AI-Assisted Screening and Diagnosis of Early-Stage Autism

Authors: Yuzhuo Yuan, Luo Fang, Luo Fang

Date: 2022-01-26T17:47:47+00:00

Abstract

Symptoms of Autism Spectrum Disorder (ASD) manifest during early infancy and toddlerhood, wherein earlier detection facilitates earlier intervention and yields superior treatment outcomes. Traditional approaches to early autism screening and diagnosis exhibit limitations in assessment methodologies and procedural workflows, rendering them inadequate to meet the demands of large-scale screening and diagnostic applications. With the rapid advancement of artificial intelligence technologies, the deployment of intelligent methods for large-scale, unobtrusive early screening and diagnosis of autism is gradually becoming feasible. Over the past decade, domestic and international research explorations into intelligent identification methods for autism have amassed substantial findings across six domains: classical task behaviors, facial expressions and emotions, eye movements, brain imaging, motor control and movement patterns, and multimodal data. Future research should concentrate on establishing a domestic intelligent medical screening and diagnostic framework for early autism, developing screening tools tailored for infant and toddler patients, constructing intelligent identification models for infants and toddlers with autism that integrate multimodal data, and formulating refined diagnostic methodologies that incorporate brain imaging technologies.

Full Text

Early Screening and Diagnosis of Autism Spectrum Disorder Assisted by Artificial Intelligence

YUAN Yuzhuo[1], **LUO Fang**[2]
[1]Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing 100875, China
[2]Faculty of Psychology, Beijing Normal University, Beijing 100875, China

Abstract

Symptoms of Autism Spectrum Disorder (ASD) manifest during early infancy, and earlier detection combined with timely intervention yields significantly better therapeutic outcomes. Traditional approaches to early autism screening and diagnosis suffer from limitations in assessment methods and procedural workflows, rendering them inadequate for large-scale screening demands. With the rapid advancement of artificial intelligence technologies, intelligent methods for conducting large-scale, non-intrusive early screening and diagnosis of autism have become increasingly feasible. Over the past decade, domestic and international research on intelligent autism detection has accumulated substantial findings across six domains: behaviors in classic diagnostic tasks, facial expressions and emotions, eye movements, brain imaging, motor control and movement patterns, and multimodal data integration. Future research should focus on establishing a domestic intelligent medical screening and diagnostic system for early autism, developing screening tools specifically for infants and toddlers, constructing intelligent recognition models for autistic infants that fuse multimodal data, and developing refined diagnostic methods that incorporate brain imaging technology.

Keywords: autism spectrum disorder, early screening and diagnosis of autism, intelligent recognition of autism, artificial intelligence, multimodal data

1. Introduction

According to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) published by the American Psychiatric Association, Autism Spectrum Disorder (ASD), also known as autism, is defined as a pervasive developmental disorder resulting from neurodevelopmental dysregulation (Hodges et al., 2020). Characterized primarily by impairments in social communication and restricted, repetitive patterns of behavior and interests, ASD predominantly affects children.

The World Health Organization' s 2019 epidemiological survey estimated that approximately 1 in 160 children worldwide suffers from autism, with prevalence rates showing a yearly increase. The U.S. Centers for Disease Control and Prevention reported autism rates of 1 in 68 in 2016, rising to 1 in 59 by 2020 (Maenner et al., 2020). In China, limited by a late start in research and a shortage of professional diagnosticians, no authoritative epidemiological data on childhood autism currently exist. The first "Report on the Development of Children with Autism in China," released in 2014, indicated a prevalence rate of approximately 1% among Chinese children. A 2020 nationwide epidemiological study assessing autism in Chinese children aged 6–12 years reported a prevalence of 0.70% (Zhou et al., 2020). According to the "Report III on the Development Status of Autism Education and Rehabilitation in China," the total number of individuals with autism in China exceeds 10 million, including over 2 million children aged 0–12, with nearly 200,000 new cases annually. Once considered a rare condition, autism now ranks as the second most common disability among young children in China, surpassed only by intellectual disability.

Autism is a lifelong condition with no known cure, causing significant personal suffering while imposing increasingly heavy burdens on families and society. A 2015 U.S. report on socioeconomic costs estimated that autism-related healthcare expenditures, non-medical costs, and productivity losses accounted for 0.994% to 2.009% of national GDP, projected to rise to 0.982% to 3.6% by 2025 (Liu et al., 2015). Chinese surveys indicate that the cost of raising a child with autism (19,582.4 RMB) significantly exceeds that for children with intellectual disabilities (6,391 RMB) or physical disabilities (16,410 RMB) (Dawson et al., 2018). The high prevalence of autism thus creates an urgent need for scientific interventions to alleviate symptoms, improve individual functioning, and reduce familial and societal burdens.

Earlier detection and intervention lead to better prognostic outcomes (Matson et al., 2008). The high neuroplasticity of infants and toddlers makes timely, appropriate early intervention particularly effective for enhancing adaptive and cognitive functioning (Xu & Yang, 2014). However, parents often do not notice significant developmental anomalies and behavioral manifestations until children reach 2–3 years of age, frequently overlooking subtle or even obvious abnormalities between six months and two years due to inexperience. Moreover, autism diagnosis relies primarily on clinical expertise, and the journey from parental concern to confirmed diagnosis involves lengthy delays, lacking convenient and objective diagnostic tools. Consequently, large-scale, non-intrusive early screening for infants and toddlers, followed by prompt diagnostic referral for at-risk cases, is essential.

Over the past decade, computer vision, speech technology, deep learning, and other AI and big data mining techniques have been successfully applied to mental health assessment, automated medical diagnosis, and intervention and rehabilitation, offering the potential for breakthroughs in early autism screening and diagnosis. Applying AI to automated and refined autism screening can lower screening barriers, enabling large-scale, non-intrusive screening of young infants in home or community settings, providing early disease warnings and accelerating intervention workflows. This paper first reviews traditional screening and diagnostic tools for infants with autism, then systematically categorizes research progress over the past decade on intelligent recognition of autism in infants and toddlers (0–3 years). Studies on intelligent recognition of autism in children and adolescents of other ages are also reviewed, as their data collection methods and intelligent recognition techniques offer valuable insights for infant autism detection. Finally, we discuss unresolved issues and future research directions, offering new perspectives for establishing an AI-assisted early screening and diagnostic system for autism in China.

2. Traditional Autism Screening and Diagnostic Methods

The earliest symptoms of autism appear within the first one to two years of life (Matson & Goldin, 2014), with 50% of parents reporting symptoms by age 2 and 90% reporting clear symptoms by age 3 (Matson et al., 2008). The typical age of diagnosis is 3 years (Gilmore et al., 2018; Pierce et al., 2019). Missing the optimal intervention window substantially reduces the effectiveness of subsequent treatments, making early diagnosis and intervention critical. In recent years, scholars worldwide have advocated for early screening at 18–24 months, with immediate referral for comprehensive screening upon detection of suspicious symptoms. Positive screening results trigger early diagnostic assessment and prompt intervention to achieve optimal rehabilitation outcomes (Hyman et al., 2020). Early screening thus forms the foundation for early diagnosis, which in turn is prerequisite for early intervention.

Traditional early screening methods fall into two categories: caregiver-report questionnaires or professional observation-based rating scales, and game-task-based observational checklists. These tools can be used from as early as 6 months of age, typically spanning at least a 6-month age range (see Table 1). Commonly used primary screening tools include the Checklist for Autism in Toddlers (CHAT), Modified CHAT (M-CHAT), Pervasive Developmental Disorder Screening Test (PDDST), and Early Screening for Autistic Traits (ESAT), which are suitable for primary healthcare settings and mostly rely on caregiver report. Among these, CHAT is the most rigorously researched and validated tool for infant autism detection (You & Yang, 2006). Secondary screening tools include the Autism Behavior Checklist (ABC), Baby and Infant Screen for Children with Autism Traits (BISCUIT), Screening Tool for Autism in Two-Year-Olds (STAT), and the Autism Observation Scale for Infants (AOSI), which generally require the child' s presence and professional observation.

The primary diagnostic criterion is DSM-5, with the "gold standard" diagnostic tools being the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) and the Autism Diagnostic Interview-Revised (ADI-R) (Akshoomoff et al., 2006; Lord et al., 1994). ADOS-2 involves direct observation of infants in standardized activities, while ADI-R is a semi-structured caregiver interview; both require assessment by trained specialists.

Table 1. Commonly Used Early Screening Tools for ASD (in months)

Tool	Authors	Age Range	Assessment Method
CHAT	Baron-Cohen et al. (1992)	18–24	Caregiver report + observation
M-CHAT	Robins et al. (2001)	16–30	Caregiver report
PDDST-II	Seigel (2004)	12–48	Caregiver report

Tool	Authors	Age Range	Assessment Method
ESAT	Dietz et al. (2006)	14–15	Caregiver report
FYI	Reznick et al. (2007)	12–24	Caregiver report
ABC	Krug et al. (1980)	18+	Caregiver report
CARS	Schopler et al. (2010)	24+	Professional observation
BISCUIT	Matson et al. (2007)	17-37	Caregiver report
STAT	Stone et al. (2000)	24–36	Play-based observation
AOSI	Bryson et al. (2000)	12–36	Semi-structured play observation
ADEC	Young (2007)	12–36	Interactive items

While some traditional tools have gained widespread acceptance, they suffer from limitations in assessment methodology and efficiency that prevent them from meeting large-scale screening needs. First, early autism symptoms and risk signals require specialist judgment, imposing professional demands on observers (Taylor et al., 2017). The evaluator’ s expertise, institutional resources, and cultural background all affect the reliability and validity of autism assessments (de Belen et al., 2020). Second, confirming an autism diagnosis involves a time-consuming and costly process of caregiver judgment, clinical interviews, observation, and evaluation (Wiggins et al., 2006). Third, autism symptom manifestations are highly heterogeneous, with some clinical features remaining unstable before ages 2–3 (Chen et al., 2011). Additionally, environmental and economic constraints limit clinicians to brief assessments rather than extended naturalistic observation, often resulting in inadequate symptom evaluation. Consequently, researchers urgently need to develop new techniques that simplify screening and diagnostic workflows while reducing time and labor costs, without compromising accuracy.

AI-assisted automated medical diagnosis has advanced rapidly. For instance, computer vision-based facial detection has enabled symptom recognition or pre-diagnosis for over 30 diseases, including psychiatric conditions such as ADHD and depression (Thevenot et al., 2017). Intelligent autism detection for infants offers several advantages: (1) it enables acquisition of naturalistic, multi-dimensional, multimodal behavioral data for comprehensive analysis, ensuring valid and objective assessments that provide reliable pre-diagnostic information to support clinical decisions; (2) computer vision can capture subtle movements in infants with autism that are unobservable to the naked eye, effectively identifying atypical behaviors or discovering new early risk markers at lower cost and with less invasiveness than manual screening, making it suitable for home or community healthcare settings.

3. Intelligent Recognition Technology for Early Autism

Although researchers have identified numerous core symptoms and early risk markers for autism, diagnosing young infants remains challenging due to substantial heterogeneity across autism subtypes, specific symptoms, and severity levels. Moreover, behavioral manifestations in infants with autism often co-occur with typical early developmental characteristics (Vyas et al., 2019) and are influenced by non-autism factors such as cognitive functioning and age (Li et al., 2019). Caregiver reports of early symptoms are prone to recall bias, while clinical observation is limited by the need for child cooperation and suffers from sampling bias. Aggregating extensive behavioral data—particularly naturalistic daily behaviors—and using objective methods to synthesize multi-source information would substantially improve screening accuracy and reliability.

AI technologies including computer vision, intelligent sensors, machine learning, and deep learning have been successfully applied to early autism warning (Hazlett et al., 2017) and robot-assisted therapy (Zheng et al., 2015). Meanwhile, the autism field generates vast amounts of data daily, with foundational datasets reaching sufficient scale to improve diagnostic efficiency through historical data utilization (Liao et al., 2021). Following PRISMA guidelines (Moher et al., 2009), we searched Web of Science, PubMed, IEEE Xplore, and ProQuest for literature published between 2010–2020 using keywords “autism spectrum disorder” (“autism”) AND “machine learning” (“deep learning,” “computer vision,” “affective computing”). After removing duplicates, 741 articles were initially retrieved. Based on our research focus of “intelligent recognition for early autism,” we applied fixed criteria: (1) human autism studies only, excluding animal research; (2) focus on intelligent technology for screening/diagnosis, not intervention/treatment; (3) exclusion of genetic or biomarker studies; (4) focus on infants, children, and adolescents, excluding adults unless particularly relevant; (5) aim of autism detection or risk behaviors in classic assessment tasks, not derived behaviors (e.g., self-injury, sleep). This yielded 576 target articles.

These studies revealed that decade-long efforts in automated autism detection have employed diverse data types, which we categorize into six subdomains: (1) classic task behavior (114 articles); (2) facial expressions and emotions (144 articles); (3) eye movement (18 articles); (4) brain imaging (169 articles); (5) motor control and movement patterns (58 articles); and (6) multimodal data (73 articles). Using citation ranking and snowball sampling, we selected 80 key articles for this review, which we present below by subdomain.

Figure 1. Literature Review Process Flowchart

3.1 Automatic Recognition Based on Classic Task Behaviors

The first step in early autism diagnosis involves screening, assessing, and processing early warning signals (Chen et al., 2011). Retrospective studies (parent reports, home video analysis), prospective studies, early screening scales, and clinical diagnoses have extensively validated early atypical behaviors, yielding numerous classic clinical assessment tasks and corresponding behavioral indicators, such as the response-to-name and visual tracking tasks in AOSI. Recent research has proposed automated detection models for atypical behaviors in these classic tasks. Researchers typically employ non-contact vision systems and sensor technologies (device front cameras, RGB cameras, Kinect 3D sensors) to collect multi-dimensional behavioral data including facial expressions, head movements, limb movements, and acoustic features during task performance, developing task-specific anomaly detection algorithms and automated assessment models to replace manual observation and improve screening efficiency. Below, we illustrate data collection techniques, procedures, and predictive models using response-to-name and visual attention tasks as examples.

Response to Name (RTN) is a classic task ubiquitous in early autism screening scales and clinical diagnosis. Infants begin responding to their names by 4-6 months, selectively turning their heads when hearing their name, demonstrating comprehension of its social significance (Imafuku et al., 2014). Traditionally, RTN requires live observation or post-hoc annotation by professionals using scoring manuals. However, atypical RTN behaviors can be quantified as computable metrics such as eye gaze, head pose changes, and response latency. For example, Bidwell et al. (2014) analyzed audio-video recordings of 50 toddlers (15-30 months) from the publicly annotated Multimodal Dyadic Behavior Dataset (MMDB). Using ceiling-mounted Kinect and front cameras with trackers, they estimated head pose changes, using yaw angle and response latency as behavioral indicators to predict positive/negative responses to social stimuli (name calling). Different classifiers achieved up to 89.4% precision and 83.3% recall. Wang et al. (2019) developed an autism-assisted screening system for RTN tasks encompassing experimental protocols, data collection, and automated assessment to reduce labor costs, particularly valuable in medically underserved regions. Their multi-sensor system (Kinect + 2 RGB cameras) simultaneously captured facial, gaze, posture, and vocal behaviors in toddlers (mean age 2 years), implementing pedestrian detection, skeleton extraction (Microsoft Kinect SDK), facial expression recognition (Baltrušaitis et al., 2015), facial landmark detection and tracking (Baltrušaitis et al., 2013), eye center localization (Wang et al., 2018), and head pose estimation (Baltrušaitis et al., 2016). Using eye center localization and head pose algorithms, they employed gaze rotation angle and fixation duration as behavioral indicators, achieving 92.7% average classification accuracy.

Atypical attention assessment is another classic task in early autism screening. Current methods can automatically identify multiple atypical attention features from audio-video and image data, such as disfluent visual tracking

(Zwaigenbaum et al., 2005), reduced face-looking frequency (Ozonoff et al., 2010), and weak attention disengagement (Elsabbagh et al., 2013). Hashemi et al. (2014) used facial detection and tracking to automatically assess two atypical visual attention tasks from AOSI: (1) attention disengagement—shifting gaze between competing visual stimuli; and (2) visual tracking—following a moving object across midline. Using GoPro Hero cameras to record multiple trials of 12 high-risk infants (5-18 months), they evaluated visual attention via yaw and pitch head movements. Automated assessments showed Cohen' s Kappa of 0.75 with expert ratings, far exceeding non-expert inter-rater reliability (0.27-0.37). Bovery et al. (2019) developed a mobile task to measure atypical attention, presenting social and non-social stimuli on split screens while recording facial dynamics of 104 toddlers (16-31 months) via front camera. They estimated head pose by computing rotation parameters between 51 facial landmarks (Hashemi et al., 2015) and a 3D canonical face model (Fischler & Bolles, 1981), combining yaw angle and iris position to estimate attention direction and measure looking time, preferences, and shifts. Campbell et al. (2019) used a similar paradigm to assess atypical attention and RTN in 16-31-month-olds, finding high concordance between automated RTN assessment and expert ratings (ICC = 0.84, 95% CI 0.67-0.91), with 96% sensitivity and 38% specificity.

Current research has integrated multiple classic clinical tasks into mobile applications combining tasks, data collection, and algorithms, creating integrated, low-cost, scalable autism screening tools applicable beyond laboratories to primary care clinics, schools, and homes. These tools have enabled automated detection of atypical emotions, social referencing, social smiling, and RTN, demonstrating promising predictive performance (Hashemi et al., 2015; Hashemi et al., 2018). However, they remain limited to relatively simple tasks; complex tasks pose greater challenges due to more diverse response patterns in young children. For instance, ADOS' s "bubble play" assesses "shared enjoyment," requiring a dimensional behavioral framework and multimodal temporal modeling of expressions, gaze, spontaneous actions, and vocalizations for effective detection.

3.2 Automatic Recognition Based on Facial Expression and Emotion Data

Social communication deficits are hallmark features distinguishing children with autism from typically developing peers, manifesting in socio-emotional reciprocity and nonverbal communication, such as impaired facial expression imitation and reduced diversity/intensity of expressions. Computer vision-based facial expression analysis overcomes human perceptual limitations, enabling rapid, objective automatic recognition of autism.

Recent advances in computer vision have propelled AI emotion recognition, primarily focused on developing algorithms to classify facial emotions from images or videos into basic emotion categories (de Belen et al., 2020). Researchers have attempted to build algorithms detecting abnormal emotion cognition and expression for automated autism identification in infants. However, limited sample sizes due to the challenges of recruiting autism populations have constrained model development. In contrast, typical-population emotion recognition has yielded numerous models and public datasets. One research approach involves transferring or adapting existing models based on typical-population facial features. For example, Han et al. (2018) extracted and compared facial expression features between typical individuals and children with autism using the FERET and Cohn-Kanade (CK+) datasets, proposing a sparse coding-based feature transfer learning algorithm that achieved over 80% accuracy in real-time emotion recognition during child-robot interactions.

Researchers have collected static facial images or dynamic videos of infants in laboratory or natural settings to construct emotion recognition or autism classification models with promising accuracy. Social smiling is an important early risk marker (Bi et al., 2020), particularly infant smiles during mother-infant interaction, which are key signals for autism detection. Automated smile recognition in clinical and home environments can enhance early screening efficiency. Tang et al. (2018) trained a Convolutional Neural Network (CNN) on 77,000 manually annotated video frames of 34 infants (6–24 months, including 11 high-risk for autism) during mother-infant interaction, achieving 87.16% average accuracy in automatic smile detection. Li et al. (2019) identified facial expressions, action units, arousal, and valence as important facial features for autism classification. They recorded facial videos of 105 children watching videos via mobile device front cameras, using CNN models pretrained on AffectNet and EmotioNet with temporal feature extraction to build a binary classification model, achieving 0.76 sensitivity and 0.69 specificity. Shukla et al. (2017) proposed an automatic developmental disorder detection method from facial images, including ASD, cerebral palsy, fetal alcohol syndrome, Down syndrome, intellectual disability, and progeria. Using a fine-tuned AlexNet CNN on over 2,000 facial images and SVM for binary classification, they achieved 93.33% average precision for autism classification, outperforming non-expert human classification.

Beyond emotion classification, research has examined the process of facial expression production in autism. Methods like Facial Action Unit coding detect and track micro-facial movements invisible to the naked eye, quantifying expression generation abilities—such as onset detection and regional muscle activation patterns—facilitating refined diagnosis and targeted intervention. Leo et al. (2019) proposed a computational framework for analyzing facial emotion expression abilities in autism, comprising four modules: face detection, landmark detection/tracking, action unit intensity estimation, and expression analysis. Using facial videos of 17 individuals with autism (6–13 years) and 10 typically developing toddlers (26–35 months) producing four basic emotions, the method accurately predicted expert-rated expression scores (binary classification) with 0.90 precision and 0.85 recall for the autism group. It also revealed that individuals with autism used both upper and lower face regions for happiness, fear, and anger, but primarily the lower face for sadness. Guha et al. (2016) used motion capture to study subtle dynamic features of facial expressions in high-functioning autism patients (9–14 years), recording 32 facial markers at

100 fps while participants imitated fixed emotion sequences. Multiscale entropy analysis revealed lower dynamic complexity and reduced variability in facial expressions, particularly in the eye region. Ahmed and Goodwin (2017) applied facial expression analysis to computer-assisted instruction for autism, measuring learning engagement through facial changes to support teaching. Using front-camera videos of autistic adolescents (mean age ~12 years) during learning, they coded action units via FACS and used CERT to obtain head orientation time and action unit activation states as engagement metrics.

Facial emotion recognition and expression features remain active research areas in autism. Automated analysis addresses limitations of manual coding, which is time-consuming and impractical for large samples or real-time analysis. Current research can be categorized by expression type (spontaneous vs. imitated), emotion-elicitation stimuli (videos, sensory input, social interaction), data type (static images vs. dynamic videos), and assessment goals (qualitative vs. quantitative). Applications include "therapy robots" that automatically recognize emotions in real-time for targeted intervention, and tools that help clinicians "read" facial expressions to improve diagnostic effectiveness. A key challenge is annotating large video datasets, as manual labeling is labor-intensive while crowdsourced annotation suffers from low inter-rater reliability. Kalantarian et al. (2019) proposed three automatic labeling algorithms for six basic emotions (disgust, neutral, surprise, fear, anger, happiness) in children' s facial videos (mean age 8.5 years), showing relatively good performance for the first four emotions.

3.3 Automatic Recognition Based on Eye Movement Data

Eye contact is a crucial nonverbal communication element indicating interest, attention, and engagement in social interaction, serving as an important indicator for language disorders, emotional states, and early autism risk markers. Substantial evidence demonstrates significant differences in gaze patterns between infants with autism and typically developing peers, including atypical fixation, eye contact, and joint attention (Chong et al., 2017), as well as differential preferences for social versus non-social images (Campbell et al., 2014; Chawarska et al., 2013; Shi et al., 2015).

Eye tracking is a common method for measuring social perception and preferences, capturing gaze trajectories ideal for studying perceptual anomalies in autism. Traditional methods include head-mounted devices requiring lengthy adaptation periods for young children, or viewpoint tracking limited to screen-based laboratory settings that cannot measure gaze in natural social contexts (Chong et al., 2017). Consequently, researchers have explored non-contact eye tracking using eye appearance from images (Lu et al., 2014) or mathematical eye models (Li & Li, 2015), while analyzing psychological factors in eye movement data. For example, Syeda et al. (2017) examined facial scanning patterns and emotion recognition in autism patients (5–17 years) using a Tobii EyeX Controller laptop eye tracker while viewing six basic emotions. They found that individuals with autism focused less on core facial features (eyes, nose, mouth), impairing emotion perception. Chrysouli et al. (2018) used a two-stream CNN model fusing optical flow between consecutive eye image frames and static spatial information to recognize engagement, boredom, or frustration states from gaze images in the MaTHiSiS dataset. For assessing eye contact during natural caregiver-child interaction, researchers increasingly use POV (point-of-view) cameras worn by adults to record children's gaze. Chong et al. (2017) developed an end-to-end deep learning framework (Pose-Implicit CNN) to detect eye contact during natural interaction using a dataset of 100 children with autism (3–6 years) and typically developing toddlers (18–36 months) comprising 156 interaction segments (22 hours), achieving 0.78 precision and 0.80 recall, outperforming other models (AlexNet, PEEC, GazeLocking).

Traditional eye movement data collection is impractical for large-scale screening due to requirements for controlled laboratory environments and sustained screen fixation unsuitable for young children. Non-invasive techniques using cameras to record facial (especially eye) and head movements enable analysis of gaze location and duration in natural social interactions. However, videos collected in homes or clinics may contain occlusions and head position offsets requiring preprocessing and correction.

3.4 Automatic Recognition Based on Brain Imaging Data

Precise autism diagnosis is critical for early intervention and treatment. International research has extensively sought behavioral, genetic, and imaging biomarkers for autism (Hong et al., 2020; Lord et al., 2020; Talbott & Miller, 2020; Wolfers et al., 2019), combining AI for objective diagnosis. However, most studies target children and adults (Dickinson et al., 2021), with infant research still nascent.

Brain imaging technology has advanced understanding of autism pathophysiology, and its integration with AI offers new opportunities for early precision diagnosis. The primary challenge in autism diagnosis is pathophysiological heterogeneity, where brain imaging excels at capturing fine-grained structural and functional information to identify subtype-specific features (Emerson et al., 2017). Consequently, brain imaging-based objective diagnosis has attracted substantial attention. Key imaging modalities include electroencephalography (EEG), structural MRI (sMRI), and functional MRI (fMRI).

sMRI can detect subtle brain structural variations in infants, showing good performance in early diagnosis. Hazlett et al. (2017) reported in *Nature* a study using sMRI from 148 infants (6–12 months), extracting cortical thickness, surface area, and brain volume features combined with deep learning to achieve 81% sensitivity and 88% specificity. EEG' s high temporal resolution enables precise characterization of abnormal spatiotemporal covariation patterns in infant brain function. Gabard-Durnam et al. (2019) used longitudinal EEG data from 171 infants (3–36 months) with logistic regression to classify autism vs. typical development (diagnosed at 36 months), finding that EEG power dynamics in the first postnatal year were most effective for early diagnosis (91% accuracy). Dickinson et al. (2021) at UCLA used EEG data from 65 infants (3 months) with support vector regression to predict autism behavior scores at 18 months, achieving a correlation of 0.76 between predicted and actual values. fMRI offers both high temporal resolution (vs. PET/SPECT) and spatial resolution (vs. EEG), providing rich information on static and dynamic brain functional activity and networks. Emerson et al. (2017) used resting-state fMRI from 59 infants (6 months), extracting functional connectivity features to build an SVM classifier predicting autism diagnosis at 24 months, achieving 81.8% sensitivity and 100% specificity.

While brain imaging holds promise for precise infant autism diagnosis, several challenges remain: (1) infant data acquisition is more difficult than for older populations; (2) successful implementation requires collaboration across medicine, neuroimaging, and computer science; (3) current research is in its infancy, focusing primarily on distinguishing autism from typical development without refined grading or subtyping. Given autism' s high heterogeneity (Elsabbagh et al., 2013), providing severity judgments and pathological subtype information would greatly aid personalized treatment planning. (4) Most studies extract relatively crude imaging features, not fully exploiting the rich information available. The brain' s complexity means autism-related functional abnormalities manifest as complex spatiotemporal patterns, yet research has relied on simple features like static functional connectivity. Recent studies suggest naturalistic fMRI is more suitable for infant brain research (Xie et al., 2021), and dynamic network properties provide richer information than static connectivity (Eslami et al., 2021). (5) Most studies use classical machine learning algorithms, underutilizing superior deep learning models. Few studies employ deep learning, though Xu et al. (2020) used fNIRS time-series data with an LSTM-CNN hybrid model to classify autism vs. typical development (mean age ~9 years), achieving 97.1% sensitivity and 94.3% specificity—an 8% improvement over previous models, demonstrating deep learning' s potential.

3.5 Automatic Recognition Based on Motor Control and Movement Pattern Data

Atypical motor control and movement patterns are early autism features. Landa et al. (2006) found that infants later diagnosed with autism showed lower fine and gross motor scores on the Mullen Scale of Early Learning at 14 and 24 months. Multiple studies report postural abnormalities, motor incoordination, and weak motor control in prone, supine, crawling, and walking positions (Esposito et al., 2009; Teitelbaum et al., 1998). These findings support using atypical motor patterns for early autism identification. Traditional motor function evaluation relies on parent reports and expert observation, with coding methods and standards specific to particular research contexts lacking validated norms (Ozonoff et al., 2008).

Current research primarily builds automated detection methods from movement videos. Dawson et al. (2018) used video-based facial detection to assess head postural control during spontaneous attention in toddlers with autism vs. typical development. They recorded facial videos of 106 toddlers (16–31 months) watching dynamic bubbles and mechanical rabbit videos, quantifying head movement by tracking displacement of facial landmarks across frames. Results showed significantly higher head movement velocity in autism, indicating difficulty maintaining midline head position during attention. Martin et al. (2018) used the same paradigm with 2.5–6.5-year-olds, employing the Zface algorithm (http://zface.org/) (Jeni et al., 2015) for real-time dense 3D facial shape reconstruction from 2D video frames, enabling 3D head tracking (pitch, yaw, roll). By computing angular displacement and velocity across axes, they found children with autism showed higher head movement levels and speeds when viewing social stimuli, suggesting they modulate social perception through head movement.

Studies have also built classification models from movement features in video sequences. Zunino et al. (2018) examined grasping behavior in children with autism (mean age ~9.8 years), analyzing video action sequences (average 83 frames) of grasping, placing, and passing water bottles. Using CNN-LSTM models, they classified autism while generating normalized attention maps from LSTM hidden states to visualize discriminative regions, providing interpretable support for clinicians. Vyas et al. (2019) used data from the NODA remote diagnostic service (https://behaviorimaging.com/), comprising 555 parent-recorded daily activity videos with expert diagnoses. They used pretrained Mask R-CNN for 15-keypoint pose estimation, applied particle filters for missing keypoint interpolation (Arulampalam et al., 2002), and represented body keypoint trajectories as RGB heatmaps (PoTion Representation) (Choutas et al., 2018) for CNN classification, achieving 72.4% accuracy, 72% precision, and 92% recall. This approach used interpretable shallow behavioral information and visualizable heatmaps to aid understanding of movement characteristics.

Advances in hardware have integrated inertial motion sensors, gyroscopes, and magnetometers into smartphones, tablets, and wearables for movement data collection. Anzulewicz et al. (2016) explored autism detection during serious gameplay, recruiting 37 children with autism (3–6 years) and 45 typically developing children. Hand movement data were recorded via touchscreen and built-in inertial sensors (3-axis accelerometer, gyroscope, magnetometer) during tablet gameplay. Using 262 features extracted from raw sensor data, they built machine learning classifiers, with Regularized Greedy Forest achieving 83% sensitivity and 85% specificity. They found significant differences in hand impact force, gesture pressure, force distribution, and tap rate between groups.

Automated detection of various atypical movement patterns has progressed substantially, covering gross postural/limb movements to fine head/hand movements, including overall body posture changes, hand movements (grasping, placing, passing), and attention-related head movements (velocity, stability). As smart sensor technology evolves, researchers can use wearables and motion-sensing devices to automatically identify atypical movements in young children, advancing research on early motor development in autism.

3.6 Automatic Recognition Based on Multimodal Data

The psychological, physiological, and cognitive states of children with autism are reflected across multiple dimensions: facial expressions, body posture, eye gaze, speech, text, and physiological signals. Due to measurement environment heterogeneity and scarce clinical samples, single-modality data often lacks sufficient information for accurate identification (Chen & Zhao, 2019). The current trend is multimodal data fusion, integrating correlated features or intermediate decisions across modalities to obtain more valuable data and higher-level information, improving prediction accuracy beyond single-modality modeling (Poria et al., 2017; de Belen et al., 2020).

For example, Chen and Zhao (2019) used photo-taking and image-viewing tasks to build autism recognition models based on atypical attention preferences. Borrowing cross-modal retrieval concepts, they fused eye movement and image data modalities to create a shared predictive model enabling feature representation and information complementarity. Multimodal modeling improved prediction performance from 76% to 84% for photo-taking and from 97% to 99% for image-viewing tasks. Liao et al. (2021) developed an intelligent recognition method for 3-6-year-olds with autism using eye movement, facial expression, cognitive scores, and reaction time data, performing feature selection through differential analysis and hierarchical fusion based on data source and temporal synchronization. Multimodal modeling showed highest consistency with Autism Behavior Checklist assessments compared to single-modality methods.

Other research uses multimodal behavioral features from child-robot interactions for autism diagnosis and intervention evaluation. Scassellati (2007) defined behavioral indicators reflecting social skills in human-robot interaction data to improve reliability of manual recording and assessment, including: (1) gaze direction and attention focus; (2) interpersonal distance and position tracking; (3) voice prosody and intonation. Researchers have evaluated engagement and participation in children with autism (mean age 3.4 years) interacting with social robots using multimodal indicators like facial orientation, relative position, and physical distance (Feil-Seifer & Matarić, 2010; Moghadas & Moradi, 2018). Online platforms now monitor social skills in autism patients using non-invasive sensors and wearables to collect multimodal daily interaction data (physical distance, posture, upper body movement, micro-expressions) as sociometer metrics, transmitting data to cloud platforms (e.g., Microsoft Azure) for storage, analysis, and visualization to guide targeted intervention design (Winoto et al., 2016).

Several publicly available multimodal datasets for autism research have been established. The Multimodal Dyadic Behavior Dataset (MMDB) contains 160+

interaction segments (3–5 minutes each) of 121 infants (15–30 months) with adult experimenters, with existing behavioral coding frameworks and manual annotations for key behaviors (attention, eye contact, social smiling, vocalizations, communicative gestures) following autism coding manuals (Rehg et al., 2013). The DE-ENIGMA dataset includes 128 children with autism in long-duration interactions with therapists/robots (13 TB), with expert annotations of valence, arousal, and posture for 50 children, enabling model training.

Most studies currently build detection models on independently collected clinical data without comparative evaluation of similar methods on identical tasks, resulting in isolated findings. Public datasets are needed as benchmarks for model performance evaluation. Increasingly available public datasets provide the scale required for machine/deep learning development, allowing researchers to pretrain models or enhance performance with different modalities to improve generalization.

4. Summary and Research Outlook

China' s large population includes a substantial and growing number of individuals with autism. The diagnostic process is time-consuming and expensive, requiring long-term special education and behavioral intervention. Earlier detection and intervention during infancy produce better rehabilitation outcomes, with the optimal treatment window before age 3; difficulty increases with age. However, China' s autism screening and diagnosis faces a "three lacks" situation: lack of diagnostic standards, professional personnel, and rehabilitation pathways. The intervention system is incomplete, overall quality is low, and treatment effectiveness is limited. Diagnosis relies primarily on clinical experience, lacking convenient and objective tools, resulting in lengthy delays from parental concern to confirmed diagnosis and causing many children to miss the optimal intervention window. Additionally, intervention methods are mostly designed for children over 3 years, with unclear efficacy and limited techniques, partly due to diagnostic delays and the challenges of intervening with infants whose behavioral and language skills are still developing.

Therefore, innovating diagnostic workflows and rehabilitation pathways to establish an intelligent-assisted early screening, diagnosis, and treatment system is necessary to reduce time pressure and labor costs. For families, earlier detection and intervention improve prognosis, promoting developmental progress, improving language, and reducing problem behaviors (Chen et al., 2011), creating lasting benefits and minimizing economic and emotional burdens. Recent technologies—computer vision, speech processing, smart wearables, and brain imaging—enable multi-modal, multi-scenario data collection. Modeling naturalistic audio-video data of infants interacting with caregivers can more authentically capture behavior-symptom relationships and detect subtle expression or movement changes unobservable manually. This paper identifies two major challenges and future research directions for establishing an intelligent, non-intrusive screening and diagnostic system for infant autism.

4.1 Lack of Effective Screening Tools for Infants and Toddlers

Despite progress in early autism screening and diagnosis, significant shortcomings remain, including difficulty with early screening, limited 普及, and imprecise assessment. Three specific issues stand out:

First, there is a lack of refined behavioral diagnostic systems for infant autism. Existing tools span large age ranges (typically $ $6 months) with inconsistent items across developmental stages. Autism assessment requires age-appropriate, environmentally contextualized tasks, yet rapid development during infancy means substantial differences emerge within just 3 months. Using the same tool across a 6-month span with uniform content and standards compromises accuracy, creating unstable sensitivity and specificity across ages (Nah et al., 2019). Precise early identification requires clear descriptions of behavioral characteristics and developmental trajectories. While studies often track features at 3-month intervals (Kaur et al., 2018), they do not comprehensively cover the critical early identification period (6–36 months), and long-term studies often use larger, uneven intervals (6–12 months). Moreover, research typically focuses on limited typical behaviors (e.g., object sharing, social smiling) rather than comprehensively covering all manifestations. Consequently, systematic, fine-grained research on infants with autism across both temporal and content dimensions is lacking, as are targeted identification systems.

Second, assessment tools inadequately evaluate emotional and affective capacities. Most screening instruments cover social interaction, language/cognitive development, and repetitive behaviors, with varying emphases by age. Some focus only on partial core symptoms (e.g., CHAT on joint attention and pretend play) (You & Yang, 2006). Even the gold-standard ADOS lacks assessment indicators for emotional/affective development. Yet infants with autism often show emotional/affective developmental deficits, including emotional blunting and difficulties in understanding/expressing emotions, considered core social deficits. Current social observation focuses only on the presence of interaction, not emotional development indicators (de Bildt et al., 2015). Emotion and affect are foundational to socio-emotional development; recognizing these features is crucial for targeted interventions and robust developmental pathways.

Third, existing screening and diagnostic methods lack innovation and integration. Diagnosis relies on medical history analysis, symptom inquiry, behavioral observation, and rating scales, supplemented by limited neuroimaging and genetic testing. This approach lacks detailed, unified standards and has inherent limitations. Semi-structured caregiver reports lack child participation and are subject to parental bias and misunderstanding, potentially over- or underestimating abilities (Johnson & Myers, 2007). Complex observational procedures require qualified professionals (Matson et al., 2011; Romero-Garcia et al., 2019).

While ADOS is the international gold standard, China has a severe shortage of trained professionals. Clinical observations are subjective, influenced by experience and training (Romero-Garcia et al., 2019), and brief doctor-child interactions cannot comprehensively evaluate multi-contextual behaviors (Fitzgerald, 2017; Zabihi et al., 2020). Complete diagnostic processes average 41 months and substantial costs, with coarse rating scales (Hyman et al., 2020), hindering early diagnosis and risking misdiagnosis due to non-standardized criteria.

In summary, traditional methods are constrained by time, personnel, and developmental considerations. There is a critical need for more detailed research on existing tools and professional standards to comprehensively construct behavioral feature systems providing precise identification criteria across ages. Developing intelligent tools to assist or replace clinicians for rapid screening in homes or community clinics while reducing subjective errors would have significant clinical value.

4.2 Lack of Intelligent Recognition Research Integrating Multimodal Data

Current intelligent autism recognition primarily targets individuals over 3 years old, with less use of clinical diagnostic and naturalistic behavioral data compared to laboratory data. Challenges vary across the six subdomains: facial expression recognition has largely used deep learning (CNN, DCNN) for emotion classification in photos/video frames (Li et al., 2019; Shukla et al., 2017), but most classifiers were developed for adults and generalize poorly to infants (Kalantarian et al., 2019). They only classify basic emotions qualitatively (happy, disgusted, angry) without assessing emotional complexity or atypicality levels (Guha et al., 2016). Eye movement research has accumulated rich evidence of gaze pattern differences and effective feature extraction methods (Liu et al., 2016; Liu et al., 2015), but data are mostly from controlled laboratory settings requiring sustained screen fixation, impractical for large-scale screening or naturalistic social interaction assessment. Motor pattern recognition has identified simple fixed movements like head motion in RTN tasks (Dawson et al., 2018) and grasping (Martin et al., 2018), but not complex task movements. Most studies use single-modality data, underutilizing multimodal information.

The trend is to obtain rich multimodal data from infants with autism, fusing facial expressions, body posture, eye gaze, speech, lip movements, and physiological signals. Exploring complementary relationships, feature transformation, and representation patterns across modalities through fusion can revolutionize screening methods and achieve diagnostic breakthroughs. International multimodal studies on other psychiatric disorders (e.g., Alzheimer' s, schizophrenia) have shown higher accuracy than single-modality approaches. However, most autism multimodal research focuses on children over 6 years, leaving a gap for infants. Available multimodal data are limited to controlled experiments, public datasets, or online videos, mostly from single-task or therapeutic contexts, with unknown applicability to large-scale infant screening. Many models predict coarse behavioral labels (simple checklist items or basic emotions) with limited clinical utility.

Building multimodal datasets is fundamental for intelligent recognition. Future research must address efficient, convenient multimodal data acquisition, noise reduction, and effective identification. While foreign public datasets have limitations (lack of infant data, single modalities, coarse labels), they have accelerated algorithm development. China' s large and growing infant autism population has dispersed case data. To build an intelligent early screening/diagnostic system, China must first clarify early diagnostic standards, construct a domestic abnormal behavior indicator system, obtain multimodal data from multiple sources, and create fine- and coarse-grained behavior annotations. Establishing large-scale autism and high-risk infant databases and behavioral feature repositories is essential for high-quality intelligent recognition research. While large datasets are being built, researchers can employ few-shot learning techniques (fine-tuning, data augmentation, transfer learning) to address the contradiction between model training data demands and scarce infant samples.

References

Bi, X., Fan, X., Mi, W., & He, H. (2020). Prospective longitudinal studies of high-risk infants and early identification of autism spectrum disorder. *Advances in Psychological Science*, 28(3), 443-455.

Chen, S., Bai, X., & Zhang, R. (2011). Symptoms, diagnosis, and intervention of autism spectrum disorders. *Advances in Psychological Science*, 19(1), 60-72.

Liao, M., Chen, J., Wang, G., & Peng, S. (2021). Intelligent recognition of autism spectrum disorder children using multimodal data fusion and its effectiveness. *Chinese Science Bulletin*, 66(20), 2618-2628.

Xiong, N., Yang, L., Yu, Y., Hou, J., Li, J., Li, Y., ···Jiao, Z. (2010). Economic burden on families of children with autism, physical disabilities, and intellectual disabilities. *Chinese Journal of Rehabilitation Theory and Practice*, 16(8), 785-788.

Xu, Y., & Yang, J. (2014). Research progress on early autism detection. *Chinese Journal of Clinical Psychology*, 22(6), 1023-1027.

Ahmed, A. A., & Goodwin, M. S. (2017, May). Automated detection of facial expressions during computer-assisted instruction in individuals on the autism spectrum. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6050-6055).

Akshoomoff, N., Corsello, C., & Schmidt, H. (2006). The role of the autism diagnostic observation schedule in the assessment of autism spectrum disorders in school and community settings. *The California School Psychologist*, 11(1), 7-19.

Anzulewicz, A., Sobota, K., & Delafield-Butt, J. T. (2016). Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism. *Scientific Reports*, 6(1), 1-13.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2), 174-188.

Baltrušaitis, T., Mahmoud, M., & Robinson, P. (2015, May). Cross-dataset learning and person-specific normalisation for automatic action unit detection. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Vol. 6, pp. 1-6). IEEE.

Baltrušaitis, T., Robinson, P., & Morency, L. P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 354-361).

Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016, March). OpenFace: An open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-10). IEEE.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months?: The needle, the haystack, and the CHAT. *The British Journal of Psychiatry*, 161(6), 839-843.

Bidwell, J., Essa, I. A., Rozga, A., & Abowd, G. D. (2014, November). Measuring child visual attention using markerless head tracking from color and depth sensing cameras. In *Proceedings of the 16th International Conference on Multimodal Interaction* (pp. 447-454).

Bovery, M. D. M. J., Dawson, G., Hashemi, J., & Sapiro, G. (2019). A scalable off-the-shelf framework for measuring patterns of attention in young children and its application in autism spectrum disorder. *IEEE Transactions on Affective Computing*.

Campbell, D. J., Chang, J., Chawarska, K., & Shic, F. (2014, March). Saliency-based Bayesian modeling of dynamic viewing of static scenes. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 51-58).

Campbell, K., Carpenter, K. L., Hashemi, J., Espinosa, S., Marsan, S., Borg, J. S., ···Dawson, G. (2019). Computer vision analysis captures atypical attention in toddlers with autism. *Autism*, 23(3), 619-628.

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry*, 74(3), 195-203.

Chen, S., & Zhao, Q. (2019). Attention-based autism spectrum disorder screening with privileged modality. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1181-1190).

Chong, E., Chanda, K., Ye, Z., Southerland, A., Ruiz, N., Jones, R. M., ··· Rehg, J. M. (2017). Detecting gaze towards eyes in natural social interactions and its use in child assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3), 1–20.

Choutas, V., Weinzaepfel, P., Revaud, J., & Schmid, C. (2018). PoTion: Pose motion representation for action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7024–7033).

Dawson, G., Campbell, K., Hashemi, J., Lippmann, S. J., Smith, V., Carpenter, K., ···Sapiro, G. (2018). Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder. *Scientific Reports*, 8(1), 1–7.

de Belen, R. A. J., Bednarz, T., Sowmya, A., & Del Favero, D. (2020). Computer vision in autism spectrum disorder research: A systematic review of published studies from 2009 to 2019. *Translational Psychiatry*, 10(1), 1–20.

de Bildt, A., Sytema, S., Zander, E., Bölte, S., Sturm, H., Yirmiya, N., ···Oosterling, I. J. (2015). Autism Diagnostic Interview-Revised (ADI-R) algorithms for toddlers and young preschoolers: Application in a non-US sample of 1,104 children. *Journal of Autism and Developmental Disorders*, 45(7), 2076–2091.

Dickinson, A., Daniel, M., Marin, A., Gaonkar, B., Dapretto, M., McDonald, N. M., & Jeste, S. (2021). Multivariate neural connectivity patterns in early infancy predict later autism symptoms. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(1), 59–69.

Dietz, C., Swinkels, S., van Daalen, E., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum disorder in children aged 14–15 months. II: Population screening with the Early Screening of Autistic Traits Questionnaire (ESAT). *Journal of Autism and Developmental Disorders*, 36(6), 713–722.

Eickhoff, S. B., Milham, M., & Vanderwal, T. (2020). Towards clinical applications of movie fMRI. *NeuroImage*, 217, 116860.

Elsabbagh, M., Fernandes, J., Webb, S. J., Dawson, G., Charman, T., Johnson, M. H., & British Autism Study of Infant Siblings Team. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, 74(3), 189–194.

Emerson, R. W., Adams, C., Nishino, T., Hazlett, H. C., Wolff, J. J., Zwaigenbaum, L., ···Piven, J. (2017). Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. *Science Translational Medicine*, 9(393).

Eslami, T., Almuqhim, F., Raiker, J. S., & Saeed, F. (2021). Machine learning methods for diagnosing autism spectrum disorder and attention-deficit/hyperactivity disorder using functional and structural MRI: A survey. *Frontiers in Neuroinformatics*, 14, 62.

Esposito, G., Venuti, P., Maestro, S., & Muratori, F. (2009). An exploration of symmetry in early autism spectrum disorders: Analysis of lying. *Brain and Development*, 31(2), 131-138.

Feil-Seifer, D., & Matarić, M. (2010, March). Using proxemics to evaluate human-robot interaction. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 143-144). IEEE.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.

Fitzgerald, M. (2017). The clinical utility of the ADI-R and ADOS in diagnosing autism. *The British Journal of Psychiatry*, 211(2), 117-117.

Fombonne, E. (2020). Epidemiological controversies in autism. *Swiss Archives of Neurology, Psychiatry and Psychotherapy*, 171(01).

Gabard-Durnam, L. J., Wilkinson, C., Kapur, K., Tager-Flusberg, H., Levin, A. R., & Nelson, C. A. (2019). Longitudinal EEG power in the first postnatal year differentiates autism outcomes. *Nature Communications*, 10(1), 1-12.

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience*, 19(3), 123.

Girdhar, R., Gkioxari, G., Torresani, L., Paluri, M., & Tran, D. (2018). Detect-and-track: Efficient pose estimation in videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 350-359).

Guha, T., Yang, Z., Grossman, R. B., & Narayanan, S. S. (2016). A computational study of expressive facial dynamics in children with autism. *IEEE Transactions on Affective Computing*, 9(1), 14-20.

Han, J., Li, X., Xie, L., Liu, J., Wang, F., & Wang, Z. (2018, November). Affective computing of children with autism based on feature transfer. In *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)* (pp. 845-849). IEEE.

Hashemi, J., Campbell, K., Carpenter, K., Harris, A., Qiu, Q., Tepper, M., ···Sapiro, G. (2015). A scalable app for measuring autism risk behaviors in young children: A technical validity and feasibility study. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare* (pp. 23-27).

Hashemi, J., Dawson, G., Carpenter, K. L., Campbell, K., Qiu, Q., Espinosa, S., ···Sapiro, G. (2018). Computer vision analysis for quantification of autism risk behaviors. *IEEE Transactions on Affective Computing.*

Hashemi, J., Tepper, M., Vallin Spina, T., Esler, A., Morellas, V., Papanikolopoulos, N., ···Sapiro, G. (2014). Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism Research and Treatment*, 2014.

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., … The IBIS Network (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, 542(7641), 348–351.

Hong, S. J., Vogelstein, J. T., Gozzi, A., Bernhardt, B. C., Yeo, B. T., Milham, M. P., & Di Martino, A. (2020). Toward neurosubtypes in autism. *Biological Psychiatry*, 88(1), 111–128.

Hyman, S. L., Levy, S. E., & Myers, S. M. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics*, 145(1).

Imafuku, M., Hakuno, Y., Uchida-Ota, M., Yamamoto, J.-i., & Minagawa, Y. (2014). "Mom called me!" Behavioral and prefrontal responses of infants to self-names spoken by their mothers. *NeuroImage*, 103, 476–484.

Jeni, L. A., & Cohn, J. F. (2016). Person-independent 3D gaze estimation using face frontalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 87–95).

Jeni, L. A., Cohn, J. F., & Kanade, T. (2015, May). Dense 3D face alignment from 2D videos in real-time. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Vol. 1, pp. 1–8). IEEE.

Johnson, C. P., & Myers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics*, 120(5), 1183–1215.

Kalantarian, H., Jedoui, K., Washington, P., Tariq, Q., Dunlap, K., Schwartz, J., & Wall, D. P. (2019). Labeling images with facial emotion and the potential for pediatric healthcare. *Artificial Intelligence in Medicine*, 98, 77–86.

Kaur, M., Srinivasan, S. M., & Bhat, A. N. (2018). Comparing motor performance, praxis, coordination, and interpersonal synchrony between children with and without Autism Spectrum Disorder (ASD). *Research in Developmental Disabilities*, 72, 79–95.

Krug, D. A., Arick, J., & Almond, P. (1980). Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. *Journal of Child Psychology and Psychiatry*, 21(3), 221–229.

Landa, R., & Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: A prospective study. *Journal of Child Psychology and Psychiatry*, 47(6), 629–638.

Lecciso, F., Levante, A., Petrocchi, S., & De Lumé, F. (2017). Basic Emotion Production Test. Technical Report.

Leigh, J. P., & Du, J. (2015). Brief report: Forecasting the economic burden of autism in 2015 and 2025 in the United States. *Journal of Autism and*

Developmental Disorders, 45(12), 4135-4139.

Leo, M., Carcagnì, P., Distante, C., Mazzeo, P. L., Spagnolo, P., Levante, A., ··· Lecciso, F. (2019). Computational analysis of deep visual data for quantifying facial expression production. *Applied Sciences*, 9(21), 4542.

Li, B., Mehta, S., Aneja, D., Foster, C., Ventola, P., Shic, F., & Shapiro, L. (2019, September). A facial affect analysis system for autism spectrum disorder. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 4549-4553). IEEE.

Li, J., & Li, S. (2015). Gaze estimation from color image based on the eye model with known head pose. *IEEE Transactions on Human-Machine Systems*, 46(3), 414-423.

Liu, W., Li, M., & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, 9(8), 888-898.

Liu, W., Yu, X., Raj, B., Yi, L., Zou, X., & Li, M. (2015, September). Efficient autism spectrum disorder prediction with eye movement: A machine learning framework. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 649-655). IEEE.

Liu, W., Zhou, T., Zhang, C., Zou, X., & Li, M. (2017, October). Response to name: A dataset and a multimodal machine learning framework towards autism study. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 178-183). IEEE.

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., ··· Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, 6(1), 1-23.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659-685.

Lu, F., Sugano, Y., Okabe, T., & Sato, Y. (2014). Adaptive linear regression for appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10), 2033-2046.

Maenner, M. J., Shaw, K. A., Baio, J., Washington, A., Patrick, M., DiRienzo, M., ···Dietz, P.M. (2020). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries*, 69(4), 1.

Martin, K. B., Hammal, Z., Ren, G., Cohn, J. F., Cassell, J., Ogihara, M., ··· Messinger, D. S. (2018). Objective measurement of head movement differences in children with and without autism spectrum disorder. *Molecular Autism*, 9(1), 1-10.

Matson, J. L., Boisjoli, J., & Wilkins, J. (2007). The baby and infant screen for children with aUtIsm traits (BISCUIT). Baton Rouge, LA: Disability Consultants, LLC.

Matson, J. L., & Goldin, R. L. (2014). Diagnosing young children with autism. *International Journal of Developmental Neuroscience*, 39, 44-48.

Matson, J. L., Rieske, R. D., & Tureck, K. (2011). Additional considerations for the early detection and diagnosis of autism: Review of available instruments. *Research in Autism Spectrum Disorders*, 5(4), 1319-1326.

Matson, J. L., Wilkins, J., & Gonzalez, M. (2008). Early identification and diagnosis in autism spectrum disorders in young children and infants: How early is too early? *Research in Autism Spectrum Disorders*, 2(1), 75-84.

Moghadas, M., & Moradi, H. (2018, October). Analyzing human-robot interaction using machine vision for autism screening. In *2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM)* (pp. 572-576). IEEE.

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

Nah, Y.-H., Young, R. L., & Brewer, N. (2019). Development of a brief version of the Autism Detection in Early Childhood. *Autism*, 23(2), 494-502.

Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., ···Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(3), 256-266. e252.

Ozonoff, S., Young, G. S., Goldring, S., Greiss-Hess, L., Herrera, A. M., Steele, J., ···Rogers, S. J. (2008). Gross motor development, movement abnormalities, and early identification of autism. *Journal of Autism and Developmental Disorders*, 38(4), 644-656.

Petric, F., Tolić, D., Miklić, D., Kovačić, Z., Cepanec, M., & Šimleša, S. (2015, August). Towards a robot-assisted autism diagnostic protocol: Modelling and assessment with POMDP. In *International Conference on Intelligent Robotics and Applications* (pp. 82-94). Springer, Cham.

Pierce, K., Gazestani, V. H., Bacon, E., Barnes, C. C., Cha, D., Nalabolu, S., ··· Courchesne, E. (2019). Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. *JAMA Pediatrics*, 173(6), 578-587.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98-125.

Rehg, J., Abowd, G., Rozga, A., Romero, M., Clements, M., Sclaroff, S., ···Ye, Z. (2013). Decoding children' s social behavior. In *Proceedings of the IEEE*

Conference on Computer Vision and Pattern Recognition (pp. 3414–3421).

Reznick, J. S., Baranek, G. T., Reavis, S., Watson, L. R., & Crais, E. R. (2007). A parent-report instrument for identifying one-year-olds at risk for an eventual diagnosis of autism: The first year inventory. *Journal of Autism and Developmental Disorders*, 37(9), 1691–1710.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 31(2), 131–144.

Romero-Garcia, R., Warrier, V., Bullmore, E. T., Baron-Cohen, S., & Bethlehem, R. A. (2019). Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism. *Molecular Psychiatry*, 24(7), 1053–1064.

Sapiro, G., Hashemi, J., & Dawson, G. (2019). Computer vision and behavioral phenotyping: An autism case study. *Current Opinion in Biomedical Engineering*, 9, 14–20.

Scassellati, B. (2007). How social robots will help us to diagnose, treat, and understand autism. In *Robotics Research* (pp. 552–563). Springer.

Schopler, E., Reichler, R. J., & Renner, B. R. (2010). The childhood autism rating scale (CARS). Los Angeles, CA, USA: WPS.

Seigel, B. (2004). PDDST-II: Pervasive Developmental Disorders Screening Test-II: Early Childhood Screener for Autistic Spectrum Disorders. Pearson.

Shi, L., Zhou, Y., Ou, J., Gong, J., Wang, S., Cui, X., ···Luo, X. (2015). Different visual preference patterns in response to simple and complex dynamic social stimuli in preschool-aged children with autism spectrum disorders. *PLoS One*, 10(3), e0122280.

Shukla, P., Gupta, T., Saini, A., Singh, P., & Balasubramanian, R. (2017, March). A deep learning framework for recognizing developmental disorders. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 705–714). IEEE.

Stone, W. L., Coonrod, E. E., & Ousley, O. Y. (2000). Brief report: Screening tool for autism in two-year-olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders*, 30(6), 607.

Syeda, U. H., Zafar, Z., Islam, Z. Z., Tazwar, S. M., Rasna, M. J., Kise, K., & Ahad, M. A. R. (2017, September). Visual face scanning and emotion perception analysis between autistic and typically developing children. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (pp. 844–853).

Talbott, M. R., & Miller, M. R. (2020). Future directions for infant identification and intervention for autism spectrum disorder from a transdiagnostic perspective. *Journal of Clinical Child & Adolescent Psychology*, 49(5), 688-700.

Tang, C., Zheng, W., Zong, Y., Cui, Z., Qiu, N., Yan, S., & Ke, X. (2018, October). Automatic smile detection of infants in mother-infant interaction via CNN-based feature learning. In *Proceedings of the Joint Workshop of the 4th Workshop on Affective Social Multimedia Computing and First Multi-Modal Affective Computing of Large-Scale Multimedia Data* (pp. 35-40).

Taylor, L. J., Eapen, V., Maybery, M., Midford, S., Paynter, J., Quarmby, L., ⋯ Whitehouse, A. J. (2017). Brief report: An exploratory study of the diagnostic reliability for autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 47(5), 1551-1558.

Teitelbaum, P., Teitelbaum, O., Nye, J., Fryman, J., & Maurer, R. G. (1998). Movement analysis in infancy may be useful for early diagnosis of autism. *Proceedings of the National Academy of Sciences*, 95(23), 13982-13987.

Thevenot, J., López, M. B., & Hadid, A. (2017). A survey on computer vision for assistive medical diagnosis from faces. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1497-1511.

The World Health Organization. (2019, June). Autism spectrum disorders. https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders

Vyas, K., Ma, R., Rezaei, B., Liu, S., Neubauer, M., Ploetz, T., ⋯Ostadabbas, S. (2019, October). Recognition of atypical behavior in autism diagnosis from video using pose estimation over time. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE.

Wang, Z., Cai, H., & Liu, H. (2018, December). Robust eye center localization based on an improved SVR method. In *International Conference on Neural Information Processing* (pp. 623-634). Springer, Cham.

Wang, Z., Liu, J., He, K., Xu, Q., Xu, X., & Liu, H. (2019). Screening early children with autism spectrum disorder via response-to-name protocol. *IEEE Transactions on Industrial Informatics*, 17(1), 587-595.

Wiggins, L. D., Baio, J., & Rice, C. (2006). Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *Journal of Developmental & Behavioral Pediatrics*, 27(2), S79-S87.

Winoto, P., Chen, C. G., & Tang, T. Y. (2016, September). The development of a Kinect-based online socio-meter for users with social and communication skill impairments: A computational sensing approach. In *2016 IEEE International Conference on Knowledge Engineering and Applications (ICKEA)* (pp. 139-143). IEEE.

Wolfers, T., Floris, D. L., Dinga, R., van Rooij, D., Isakoglou, C., Kia, S. M., ⋯

Beckmann, C. F. (2019). From pattern classification to stratification: Towards conceptualizing the heterogeneity of Autism Spectrum Disorder. *Neuroscience & Biobehavioral Reviews*, 104, 240–254.

Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., & Yu, S. (2021). A survey on incorporating domain knowledge into deep learning for medical image analysis. *Medical Image Analysis*, 101985.

Xu, Y., & Yang, J. (2014). Research progress on early autism detection. *Chinese Journal of Clinical Psychology*, 22(6), 1023–1027.

Young, R. L. (2007). Autism Detection in Early Childhood: ADEC. Victoria, Australia: Australian Council of Educational Research.

Zabihi, M., Floris, D. L., Kia, S. M., Wolfers, T., Tillmann, J., Arenas, A. L., ···Marquand, A. (2020). Fractionating autism based on neuroanatomical normative modeling. *Translational Psychiatry*, 10(1), 1–10.

Zheng, Z., Young, E. M., Swanson, A. R., Weitlauf, A. S., Warren, Z. E., & Sarkar, N. (2015). Robot-mediated imitation skill training for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(6), 682–691.

Zhou, H., Xu, X., Yan, W., Zou, X., Wu, L., Luo, X., ···Wang, Y. (2020). Prevalence of autism spectrum disorder in China: A nationwide multi-center population-based study among children aged 6 to 12 years. *Neuroscience Bulletin*, 36(9), 961–971.

Zunino, A., Morerio, P., Cavallo, A., Ansuini, C., Podda, J., Battaglia, F., ··· Murino, V. (2018, August). Video gesture analysis for autism spectrum disorder detection. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 3421–3426). IEEE.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, 23(2-3), 143–152.

Figures

Source: ChinaXiv —Machine translation. Verify with original.

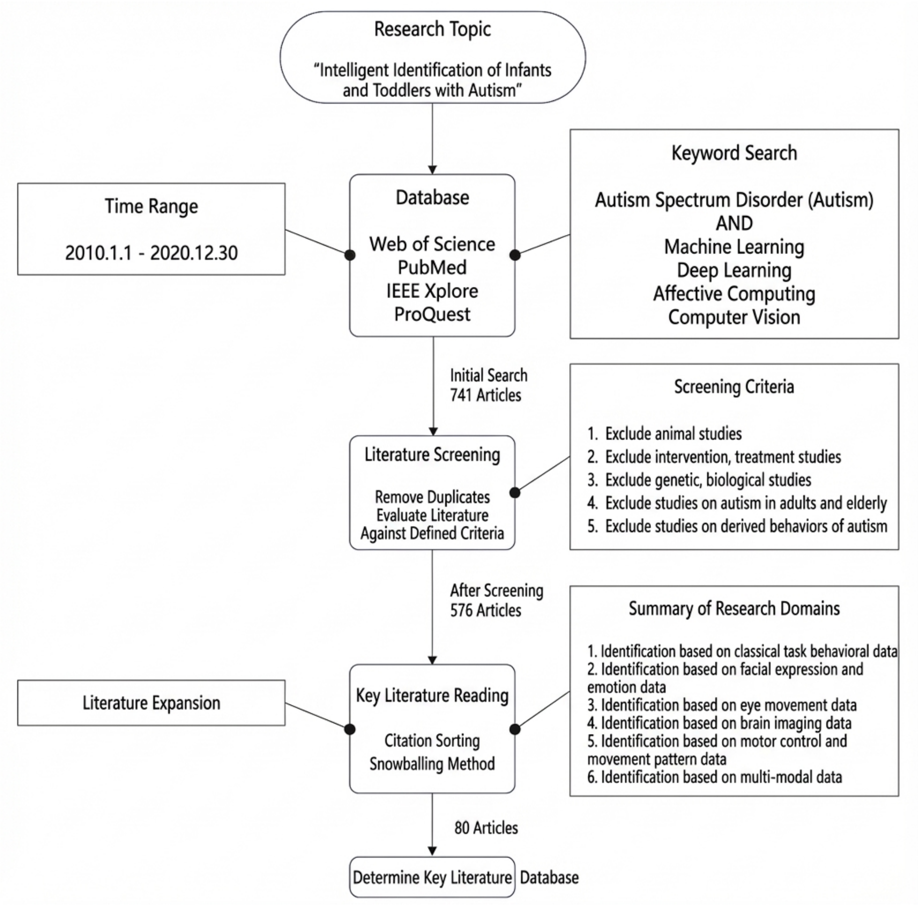
Research Topic
"Intelligent Identification of Infants and Toddlers with Autism"
Time Range
2010.1.1 - 2020.12.30
Database
Web of Science
PubMed
IEEE Xplore
ProQuest
Keyword Search
Autism Spectrum Disorder (Autism)
AND
Machine Learning
Deep Learning
Affective Computing
Computer Vision
Initial Search
741 Articles
Literature Screening
Remove Duplicates
Evaluate Literature
Against Defined Criteria
Screening Criteria
1. Exclude animal studies
2. Exclude intervention, treatment studies
3. Exclude genetic, biological studies
4. Exclude studies on autism in adults and elderly
5. Exclude studies on derived behaviors of autism
After Screening
576 Articles
Summary of Research Domains
1. Identification based on classical task behavioral data
2. Identification based on facial expression and emotion data
3. Identification based on eye movement data
4. Identification based on brain imaging data
5. Identification based on motor control and movement pattern data
6. Identification based on multi-modal data
Literature Expansion
Key Literature Reading
Citation Sorting
Snowballing Method
80 Articles
Determine Key Literature
Database

Figure 1: Figure 1