Ilmiy tadqiqotlar va ularning yechimlari jurnali
Aoka guvohnoma № 323759 ISSN: 3060-4818

ARTIFICIAL INTELLIGENCE–ASSISTED EARLY AUTISM SCREENING: INTEGRATING M-CHAT-R/F WITH DIGITAL BEHAVIORAL MARKERS

Shodmonkulova Dilorom
Department of Neurology, Tashkent State Medical University,
Tashkent, Uzbekistan
Article type: Narrative review

Abstract

Background: Early recognition of autism spectrum disorder (ASD) remains uneven because caregiver recognition, screening fidelity, specialist capacity, and referral completion vary across health systems. The Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F) is scalable, while artificial intelligence (AI) can analyze questionnaire, video, gaze, speech, movement, and clinical-record data. Neither approach independently establishes diagnosis.

Objective: To critically evaluate how AI may augment M-CHAT-R/F-based early screening and to define evidence and governance requirements for safe clinical implementation.

Methods: A structured narrative search of PubMed/MEDLINE and authoritative guidance covered English-language publications from 1 January 2014 to 4 September 2026, supplemented by landmark earlier studies. Priority was assigned to validation studies, systematic reviews, meta-analyses, prospective diagnostic studies, and clinical guidelines. Evidence was compared by intended use, reference standard, external validation, calibration, subgroup performance, and clinical utility. No new meta-analysis was performed.

Results: Meta-analysis of 50 M-CHAT(-R/F) studies reported pooled sensitivity of 0.83 and specificity of 0.94, although performance varied substantially across settings. The follow-up interview reduces false-positive referrals and must be preserved. AI models can extract complementary signals from home video and eye tracking and can shorten structured observation; however, many studies use enriched clinical samples, limited external testing, and incomplete reporting of calibration and subgroup errors. High discrimination does not guarantee improved access or outcomes. The strongest near-term design is a sequential pathway in which standardized screening is followed by locally validated AI-supported triage and clinician-led diagnostic assessment.

Conclusions: AI should function as a transparent, auditable adjunct rather than an autonomous diagnostic authority. Prospective multisite validation, culturally representative data, privacy safeguards, monitoring of missed cases, and evidence of pathway benefit are required before routine adoption.

Keywords: autism spectrum disorder; artificial intelligence; M-CHAT-R/F; early screening; digital biomarkers; eye tracking; home video; clinical validation

Highlights

• M-CHAT-R/F performance depends on correct scoring, structured follow-up, population, and referral completion.

• AI can integrate complementary behavioral signals but reported accuracy is vulnerable to spectrum bias and dataset shift.

• A low AI score must not close a referral when caregiver or clinician concern persists.

• Clinical utility, calibration, equity, privacy, and post-deployment monitoring are prerequisites for implementation.

1. Introduction

ASD is a heterogeneous neurodevelopmental condition characterized by persistent differences in social communication and interaction together with restricted or repetitive behaviors, interests, or sensory features [1,3]. Its presentation varies with age, language, intellectual ability, co-occurring conditions, sex, culture, and environmental demands. Because signs unfold over time, early detection is not a single test event but a longitudinal process that links surveillance, standardized screening, referral, diagnostic assessment, and individualized support [5,6].

Caregiver-report tools are feasible at scale but may be influenced by language, health literacy, expectations, and access to follow-up. Specialist observation offers richer information yet requires time and trained personnel. AI has therefore been proposed as a method for extracting reproducible features from brief behavioral samples or routinely collected data. The relevant clinical question is not whether an algorithm can separate two research groups, but whether it improves decisions in the population and workflow in which it will be used.

1.1 Objective and novelty

This review examines M-CHAT-R/F and AI as components of one pathway rather than competing tests. Its contribution is a readiness framework connecting diagnostic accuracy to calibration, equity, workflow impact, and governance. It also specifies failure-safe rules that preserve clinical referral when algorithmic and human judgments disagree.

2. Methods

2.1 Information sources and eligibility

PubMed/MEDLINE was searched using combinations of “autism,” “early screening,” “M-CHAT,” “M-CHAT-R/F,” “artificial intelligence,” “machine learning,” “home video,” “computer vision,” “eye tracking,” “speech,” “movement,” “digital biomarker,” “diagnostic accuracy,” and “external validation.” English-language evidence published between 1 January 2014 and 4 September 2026 was considered, with landmark earlier work and current DSM, WHO, AAP, and NICE guidance added. Priority evidence included prospective diagnostic-accuracy studies, external validations, systematic reviews, meta-analyses, and clinically relevant guidelines. Opinion-only papers and unvalidated commercial performance claims were excluded as primary support.

2.2 Appraisal and synthesis

Evidence was organized by screening stage and data modality. Appraisal considered recruitment setting, participant spectrum, reference standard, handling of indeterminate cases and missing data, separation of training and testing, external validation, calibration, subgroup performance, and consequences of false-positive and false-negative results. As a narrative review, screening was not duplicated, a formal risk-of-bias tool was not applied to every study, and effect estimates were not re-pooled.

3. M-CHAT-R/F as the screening foundation

3.1 Structure and interpretation

M-CHAT-R/F is a two-stage parent-report screener intended for children aged 16–30 months [15]. The initial 20-item checklist is followed, for eligible intermediate-risk results, by a structured interview that clarifies whether a reported behavior is truly absent or simply misunderstood. This second stage is not optional administrative detail: it materially improves specificity and limits avoidable referral burden. A positive screen indicates elevated likelihood and a need for evaluation; it does not diagnose ASD.

A negative screen cannot outweigh persistent concern, developmental regression, or clinician observation. Children with communication delay or other developmental differences may require evaluation regardless of ASD screening classification. Screening quality therefore includes not only instrument accuracy but also correct age, validated translation, staff training, follow-up completion, referral timeliness, and documentation of outcomes.

3.2 Accuracy and heterogeneity

A systematic review and meta-analysis of 50 M-CHAT(-R/F) studies estimated pooled sensitivity at 0.83 (95% CI 0.77–0.88) and specificity at 0.94 (95% CI 0.89–0.97) [16]. These summary values conceal substantial heterogeneity related to risk setting, age, translation, scoring rules, completion of the follow-up interview, and whether screen-negative children received a reference assessment. Verification limited to screen-positive participants can inflate sensitivity. Predictive values also change with prevalence, making results from referred clinics unsuitable for direct transfer to universal population screening.

Element	Evidence-based role	Frequent failure	Quality indicator
Developmental surveillance	Longitudinally identifies concern and change	Treating a single normal visit as definitive	Documented concern and milestone review
M-CHAT-R initial screen	Scalable standardized risk estimation	Using outside validated age/language context	Eligible-population coverage
Structured follow-up	Clarifies intermediate-risk responses	Omission or informal substitution	Follow-up completion rate
Referral and assessment	Determines diagnosis and needs	Delay or loss after positive screen	Time to evaluation and referral uptake

4. AI-enabled screening modalities

4.1 Questionnaire-based machine learning

Machine-learning models can reweight questionnaire items, identify interactions, and combine parent responses with demographic or developmental variables. Their low hardware burden is attractive for primary care. Yet models trained on case-control datasets may learn referral patterns rather than ASD-specific signals. Feature reduction may shorten administration, but omitted items can remove

clinically meaningful information or reduce sensitivity in underrepresented phenotypes. Any revised algorithm requires independent validation; it cannot inherit the evidence base of the original scoring rule.

4.2 Home video and computer vision

Short caregiver-recorded videos provide behavior in naturalistic settings and may support remote triage. Prospective work has combined human-coded or computational features with machine learning to distinguish autistic from non-autistic children [18,19]. Benefits include reach and ecological relevance. Limitations include variable lighting, framing, activity selection, device quality, caregiver prompting, privacy risk, annotation cost, and systematic exclusion of families with limited connectivity. Models may also capture home context or recording behavior instead of neurodevelopmental phenotype.

4.3 Eye tracking and social visual engagement

Eye tracking quantifies allocation of visual attention to people, objects, and dynamic social scenes. Prospective studies have reported associations between social visual engagement metrics and expert diagnosis in referred and primary-care pathways [20,21]. The technology provides continuous, objective measurement, but performance depends on stimulus design, usable gaze data, age, attention, vision, motor ability, and clinical spectrum. A 2023 study reported sensitivity of 71.0% and specificity of 80.7% for categorical diagnosis, illustrating that an objective biomarker can be informative without being sufficient as a stand-alone test [20].

4.4 Speech, movement, neurophysiology, and multimodal models

Speech analytics may quantify prosody, vocal turn-taking, phonation, and lexical features, while movement models analyze gesture, posture, or motor synchrony. EEG and neuroimaging models may reveal mechanistic signatures but are less scalable because of cost, motion artifacts, specialist infrastructure, and uncertain transportability [22–25]. Multimodal systems may be more robust because missing or weak information in one channel can be supplemented by another. Conversely, they introduce complex missing-data patterns, higher burden, greater privacy exposure, and more opportunities for hidden bias.

Modality	Primary advantage	Main validity threat	Current plausible use
Questionnaire ML	Low cost and rapid scoring	Dataset shift; loss of original validation	Decision support after standard administration
Home video	Remote naturalistic behavior	Recording and selection bias	Specialist triage with consent
Eye tracking	Objective continuous attention measure	Equipment and spectrum dependence	Adjunctive measurement in validated pathways
Speech/movement	Brief digital phenotyping	Language, age, and device dependence	Research or monitored pilot
EEG/neuroimaging	Biological and stratification signals	Cost and limited generalizability	Research; not population screening

Multimodal AI	Combines complementary signals	Complexity, missingness, privacy	Prospective evaluation before deployment

5. Evaluating clinical validity and utility

5.1 Beyond accuracy

Discrimination metrics describe ranking, not whether predicted probabilities are correct. Calibration must be assessed overall and across clinically important groups. Sensitivity and specificity should be reported with confidence intervals at a prespecified threshold; positive and negative predictive values should be interpreted at the target prevalence. Decision-curve analysis can estimate whether a model produces net benefit across referral thresholds. A clinically successful tool should reduce waiting time or missed follow-up without increasing inequitable access, family distress, or unnecessary specialist burden.

5.2 External and prospective validation

Random division of one dataset is not equivalent to external validation. Participants from the same family, clinician, site, or recording session must not leak across training and testing. Strong validation uses a locked model at a new site and prospectively enrolls consecutive eligible children. The reference standard should be independent of the index test and should account for differential verification. Impact evaluation then compares AI-assisted care with current practice, including clinician overrides and downstream outcomes.

5.3 Equity and transportability

ASD may be recognized differently in girls, multilingual children, racial or ethnic minority populations, and children with intellectual, motor, hearing, or visual disabilities. Aggregate performance can hide clinically serious subgroup errors. Development and validation samples should report demographics and co-occurring conditions, and prespecified analyses should evaluate sensitivity, specificity, calibration, missing-data rates, and abstention by subgroup. Local validation is necessary because language, referral rules, prevalence, devices, and service capacity determine performance.

6. Proposed integrated clinical pathway

Stage	Required action	Role of AI	Safety rule
1. Surveillance	Elicit concerns and monitor development	None or structured documentation	Concern can directly trigger referral
2. M-CHAT-R/F	Use validated language and complete follow-up	Assist data quality; do not alter scoring silently	Positive result advances pathway
3. AI-supported triage	Analyze consented behavioral data with locked model	Prioritize timing or additional observation	Low score cannot cancel persistent concern
4. Diagnostic assessment	Comprehensive history, observation, and differential diagnosis	Summarize measurements for clinician	Clinician retains accountable decision

5. Monitoring	Link predictions to outcomes and audit drift	Detect calibration or subgroup deterioration	Pause or revise unsafe model

The integrated pathway treats AI as an additional measurement layer. It may prioritize children, identify inadequate data, or recommend more observation, but it should not issue an unreviewed diagnosis. The system should permit abstention when inputs are low quality or outside the validated population. Families need understandable information about purpose, data use, limitations, and consequences, plus a route to human review.

7. Governance, privacy, and implementation

Children's videos, voices, gaze traces, and medical histories are sensitive and potentially identifying. Data minimization, purpose limitation, secure storage, controlled access, retention limits, and documented consent are essential. Secondary model training should not be assumed from clinical consent. Procurement should require model documentation, intended-use statements, training-population description, version control, cybersecurity assessment, and responsibility for updates. Post-deployment surveillance must detect drift after changes in devices, populations, or workflow.

Implementation should begin with a silent prospective validation or monitored pilot. Predefined stop criteria may include loss of subgroup sensitivity, excessive abstention, unexplained calibration shift, or referral overload. Staff need training in communicating probabilistic results and overriding the tool. Patient and autistic-community participation should inform acceptable outcomes and language.

8. Research priorities

Future research should preregister intended use, eligibility, outcomes, and thresholds; recruit multisite community cohorts; compare AI-assisted pathways with current care; evaluate calibration and net benefit; publish subgroup and missing-data analyses; quantify time, cost, and family burden; test privacy-preserving methods; and report negative results and model updates. Trials should examine whether earlier triage leads to earlier needs-based support and improved family experience—not merely a higher area under the curve.

9. Limitations

This was a structured narrative review rather than a registered systematic review. Study selection and appraisal were not independently duplicated, and no new risk-of-bias assessment or meta-analysis was performed. Restriction to English-language literature may omit regional evidence. Rapid technical evolution can make model-specific findings obsolete, and heterogeneous datasets preclude a universal performance estimate for AI-assisted screening.

10. Conclusion

M-CHAT-R/F remains a practical foundation for early ASD screening when the structured follow-up and referral pathway are implemented faithfully. AI can add objective behavioral information and support triage, but internal accuracy is insufficient evidence for clinical deployment. The appropriate near-term model is sequential, locally validated, clinician-supervised, and failure-safe. Evaluation must

prioritize calibration, equity, privacy, workflow benefit, and monitoring of missed cases. AI should expand access to careful assessment, never become a barrier to it.

Declarations

Ethics approval and consent to participate: Not applicable; no new participant data were collected.

Consent for publication: Not applicable. Availability of data and materials: No new datasets were generated or analyzed.

Competing interests: The author declares no competing interests. Funding: No specific funding was received.

Author contributions: Shodmonkulova Dilorom conceived the review, synthesized and critically interpreted the literature, and prepared and approved the manuscript. Acknowledgements: None.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed., text rev. Washington, DC: APA; 2022.
2. World Health Organization. Autism. Fact sheet. Geneva: WHO; 2025.
3. Lord C, Brugha TS, Charman T, et al. Autism spectrum disorder. Nat Rev Dis Primers. 2020;6:5. doi:10.1038/s41572-019-0138-4.
4. Zeidan J, Fombonne E, Scorah J, et al. Global prevalence of autism: a systematic review update. Autism Res. 2022;15:778–790. doi:10.1002/aur.2696.
5. Hyman SL, Levy SE, Myers SM. Identification, evaluation, and management of children with autism spectrum disorder. Pediatrics. 2020;145:e20193447. doi:10.1542/peds.2019-3447.
6. Zwaigenbaum L, Bauman ML, Fein D, et al. Early screening of autism spectrum disorder: recommendations for practice and research. Pediatrics. 2015;136(Suppl 1):S41–S59. doi:10.1542/peds.2014-3667D.
7. Sandin S, Lichtenstein P, Kuja-Halkola R, et al. The heritability of autism spectrum disorder. JAMA. 2017;318:1182–1184. doi:10.1001/jama.2017.12141.
8. Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–444. doi:10.1038/s41588-019-0344-8.
9. Satterstrom FK, Kosmicki JA, Wang J, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell. 2020;180:568–584.e23. doi:10.1016/j.cell.2019.12.036.
10. Qiu S, Qiu Y, Li Y, Cong X. Genetics of autism spectrum disorder: an umbrella review of systematic reviews and meta-analyses. Transl Psychiatry. 2022;12:249. doi:10.1038/s41398-022-02009-6.
11. Modabbernia A, Velthorst E, Reichenberg A. Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. Mol Autism. 2017;8:13. doi:10.1186/s13229-017-0121-4.
12. Kim JY, Son MJ, Son CY, et al. Environmental risk factors and biomarkers for autism spectrum disorder: an umbrella review of the evidence. Lancet Psychiatry. 2019;6:590–600. doi:10.1016/S2215-0366(19)30181-6.

13. Gardener H, Spiegelman D, Buka SL. Prenatal risk factors for autism: comprehensive meta-analysis. Br J Psychiatry. 2009;195:7–14. doi:10.1192/bjp.bp.108.051672.
14. Love C, Sominsky L, O'Hely M, et al. Prenatal environmental risk factors for autism spectrum disorder and their potential mechanisms. BMC Med. 2024;22:393. doi:10.1186/s12916-024-03617-3.
15. Robins DL, Casagrande K, Barton M, et al. Validation of the Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F). Pediatrics. 2014;133:37–45. doi:10.1542/peds.2013-1813.
16. Wieckowski AT, Williams LN, Rando J, Lyall K, Robins DL. Sensitivity and specificity of the Modified Checklist for Autism in Toddlers (original and revised): a systematic review and meta-analysis. JAMA Pediatr. 2023;177:373–383. doi:10.1001/jamapediatrics.2022.5975.
17. Schjølberg S, Shic F, Volkmar FR, et al. What are we optimizing for in autism screening? Examination of algorithmic changes in the M-CHAT. Autism Res. 2022;15:296–304. doi:10.1002/aur.2643.
18. Tariq Q, Daniels J, Schwartz JN, et al. Mobile detection of autism through machine learning on home video: a development and prospective validation study. PLoS Med. 2018;15:e1002705. doi:10.1371/journal.pmed.1002705.
19. Abbas H, Garberson F, Glover E, Wall DP. Machine learning approach for early detection of autism by combining questionnaire and home video screening. J Am Med Inform Assoc. 2018;25:1000–1007. doi:10.1093/jamia/ocy039.
20. Jones W, Klaiman C, Richardson S, et al. Eye-tracking-based measurement of social visual engagement compared with expert clinical diagnosis of autism. JAMA. 2023;330:854–865. doi:10.1001/jama.2023.13295.
21. Keehn B, Kinnear M, Shic F, et al. Eye-tracking biomarkers and autism diagnosis in primary care. JAMA Netw Open. 2024;7:e2411190. doi:10.1001/jamanetworkopen.2024.11190.
22. Moon SJ, Hwang J, Kana R, Torous J, Kim JW. Accuracy of machine learning algorithms for the diagnosis of autism spectrum disorder: systematic review and meta-analysis of brain magnetic resonance imaging studies. Psychiatry Investig. 2019;16:129–140. doi:10.30773/pi.2018.11.12.1.
23. Solek P, Kaszyński J, Słowik A, et al. The role of artificial intelligence for early diagnostic tools of autism spectrum disorder: a systematic review. J Clin Med. 2025;14:1792. doi:10.3390/jcm14061792.
24. Thabtah F. Machine learning in autistic spectrum disorder behavioral research: a review and ways forward. Inform Health Soc Care. 2019;44:278–297. doi:10.1080/17538157.2017.1399132.
25. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019;1:e271–e297. doi:10.1016/S2589-7500(19)30123-2.
26. Wolff JJ, Piven J. Predicting autism in infancy. J Am Acad Child Adolesc Psychiatry. 2021;60:958–967. doi:10.1016/j.jaac.2020.07.910.
27. Dawson G, Bernier R. A quarter century of progress on the early detection and treatment of autism spectrum disorder. Dev Psychopathol. 2013;25:1455–1472. doi:10.1017/S0954579413000710.

28. Chawarska K, Macari S, Shic F. Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. Biol Psychiatry. 2013;74:195–203. doi:10.1016/j.biopsych.2012.11.022.
29. Ozonoff S, Young GS, Landa RJ, et al. Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. J Child Psychol Psychiatry. 2015;56:988–998. doi:10.1111/jcpp.12421.
30. National Institute for Health and Care Excellence. Autism spectrum disorder in under 19s: recognition, referral and diagnosis (CG128). London: NICE; updated 2017.