLE 2. (Continued.) Shortlisted studies (N=35) participants, evaluation duration, hardware, software, limitations and future Directions summary.

Checklist; CCMD-III 2001 – Chinese Classification and Diagnostic Criteria of Mental Disorders; CG – Control group; CSBS-DP – Communication and Symbolic Behavior Scales Developmental Profile; DSM-IV – Diagnostic Statistical Manual 4; DSM-V – Diagnostic Statistical Manual 5; DQ – Development Quotient; GARS-2 – Gilliam Autism Rating Scale-Second Edition; GDSQ – Gesell Developmental Scale; K-CARS – The Korean version of the Childhood Autism Rating Scale; M-CHAT-R – Modified Checklist for Autism in Toddlers, Revised, M-CHAT-R/F – The Modified Checklist for Autism in Toddlers, Revised with Follow-Up; MSEL – Mullen Scales of Early Learning; SCQ – Social Communication Questionnaire; SRS-2 – Social responsiveness Scale-2; VABS – Vineland Adaptive Behavior Scale; WSI – Wechsler Intelligence Scale; GMDS – Griffiths Mental Development Scale	
Misc	ADHD – Attention deficit hyperactive disorder; CG – Control Group; EG – Experimental Group; FF – Face to Face; HR – High Risk; HF – High Functioning; LR – Low-Risk; ODD – Other Development Delays; SF – Still-Face; SLC – Speech and Language condition; TD – Typically Developing; WS – West Syndrome

TABLE 3. Rating scales and quality evaluation.

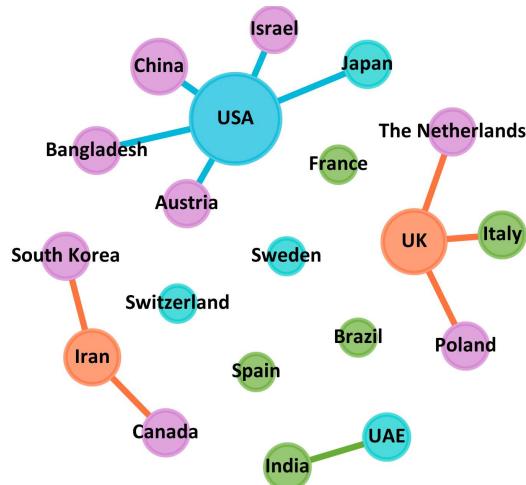
Study type	High quality	Moderate quality	Low quality
Cohort studies	≥ 22	16-21	≤ 15
Case-control or other studies	≥ 20	16-19	≤ 15

**FIGURE 2.** Yearly publication count per Data category.

studies for the review; 1) selecting studies focussing on ASD detection rather than an intervention that has seen higher technological adoption, 2) including only behavioral-based detection methods and excluding EEG, MRI, and genetic methods that have incorporated technology since 2011, 3) selecting studies with participants of less than six years, and 4) skewed temporal adoption of technology in ASD detection.

2) COAUTHORSHIP PATTERNS

The publication pattern shown in Figure 3 depicts the country as a node and its size as the publication frequency from the country's authors. The country node's edge strength indicates collaboration between co-authors from multiple countries. The significant country-level contributions are from the United States of America (USA), whose researchers

**FIGURE 3.** Coauthorship pattern based on author's country of origin and respective ties.

co-authored with researchers from Austria, Bangladesh, China, Japan, and Israel. Authors from Iran, Canada, South Korea, the United Kingdom (UK), the Netherlands, Poland, Italy, the UAE, and India are the other countries with co-authorship collaborations. Authors from Brazil, France, Sweden, Spain, and Switzerland collaborated with other co-authors from the same country. The analysis highlights that most research initiatives and partnerships are from developed economies that have formed partnerships with selected developing economies.

3) KEYWORDS ASSOCIATION

Figure 4 depicts the most important and frequently used keywords in the shortlisted studies. The size of the keyword nodes represents the frequency of occurrences in the shortlisted studies, with the edge weights indicating their simultaneous occurrence in other studies as shown in Figure 4.

Keywords with substantial edge weights and dense connections share a semantic relationship. The most frequently used keywords in research papers are “Autism spectrum disorder,” “child,” “age,” “clinician,” “joint attention,” “behavior,” “time,” “analysis,” “deep neural network,” “td child,” “machine learning,” “development,” “study,” “gaze,” “asd child,” “screening tool,” “feature,” “classifier,” “symptom,” “infants,” and “pattern.” These keywords suggest that most ASD solutions used ML and DL classification methods on multimodal eye-gaze, behavior, and joint-attention data.

4) JOURNAL PUBLICATIONS

The frequency distribution of 35 review articles was as follows: four in the Journal of Autism and Developmental Disorders, three in Scientific Reports, two in the Journal of Medical Internet Research, and the remaining publications were published in different journals. The breadth of studies published in various journals suggests the adaptability and validation of a wide range of technology-based ASD detection innovations, with multi-country authorships and multimodal data types.

C. RQ2 HOW DO RESEARCHERS USE THE VARIOUS BIO-BEHAVIORAL MARKERS TO DETECT ASD?

Each shortlisted study is assigned to one of the seven data categories shown in Figure 5, also referred to as bio behavior. Listed below are the study counts for each data category.

- 1) Stereotypical behavior (Nine Studies)
- 2) Eye gaze (Six Studies)
- 3) Facial expressions (Three Studies)
- 4) Postural analysis (Three Studies)
- 5) Motor control and movements (Four Studies)
- 6) Auditory data (Three Studies) and
- 7) Assessments and electronic health record data (Seven Studies).

We summarize shortlisted studies in seven data categories in the subsections below.

1) STEREOTYPED ASD BEHAVIORS

In this review section, nine studies [47], [48], [49], [50], [51], [52], [53], [54], [55] extracted and classified the ASD deterministic behaviors such as tantrums, self-stimulatory and injurious behavior, non-compliance, the objects lining, and poor communication, eye contact, or social skills from videos to perform ASD detection.

[47] developed ML models through a two-stage process: (1) feature selection and (2) ASD and TD classification. They trained ML models using historical ADI-R and ADOS-2 records, shortlisted 20 critical features using the DF (Decision Forest) algorithm, and incorporated them in the parental questionnaire (PQ) and annotation-based video-tagging module. In the second stage, researchers integrated responses of both modules applying L2-regularized logistic regression (LR) [88], whose psychometric outcomes outperformed those of M-CHAT, CBCL (Child Behavior Checklist), standalone questionnaire, and video modules. [48] enhanced their previous work by introducing a third

clinician questionnaire module. The three-module screener implemented in 8-10 minutes outperformed earlier psychometric outcomes using the GBDT (Gradient Boosted Decision Tree) algorithm.

[49] collected one to five-minute home-based videos rated by non-experts generating a feature set analyzed by eight ML classifiers previously trained on ADI-R and ADOS datasets. All classifiers had a sensitivity above 0.945, but only three had a specificity above 0.5. The LR5 (LR model with five shortlisted features) outperformed other ML models. [50] validated their previous work [49] on Bangladeshi children, including those with SLC (speech-language conditions). Non-expert US raters, after one-hour training, reviewed videos and responded to 31 multiple-choice questions, generating a feature set from the responses. The LR with Elastic Net penalty [89] and LR5 were the best performing ML models on the feature set with sensitivity, specificity, AUC, and accuracy for ASD vs. TD as 0.76, 0.58, 0.76, and 0.70 and ASD vs. ODD as 0.76, 0.77, 0.85 and 0.76 respectively.

[51] video-recorded mother and child social interactions of HR (High-Risk) toddlers aged 9-12 months in three social situations. 1) Face-to-Face (FF) mother-child interactions, 2) mother’s unresponsive Still-face (SF), followed by 3) usual mother and child interactions. The SVM classification model outperformed NB (Naive Bayes) and RF (Random forest) in the ASD detection and classification.

Further, [52] developed a Video-referenced Infant Rating System for Autism (VIRSA). The system algorithm proposed a series of parent-infant interactive age-matched videos, with parents choosing the most appropriate ones matching their child, resulting in a score computation. At ages 6, 9, 12, and 18 months, children were clinically examined, diagnosed, and rated on the VIRSA. The statistical analysis of VIRSA scores predicted 100% ASD in children at 18 months and 78% at 36 months compared to diagnostic established using gold-standard tools. This study is a first step towards creating a novel video-based online rating system for detecting ASD in children with robust psychometric properties.

[53] developed a smartphone application, NODAsmart-Capture empowering parents to record home videos of child’s behavior and label social dialogue, play, and problematic behavior in four social scenarios. Diagnosticians annotated the videos with built-in tags designed on DSM criteria such as “no eye contact” or “repetitive play,” matching ninety-one percent of their recommendations with the ground truth diagnosis recorded at study enrolment.

West syndrome (WS) disorder [90], diagnosed in 0.06% of infants and children, is characterized by epileptic spasms, often leading to mental impairment in children. [54] implemented ML to predict the onset of ASD/ID (Intellectual disability) in high-risk 9–12-months with WS. Researchers captured three video-recorded social engagement scenarios, and out of SVM, J48, and RF [91], the DS (Decision stump) [92] algorithm predicted WS vs. TD with 0.765 and WS+ vs. WS- with 0.812 accuracy using multimodal audio and video data.

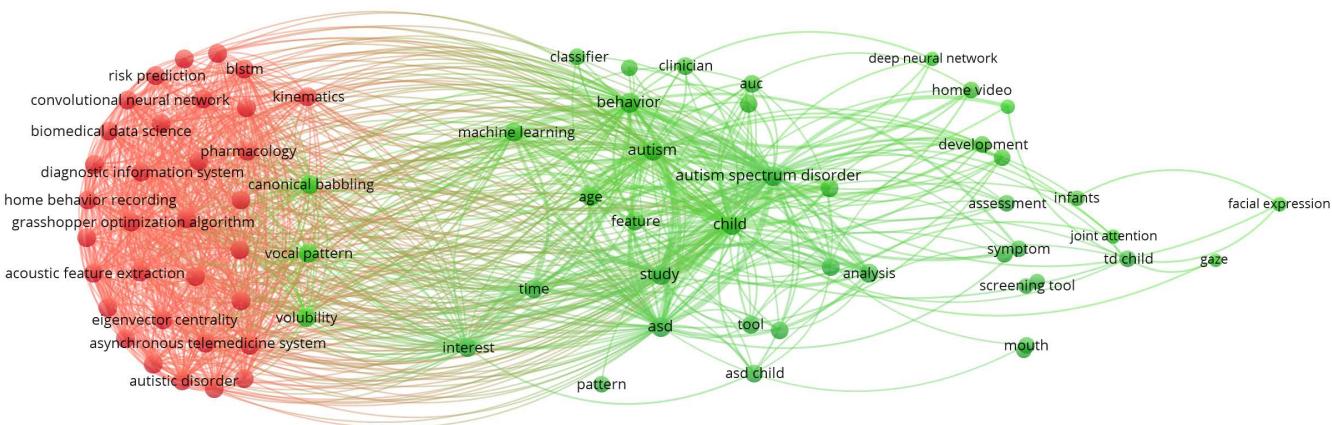


FIGURE 4. Shortlisted studies (N=35) keywords association.

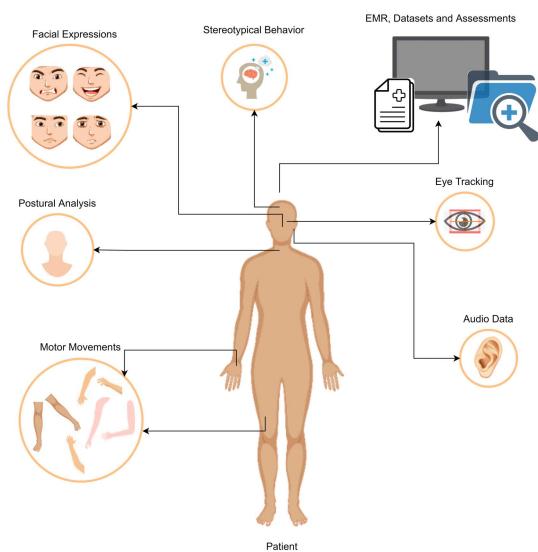


FIGURE 5. Studies distribution -Seven multimodal data categories.

Based on a movie stimuli [55] elicited and engaged the child's attention, video recorded behavioral and social reactions of children. They analyzed the scenes using computer vision to decipher children's emotional, behavioral codings, and head positions and classified ASD children with 85-95% accuracy.

2) EYE-TRACKING

Eye-tracking is a non-invasive method for examining an individual's attention and mental processing abilities, which serve as proxies for cognitive and neurological functioning.

In this review, six studies, [56], [57], [58], [59], [60], [61] used eye-tracking and gaze analysis to measure fixation frequency, duration, and AOI (Area of Interest) responses from children's gaze towards social and nonsocial stimuli

in images and videos. The studies hypothesized that children with ASD prefer circumscribed interests (CIs) [93], preferring specific animated characters, toys, or activities. Researchers use the gaze preference of children on the content of images or AOI to make ASD vs. TD classification

In the experiment by [56], TD and ASD groups observed six scenic images with social (e.g., people) or without social cues (e.g., bowl). The researchers extended the experiment with twelve images, half with CI (e.g., a toy car) and another half without CI (e.g., a plant). Within-subjects CI and non-CI eye-gaze data for ASD and TD groups using T-tests suggested poor social attention processing abilities for the ASD group.

The study [57] recruited children from ASD and TD groups who were similar in age and gender. For 10 seconds, participants observed a female speak the English alphabet, and their fixation data on various facial and body areas were collected. They applied DA (Discriminant analysis) to mouth and body AOI fixation data and classified ASD and TD children.

[58] studied six-month-old preterm children's gaze and fixation on social figures, suggesting that children preferred looking at the eyes or lips of social figures over nonsocial images. However, at 18 months, each subject tested negative for ASD when evaluated on M-CHAT and without CG (control group) presence; the results provided weak evidence to detect and classify ASD and ODD.

[59] recorded participants' eye movements while viewing eleven photographs and constructed a virtual network graph using temporal gaze patterns and fixation time on the seven AOI on the human face. Betweenness centrality at four face features, under the right and left eye, left eye, and mouth, was lower in ASD children than TD children by 27, 53, 42, and 61%, respectively, forming a basis of ASD detection.

[60] captured the gaze modulation of children with ASD and TD children using an eye tracker as they played a variant of the Go/No-Go game. AdaBoost's meta-learning algorithm could distinguish ASD and non-ASD participants with an accuracy of 88.6% based on gaze patterns.

[61] evaluated if an impaired response to joint attention (RJA) in infancy is a critical ASD marker. The infant eye gaze was recorded in a 10-minute session of several IJA (Initiation to joint attention) tasks. Since newborns utilize their gaze for RJA and IJA, this method can be used to quantify children's social cognition milestones at an early development age of 10–18 months.

3) FACIAL EXPRESSIONS

Children with ASD struggle to produce and perceive facial expressions that express a range of emotions and display affection [94], impacting their social functions. Deep learning (DL) models can identify facial expressions from images or videos in three steps, 1) preprocessing the image or videos, 2) extracting facial expression features, and 3) classifying the extracted features to various emotions.

In this review section, two studies [62], [63] used publicly available facial expressions datasets to train the DL models and extracted and analyzed facial expressions from EG (experiment group) and CG (control group) to make an ASD diagnosis.

[62] trained CVA model on Binghamton University 3D Facial Expression database [95] to extract facial landmarks that SVM classified into positive, neutral, and other categories. They observed that children with ASD had more neutral expressions than children without ASD. The AUROC with age-covariates ranged between 0.75 to 0.83 for five movies that children with ASD and TD watched.

Imitation of facial expressions is a critical measure of social interaction skills. Studies demonstrate that children with ASD on prompted stimuli usually perform imitation slower than TD children [96]. [63] trained the DL model to recognize facial expressions using FER2013 [97], CK databases [98], and augmented the model learning with sixteen Chinese children's facial expressions. The participants imitated seven facial expressions, and their responses were video-recorded. For the ASD group, average expression imitation was lower than 60%, compared with TD, a critical ASD deterministic threshold.

[64] studied facial expressions using the Facial Action Coding System (FACS). An OpenFace software extracted the subtle dynamics of social smiles of ASD and TD children from their home recordings. The results suggested that ASD children display happy facial expressions less intensely than their TD counterparts during the first year of life.

4) POSTURAL AND HEAD MOVEMENT DATA

Children with ASD demonstrate a diminished capacity for postural stability [99] and functional balance [100].

Two studies, [65], [66], used CVA (computer vision analytics) from recorded videos to measure head postural control in study participants to distinguish ASD and TD groups.

[65] induced social and nonsocial stimuli by asking study participants to watch five movies comprising animated and complex characters and recorded participant's rate of head movements using CVA. After adjusting for age, ethnic

origin/race, and sex, the ASD group had a faster head movement rate in four of five movies with complex stimuli. By removing the ODD (other developmental delays) group from the non-ASD group, the 95% CI level adjusted rate ratios to distinguish ASD vs. TD were significant.

Reinforcement learning is a subfield of AI (Artificial Intelligence) that guides intelligent entities' behavior based on a reward-based environment [101]. [66] in multiple stimuli, single Child-Robot Interaction (CRI) session measured head postures, joint-attention, and eye-gaze data [102] using RGBD sensors and cameras. They used CNN (Convolutional Neural Network), CVA, and CLNF (Constrained Local Neural Field), differentiating TD and ASD children. The TD group had good adherence to IJA (Initiation of Joint Attention) and RJA (Responding to Joint Attention) with the therapist and robot than the ASD group. However, the children with ASD displayed higher comfort and engagement with robots and a high IJA towards the therapist during the transition.

In addition, [67] developed and validated a deep neural network (CNN-LSTM architecture) trained on the non-verbal aspects of social interaction from video recordings captured during ADOS-2 assessments that distinguished ASD and TD peers with an accuracy of 80.9%.

5) MOTOR MOVEMENTS

Children with ASD have varying degrees of fine and gross motor skills. Gross motor deficits in children with ASD can impair body balance and make it challenging to participate in sports or do daily tasks [103]. Difficulties with fine motor skills might limit participation in activities that demand hand muscle movements [104]. In this review section, we covered four studies [68], [69], [70], [71] that used motor data to classify ASD children.

[68] used a smart tablet with touch-sensitive screens and inertial movement sensors to capture the study participant's contact impact data patterns while playing games. They applied the Kolmogorov-Smirnov test (KST) [105] on the sensor dataset, shortlisting the ten most significant features from 262 and classified ASD vs. TD using RGF2 (Regularized Greedy Forest) [106] algorithm computing AUC, sensitivity, and specificity scores. [69] analyzed participant's upper-limb movements in a reach-to-drop task exercise. The participant reached the ball, placed it in the support, and transferred it to the target box hole. They shortlisted seven discriminating features out of seventeen using Fisher discriminant ratio (FDR) [107] for both EG and gender and mental age-matched CG. They used the SVM algorithm to identify ASD children using seven features. [70] on three real-world virtual reality imitation tasks collected participant's body movements in response to visual, auditory, and olfactory stimuli. They identified joint motions using the DL (Deep Learning) OpenPose and shortlisted critical and extensive body part movements using PCA (Principal component analysis) [108] to detect children with ASD. The SVM algorithm classified ASD children with an accuracy of 0.893 using

five joint movements (head, trunk, arms, legs, and feet) in response to visual stimuli. Inter-joint coordination and motor synergies [71] can be potential substrates of ASD markers. Researchers asked ASD and TD participants to engage in a motor task behavior by manipulating a felt-tip pen to draw on a sheet of paper. At the same time, an optoelectronic motion capture system recorded their movement kinematics that was analyzed by the SVM algorithm to classify ASD and TD participants with a 94.7 percent accuracy. The analysis implies that an ecologically valid autism motor signature can predict ASD risk in children.

6) ASSESSMENTS, DATASETS, AND EMR ANALYSIS

We discussed seven papers in this section. Two studies [72], [73] incorporated natural language processing (NLP), [74] used Electronic Medical Records (EMR). Another two publications analyzed Q-CHAT [77], [78], one study VABS (Vineland Adaptive Behavior Scales) [76] and [75] used customized ASD assessment dataset to classify ASD and TD children.

Word2Vec algorithms [109] convert words to vectors, evaluate similarities, and group words logically, allowing the processing of sizeable unstructured text repositories. In addition, LDA (Latent Dirichlet Allocation) [110] uses a prior Dirichlet distribution [111] matching word distributions with logical topics. Combining LDA and Word2Vec, both parts of NLP can generate discriminative features for a topic based on contextual associations.

[72] analyzed unstructured ASD evaluation referrals by scanning, preprocessing, physical records, and reading through OCR (Optical character reader). The dataset was upsampled [112] by adding two simulated positive samples for each positive case and feature reduced using L1 and L2 regularizations [88] using SVM. Word2Vec predicted ASD risk with precision, recall, and F2 scores of 0.646, 0. 911, and 0.842, respectively, outperforming LDA.

[73] predicted ASD risk by asking families of HR children to state social-communication developmental concerns in a sentence. A regression tree algorithm analyzed the textual responses that either suggested ASD risk or presented an additional M-CHAT-R [7] or ASQ [113] question and, after processing, suggested ASD risk. The ML model AUC with text-only analysis ranged between 0.36 to 0.54, and for text and with M-CHAT-R [7] questionnaire between 0.74 to 0.88.

The EMR [114] is usually implemented in clinicians' offices, clinics, and hospitals to capture notes, assessments, and treatment records cross-sectionally and longitudinally for diagnosis and treatment. [74] extracted 89 features from longitudinal retrospective EMR data and shortlisted 20 features using RF Gini impurity [115] scores. They used SMOTE [115] to upsample and overcome the class imbalance in the ASD dataset. The LR predicted ASD risk with an AUC of 0.727. Researchers obtained ground truth labels for patients (ASD or non-ASD) in the studies [116] from the clinical reports.

In addition to analyzing and classifying multimodal data, few studies focused on enhancing ML performance. [75] used Grasshopper Optimization Algorithm (GOA) [117] on three datasets [79] and predicted ASD with near 100% accuracy.

[76] assessed HR and low-risk infants at eight, fourteen, twenty-four, and thirty-six months. The best ML classifier was SVM (AUC of 0.713) trained on VABS [118] daily living module [119] records that were captured at 14 months, normalized and z-scored. [77] used ML to investigate the Q-CHAT [120] assessment records to distinguish between ASD and non-ASD children. Of five ML algorithms: RF, NB, SVM, LR, and KNN, the SVM achieved the highest accuracy of 95%. [78] used the Q-CHAT and Q-CHAT-10 (Q-CHAT with ten features) datasets to develop two 5-layer DNNs to detect children with ASD. They compared the performance of both the models and observed that the Q-CHAT-10 model reported higher AUROC, sensitivity, and specificity than the outcome of SVM and DNN algorithms processed on Q-CHAT data [78]. The findings confirm the role of ML models in reducing the assessment features and predicting an ASD condition.

7) AUDIO DATA

DL models can identify distinctive vocal patterns by analyzing the production of canonical syllables and speech volubility [121]. Canonical syllables [122] have a consonant and a vowel-like component that emerge by the second half-year of life and not later than ten months in TD children. Volubility refers to syllable production frequency and is usually limited in children with ASD [123]. In the review, three studies analyzed audio data; two used syllable production, speech patterns, and canonical babbling [44], [45], and the third used crying patterns [46] to detect ASD.

[44] used a pre-trained feature extraction auto-encoder integrated with a joint optimization method, and trained four ML models on eGeMAPS (Geneva minimalistic acoustic parameter set) dataset [124]. The ML models: SVM, BLSTM (88 features) [125], BLSTM (54 features), and optimized AE BLSTM were tested on 95 ASD and 130 TD utterances across five vocalizations categories: syllables, canonical babbling, calling mother or father, screaming and crying. The BLSTM AE model outperformed other ML models with precision, recall, and F1 scores of 0.4526, 0.6869, and 0.5457.

[45] conducted a retrospective study examining the vocalizations of 37 infants from two 5-mins videos in the 9–12 months and the 15–18 months age range; that included family play, vacations, and familiar routines (e.g., mealtimes). The video recordings were annotated on canonical babbling, syllables production, and speech volubility features. The LR model trained on the canonical babbling ASD features was the strongest predictor to classify 90% of ASD and 63% of TD infants at 9–12 months. Further, Log odds ratios (log OR) confirmed that TD infants reached the canonical babbling [123] stage earlier than other infants who were later diagnosed with ASD.

[46] for ten ASD and TD children collected crying samples (300 ms to 3-sec clips), preprocessed and cleansed them by removing screaming, babbling, or vocalizations instances with a closed or non-empty mouth. They used phonation and vocal quality features from Belalcazar-Bolaños dataset [126] created from audios of Parkinson's patients. To minimize misclassification, they used a novel SubSet Instance (SSI) method using unsupervised and supervised methods. They shortlisted two discriminative speech features, i.e., an MFCC and SONE coefficient, to measure tone's timbre and loudness with temporal difference variance to form a basis to screen children with ASD.

D. RQ3 WHAT DEMOGRAPHIC CATEGORIES, DATABASES, CONTROLS, COMPARATORS, AND ASSESSMENT INSTRUMENTS ARE A PART OF THE TECHNOLOGY-FACILITATED ASD DETECTION PROCESS?

This section identifies various participant counts, datasets, experiment and control groups, assessment instruments, locations, and durations.

1) PARTICIPANT COUNTS

The study participant counts are reported in Table 2 in the column 'Participant Count'. The majority of studies reported limited Participant enrollment in the study. There are only three studies that reported greater than 250 participants [74], [77], [78], six studies between 150-250 [47], [48], [49], [50], [72], [76] and another six studies between 100-150 participants [52], [61], [62], [65], [67], [73]. Seven studies [46], [51], [54], [56], [57], [60], [68] reported enrolment between 50-100 participants and nine studies [44], [45], [55], [58], [59], [64], [69], [70], [71] represented between 10-50 enrollments. Two studies reported less than 10 participants [53], [66]. The remaining studies reported using datasets for analysis.

2) DATASETS

Large-scale datasets give researchers the motivation and necessary sample size to develop, collaborate, and benchmark the performance of ML and DL algorithms. The studies reported use of datasets in the audio [126], [127], assessments [79], facial expression [95], [97], [98], and EMR category [74]. The studies reported challenges such as data preprocessing, cleaning, and augmenting the audio and EMR datasets as they were neither age-matched [62] nor culturally relevant to the experiment data [63]. Additional datasets are listed in Appendix A, allowing researchers to collaborate and develop ASD detection innovations and improve the current ASD detection process.

3) CONTROLS AND COMPARATORS

While the majority of studies focused on categorizing children with ASD and TD, a few studies included children with ODD as well. For example, [54] performed classification between WS and TD and WS+ vs. WS-. In addition [46],

[47], [48], [49], [50], [52], [54], [62], [62], [65], [65], [68], [76] included HR, ASD and ODD children in the control group to perform classification tasks.

4) ASSESSMENT TOOLS

Out of the seventeen different psychometric tools, six of the most widely used in the review were ADOS, ADI-R, M-CHAT-R, MSEL, CARS, and DSM. The ADOS, ADI-R, CARS-2, and DSM are gold-standard ASD diagnostic tools. The outcomes of these tools are matched with the outcomes of technology-based tools to calculate psychometric properties. The MSEL measures children's cognitive development and ensures that the controls and comparators recruited in the study are age and IQ matched.

5) ASSESSMENT DURATION AND LOCATION

The assessment duration is reported in Table 2 in the column 'Evaluation duration'. Most assessments lasted less than 10 minutes. In addition, studies capturing bio-behavioral data involved parents and non-experts to perform annotations were conducted in-home setting [47], [48], [49], [50], [51], [52], [53], [54], [73]. Studies that used eye-contact, postural measures and facial expressions required extensive set up of sensors and cameras and were conducted in clinic or hospitals [44], [45], [56], [57], [58], [59], [60], [61], [63], [65], [66], [67], [69], [71], [72], [74], [76], [77], [78]. Few studies that captured motor, behavioral and postural data captured information in home and clinic settings [46], [48], [55], [62], [64], [68]. Studies that used virtual reality framework to capture motor data required a dedicated VR room [70].

E. RQ4 HOW HAVE RESEARCHERS GATHERED AND PROCESSED MULTIMODAL DATA? HOW DO TECHNOLOGICAL INNOVATION'S RESULTS COMPARE TO CONVENTIONAL ASD DETECTION METHODS?

This section list various data collection methods, shortlisted ASD markers post feature reduction, the performance of ML and DL models, and their psychometric outcomes.

1) DATA COLLECTION METHODS

The review selected a variety of approaches to capture unstructured data. Eye trackers such as Tobii [56], [61], BeGaze software [57], SensoMotoric with infrared tech sensors [59] and Mirametrix S2 Eye Tracker [58] captured eye-gaze data. The inertial sensors, tri-axial accelerometers, and gyroscopes captured motor motions [68], [69], [70] data.

Further, cameras and RGBD sensors; captured videos that were analyzed using CVA to classify facial expressions [62], [63], [64], postural and head movements [55], [65], [66], and socio-emotional behaviors and head positions [55]. In addition, cameras recorded ASD-specific behavior markers [47], [48], [49], [50], [51], [52], [53], [54] and the video frames were manually annotated to generate feature sets on which ML and DL models were trained. The ML models were trained on structured assessments such as VABS [76],

Q-CHAT [77], [78], ADOS-2 [47], [49], and ADI-R [47], [49]. The scanned referrals [72] were processed using OCR (optical character readers) and classified ASD and TD cases by training ML models with ground truth diagnosis as labels.

2) SHORTLISTED ASD MARKERS

Numerous studies incorporated feature reduction methods, marked in column “reduced features” in Table 1, and short-listed critical ASD deterministic landmarks. Researchers trained ML models on these features to perform ASD vs. TD classification using a supervised learning method shown in figure 6. For example, [72] applied feature reduction on the scanned referral records and shortlisted behavioral patterns such as vocal vowel sounds and mood swings as critical ASD deterministic markers. Additionally, [74] shortlisted parental age, medication use, treatment, and dietary patterns as significant predictors of ASD.

Further, [63] highlighted that children with ASD can comprehend and imitate facial expressions such as happiness and sadness but struggle with complicated facial expressions such as neutrality, aversion, disgust, and surprise. [62] reported that non-ASD children, while watching movies, often raised eyebrows and an open mouth, a characteristic of normal development and a feature not displayed by ASD children. The social communication deficit is a critical marker for ASD. [49], [51], [67] highlighted speech patterns, communicative engagement, language understanding, emotional expression, sensory seeking, responsive social smile, and stereotyped speech as critical markers for ASD. Further, [50], [53], [67], [76] highlighted the child’s stereotyped behaviors, repetitive interests, and poor eye contact as important markers for ASD risk determination. In addition, [55] suggested name-call responses and emotional state analysis as an enabler for early ASD warning flags in children. [51], [52], [53] emphasized shorter duration and lower frequencies of eye contact, lack of social smiling, and poor social engagement as ASD risk markers. However, [76] revealed that poor eye contact and repetitive hand movements alone did not accurately diagnose ASD. Individual behaviors such as daily living skills impairments and compliance within the household must be considered in conjunction with other behaviors to suggest predictive accuracy of ASD. Further, [54] used PCA to identify stereotypical hand motions (HM), mother-child communication exchange, and speech analysis as essential behavioral and auditory markers for ASD among children with WS. Thus, ML models can analyze facial expressions, gestural patterns, stereotypical behavior, and communication exchanges to predict ASD risk with high confidence.

While measuring joint attention skills, [61] reported that infants later diagnosed with ASD exhibit considerable atypical JJA but not RJA. In addition, the prevalence of atypical nonverbal behaviors manifested by displaying uncommon, limited gestural postures decoupled from visual contact, facial affect, and speech in ASD children [67] can lead to ASD identification.

Atypical motor movements can predict the risk for ASD. In an experiment, [68], researchers observed that while playing tablet games, gesture velocity was more significant in the ASD group, while the time to tap a screen was shorter than in the control group. In another study, a ball drop task [69] indicated an improper wrist angle position, hand inclination, and slower, fragmented movement as critical criteria for ASD and TD classification. Similar findings were reported by [71] in a reaching-grasping paradigm in which children with ASD displayed decreased coupling between DoF (degree of freedom), which correlated with the severity of their socio-communicative symptoms. During a virtual reality, motor movement task [70] could classify ASD and TD groups with 82.98% using only head movements, 74.47%, and 72.34% accurately using arms and legs movements, respectively. The findings corroborate the literature suggesting that head spinning and banging, body rocking, and foot-stomping are three major stereotypes and repetitive motions associated with ASD.

[77], [78] findings indicated that ML algorithms could detect ASD with an accuracy greater than 90 percent from a selection of 14 feature items and greater than 80 percent using only three items of Q-CHAT. In addition, VABS (Vineland Adaptive Behavior Scale) [118] daily living normalized z-scored [119] assessment scores at 14 months reported AUC of 0.713 [76] for ASD detection.

A study by [60] using the eye gaze reported that ASD children exhibited more unstable gaze modulation and demonstrated significantly shorter initial, average, and total fixation durations for social stimuli [56]. Further, [57] suggested that children with ASD show reduced fixation time at the eyes, mouth, and nose, affirming the critical role of fixation on the eyes in detecting autism via eye-tracking. However, findings of [58] suggested quite the opposite, as preterm children preferred to glance at the eyes or lips of social images or people. Therefore the ability to process social cues by analyzing the fixation duration at various body parts can predict the severity of ASD in children. [45] presented the vocal analysis of the children and confirmed that at 9–12 months, TD infants reached the canonical babbling [123] stage earlier than other infants later diagnosed with ASD. They further confirmed that infants diagnosed later with ASD produced fewer words per minute than those diagnosed with TD. Therefore canonical ability and syllables production in younger years can confirm the risk of ASD.

3) MULTIMODAL DATA PROCESSING

The research utilized seventeen ML algorithms listed in the column “Algorithms” of Table 2. Decision trees, random forests, and support vector machines were the most often used machine learning models. CNN algorithm is utilized approximately 80% of the time when deep learning methods are employed. The review employed six statistical methods, with the ANOVA, T-test, and Chi-squared test being the most often utilized. Ensemble decision trees [128] performed the best on structured data generated from the video annotation of

ASD-relevant behaviors. In eye-tracking, statistical and discriminant analysis were the most effective algorithms. CVA ranked highest in analyzing unstructured facial expressions and postural and head movement data. Additionally, SVM scored admirably in the structured feature reduced data captured from the motor movements. MGOA and word2vector algorithms outperformed all other algorithms in Assessments, Datasets, and EMR Analysis. Finally, BLSTM (Bidirectional LSTM), AE, and SVM effectively classified audio data to detect ASD conditions.

4) MACHINE LEARNING VS. DEEP LEARNING

A ML model trained on multimodal data can classify ASD and TD children at the current state of the art. However, DL outperformed ML methods in feature extraction and classification tasks on unstructured data. For instance, researchers captured features of interest for ASD classification using DL from facial expressions [62], [63], postural and head movements [65], [66], [67], text analysis using NLP [72], [74], [78], motor movements [68], [69], [70], and audio recordings [44], [45], [46] and incorporated supervised learning techniques as shown in figure 6 to classify ASD children. As a result, a conclusive DL model trained on multimodal data sourced from one or more of the seven categories can make the ASD diagnosis procedure efficient.

5) PSYCHOMETRIC OUTCOMES

The robustness of the technological solutions can be measured on psychometric properties such as sensitivity, specificity, and accuracy listed in Table 1. The following eight studies [46], [47], [48], [49], [53], [69], [74], [75] reported psychometric properties of greater than 0.9 on any one of the sensitivity, specificity, and accuracy measures. Seventeen studies [45], [50], [51], [52], [54], [55], [57], [60], [67], [68], [70], [71], [72], [73], [76], [77], [78] reported psychometric outcomes between 0.7 to 0.9, one study [44] less than 0.7 and ten studies [56], [58], [59], [61], [62], [63], [64], [65], [66] did not report any outcomes.

IV. DISCUSSION

The scoping review shortlisted 35 studies after eligibility and inclusion-exclusion assessments. The review analyzes technology's viability, application, division, and outcomes for the following seven bio-behaviors; (a) Stereotypical behavior; (b) facial expressions; (c) eye gaze; (d) motor movements; (e) postural analysis; (f) assessments and EHR datasets; and (g) auditory data. The review data summary is populated in table Table 1 which includes multimodal input data, feature reduction steps, environment setting, data processing algorithms, and psychometric outcomes, i.e., sensitivity, specificity, and accuracy. Table 2 lists enrolment counts, software or hardware devices used, assessment tools, assessment duration, limitations, and future directions. The review uses table data and answers four research questions on technology usability, multimodal data capture, data analysis, quality evaluation, limitations, and strengths.

The review contributes to the literature by

- 1) Shortlisting various multimodal bio-behavioral markers for ASD detection;
- 2) Analyzing automatic multimodal data extraction, feature optimization, and data processing methods;
- 3) Highlighting psychometric outcomes from technological innovations and comparing them with traditional methods, and
- 4) Identifying relevant datasets for researchers to collaborate and co-create ASD and ODD detection innovations bringing efficacy to the detection process. The review highlights that ASD detection ML and DL methods can be applied to identify children at risk of ODD, including speech and developmental delays and hyperactive challenges. Researchers can shortlist specific feature sets for each condition and train machine learning models with statistically significant data volume. The outcomes of the machine learning models can be measured based on psychometric properties calculated by comparing predicted diagnoses with gold-standard tools.

The subsections below detail the role of technology in ASD detection and internal and external validity threats.

A. PROMISING ROLE OF TECHNOLOGY IN ASD DETECTION

The analysis highlights an upward trend in adopting technology-based ASD detection solutions during 2018-2021 attributed to multiple factors.

- 1) The demand for low-cost diagnoses [129], universal screening [130], and the availability of research funding [131] have promoted research initiatives to develop technology-based ASD screening innovations.
- 2) The high penetration of mobile devices, low-cost cameras, and Micro-Electro-Mechanical System (MEMS) sensors such as accelerometers and gyroscopes [132] have enabled the real-time capturing of vast volumes of structured and unstructured data. In clinical situations, cameras and sensors are more practical, less expensive, and less invasive technologies than fMRI and EEG.
- 3) With technology maturing, the generation and multimodal data processing is automated by researchers by building data pipelines on a low-cost cloud infrastructure. Integrating data pipelines with technologies such as AI, ML, and DL has expedited the development of cost-effective and superior detection and on-risk identification of ASD and ODD population.

However, the traditional ASD diagnostic services are not always accessible, affordable, or data-driven [27], [133]. The review findings suggest that technology-based ASD methods can be extrapolated to the ODD population and can effectively, efficiently, rapidly, and potentially serve larger population groups with improved quality, access, and affordability [27].

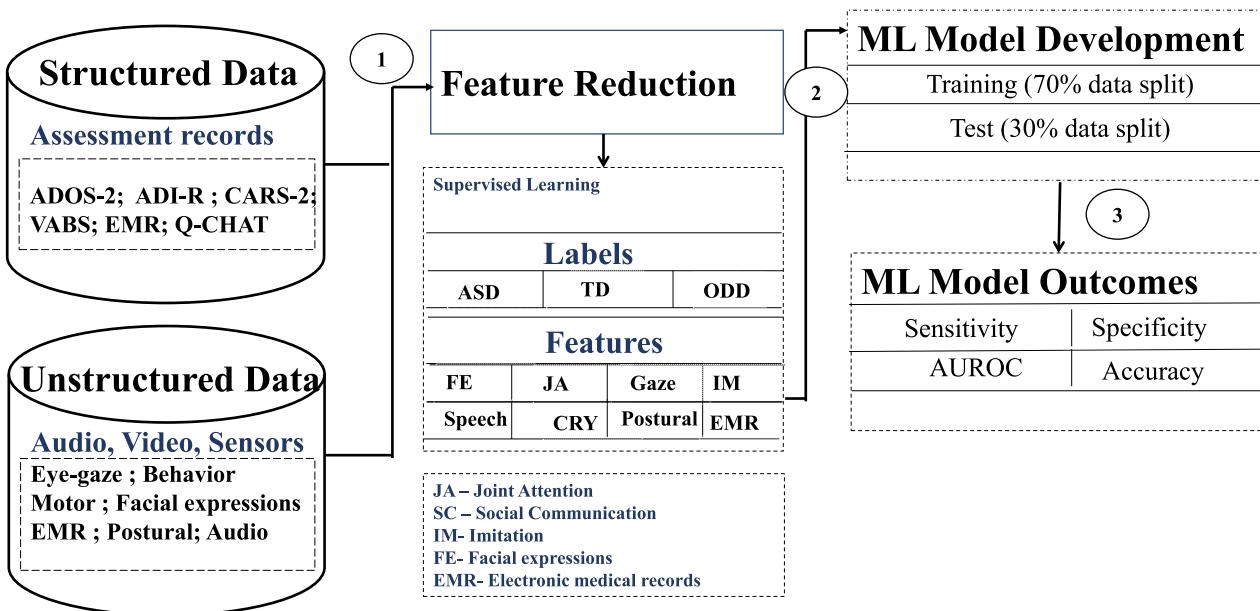


FIGURE 6. Process flow of Supervised Machine Learning methods.

Further, the technology-facilitated innovations are expected to supplement traditional detection methods because of the following reasons:

- 1) Diagnostic methods based on ML and DL can be trained on a large volume of involuntary generated multimodal data from various bio-behaviors to detect children with ASD and ODD risk.
- 2) Traditional ASD screening methods can misdiagnose children with borderline ASD or with speech delay or ODD as ASD. These limitations can be overcome using technological innovations such as an inconclusive ML classifier developed by [47] trained solely on misclassified data instances. The method reduces misdiagnosis of comorbid conditions with an implementation time of under ten minutes by suggesting borderline or ODD instances into an inconclusive class and recommending users for further evaluation by a clinician. Thus, ML technologies can potentially alleviate the misdiagnosis of detecting comorbid, ASD, or ODD borderline conditions such as speech and developmental delays with increased accuracy.
- 3) A few gold-standard tools, such as CARS-2, can diagnose children only beyond two years. Also, children's social communication, language, and other critical milestones do not develop until the second or third year of life. Therefore, evaluating ASD risk in children under two years can give conflicting results by an inexperienced clinician. The review emphasized the extraction and analysis of ASD and ODD landmarks from

behavioral [51], [52], [54], eye gaze [58], audio [44], [45], [46], Facial expressions [62], postural [65] and assessments [73], [74], [76] data to identify children at risk of ASD between 6-18 months, circumventing traditional diagnostic instrument's age constraints. These improvements can advance the field by promoting early identification, improving clinician's capacity, and thereby improving access to early intervention [134] services.

Even though the review highlights that the demand for technology-based detection methods has grown from 2018-2021, the actual adoption of these innovations has been minimal. These innovations should ideally be used by non-specialists, available on mobile applications (to ensure widespread adoption), and able to identify TD, ASD, and ODD (speech delay, development delay, Intellectual delay) based on well-defined minimal distinguishable features, in the first three years [135]. The adoption of these technologies can be supported through controlled pilots through the participation of stakeholders such as parents, clinicians, and schools, digitizing downstream detection processes, assessments, and treatments [136]. A digital human-supported ASD and ODD detection and management framework can be initiated and transition to an autonomous and need-based blended digital model, optimizing cost and maximizing scale [137].

Further, adopting these technologies can be supported with vernacular massive online open courses" (MOOCs), training websites, and brief knowledge content, including text-based training procedures with video clips [137].

TABLE 4. Internal and external validity threats and mitigation measures.

Bias / Challenges	Bias and challenges explanation	Internal threats	Validity	External Validity threats	Mitigation measures
Sample bias	Limited participants; lack of control group; non-inclusion of participants with ODD such as speech, intellectual and development delay. The control (CG) and experimental group (EG) participants were not age, IQ, or gender-matched. Dataset inconsistencies, class imbalances, and non-scientific methods to upscale datasets.	Studies limited to ASD and TD identification on limited statistically participant group-size; Study design not capturing identification of children with ODD.	Risk of tree-based ML models overfitting on a small sample [128] producing non-generalizable results.	Recruit statistically relevant participants; deploy simpler probabilistic ML models such as LR; reduce DL model layers; incorporate regularization methods such as pruning decision trees or dropout, and design ensemble boosting and bagging models. Recalibrating ML and DL models using statistical methods to match CG and EG sizes that are age and gender-matched.	
Validation challenges	Several studies did not report psychometric properties such as sensitivity, specificity, and accuracy or reported only low values for these variables (subsection III-E5 , Table 1)	Poor validation results	Generalization challenges due to internal validity constraints	Robust study design, recruit statistically significant age and intelligence-matched control groups with ASD, TD, and other ODD participants, compute outcomes on established psychometric parameters that are part of study design and protocols.	
Infrastructure and location bias	Studies using eye contact, postural measures, and facial expressions required extensive infrastructure setups such as mounting sensors and cameras and were conducted in clinics or hospitals.	Compare setup recreation time,cost, and repeatability with those of traditional methods.	Need results replication in multiple settings, locations, and with broader population groups	In study design introduce locational variability, sufficient measurement recordings, and study locations outside the laboratory, i.e. in the home,clinics and schools.	
Maturation effect	Independent variables might influence the study outcomes temporally. For example, children with ASD or ODD can improve over time, either with the effect of the treatment or with natural age progression.	In the longitudinal studies [45], [45], [52], [54], [72], [74], [76]–[78] the maturation effect may have led to biasedness in the training of ML and DL models.	Generalization challenges due to internal validity constraints	Define maturation and attrition management and mitigation measures as part of study protocols in longitudinal studies.	
Confounding bias	Studies participants with specific demographics were recruited from selected schools and centers, inducing confounding biases with respect to socioeconomic status, gender, and age. The diagnosis status of the participants was known to the experimenter.	Biased outcomes of statistical, ML, or DL models due to skewed and biased data.	Generalization challenges due to internal validity constraints	Recruit participants from multiple locations and sources, Experimenter should be blinded to participant's diagnostic status.	
Testing bias	Repeated testing may prime the participants, endanger the internal validity, and lead participants to respond differently than they would have otherwise.	Biased outcomes of statistical, ML, or DL models	Generalization challenges due to internal validity constraints	Manage testing biases, repeat testing should be done as part of study design and protocols	
Regression effect	Most studies reported assessment duration between 8-20	The short assessments can lead borderline	Generalization challenges due	Experiment execution should be given sufficient time. The assess-	

TABLE 4. (Continued.) Internal and external validity threats and mitigation measures.

	minutes.	ASD, TD, or ODD participants with misdiagnosis or missed diagnosis.	to internal validity constraints	ment duration should be defined separately for clinic and field trials only.
Experimental biases	The study [50], reported experimental bias where the non-expert US origin annotators displayed cultural and demographic biases towards the US population groups.	Annotators identified social interaction and stereotypical speech for the US children and eye contact for the Bangladeshi children due to cultural and language unfamiliarity.	Generalization challenges due to internal validity constraints	Train experimenters and annotators on cultural and linguistic nuances to overcome biases.
Data quality challenges	Studies reported 1) low-quality of home-based videos due to poor lighting [63], and 2) occultations concealing critical ASD behavioral markers and body part movements [54], [55], [66]. In addition, the ML classifiers [47]–[51], [53] were developed on annotations performed by parents and non-experts who had limited understanding of ASD-specific behavioral landmarks.	The high inter-rater agreement (IRA) on completed annotations was reported by [55]; in contrast, [49] reported 75% IRA across all ages, except for children younger than two years, where a higher discrepancy rate was reported. The low IRA score and poor annotations quality can pose an internal validity threat to the ML model's performance.	Poor generalization of ML, DL, and statistical models on new unseen cross-cultural datasets.	Introduce noise as a part of data in the experiment design. ML and DL models can be re-calibrated using statistical methods to correct external validity-related issues such as unequal CG and EG and matching age and gender. The non-experts should be trained in performing annotations. The IRA scores should be computed during training and real-life scenarios, and high discrepancy scores should be evaluated for root cause analysis and training needs.
Feature engineering challenges	The studies [44], [46]–[51], [54], [61], [63], [72], [74], [78] implemented dimensionality reduction methods on assessments, EMR data and annotation records.	Feature engineering can reduce feature count, leading to rapid diagnostic assessment in minimum time with a risk of overstating ML performance on a reduced feature set.	Limited generalization of ML models on unseen real-life datasets and population samples threatening external validity.	Study protocol and design with clear feature engineering and the data-cleansing rule should mitigate the constraints and risk imposed by feature engineering.
Data transformation challenges	Training and testing machine learning models on transformed, normalized data using z-score [31] methods can improve outcome metrics on a small dataset.	The data transformation can impair the original data's relevance and overstate the ML, DL, and statistical psychometric outcomes [140].	The sensitivity, specificity, and AUROC scores from small samples may not generalize to a bigger cross-cultural population.	Study protocol and design should mitigate the constraints and risk imposed by data transformation and normalization.
Studies replicability	Clinical protocol and dataset reusability can generalize outcomes and improve external validity. For example, [62] assessed facial expressions using similar methods to those used by [65] (using postural	Studies reported dataset preprocessing, cleaning, and augmentation challenges [95], [97], [98], [126]. The datasets were neither	Poor generalization of ML, DL, and statistical models on new unseen datasets.	Clinical protocol standardization, its reusability, datasets reusability (listed in Appendix A), collaborative research models, and inclusive stakeholder framework with parents, schools, and clinicians

TABLE 4. (Continued.) Internal and external validity threats and mitigation measures.

	and head movement data) to predict ASD risk.	age-matched [62] nor culturally relevant to the experiment data [63], posing internal validity threats to the outcomes of ML and DL models.	datasets.	in requirement design can result in robust solution development, validation and [141], generalization. Crowd-sourced video acquisition and feature tagging methods can fast-track the creation of public datasets for ASD and other disorders (ODD).
Cross-cultural validations	[50] constructed ML models using a U.S. video dataset of children and validated on Bangladeshi children with moderate psychometric outcomes because of two reasons. Firstly, ML models trained on ASD features of U.S. children but were validated on Bangladeshi children with ASD and SLC (Speech Language condition). Secondly, cultural, linguistic, and demographic biases were observed during annotations performed by the U.S. non-experts of U.S. and Bangladeshi video data.	The reported biasedness in data and annotation handling poses internal validity threats.	Poor generalization of ML, DL, and statistical models on new unseen Cross-cultural datasets.	A collaborative research framework between developed and developing countries to build new datasets and enhance the existing ones listed in Appendix A can bring efficiency, affordability, access, and percolate research outcomes globally. The initiative can make fair performance comparisons of technological solutions, execute collaborative experiments, and ensure global rigorous and systematic empirical validation of experiments.

B. VALIDITY THREATS

Internal [138] and external validity [139] threats need to be reviewed and managed to ensure the reliability and robustness of the study's research methods and their outcomes. Internal validity evaluates study appropriateness concerning its method, rigor of an experiment, protocol, structure, study variables, and execution. External validity confirms study findings in the real world and leads to broader adoption. While rigorous research procedures can ensure the study's internal validity, they may limit its generalization, application, and external validation. Below Table 4 list internal and external validity threats and suggest methods to overcome those.

V. LIMITATIONS

The access and reach of technology-based ASD detection methods depend on the availability of computers, mobile phones, and the internet. A lack of internet coverage may disproportionately disadvantage those in rural and underserved locations hampered by sluggish internet speeds, poor quality, unstable connectivity, and persons' lack of technological ability and trust in technology. In addition, the internal and external validity threats listed above limit the acceptance and generalization of the innovations.

Further, although the scoping review eligibility to shortlist studies was for children between two and six years, the

following studies had overlapping and higher age ranges. Three studies recruited children between two to eight years [48], [63], [70], and two studies included adolescents, teens, and adults age-group [72], [75]. The mismatch between study eligibility definition and study selection can limit the validity of the scoping review.

VI. FUTURE DIRECTIONS

The review highlighted the presence of a sophisticated technical stack, which produced promising but non-generalizable results with privacy, legal, cultural, and ethical challenges [142]. Therefore future studies should:

- 1) Future research should focus on establishing trust with the “vulnerable” population and their families by addressing ethical, legal, and cultural obstacles in addition to the internal and external validity risks discussed in the previous section.
- 2) Initiate a framework for collaborative research to create new datasets and improve the existing ones described in Appendix A to disseminate research results globally. This can encourage collaborations between academics from developing and developed nations, enabling them to conduct fair performance comparisons of technical solutions, conduct joint experiments, and assure thorough worldwide and systematic empirical validation of studies.

- 3) Researchers should focus on policy-level activities involving stakeholders in developing study designs based on the FATE (Fairness, Accountability, Transparency, and Ethics) framework for using AI for ASD detection. In addition, adopting and enforcing strong regulations, policies, and data protection and privacy legislation to prevent inadvertent data leaking can inspire confidence among stakeholders [143].

In addition, the development of technology-facilitated early ASD and ODD detection solutions should be supported downstream with reliable referral and intervention infrastructure, improving the healthcare system's efficiency, capacity, and efficacy [136].

Most studies in the review captured data from a single bio-behavior to develop ASD detection innovations. Future improvements should include capturing multimodal data from diverse bio-behavior categories. The feature engineering methods can assign weights to multimodal data originating from more than one of the seven categories, such as eye contact, stereotyped behaviors, postural demonstration, and speech, to develop ML and DL models to predict ASD and ODD risk and their severity levels for broader age groups. These improvements can be offered as a service on a mobile application to improve its adoption and usability.

Finally, future studies should focus on preventive methods incorporating genetic approaches. For example, [84] used Hidden Markov Models and genetics to examine the risk of having an ASD offspring, as the ASD risk is multiplied by 40 to 65 times in parents with an ASD diagnosis or carrying a risk gene. Therefore, future genetic-focused trials can preempt the risk of ASD in children and empower parents to decide on starting families with possible risk exposure. In addition, technological innovations using trained robots to treat and diagnose ASD in young children, using POMDP (Partially Observable Markov Decision Process) [144], [145] can significantly automate the ASD detection process and should be a focus for future research.

VII. CONCLUSION

The review comprised 35 studies grouped into seven multimodal data categories: (a) stereotyped behavior, (b) facial expressions, (c) eye gaze, (d) motor control and movements, (e) postural analysis, (f) auditory data, and (g) assessments and electronic health record data. A scoping review based on PRISMA guidelines revealed a rising trend of technology-based ASD detection tools incorporating multimodal data analyzed through ML and DL methods and supports the role and effectiveness of technology applications in improving current ASD screening and diagnosis methods. The review reported internal and external validity challenges with ethical, legal, dataset, and restricted participant and controls as critical challenges. In addition, most solutions reported outcomes limited to the laboratory with non-generalizable outcomes. Therefore, additional cross-cultural intensive trials with large population groups with various other disorders are needed to examine the field preparedness,

ethical, legal, and adoption challenges of technological solutions in real-world scenarios. The review can aid academics, clinicians, and practitioners by offering vital inputs for developing technologically-based ASD screening and diagnostic solutions that are efficient, cost-effective, and data-driven and can address the current constraints of the industry.

APPENDIX A

Appendix lists important datasets in autism research:

JAFFE [146] – The database of 213 images containing the facial expressions of ten Japanese women. There are seven distinct facial expressions: neutral, happy, smiling, sad, surprised, anger, disgust, and fear.

CK+ [147] – The expression database created in the laboratory includes 593 expression sequences from 123 individuals, 69% female and 31% male from African Americans, Asians, and South Americans. It comprises seven facial expressions: disdain, disgust, fear, happiness, sadness, and surprise.

2013 FER [148] – The library contains 35,887 facial photos in gray-scale representing seven different facial expressions: angry, disgusted, fearful, pleased, sad, surprised, and neutral.

MMI [149] – The expression database is broken into two sections: the first is a dynamic data set containing over 2,900 video sequences; the second is a static data set containing over 2,900 video sequences. The second component is a static data collection consisting of many high-resolution photographs. The collection contains seven distinct types of expressions.

AFEW [150] – All of the facial photos in the database were edited from movies and included seven fundamental facial expressions

SFEW [151] – The expression library consists of a static frame image from the AFEW data set containing seven fundamental expressions.

eGeMAPS [152] – A set of acoustic parameters suitable for use in various areas of automatic voice analysis, including para-linguistic and clinical speech analysis. The set is designed to serve as a single reference point for future research evaluations and prevent discrepancies produced by separate parameter sets or even by different implementations of the same parameter

The Simons Simplex Collection (SSC) [153] is a resource of the Simons Foundation Autism Research Initiative (SFARI). The SSC established a permanent repository of genetic samples from 2,600 simplex families, each of which has one child affected with an autism spectrum disorder, and unaffected parents and siblings.

Binghamton University 3D Facial Expression database [95] has currently, 100 participants (56% female, 44% male), ranging in age from 18 to 70 years and representing a diversity of ethnic/racial ancestries. Each person made seven different facial expressions in front of the 3D face scanner. Except for the neutral emotion, each of the six prototypical expressions (happiness, disgust, fear, anger, surprise, and sadness) has four intensity levels.

TABLE 5. Explanation of technical terms used in the review.

Technical terms	Explanation
Supervised ML [154]	ML models are trained on datasets to derive a mapping function between the input (x) and labeled output (y) and predict outcome on new unseen data.
Feature reduction [155]	Feature reduction reduces independent variables in a computationally intensive operation without sacrificing the accuracy of a statistical or ML model. The current review deployed various feature reduction techniques such as discriminant analysis (DA), autoencoders (AE), non-negative matrix factorization, and principal component analysis (PCA).
Data normalization [156]	Normalization ensures that the critical small scale features retain their importance in multiple large scale feature datasets. Z-score, Log Odds ratio (LOGOR) are some examples of data normalization methods used in the review.
Fixation duration [157]	The duration of eye contact when ASD and TD (typically developing) children observe social (human face) or nonsocial (object) visual stimuli.
Area of interest (AOI)[158]	An AOI is a selected region, for example, different parts of the face from where gaze and fixation metrics are extracted.
Key metrics [159]	The sensitivity represents the ability of the test to identify ASD from the group of children who have ASD. The specificity reflects the ability to confirm children from the group who do not have ASD. Accuracy is a ratio of correct predictions (with or without ASD condition) to total predictions. AUROC curves illustrate the relationship between sensitivity and specificity for tests and suggest efficacy, from 0.5 (non-confirmed) to 1.0 (perfect results).
Network and centrality[160]	A node represents an entity in a network graph, and edges represent their connections. The eye-gaze frequency and intensity on the various body and face parts (AOI) can generate a virtual graph on which centrality scores can identify influential nodes. For example, nodes with a high betweenness centrality control the information flow with other nodes in a network.

APPENDIX B

See Table 5.

APPENDIX C

PRISMA Checklist: PRISMA-ScR checklist for studies. CASP Evaluation Sheet: The results of the CASP quality assessment tool for studies. (Microsoft Excel Open XML Spreadsheet (XLSX))

ACKNOWLEDGMENT

Study Registration: Similar to most other scoping reviews, the current scoping study was not registered. Author Contributions: Manu Kohli: conceptualization, methodology, writing (original draft), writing (review and editing), software, formal analysis, investigation, resources, data curation, visualization; Arpan Kumar Kar: conceptualization, methodology, writing (review and editing), supervision, validation, and investigation; and Shuchi Sinha: conceptualization, methodology, writing (review and editing), and supervision.

REFERENCES

- CDC. *Data and Statistics on Autism Spectrum Disorder*. Accessed: Mar. 2, 2022. [Online]. Available: <https://www.cdc.gov/ncbddd/autism/data.html>
- J. Baio, "Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010," *Centers Disease Control MMWR Office*, vol. 69, no. 4, pp. 1–12, Mar. 2014.
- B. Mozolic-Staunton, M. Donelly, J. Yoxall, and J. Barbaro, "Early detection for better outcomes: Universal developmental surveillance for autism across health and early childhood education settings," *Res. Autism Spectr. Disorders*, vol. 71, Mar. 2020, Art. no. 101496.
- Early Childhood Development and Disability: A Discussion Paper*, World Health Org., Geneva, Switzerland, 2012.
- R. C. Sheldrick, S. Marakovitz, D. Garfinkel, A. S. Carter, and E. C. Perrin, "Comparative accuracy of developmental screening questionnaires," *JAMA Pediatrics*, vol. 174, no. 4, pp. 366–374, 2020.
- M. Khowaja, D. L. Robins, and L. B. Adamson, "Utilizing two-tiered screening for early detection of autism spectrum disorder," *Autism*, vol. 22, no. 7, pp. 881–890, Oct. 2018.
- D. L. Robins, K. Casagrande, M. Barton, C.-M.-A. Chen, T. Dumont-Mathieu, and D. Fein, "Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F)," *Pediatrics*, vol. 133, no. 1, pp. 37–45, Jan. 2014.
- S. Ravi, V. Chandrasekaran, S. Kattimani, and M. Subramanian, "Maternal and birth risk factors for children screening positive for autism spectrum disorders on M-CHAT-R," *Asian J. Psychiatry*, vol. 22, pp. 17–21, Aug. 2016.
- C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism," *J. Developmental Disorders*, vol. 30, no. 3, pp. 205–223, Jun. 2000.
- C. Lord, M. Rutter, and A. Le Couteur, "Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders," *J. Autism Devol. Disorders*, vol. 24, no. 5, pp. 659–685, Oct. 1994.
- K. Pierce, V. H. Gazestani, E. Bacon, C. C. Barnes, D. Cha, S. Nalabolu, L. Lopez, A. Moore, S. Pence-Stophaeros, and E. Courchesne, "Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months," *JAMA Pediatrics*, vol. 173, no. 6, pp. 578–587, 2019.
- S. Jayanath and S. Ozonoff, "First parental concerns and age at diagnosis of autism spectrum disorder: A retrospective review from Malaysia," *Malaysian J. Med. Sci.*, vol. 27, no. 5, p. 78, 2020.
- M. J. Maenner, K. A. Shaw, J. Baio, A. Washington, M. Patrick, M. DiRienzo, D. L. Christensen, L. D. Wiggins, S. Pettygrove, J. G. Andrews, and M. Lopez, "Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2016," *MMWR Surveill. Summaries*, vol. 69, no. 4, p. 1, 2020.
- H. Manohar, P. Kandasamy, V. Chandrasekaran, and R. P. Rajkumar, "Early diagnosis and intervention for autism spectrum disorder: Need for pediatrician-child psychiatrist liaison," *Indian J. Psychol. Med.*, vol. 41, no. 1, pp. 87–90, Jan. 2019.
- L. Klinwall and S. Eikeseth, "Early and intensive behavioral intervention (EIBI) in autism," *Comprehensive Guide Autism*, vol. 1, pp. 117–137, Jun. 2014.
- C. Luciano and R. Keller, "Misdiagnosis of high function autism spectrum disorders in adults: An Italian case series," *Autism-Open Access*, vol. 4, no. 131, p. 2, 2014.
- C. Gesi, G. Migliarese, S. Torriero, M. Capellazzi, A. C. Omboni, G. Cerveri, and C. Mencacci, "Gender differences in misdiagnosis and delayed diagnosis among adults with autism spectrum disorder with no language or intellectual disability," *Brain Sci.*, vol. 11, no. 7, p. 912, Jul. 2021.
- S. Aggarwal and B. Angus, "Misdiagnosis versus missed diagnosis: Diagnosing autism spectrum disorder in adolescents," *Australas. Psychiatry*, vol. 23, no. 2, pp. 120–123, Apr. 2015.
- Y. D. Keller-Bell, "Disparities in the identification and diagnosis of autism spectrum disorder in culturally and linguistically diverse populations," *Perspect. ASHA Special Interest Groups*, vol. 2, no. 14, pp. 68–81, Jan. 2017.
- M.-C. Lai and S. Baron-Cohen, "Identifying the lost generation of adults with autism spectrum conditions," *Lancet Psychiatry*, vol. 2, no. 11, pp. 1013–1027, Nov. 2015.
- M. Kohli, A. K. Kar, A. Bangalore, and P. Ap, "Machine learning-based ABA treatment recommendation and personalization for autism spectrum disorder: An exploratory study," *Brain Informat.*, vol. 9, no. 1, p. 16, Jul. 2022, doi: [10.1186/s40708-022-00164-6](https://doi.org/10.1186/s40708-022-00164-6).

- [22] E. Schopler, R. J. Reichler, and B. R. Renner, *The Childhood Autism Rating Scale (CARS)*, 2nd ed. Los Angeles, CA, USA: Western Psychological Services, 2010.
- [23] Y. Pang, C. M. Lee, M. Wright, J. Shen, B. Shen, and J. Bo, "Challenges of case identification and diagnosis of autism spectrum disorders in China: A critical review of procedures, assessment, and diagnostic criteria," *Res. Autism Spectr. Disorders*, vol. 53, pp. 53–66, Sep. 2018.
- [24] N. J. Hidalgo, L. L. McIntyre, and E. H. McWhirter, "Sociodemographic differences in parental satisfaction with an autism spectrum disorder diagnosis," *J. Intellectual Develop. Disability*, vol. 40, no. 2, pp. 147–155, Apr. 2015.
- [25] N. Malik-Soni, A. Shaker, H. Luck, A. E. Mullin, R. E. Wiley, M. Lewis, J. Fuentes, and T. W. Frazier, "Tackling healthcare access barriers for individuals with autism from diagnosis to adulthood," *Pediatric Res.*, vol. 91, no. 5, pp. 1028–1035, 2021.
- [26] A. J. Kumm, M. Viljoen, and P. J. de Vries, "The digital divide in technologies for autism: Feasibility considerations for low- and middle-income countries," *J. Autism Develop. Disorders*, vol. 52, no. 5, pp. 2300–2313, May 2022.
- [27] N. Malik-Soni, A. Shaker, H. Luck, A. E. Mullin, R. E. Wiley, M. E. S. Lewis, J. Fuentes, and T. W. Frazier, "Tackling healthcare access barriers for individuals with autism from diagnosis to adulthood," *Pediatric Res.*, vol. 91, no. 5, pp. 1028–1035, Apr. 2022.
- [28] M. P. Kelly and P. Reed, "Examination of stimulus over-selectivity in children with autism spectrum disorder and its relationship to stereotyped behaviors and cognitive flexibility," *Focus Autism Other Develop. Disabilities*, vol. 36, no. 1, pp. 47–56, Mar. 2021.
- [29] M. Rutter, A. L. Couteur, and C. Lord, *Autism Diagnostic Interview-Revised*, vol. 29. Los Angeles, CA, USA: Western Psychological Services, 2003, p. 30.
- [30] R. M. Jones, S. Risi, D. Wexler, D. Anderson, C. Corsello, A. Pickles, and C. Lord, "How interview questions are placed in time influences caregiver description of social communication symptoms on the ADI-R," *J. Child Psychol. Psychiatry*, vol. 56, no. 5, pp. 577–585, May 2015.
- [31] T. Akter, M. Shahriare Satu, M. I. Khan, M. H. Ali, S. Uddin, P. Lió, J. M. W. Quinn, and M. A. Moni, "Machine learning-based models for early stage detection of autism spectrum disorders," *IEEE Access*, vol. 7, pp. 166509–166527, 2019.
- [32] P. Pandey, P. Ap, M. Kohli, and J. Pritchard, "Guided weak supervision for action recognition with scarce data to assess skills of children with autism," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 1, pp. 463–470.
- [33] A. V. Dahiya, E. DeLucia, C. G. McDonnell, and A. Scarpa, "A systematic review of technological approaches for autism spectrum disorder assessment in children: Implications for the COVID-19 pandemic," *Res. Develop. Disabilities*, vol. 109, Feb. 2021, Art. no. 103852.
- [34] J. Peterson, P. F. Pearce, L. A. Ferguson, and C. A. Langford, "Understanding scoping reviews: Definition, purpose, and process," *J. Amer. Assoc. Nurse Practitioners*, vol. 29, no. 1, pp. 12–16, 2017.
- [35] Y. Mintzker, D. Blum, and L. Adler, "Replacing PICO in non-interventional studies," *BMJ Evidence-Based Med.*, Jan. 2022, doi: 10.1136/bmjebm-2021-111889.
- [36] H. J. Luijendijk, "How to create PICO questions about diagnostic tests," *BMJ Evidence-Based Med.*, vol. 26, no. 4, pp. 155–157, Aug. 2021.
- [37] A. C. Tricco, "PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation," *Ann. Internal Med.*, vol. 169, no. 7, pp. 467–473, Oct. 2018.
- [38] S. B. Mukherjee, "Autism spectrum disorders-diagnosis and management," *Indian J. Pediatrics*, vol. 84, no. 4, pp. 307–314, 2017.
- [39] X. X. Wang Li and V. C. M. Leung, "Artificial intelligence-based techniques for emerging heterogeneous network: State of the arts, opportunities, and challenges," *IEEE Access*, vol. 3, pp. 1379–1391, 2015.
- [40] J.-H. Huh and Y.-S. Seo, "Understanding edge computing: Engineering evolution with artificial intelligence," *IEEE Access*, vol. 7, pp. 164229–164245, 2019.
- [41] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [42] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, pp. 1–16, 2016.
- [43] M. Hilbert and P. López, "The world's technological capacity to store, communicate, and compute information," *Science*, vol. 332, no. 6025, pp. 60–65, Apr. 2011.
- [44] J. H. Lee, G. W. Lee, G. Bong, H. J. Yoo, and H. K. Kim, "Deep-learning-based detection of infants with autism spectrum disorder using auto-encoder feature representation," *Sensors*, vol. 20, no. 23, p. 6762, Nov. 2020.
- [45] E. Patten, K. Belardi, G. T. Baranek, L. R. Watson, J. D. Labban, and D. K. Oller, "Vocal patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency," *J. Autism Develop. Disorders*, vol. 44, no. 10, pp. 2413–2428, Oct. 2014.
- [46] A. Khozaei, H. Moradi, R. Hosseini, H. Pouretemad, and B. Eskandari, "Early screening of autism spectrum disorder using cry features," *PLoS ONE*, vol. 15, pp. 1–21, Dec. 2020. [Online]. Available: <https://doi.org/10.1371/journal.pone.0241690>
- [47] H. Abbas, F. Garberson, E. Glover, and D. P. Wall, "Machine learning approach for early detection of autism by combining questionnaire and home video screening," *J. Amer. Med. Informat. Assoc.*, vol. 25, no. 8, pp. 1000–1007, 2018.
- [48] H. Abbas, F. Garberson, S. Liu-Mayo, E. Glover, and D. P. Wall, "Multi-modular AI approach to streamline autism diagnosis in young children," *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, Dec. 2020.
- [49] Q. Tariq, J. Daniels, J. N. Schwartz, P. Washington, H. Kalantarian, and D. P. Wall, "Mobile detection of autism through machine learning on home video: A development and prospective validation study," *PLOS Med.*, vol. 15, no. 11, pp. 1–20, Nov. 2018. [Online]. Available: <https://doi.org/10.1371/journal.pmed.1002705>
- [50] Q. Tariq, S. L. Fleming, J. N. Schwartz, K. Dunlap, C. Corbin, P. Washington, H. Kalantarian, N. Z. Khan, G. L. Darmstadt, and D. P. Wall, "Detecting developmental delay and autism through machine learning models using home videos of Bangladeshi children: Development and validation study," *J. Med. Internet Res.*, vol. 21, no. 4, Apr. 2019, Art. no. e13822.
- [51] N. Qiu, C. Tang, M. Zhai, W. Huang, J. Weng, C. Li, X. Xiao, J. Fu, L. Zhang, T. Xiao, and H. Fang, "Application of the still-face paradigm in early screening for high-risk autism spectrum disorder in infants and toddlers," *Frontiers Pediatrics*, vol. 8, p. 290, Jun. 2020.
- [52] G. S. Young, J. N. Constantino, S. Dvorak, A. Belding, D. Gangi, A. Hill, M. Hill, M. Miller, C. Parikh, A. J. Schwichtenberg, E. Solis, and S. Ozonoff, "A video-based measure to identify autism risk in infancy," *J. Child Psychol. Psychiatry*, vol. 61, no. 1, pp. 88–94, Jan. 2020.
- [53] N. Nazneen, A. Rozga, C. J. Smith, R. Oberleitner, G. D. Abowd, and R. I. Arriaga, "A novel system for supporting autism diagnosis using home videos: Iterative development and evaluation of system design," *JMIR mHealth uHealth*, vol. 3, no. 2, p. e68, Jun. 2015.
- [54] L. Ouss, G. Palestra, C. Saint-Georges, M. Leitgel Gille, M. Afshar, H. Pellerin, K. Baily, M. Chetouani, L. Robel, B. Golse, R. Nababout, I. Desguerre, M. Guergova-Kuras, and D. Cohen, "Behavior and interaction imaging at 9 months of age predict autism/intellectual disability in high-risk infants with west syndrome," *Transl. Psychiatry*, vol. 10, no. 1, pp. 1–7, Dec. 2020.
- [55] J. Hashemi, G. Dawson, K. L. H. Carpenter, K. Campbell, Q. Qiu, S. Espinosa, S. Marsan, J. P. Baker, H. L. Egger, and G. Sapiro, "Computer vision analysis for quantification of autism risk behaviors," *IEEE Trans. Affect. Comput.*, vol. 12, no. 1, pp. 215–226, Jan. 2021.
- [56] X. Wang, L. Chen, P. Liu, R. J. Polk, and T. Feng, "Orientation to and processing of social stimuli under normal and competitive conditions in children with autism spectrum disorder," *Res. Autism Spectr. Disorders*, vol. 78, Oct. 2020, Art. no. 101614.
- [57] G. Wan, X. Kong, B. Sun, S. Yu, Y. Tu, J. Park, C. Lang, M. Koh, Z. Wei, Z. Feng, Y. Lin, and J. Kong, "Applying eye tracking to identify autism spectrum disorder in children," *J. Autism Developmental Disorders*, vol. 49, pp. 209–215, Jan. 2019.
- [58] V. R. G. Lederman, A. L. Goulart, J. G. Negrão, D. H. F. da Cunha, A. M. N. dos Santos, and J. S. Schwartzman, "Visual scanning preferences in low birth weight preterm infants," *Trends Psychiatry Psychotherapy*, vol. 41, no. 4, pp. 334–339, Oct. 2019.
- [59] M. Sadria, S. Karimi, and A. T. Layton, "Network centrality analysis of eye-gaze data in autism spectrum disorder," *Comput. Biol. Med.*, vol. 111, Aug. 2019, Art. no. 103332.
- [60] P. U. Putra, K. Shima, S. A. Alvarez, and K. Shimatani, "Identifying autism spectrum disorder symptoms using response and gaze behavior during the Go/NoGo game CatChicken," *Sci. Rep.*, vol. 11, no. 1, p. 22012, Nov. 2021, doi: 10.1038/s41598-021-01050-7.

- [61] P. Nyström, E. Thorup, S. Bölte, and T. Falck-Ytter, "Joint attention in infancy and the emergence of autism," *Biol. Psychiatry*, vol. 86, no. 8, pp. 631–638, Oct. 2019, doi: [10.1016/j.biopsych.2019.05.006](https://doi.org/10.1016/j.biopsych.2019.05.006).
- [62] K. L. H. Carpenter, J. Hahemi, K. Campbell, S. J. Lippmann, J. P. Baker, H. L. Egger, S. Espinosa, S. Vermeer, G. Sapiro, and G. Dawson, "Digital behavioral phenotyping detects atypical pattern of facial expression in toddlers with autism," *Autism Res.*, vol. 14, no. 3, pp. 488–499, Mar. 2021.
- [63] W. Zhao and L. Lu, "Research and development of autism diagnosis information system based on deep convolution neural network and facial expression data," *Library Hi Tech*, vol. 38, no. 4, pp. 799–817, Mar. 2020.
- [64] G. Alvari, C. Furlanello, and P. Venuti, "Is smiling the key? Machine learning analytics detect subtle patterns in micro-expressions of infants with ASD," *J. Clin. Med.*, vol. 10, no. 8, p.1776, 2021. [Online]. Available: <https://www.mdpi.com/2077-0383/10/8/1776>
- [65] G. Dawson, K. Campbell, J. Hashemi, S. J. Lippmann, V. Smith, K. Carpenter, H. Egger, S. Espinosa, S. Vermeer, J. Baker, and G. Sapiro, "Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder," *Sci. Rep.*, vol. 8, no. 1, pp. 1–7, Dec. 2018.
- [66] A. A. Ramírez-Duque, A. Frizera-Neto, and T. F. Bastos, "Robot-assisted autism spectrum disorder diagnostic based on artificial reasoning," *J. Intell. Robotic Syst.*, vol. 96, no. 2, pp. 267–281, Nov. 2019.
- [67] N. Kojovic, S. Natraj, S. P. Mohanty, T. Maillart, and M. Schaefer, "Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children," *Sci. Rep.*, vol. 11, no. 1, p. 15069, Jul. 2021, doi: [10.1038/s41598-021-94378-z](https://doi.org/10.1038/s41598-021-94378-z).
- [68] A. Anzulewicz, K. Sobota, and J. T. Dellafield-Butt, "Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism," *Sci. Rep.*, vol. 6, no. 1, pp. 1–13, Aug. 2016.
- [69] A. Crippa, C. Salvatore, P. Perego, S. Forti, M. Nobile, M. Molteni, and I. Castiglioni, "Use of machine learning to identify children with autism and their motor abnormalities," *J. Autism Develop. Disorders*, vol. 45, no. 7, pp. 2146–2156, 2015.
- [70] M. Alcañiz Raya, J. Marín-Morales, M. E. Minissi, G. T. Garcia, L. Abad, and I. A. Chicchi Giglioli, "Machine learning and virtual reality on body movements' behaviors to classify children with autism spectrum disorder," *J. Clin. Med.*, vol. 9, no. 5, p. 1260, Apr. 2020.
- [71] M. Emanuele, G. Nazzaro, M. Marini, C. Veronesi, S. Boni, G. Polletta, A. D'Ausilio, and L. Fadiga, "Motor synergies: Evidence for a novel motor signature in autism spectrum disorder," *Cognition*, vol. 213, Aug. 2021, Art. no. 104652.
- [72] J. Yuan, C. Holtz, T. Smith, and J. Luo, "Autism spectrum disorder detection from semi-structured and unstructured medical data," *EURASIP J. Bioinf. Syst. Biol.*, vol. 2017, no. 1, pp. 1–9, Dec. 2016.
- [73] A. Ben-Sasson and E. Yom-Tov, "Online concerns of parents suspecting autism spectrum disorder in their child: Content analysis of signs and automated prediction of risk," *J. Med. Internet Res.*, vol. 18, no. 11, p. e300, 2016.
- [74] R. Rahman, A. Kodesh, S. Z. Levine, S. Sandin, A. Reichenberg, and A. Schlessinger, "Identification of newborns at risk for autism using electronic medical records and machine learning," *Eur. Psychiatry*, vol. 63, no. 1, pp. 1–7, 2020.
- [75] N. Goel, B. Grover, Anuj, D. Gupta, A. Khanna, and M. Sharma, "Modified grasshopper optimization algorithm for detection of autism spectrum disorder," *Phys. Commun.*, vol. 41, Aug. 2020, Art. no. 101115.
- [76] G. Bussu, E. J. H. Jones, T. Charman, M. H. Johnson, J. K. Buitelaar, and BASIS Team, "Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: A longitudinal cross-domain classifier analysis," *J. Autism Develop. Disorders*, vol. 48, no. 7, pp. 2418–2433, 2018.
- [77] G. Tartarisco, G. Cicceri, D. Di Pietro, E. Leonardi, S. Aiello, F. Marino, F. Chiarotti, A. Gagliano, G. M. Arduino, F. Apicella, F. Muratori, D. Bruneo, C. Allison, S. B. Cohen, D. Vagni, G. Pioggia, and L. Ruta, "Use of machine learning to investigate the quantitative checklist for autism in toddlers (Q-CHAT) towards early autism screening," *Diagnostics*, vol. 11, no. 3, p. 574, Mar. 2021. [Online]. Available: <https://www.mdpi.com/2075-4418/11/3/574>
- [78] K. K. Mujeeb Rahman and M. Monica Subashini, "A deep neural network-based model for screening autism spectrum disorder using the quantitative checklist for autism in toddlers (QCHAT)," *J. Autism Develop. Disorders*, vol. 52, no. 6, pp. 2732–2746, Jun. 2021, doi: [10.1007/s10803-021-05141-2](https://doi.org/10.1007/s10803-021-05141-2).
- [79] D. Dua and C. Graff, "UCI machine learning repository," School Inf. Comput. Sci., Univ. California, Irvine, CA, USA, 2017. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [80] CASP. (2022). *Critical Appraisal Skills Programme (CASP)*. CASP Checklists. [Online]. Available: <https://casp-uk.net/casp-tools/checklists/>
- [81] B. M. Fullen, G. D. Baxter, B. G. G. O'Donovan, C. Doody, L. Daly, and D. A. Hurley, "Doctors' attitudes and beliefs regarding acute low back pain management: A systematic review," *Pain*, vol. 136, no. 3, pp. 388–396, Jun. 2008.
- [82] A. J. Ricoy-Cano, E. Obrero-Gaitán, F. Caravaca-Sánchez, and Y. M. D. L. Fuente-Robles, "Factors conditioning sexual behavior in older adults: A systematic review of qualitative studies," *J. Clin. Med.*, vol. 9, no. 6, p. 1716, Jun. 2020.
- [83] A. Ben-Sasson, D. L. Robins, and E. Yom-Tov, "Risk assessment for parents who suspect their child has autism spectrum disorder: Machine learning approach," *J. Med. Internet Res.*, vol. 20, no. 4, p. e134, Apr. 2018. [Online]. Available: <http://www.jmir.org/2018/4/e134/>
- [84] E. A. Carvalho, C. P. Santana, I. D. Rodrigues, L. Lacerda, and G. S. Bastos, "Hidden Markov models to estimate the probability of having autistic children," *IEEE Access*, vol. 8, pp. 99540–99551, 2020.
- [85] J. A. Kientz, M. S. Goodwin, G. R. Hayes, and G. D. Abowd, "Interactive technologies for autism," *Synth. Lectures Assistive, Rehabilitative, Health-Preserving Technol.*, vol. 2, no. 2, pp. 1–177, 2013.
- [86] H. Meadan, L. E. Meyer, M. R. Snodgrass, and J. W. Halle, "Coaching parents of young children with autism in rural areas using internet-based technologies: A pilot program," *Rural Special Educ. Quart.*, vol. 32, no. 3, pp. 3–10, Sep. 2013.
- [87] N. Bauminger-Zviely, S. Eden, M. Zancanaro, P. L. Weiss, and E. Gal, "Increasing social engagement in children with high-functioning autism spectrum disorder using collaborative technologies in the school environment," *Autism*, vol. 17, no. 3, pp. 317–339, May 2013.
- [88] L. Connelly, "Logistic regression," *Medsurg Nursing*, vol. 29, no. 5, pp. 353–354, 2020.
- [89] X.-Y. Liu, Y. Liang, S. Wang, Z.-Y. Yang, and H.-S. Ye, "A hybrid genetic algorithm with wrapper-embedded approaches for feature selection," *IEEE Access*, vol. 6, pp. 22863–22874, 2018.
- [90] Q. Shaukat, J. Hertecant, A. W. El-Hattab, B. R. Ali, and J. Suleiman, "West syndrome, developmental and epileptic encephalopathy, and severe CNS disorder associated with WWOX mutations," *Epileptic Disorders*, vol. 20, no. 5, pp. 401–412, Oct. 2018.
- [91] I. Ahmad, M. Basher, M. J. Iqbal, and A. Raheem, "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection," *IEEE Access*, vol. 6, pp. 33789–33795, 2018.
- [92] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019.
- [93] A. Klin, J. H. Danovitch, A. B. Merz, and F. R. Volkmar, "Circumscribed interests in higher functioning individuals with autism spectrum disorders: An exploratory study," *Res. Pract. Persons Severe Disabilities*, vol. 32, no. 2, pp. 89–100, Jun. 2007.
- [94] I. Gordon, M. D. Pierce, M. S. Bartlett, and J. W. Tanaka, "Training facial expression production in children on the autism spectrum," *J. Autism Developmental Disorders*, vol. 44, no. 10, pp. 2486–2498, Oct. 2014.
- [95] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGF)*, 2006, pp. 211–216.
- [96] H. Drimalla, I. Baskow, B. Behnia, S. Roepke, and I. Dziobek, "Imitation and recognition of facial emotions in autism: A computer vision approach," *Mol. Autism*, vol. 12, no. 1, pp. 1–15, Dec. 2021.
- [97] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–10.
- [98] H. Amiri, K. Lari, and P. J. Coelho, "Comparison of CK model and line by line method using old and updated parameters/databases," *Int. J. Thermal Sci.*, vol. 118, pp. 448–460, Aug. 2017.
- [99] N. Nurwulan, B. Jiang, and V. Novak, "Development of postural stability index to distinguish different stability states," *Entropy*, vol. 21, no. 3, p. 314, Mar. 2019.
- [100] J. C. Leite, J. C. D. J. Neves, L. G. V. Vitor, and D. S. Fujisawa, "Postural control in children with down syndrome: Evaluation of functional balance and mobility," *Revista Brasileira de Educação Especial*, vol. 24, pp. 173–182, Apr. 2018.

- [101] R. Li, Z. Zhao, Q. Sun, C.-L. I, C. Yang, X. Chen, M. Zhao, and H. Zhang, "Deep reinforcement learning for resource management in network slicing," *IEEE Access*, vol. 6, pp. 74429–74441, 2018.
- [102] T. Gargot, D. Archambault, M. Chetouani, D. Cohen, W. Johal, and S. M. Anzalone, "Automatic assessment of motor impairments in autism spectrum disorders: A systematic review," *Cognit. Comput.*, vol. 14, no. 2, pp. 624–659, Jan. 2022.
- [103] H. D. Pusponegoro, P. Efar, Soedjatmiko, A. Soebadi, A. Firmansyah, H.-J. Chen, and K.-L. Hung, "Gross motor profile and its association with socialization skills in children with autism spectrum disorders," *Pediatrics Neonatol.*, vol. 57, no. 6, pp. 501–507, Dec. 2016.
- [104] T. Talkar, J. R. Williamson, D. J. Hannon, H. M. Rao, S. Yuditskaya, K. T. Claypool, D. Sturim, L. Nowinski, H. Saro, C. Stamm, M. Mody, C. J. Mcdoogle, and T. F. Quatieri, "Assessment of speech and fine motor coordination in children with autism spectrum disorder," *IEEE Access*, vol. 8, pp. 127535–127545, 2020.
- [105] M. Aslam, "Introducing Kolmogorov-Smirnov tests under uncertainty: An application to radioactive data," *ACS Omega*, vol. 5, no. 1, pp. 914–917, Jan. 2020.
- [106] B. Wundervald, A. C. Parnell, and K. Domijan, "Generalizing gain penalization for feature selection in tree-based models," *IEEE Access*, vol. 8, pp. 190231–190239, 2020.
- [107] H. Sifaou, A. Kammoun, and M.-S. Alouini, "High-dimensional quadratic discriminant analysis under spiked covariance model," *IEEE Access*, vol. 8, pp. 117313–117323, 2020.
- [108] D. Granato, J. S. Santos, G. B. Escher, B. L. Ferreira, and R. M. Maggio, "Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective," *Trends Food Sci. Technol.*, vol. 72, pp. 83–90, Feb. 2018.
- [109] M. Habib, M. Faris, A. Alomari, and H. Faris, "AltabbiVec: A word embedding model for medical and health applications in the Arabic language," *IEEE Access*, vol. 9, pp. 133875–133888, 2021.
- [110] H. Shao, Y. Li, Y. Ding, Q. Zhuang, and Y. Chen, "Land use classification using high-resolution remote sensing images based on structural topic model," *IEEE Access*, vol. 8, pp. 215943–215955, 2020.
- [111] H. Chen, L. Chen, Q. Jiang, and S. Guo, "Prior Dirichlet distribution based feature selection in naive Bayes on high dimensional data classification," in *Proc. 6th Int. Conf. Big Data Inf. Analytics (BigDIA)*, Dec. 2020, pp. 415–420.
- [112] X. Wang, L. Liang, H. Yan, X. Wu, W. Lu, and J. Cai, "Duplex restricted network with guided upsampling for the semantic segmentation of remotely sensed images," *IEEE Access*, vol. 9, pp. 42438–42448, 2021.
- [113] S. K. Berument, M. Rutter, C. Lord, A. Pickles, and A. Bailey, "Autism screening questionnaire: Diagnostic validity," *Brit. J. Psychiatry*, vol. 175, no. 5, pp. 444–451, Nov. 1999.
- [114] M. Udhtha, K. Nomie, E. Yu, and J. Sanner, "Novel and emerging strategies for longitudinal data collection," *J. Nursing Scholarship*, vol. 47, no. 2, pp. 152–160, Mar. 2015.
- [115] H. Zhou, K.-M. Yu, Y.-C. Chen, and H.-P. Hsu, "A hybrid feature selection method RFSTL for manufacturing quality prediction based on a high dimensional imbalanced dataset," *IEEE Access*, vol. 9, pp. 29719–29735, 2021.
- [116] G. Leroy, Y. Gu, S. Pettygrove, M. K. Galindo, A. Arora, and M. Kurzus-Spencer, "Automated extraction of diagnostic criteria from electronic health records for autism spectrum disorders: Development, evaluation, and application," *J. Med. Internet Res.*, vol. 20, no. 11, Nov. 2018, Art. no. e10497.
- [117] S. Z. Mirjalili, S. Mirjalili, S. Saremi, H. Faris, and I. Aljarrah, "Grasshopper optimizer algorithm for multi-objective optimization problems," *Appl. Intell.*, vol. 48, no. 4, pp. 805–820, 2018.
- [118] B. Freeman, M. Del'Homme, D. Guthrie, and F. Zhang, "Vineland adaptive behavior scale scores as a function of age and initial IQ in 210 autistic children," *J. Autism Develop. Disorders*, vol. 29, no. 5, pp. 379–384, 1999.
- [119] Q. Hu, Q. Zhang, X.-S. Si, A.-S. Qin, and Q.-H. Zhang, "Fault diagnosis based on multi-scale redefined dimensionless indicators and density peak clustering with geodesic distances," *IEEE Access*, vol. 8, pp. 84777–84791, 2020.
- [120] C. Allison, S. Baron-Cohen, S. Wheelwright, T. Charman, J. Richler, G. Pasco, and C. Brayne, "The Q-CHAT (Quantitative Checklist for autism in Toddlers): A normally distributed quantitative measure of autistic traits at 18–24 months of age: Preliminary report," *J. Autism Develop. Disorders*, vol. 38, no. 8, pp. 1414–1425, Sep. 2008.
- [121] E. Schoen, R. Paul, and K. Chawarska, "Phonology and vocal behavior in toddlers with autism spectrum disorders," *Autism Res.*, vol. 4, no. 3, pp. 177–188, Jun. 2011.
- [122] A. S. Warlaumont and H. L. Ramsdell-Hudock, "Detection of total syllables and canonical syllables in infant vocalizations," in *Proc. Interspeech*, Sep. 2016, pp. 2676–2680.
- [123] S. Lang, K. D. Bartl-Pokorny, F. B. Pokorny, D. Garrido, N. Mani, A. V. Fox-Boyer, D. Zhang, and P. B. Marschik, "Canonical babbling: A marker for earlier identification of late detected developmental disorders?" *Current Develop. Disorders Rep.*, vol. 6, no. 3, pp. 111–118, Sep. 2019.
- [124] S. Latif, J. Qadir, and M. Bilal, "Unsupervised adversarial domain adaptation for cross-lingual speech emotion recognition," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2019, pp. 732–737.
- [125] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
- [126] E. A. Belalcázar-Bolanos, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, T. Haderlein, and E. Nöth, "Glottal flow patterns analyses for Parkinson's disease detection: Acoustic and nonlinear approaches," in *Proc. Int. Conf. Text, Speech, Dialogue*. Cham, Switzerland: Springer, 2016, pp. 400–407.
- [127] N. Cummins, B. Vlasenko, H. Sagha, and B. Schuller, "Enhancing speech-based depression detection through gender dependent vowel-level formant features," in *Proc. Conf. Artif. Intell. Med. Eur.* Cham, Switzerland: Springer, 2017, pp. 209–214.
- [128] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "A comparison of decision tree ensemble creation techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 173–180, Jan. 2006.
- [129] T. Yuen, M. T. Carter, P. Szatmari, and W. J. Ungar, "Cost-effectiveness of universal or high-risk screening compared to surveillance monitoring in autism spectrum disorder," *J. Autism Develop. Disorders*, vol. 48, no. 9, pp. 2968–2979, Sep. 2018.
- [130] S. L. Hyman, S. E. Levy, and S. M. Myers, "Identification, evaluation, and management of children with autism spectrum disorder," *Pediatrics*, vol. 145, no. 1, Jan. 2020, Art. no. e20193447.
- [131] K. Norris and J. Vaizey, *The Economics of Research and Technology*. Evanston, IL, USA: Routledge, 2018.
- [132] J.-J. Cabibihan, H. Javed, M. Aldosari, T. Frazier, and H. Elbashir, "Sensing technologies for autism spectrum disorder screening and intervention," *Sensors*, vol. 17, no. 12, p. 46, Dec. 2016.
- [133] J. Kornack, A. Persicke, P. Cervantes, J. Jang, and D. Dixon, "Economics of autism spectrum disorders: An overview of treatment and research funding," in *Handbook of Early Intervention for Autism Spectrum Disorders*. New York, NY, USA: Springer, 2014, pp. 165–178, doi: 10.1007/978-1-4939-0401-3_9.
- [134] C. M. Conner, R. D. Cramer, and J. J. McGonigle, "Examining the diagnostic validity of autism measures among adults in an outpatient clinic sample," *Autism Adulthood*, vol. 1, no. 1, pp. 60–68, 2019.
- [135] I. Dubey, R. Bishain, J. Dasgupta, S. Bhavnani, M. K. Belmonte, T. Gliga, D. Mukherjee, G. L. Estrin, M. H. Johnson, S. Chandran, and V. Patel, "Using mobile health technology to assess childhood autism in low-resource community settings in India: An innovation to address the detection gap," *medRxiv*, Jan. 2021, doi: 10.1101/2021.06.24.21259235.
- [136] C. Lord, "The lancet commission on the future of care and clinical research in autism," *Lancet*, vol. 399, no. 10321, pp. 271–334, Jan. 2022.
- [137] C. G. Fairburn and V. Patel, "The impact of digital technology on psychological treatments and their dissemination," *Behaviour Res. Therapy*, vol. 88, pp. 19–25, Jan. 2017.
- [138] K. J. Flannelly, L. T. Flannelly, and K. R. B. Jankowski, "Threats to the internal validity of experimental and quasi-experimental research in healthcare," *J. Health Care Chaplaincy*, vol. 24, no. 3, pp. 107–130, Jul. 2018.
- [139] M. Viswanathan, C. D. Patnode, N. D. Berkman, E. B. Bass, S. Chang, L. Hartling, M. H. Murad, J. R. Treadwell, and R. L. Kane, "Assessing the risk of bias in systematic reviews of health care interventions," in *Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]*. Agency for Healthcare Research and Quality (US), 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/sites/books/NBK519366/>

- [140] G. R. DeVore, "Computing the Z score and centiles for cross-sectional analysis: A practical approach," *J. Ultrasound Med.*, vol. 36, no. 3, pp. 459–473, Mar. 2017.
- [141] A. L. Beam, A. K. Manrai, and M. Ghassemi, "Challenges to the reproducibility of machine learning models in health care," *Jama*, vol. 323, no. 4, pp. 305–306, 2020.
- [142] A. Kumar, B. Finley, T. Braud, S. Tarkoma, and P. Hui, "Sketching an AI marketplace: Tech, economic, and regulatory aspects," *IEEE Access*, vol. 9, pp. 13761–13774, 2021.
- [143] H. Jin, Y. Luo, P. Li, and J. Mathew, "A review of secure and privacy-preserving medical data sharing," *IEEE Access*, vol. 7, pp. 61656–61669, 2019.
- [144] A. Mazel and S. Matu, "DREAM lite: Simplifying robot assisted therapy for ASD," 2021, *arXiv:2104.08034*.
- [145] F. Petric, D. Tolić, D. Miklić, Z. Kovačić, M. Cepanec, and S. Šimleša, "Towards a robot-assisted autism diagnostic protocol: Modelling and assessment with POMDP," in *Proc. Int. Conf. Intell. Robot. Appl.* Cham, Switzerland: Springer, 2015, pp. 82–94.
- [146] M. Lyons, M. Kamachi, and J. Gyoba, "The Japanese female facial expression (JAFFE) dataset," Zenodo, Japan, Apr. 1998. [Online]. Available: <https://zenodo.org/record/3451524>; doi: [10.5281/zenodo.3451524](https://doi.org/10.5281/zenodo.3451524).
- [147] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops*, Jun. 2010, pp. 94–101.
- [148] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," in *Proc. Adv. Hybridization Intell. Methods*. Cham, Switzerland: Springer, 2018, pp. 1–16.
- [149] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2005, p. 5.
- [150] J. Kossaifi, G. Tzimiropoulos, S. Todorovic, and M. Pantic, "AFEW-VA database for valence and arousal estimation in-the-wild," *Image Vis. Comput.*, vol. 65, pp. 23–36, Sep. 2017.
- [151] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 2106–2112.
- [152] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. Andrè, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, "The Geneva minimalist acoustic parameter set (GeMAPS) for voice research and affective computing," *IEEE Trans. Affect. Comput.*, vol. 7, no. 2, pp. 190–202, Jun. 2016.
- [153] J. S. Son, L. J. Zheng, L. M. Rowehl, X. Tian, Y. Zhang, W. Zhu, L. Litcher-Kelly, K. D. Gadow, G. Gathungu, C. E. Robertson, D. Ir, D. N. Frank, and E. Li, "Comparison of fecal microbiota in children with autism spectrum disorders and neurotypical siblings in the simons simplex collection," *PLoS ONE*, vol. 10, no. 10, Oct. 2015, Art. no. e0137725.
- [154] Y. Lin, K. Zhou, and J. Li, "Prediction of slope stability using four supervised learning methods," *IEEE Access*, vol. 6, pp. 31169–31179, 2018.
- [155] X. Chen and X. Hao, "Feature reduction method for cognition and classification of IoT devices based on artificial intelligence," *IEEE Access*, vol. 7, pp. 103291–103298, 2019.
- [156] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 105524.
- [157] X. Zhang and S.-M. Yuan, "An eye tracking analysis for video advertising: Relationship between advertisement elements and effectiveness," *IEEE Access*, vol. 6, pp. 10699–10707, 2018.
- [158] B. Banire, D. Al-Thani, M. Qaraqe, K. Khawaja, and B. Mansoor, "The effects of visual stimuli on attention in children with autism spectrum disorder: An eye-tracking study," *IEEE Access*, vol. 8, pp. 225663–225674, 2020.
- [159] F. Shabani, L. Kumar, and M. Ahmadi, "Assessing accuracy methods of species distribution models: AUC, specificity, sensitivity and the true skill statistic," *Glob. J. Hum. Soc. Sci.*, vol. 18, no. 1, pp. 6–18, 2018.
- [160] X. Deng, Y. Dou, T. Lv, and Q. V. H. Nguyen, "A novel centrality cascading based edge parameter evaluation method for robust influence maximization," *IEEE Access*, vol. 5, pp. 22119–22131, 2017.



MANU KOHLI (Member, IEEE) is currently pursuing the Ph.D. degree with the Indian Institute of Technology Delhi (IIT-Delhi). He has 18 years of experience in executing large-scale business and digital transformation projects in multiple countries. He has undertaken leadership positions in Fortune 500 organizations globally and has worked in multiple technologies, such as SAP, SaaS, and machine learning. He is also the Chief Technology Officer at CogniAble, where he has

developed innovative artificial intelligence solutions for detecting and managing developmental challenges, including autism, with outstanding psychometric properties. He has led the formation of new technology-enabled businesses and ensured their commercialization. He has authored multiple publications in peer-reviewed journals and books by SAP PRESS. His research interests include developing cutting-edge machine learning, deep learning, and computer vision methods to solve complex business and healthcare problems. He has received numerous honors, including the UNICEF blue ribbon, AI Gamechangers, and cash prizes from Lockheed Martin, Tata-Trusts, Western Digital, NASSCOM, NTT-Data, and the Ministry of Electronics for his innovations.



ARPAN KUMAR KAR is currently a Chair Professor in information systems with IIT Delhi. He has published over 170 articles, which includes over 50 ABS 3/ABDC A and WoS Q1 Publications. He is the Editor-in-Chief of *International Journal of Information Management Data Insights* (Elsevier). He is also an Associate Editor in journals, such as *Journal of Public Affairs*, *Global Journal of Flexible Systems Management*, and *Electronic Journal of E-Government Research*,

and the Coordinating Editor of *Information Systems Frontiers*. He is on the editorial board of 12 other journals. He received the Research Excellence Award by Clarivate Analytics for highest WoS citations, from 2015 to 2020, in India. He received the Basant Kumar Birla Distinguished Researcher Award for the count of ABDC A*/ABS four publications, from 2014 to 2019. He received the Best Seller Award from Ivey/Harvard Business Publishing Cases, in 2020. He has received multiple other awards, such as the three IFIP Best Papers, the ACM ICEGOV Best Paper, the Tata Consultancy Services Gold Medal, the Project Management Institute Research Award, the AIMS JL Batra Research Award, the IIT Delhi Teaching Excellence Award, six Outstanding Reviewer Awards from Elsevier, and many more best paper awards.



SHUCHI SINHA received the Ph.D. degree from the University of London.

She has taught and researched in the U.K. and India. She has also worked as a Postdoctoral Research Fellow with the National Health Service (NHS) South West Strategic Health Authority, U.K. She is currently an Associate Professor at the Indian Institute of Technology Delhi. Her research interests include workplace deviance, identity work, future of work, and positive work behaviors. She has completed several consulting and multi-year research projects, and delivered invited sessions for prestigious organizations in India. Her research has been published in international journals of repute, such as the *Journal of International Business Studies*, *Journal of Business Research*, *Human Performance*, *Culture and Organization*, *Journal of Management*, *Spirituality & Religion*, and among others.

Dr. Sinha is a reviewer for several journals of repute. She is currently an Editorial Board Member of *Human Relations* (which is rated A* in ABDC list of management journals and included in FT50 list).