International Journal of Medical Informatics 214 (2026) 106417

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Review

Challenges in translating AI-driven ASD/ADHD diagnosis: A methodological systematic review

Abdur Rasool[a,b,*], Faizan Ahmad[c], Chayut Bunterngchit[d], Saba Aslam[e]

[a] Department of Information and Computer Sciences, University of Hawaii at Manoa, Honolulu, HI 96822, USA
[b] Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518055, China
[c] College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518122, China
[d] Division of Industrial and Logistics Engineering Technology, Faculty of Engineering and Technology, King Mongkut's University of Technology North Bangkok, Rayong Campus, Rayong 21120, Thailand
[e] Research Center for Biomedical Information Technology, Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

HIGHLIGHTS

- Systematic review of 43 AI studies for ASD/ADHD diagnosis (2020–2024).
- APPRAISE-AI reveals strong clinical relevance (72.8%) but low reproducibility (41.0%).
- Multimodal fusion dominates (38% of studies), outperforming unimodal approaches.
- EEG identified as a central integration hub for multimodal NDDs diagnostic pipelines.
- Proposes a translational framework addressing data heterogeneity, privacy, and interpretability.

ARTICLE INFO

Keywords:
Neurodevelopmental disorders
Multimodal data integration
Deep learning
Computer vision
Clinical decision support
ASD
ADHD

ABSTRACT

Background: Early and accurate diagnosis of neurodevelopmental disorders (NDDs), including autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD), remains a critical challenge in pediatric care. Traditional methods rely on subjective behavioral assessments that are time-intensive and prone to bias.
Objective: This systematic review synthesizes biomedical informatics methodologies using deep learning-driven computer vision to enable objective, data-driven diagnostic decision support for pediatric NDDs.
Methods: Following PRISMA guidelines, we searched Web of Science and Scopus (2020–2024), identifying 43 Q1/Q2 studies. Four informatics-focused research questions were addressed: multimodal feature extraction, deep learning architectures, high-performing strategies, and robust data integration challenges. Methodological quality and bias were assessed using the APPRAISE-AI framework.
Results: Multimodal fusion and hybrid informatics pipelines dominated (38% of studies), outperforming unimodal approaches by integrating complementary streams—facial imaging (high specificity), EEG/fMRI (superior sensitivity). Transfer learning and fusion techniques were prevalent, but federated learning and explainable AI were underutilized. APPRAISE-AI revealed strong clinical relevance (72.8%) and reporting quality (66.1%), yet substantial gaps in reproducibility (41.0%) and result robustness (45.1%).
Conclusions: AI-driven biomedical informatics holds significant potential to reduce diagnostic delays and costs in NDDs. However, reproducibility, interpretability, and ethical data integration must be improved through standardized, privacy-preserving, and auditable frameworks to enable scalable clinical deployment.

* Corresponding author at: Department of Information and Computer Sciences, University of Hawaii at Manoa, Honolulu, HI 96822, USA.
Email address: abdur@hawaii.edu (A. Rasool).

https://doi.org/10.1016/j.ijmedinf.2026.106417
Received 7 November 2025; Received in revised form 26 March 2026; Accepted 29 March 2026
Available online 4 April 2026

1. Introduction

The diagnosis of NDDs, such as ASD and ADHD, remains a core challenge in biomedical informatics, requiring integration of heterogeneous data for timely, objective clinical decision-making. Traditional workflows rely on subjective behavioral assessments that are labor-intensive, prone to inter-rater variability, and limited by access to standardized data [1–8]. In the United States, 1 in 36 children receives an ASD diagnosis, while ADHD affects 5%–7% of children worldwide [9,10]. Delayed diagnosis incurs lifetime costs exceeding $1.4 million per ASD individual, highlighting the need for scalable, reproducible informatics-driven diagnostic support deployable across diverse settings [2,7,9,11–16].

Advances in deep learning and computer vision enable automated extraction, fusion, and interpretation of multimodal neurobehavioral data [2,8,17–28]. For instance, 3D CNNs (convolutional neural networks) detect millisecond-level behavioral markers in video with 93% accuracy on the SSBD dataset [29]; transformers identify ASD-specific gait with >92% accuracy [30]. Such pipelines reduce diagnostic delays of 18–24 months in high-income settings—and longer in low-resource areas—by up to 60% via mobile-deployable tools [31,32], promoting equitable early screening.

However, unimodal pipelines (e.g., video or eye-tracking) lack generalizability in comorbid or real-world scenarios [5,33–41]. Ahmed et al. [33] achieved 99.8% accuracy with eye-tracking alone, but performance dropped under variability. Multimodal integration yields 15%–20% accuracy gains yet increases complexity in data alignment, privacy, and validation [5,36,42–44]. Han et al. [36] reported a 95.56% accurate EEG (electroencephalography)–eye-tracking model requiring impractical synchronized acquisition. Key questions include: Can self-supervised pre-training address data scarcity? Do attention-based fusion layers provide interpretable rationales? How can pipelines robustly handle noise, bias, and missing modalities in federated networks?

This PRISMA-guided review uses the APPRAISE-AI framework to quantify gaps in reproducibility and robustness despite strong clinical relevance. We address four informatics-centric Research Questions (RQs):

- **RQ1.** How do informatics-driven feature extraction strategies handle unimodal versus multimodal neurobehavioral data to enhance diagnostic fidelity?
- **RQ2.** How do deep learning architectures leverage computer vision within end-to-end informatics pipelines for pediatric behavioral data analysis?
- **RQ3.** Which learning strategies optimize performance while preserving reproducibility and interpretability in vision-based diagnostic models?
- **RQ4.** What informatics challenges—data heterogeneity, privacy, bias, and scalability—impede multimodal integration, and which emerging solutions enable robust screening?

We synthesize why fusion dominates (complementarity reduces modality bias), benchmark underused strategies (e.g., federated learning), and propose standardized frameworks—XAI (eXplainable Artificial Intelligence) dashboards, ontology-aligned registries, containerized pipelines—for EHR (Electronic Health Records) integration. Prioritizing reproducibility, ethical governance, and clinician-in-the-loop validation, we advance AI-augmented decision support to accelerate diagnosis, lower costs, and serve global populations equitably.

2. Statement of significance

Problem: AI-driven diagnosis of ASD/ADHD lacks rigorous methodological quality assessment, hindering clinical translation of multimodal computer vision approaches.

What is Already Known: Existing reviews focus on unimodal data or technical metrics without systematic quality evaluation, missing reproducibility and bias analysis across modalities. A comparative summary is in Table 1.

What this Paper Adds: This review introduces three methodological innovations: using the APPRAISE-AI framework for quality assessment, feature co-occurrence network analysis for pipeline design, and modality-specific bias evaluation, providing a systematic quantification of reproducibility gaps (41.0%) in NDDs diagnostics.

Who would benefit: Biomedical informatics researchers, clinical practitioners, and healthcare organizations implementing scalable, bias-aware screening systems for pediatric NDDs.

3. Methodology

This systematic review followed PRISMA guidelines for transparent, reproducible study identification, selection, and synthesis. It comprised four phases: search strategy, eligibility criteria, study selection, and quality assessment.

3.1. Search strategy

We searched Web of Science (WoS) and Scopus (2020–2024) to ensure inclusion of high-impact, peer-reviewed research at the intersection of deep learning, computer vision, and NDDs. The query combined four categories using Boolean operators: (1) deep learning techniques, (2) neurodevelopmental conditions (ASD, ADHD), (3) multimodal behavioral data, and (4) feature extraction/fusion methods—additionally ANDed with (“child” OR “pediatr*” OR “adolescen*” OR “minor”).

3.1.1. Web of science (WoS) query

```
TS=(("deep learning" OR "machine learning" OR "convolutional
neural network" OR "CNN" OR "recurrent neural network" OR "RNN" OR
"transformer" OR "vision transformer" OR "VT" OR "multimodal model")
AND ("autism spectrum disorder" OR "autistic" OR "ASD" OR "attention-
deficit/hyperactivity disorder" OR "ADHD" OR "neurodevelop*" OR
"neurolo* disorder") AND ("video" OR "image" OR "eye-track*" OR
"voice" OR "facial expression" OR "body move*" OR "behavi* analysis")
AND ("feature extraction" OR "behavi* feature extraction" OR
"multimodal fusion") AND ("child" OR "pediatr*" OR "adolescen*" OR
"minor"))
```

Table 1
Summary of Existing Systematic Reviews.

Study	ML Techniques	Features Extraction	DL Techniques	Vision Features	Unimodal Techniques	Multimodal Fusion	APPRAISE-AI / Co-occurrence	Specific Disorders
[45]	✓	✓	✓	✓	✓			NDDs
[46]	✓	✓			✓			ASD
[47]	✓	✓			✓	✓		ASD, ADHD
[48]	✓	✓	✓	✓	✓	✓		Genetic, NDDs
[49]	✓	✓	✓		✓			Psychiatric, Mental Health
[50]	✓	✓	✓		✓	✓		Mental Health
[51]	✓	✓	✓			✓		Mental Health
[52]	✓	✓	✓			✓		Mental Illness
[53]	✓	✓	✓		✓	✓		ASD
Our Study	✓	✓	✓	✓	✓	✓	✓/✓	ASD, ADHD, and NDDs

3.1.2. Scopus query

```
TITLE-ABS-KEY (("deep learning" OR "machine learning" OR
"convolutional neural network" OR "CNN" OR "recurrent neural
network" OR "RNN" OR "transformer" OR "vision transformer" OR "VT"
OR "multimodal model") AND ("autism spectrum disorder" OR "autistic"
OR "ASD" OR "attention-deficit/hyperactivity disorder" OR "ADHD" OR
"neurodevelop*" OR "neurolo* disorder") AND ("video" OR "image" OR
"eye-track*" OR "voice" OR "facial expression" OR "body move*" OR
"behavi* analysis") AND ("feature extraction" OR "behavi* feature
extraction" OR "multimodal fusion") AND ("child" OR "pediatr*" OR
"adolescen*" OR "minor")) AND PUBYEAR > 2018 AND (LIMIT-TO (SUBJAREA,
"COMP") OR LIMIT-TO (SUBJAREA, "MEDI") OR LIMIT-TO (SUBJAREA, "ENGI")
OR LIMIT-TO (SUBJAREA, "PSYC")) AND (LIMIT-TO (DOCTYPE, "ar")) AND
(LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (OA, "all")))
```

3.2. Eligibility criteria

3.2.1. Inclusion

(i) Studies using feature engineering + classical machine learning (ML) or deep learning (e.g., Convolutional Neural Network (CNN) / Recurrent Neural Network (RNN), transformer, multimodal models) for detection of neurodevelopmental conditions (ASD/ADHD).
(ii) Video/image analysis via computer vision.
(iii) Q1/Q2 peer-reviewed journal/conference articles (2020–2024), in English, from WoS/Scopus.

3.2.2. Exclusion

(i) Non-deep-learning classification studies.
(ii) Reviews, meta-analyses, editorials, book chapters, letters, or abstracts without full text.
(iii) Pre-2020 publications.

3.3. Study selection and screening

The PRISMA flow is shown in Fig. 1 and its checklist is provided in Supplementary File S1. WoS yielded 87 records; after applying article/index filters, 56 remained; Q1/Q2 filtering left 30. Scopus returned 35; after deduplication and filtering, 13 remained. Full-text review against the criteria selected 43 out of 122 studies for synthesis.

3.4. Methodological quality and bias assessment protocol

We used APPRAISE-AI [54] to evaluate 43 studies across six domains: Clinical Relevance, Data Quality, Methodological Conduct, Robustness of Results, Reporting Quality, and Reproducibility. Two authors (A.R., F.A.) independently extracted data using a standardized form and scored each study; inter-rater agreement was strong (Cohen's $\kappa = 0.84$), with discrepancies resolved by consensus (C.B., S.A.). Each study was scored item-by-item; domain sums yielded overall quality (0-100): Very High (80-100), High (60-79), Moderate (40-59), Low (20-39), Very Low (0-19). Item-level scores for all studies are provided in the Supplementary File S2 to ensure standardized, reliable evaluation of clinical applicability and reproducibility. This review was not prospectively registered. Following the APPRAISE-AI framework, we define: (1) reproducibility as the availability of code, data, and preprocessing details enabling independent verification; (2) robustness as consistent model performance across error analysis, bias assessment, and sensitivity testing; and (3) generalizability as external validity across populations, settings, and modalities.

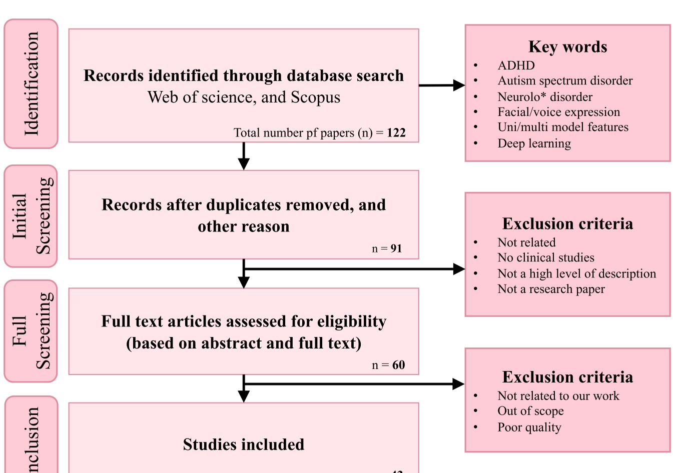

Fig. 1. PRISMA (Preferred Reporting Items for systematic Reviews and Meta-Analyses) 2020 Flow Diagram illustrating the study selection process.

4. Results

Diagnostic informatics pipelines integrate multimodal sources—facial/MRI (Magnetic Resonance Imaging), EEG/fNIRS signals, and video tracking—via standardized extraction (PCA (Principal Component Analysis), wavelets, or CNNs/RNNs/hybrids). Features undergo early/late fusion to boost accuracy. Transfer learning, attention, and continual adaptation enhance generalization across ASD, ADHD, and cognitive profiles (Fig. 2).

Study outcomes are synthesized in Table 2, covering data provenance, feature engineering, architecture, strategies, validation, clinical findings, informatics contributions, and gaps—addressing RQs on design, interpretability, reproducibility, and scalability.

4.1. Quantitative synthesis

Feature development analysis (Fig. 3a) shows a dominant preference for fusion and hybrid methods (38%), emphasizing multimodal data integration. Temporal/spatial and statistical/texture-based features are comparably adopted, reflecting their complementary roles in capturing dynamic behaviors and biomarkers. Image-based extraction shows moderate use, while dimensionality reduction and graph-based methods are underutilized (4 studies), representing missed opportunities for simplifying high-dimensional neuroimaging data. Diagnostic focus distribution (Fig. 3b) reveals a pronounced emphasis on ASD (47% of studies), compared to ADHD (19%) and both conditions (16%), highlighting a gap in unified frameworks for shared neurological markers across NDDs.

Performance across modalities (Fig. 3c) shows data fusion consistently outperforming unimodal approaches in accuracy and AUC. These trends derive from narrative synthesis of direct unimodal versus multimodal comparisons; cross-study comparability is limited by heterogeneity in datasets, metrics, and experimental conditions, precluding formal meta-analysis. While facial features achieve high specificity and EEG/fMRI (functional MRI) superior sensitivity, their combination addresses individual weaknesses. Learning strategy distribution (Fig. 3d) shows transfer learning dominating (14 papers), followed by fusion (11) and ensembling (10). Advanced methods—federated learning, reinforcement learning, and chaotic butterfly optimization—are severely underutilized (1 paper each), revealing critical gaps in privacy-preserving collaboration, adaptive learning, and hyperparameter optimization.

4.2. Feature co-occurrence on multimodal integration

To construct the co-occurrence network (Fig. 4), we systematically extracted all modality pairs from the 43 reviewed studies and computed edge weights as the frequency with which each pair co-occurred in the same pipeline. Only pairs with co-occurrence frequency ≥2 were included to ensure meaningful synergy patterns. This quantitative approach reveals empirically derived synergy patterns guiding standardized NDDs pipeline design [51,55]. The directed graph identifies EEG as the central hub, with edge weights reflecting co-occurrence frequency driven by complementary diagnostic value and data availability [56].

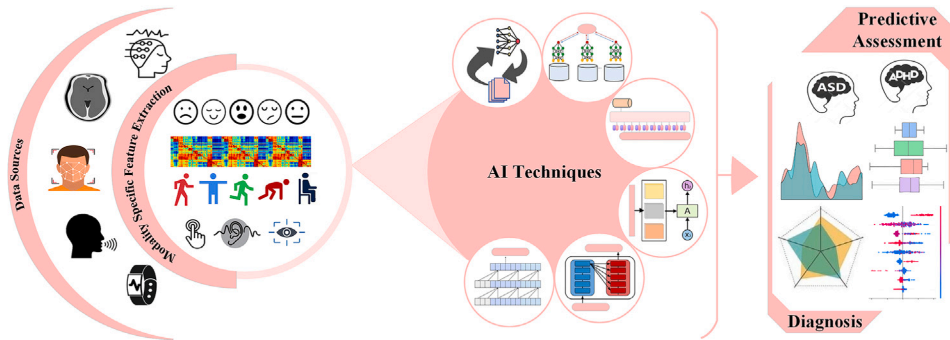

Fig. 2. End-to-end biomedical informatics pipeline AI-driven ASD/ADHD diagnosis: multimodal data processing through feature extraction, fusion, and learning strategies to support robust and integrated NDDs diagnosis.

(1) **EEG–fMRI (weight 5.0)**: This strongest link combines EEG's millisecond temporal resolution with fMRI's spatial mapping of functional connectivity, enabling joint modeling of oscillatory and network-level disruptions [57–59]. fMRI provides crucial anatomical grounding for EEG source localization, enhancing clinical interpretation [60].
(2) **EEG–Eye-Tracking (weight 4.0)**: High co-occurrence stems from shared attention circuitry—EEG detects neural correlates of sustained attention while eye-tracking quantifies behavioral output. Fusion yields superior sensitivity to joint attention deficits in ASD [7,36,61].
(3) **Facial Features–fMRI (weight 3.0)**: Links facial dynamics with fMRI hypoactivation in social-emotional regions, enabling dual-level phenotyping beyond EEG's scope [17,54,62,63]. This supports ontology-mapped fusion for explainable models [51,55].

The network reveals critical informatics insights:

- EEG hub dominance, while reflecting acquisition advantages, risks electrophysiological bias [56].
- Sparse peripheral nodes (fNIRS, Behavioral) indicate standardization gaps that limit federated learning [18].
- Missing links (e.g., Facial–fNIRS) suggest untapped synergy for modeling emotional regulation [51].

For translational impact, this analysis mandates evidence-based modality selection: prioritize EEG-inclusive triplets for comorbidity, enforce Brain Imaging Data Structure (BIDS)-compliant fNIRS integration, and develop ontology-linked feature registries for automated, reproducible multi-site pipelines [51,55,56,64,65].

4.3. Synthesis of methodological quality and bias

The APPRAISE-AI evaluation of 43 studies reveals critical gaps in translational readiness (Fig. 5). While Clinical Relevance scored highest (72.8% of maximum possible score across all studies), indicating clear diagnostic objectives, and Reporting Quality was adequate (66.1%), fundamental methodological weaknesses prevail. Data Quality (54.5%) and Methodological Conduct (51.0%) were moderate, reflecting inconsistent validation and data provenance. Most critically, Robustness of Results (45.1%) and Reproducibility (41.0%) scored lowest, exposing pervasive deficiencies in error analysis, bias mitigation, and—most severely—the availability of open-source code and datasets that prevent external validation. Analysis of specific reproducibility dimensions reveals that complete source code was available for only 18.6% of studies, data accessible via repository in 37.2%, preprocessing details reported in 72.1%, and only 7.0% provided executable end-to-end replication packages. These findings underscore that reproducibility, not clinical relevance, remains the primary barrier to deploying AI tools in clinical informatics pipelines for NDDs diagnosis. Individual scores are in the Supplementary File S2.

Analysis of bias resilience across modalities (Table 3) reveals critical vulnerabilities. Facial imaging shows 0% diversity across bias indicators, creating substantial cultural overfitting risks. While fMRI demonstrates better geographic representation (71% multi-country), it lacks low/middle-income cohort inclusion. EEG balances moderate universality (44% multi-country) with poor explainability (33% XAI adoption). Multimodal fusion, while reducing single-modality bias, maintains significant opacity (60% error analysis, 20% XAI). These findings necessitate standardized, bias-aware informatics pipelines that integrate XAI, federated learning, and ontology-driven governance to ensure equitable diagnostic performance.

5. Research questions analysis

5.1. Feature extraction in unimodal vs. multimodal informatics pipelines (RQ1)

Computer vision-based feature extraction serves as a foundational component of biomedical informatics pipelines for NDDs diagnosis, converting heterogeneous inputs into structured, ontology-aligned representations for reproducible clinical decision support [51,77–79]. Unimodal approaches offer computational efficiency and targeted biomarker discovery but suffer from poor generalization across diverse populations [33,55,80]. For instance, facial analysis achieves high specificity in controlled settings [17,33] but degrades due to cultural biases in expression norms and sensor variability [31,81–83].

Multimodal fusion addresses these limitations by leveraging complementary data streams—spatial-temporal dynamics from computer vision, electrophysiological signals from EEG/fNIRS, and functional connectivity from fMRI—to construct robust, bias-mitigated embeddings [36,51,55,61,84]. This integration exploits fundamental modality synergy: facial imaging provides high-resolution behavioral phenotyping while EEG captures millisecond-level neural oscillations, collectively reducing false positives/negatives in heterogeneous NDDs presentations [66]. The success of fusion approaches, demonstrated by 15–20% improvement in ASD detection [61] (derived from narrative synthesis of studies with direct unimodal comparisons; see Section 4.1) and 94.0% comorbidity screening accuracy [66], stems from systematic data harmonization including temporal alignment, standardized preprocessing, and clinical ontology mapping [51,67,72,85,86].

However, multimodal pipelines introduce critical informatics barriers: synchronized data acquisition remains clinically impractical, high-dimensional inputs strain computational resources in low-resource settings, and inconsistent feature normalization undermines reproducibility—reflected in the low 41.0% APPRAISE-AI robustness score due to absent data dictionaries and code repositories [55, 56]. Future frameworks must therefore incorporate federated feature

Table 2
Summary of included studies employing deep learning-driven computer vision methodologies in biomedical informatics for the diagnosis of ASD and ADHD: datasets, feature spaces, models, learning strategies, performance metrics, key findings, and research gaps.

Article	Dataset	Feature Space	Model	Learning Strategies	Performance	Findings	Research Gap
[66]	fNIRS (functional Near-Infrared Spectroscopy)	Signals from the 16-channel	CNN-BiLSTM	Fusion	Accuracy = 94.0% Sensitivity = 89.7% Specificity = 97.8% F1-score = 93.3% AUC = 0.938	Novel hybrid CNN-BiLSTM model integrating spatial and temporal features from fNIRS	limited to ADHD-ASD comorbidity, needs expansion to other comorbidities.
[80]	Gaze Capture	Eye movement tracking	CNN	Fusion	Accuracy = 97.53% Sensitivity = 96.03%	Novel application of AI for ADHD screening, comprehensive analysis of eye movement	Longitudinal studies, multimodal data integration for enhanced accuracy.
[17]	Autism Image Data	Facial expression	Hybrid CNN	Fusion	Accuracy = 98.8% Sensitivity = 99% Specificity = 99.1% AUC = 99.25% Precision = 98.9%	Early ASD detection using hybrid CNN models through facial expressions.	The study addresses the lack of effective fusion techniques for combining features from multiple CNN models to enhance early autism spectrum disorder detection.
[29]	Self-Stimulatory Behavior (SSBD)	Spatiotemporal features	VGG-16-LSTM	Transfer Learning	Accuracy = 93% Sensitivity = 93% Specificity = 100%	Deep learning for real-time behavior analysis to detect ASD-related behaviors (e.g., hand flapping)	Performance may degrade with low-quality videos or unseen behavioral variants.
[11]	Eye-tracking Scan Path (ETSP)	Eye-tracking Visual patterns	CNN	Ensembling	Accuracy = 96% Sensitivity = 85.4% F1-score = 82.2%	Bayesian optimization with CNN for ASD detection using eye-tracking data.	Model robustness to data quality and environmental factors
[4]	ADHD-200 Consortium	fMRI time series	Auto-Encoder	Ensembling	Accuracy = 99.8% Sensitivity = 99.7% Specificity = 99.9%	Used identical architecture and training settings for all four datasets, requiring no site-specific tuning to achieve high accuracy	Feature selection is still hand-crafted, using SVM rather than end-to-end deep learning, which may miss non-linear patterns
[91]	Human Connectome Project (HCP)	T1-weighted MRI	2D CNN	Transfer Learning	Accuracy = 95.06% Sensitivity = 95.87% Specificity = 94.08% AUC = 0.985	Transfer learning significantly boosts performance	Only structural MRI was used; extending to fMRI or multimodal data could further improve the performance
[85]	Multi-site datasets	3D T1-weighted MRI volumes	3D CNN	Transfer Learning	Accuracy = 97% Recall = 98.50% Specificity = 98% F1-score = 96.24%	Brain segmentation with no significant performance drop on unseen data	Pediatric under-2 brains: anatomical priors may not generalize to infants < 2 years (distinct morphology).
[64]	In-house ASD dataset	Imaging data	VGG16 (Visual Geometry Group)-CNN	Transfer Learning	Accuracy = 81% AUC = 89%	The model outperformed human ASD experts, highlighting a unique capability to discern ASD-related visual attention deficits	The study had a small sample size, with only 16 participants with ASD, which limits generalizability.
[95]	Self-reported Mental Health Diagnoses (SHMD)	Linguistic features	BiLSTM	Fusion	Accuracy = 68.02% F1-score = 67.12% Precision = 67.77% Recall = 66.58%	The model trained on interpretable features outperforms the transformer-based MentalRoBERTa model in terms of model transparency	The study acknowledges that the lack of sociodemographic factors in the datasets might limit the models' performance and generalizability.
[67]	ABIDE-I ABIDE-II (Autism Brain Imaging Data Exchange)	fMRI images	CNN	Transfer Learning	Accuracy = 79.09% Sensitivity = 80.71% Specificity = 78.71%	The study developed a deep learning-based tool for early detection of ASD using brain imaging data.	While progress in ASD detection has been made, challenges remain, especially in developing models that can effectively handle diverse age groups and resources.
[5]	In-house	Head Pose + Gaze Parameters	Machine Learning	Ensembling	Accuracy = 98% F1-score = 98% Precision = 98% Recall = 99%	The model using head pose and gaze parameters effectively differentiates between attention and inattention in children with ASD and neurotypical children	The model is primarily designed for neurotypical children and is not directly applicable to children with ASD due to different attention dynamics and frequent head movements
[96]	In-house	Gaze fixation patterns	LSTM	Transfer Learning	Precision = 59.5% Recall = 65.6%	Gaze fixation and scanning pattern differences indicate ASD and may aid in diagnosis.	Poor lighting and camera movement in video data challenge accurate gaze analysis.
[94]	In-house	Eye movement tracking	LSTM	Transfer Learning	Accuracy = 92.6% Sensitivity = 91.9% Specificity = 93.4%	The model outperformed traditional machine learning methods	Extending the dataset with multimodal data is suggested for better feature extraction.
[76]	ADNI (Alzheimer's Disease Neuroimaging Initiative) and ABIDE dataset	MRI volumes	3D CNN	Transfer Learning	Accuracy = 85%	The method identified key brain regions critical to classification performance	Incorporating multimodal neuroimaging data, like combining functional and structural MRI, will enhance performance

(continued on next page)

Table 2 *(continued)*

Article	Dataset	Feature Space	Model	Learning Strategies	Performance	Findings	Research Gap
[35]	In-house	Facial expression	LSTM	Fusion	Accuracy = 88.74%	The study shows that the hybrid model greatly improves autism identification accuracy compared to a coarse-grained approach	Traditional facial expression analysis methods have limited representation capacity and struggle with intraclass variability in autism
[3]	ChAMP (Childhood Assessment and Management of Digital Phenotypes)	Behavioral & demographic features	Machine Learning	Transfer Learning	Accuracy = 73% Sensitivity = 53% Specificity = 92% AUC = 77%	Combining objective physiological features and behaviors across tasks can enhance the classification of children's mental health conditions like anxiety, depression, and ADHD.	Refine features, integrate physiological signals, and explore advanced modeling for optimizing early childhood mental health diagnosis.
[58]	ABC-CT (Autism Biomarkers Consortium for Clinical Trials) resting EEG	Brain network features	Hybrid CNN	Fusion	Accuracy = 87% Sensitivity = 85.32%	Rest-HGCN (Hybrid Graph Convolutional Network) effectively captures connectivity patterns in resting EEG and outperforms previous EEG-based ASD detectors, showing its potential for clinical use.	Integration of resting and task-state EEG for abnormal brain state modeling.
[36]	In-house	EEG Features	Auto-Encoder	Fusion	Accuracy = 95.56% Sensitivity = 92.5% Specificity = 98% AUC = 98.4%	The MMSDAE (Multimodal Stacked Denoising Autoencoder) model outperforms unimodal methods (EEG-SDAE, ET-SDAE) and the feature-level fusion method (CONCAT-SDAE) in identifying ASD in children.	Simultaneous EEG and ET data availability for model use.
[69]	In-house	Facial Expressions	DCNN	Fusion	Accuracy = 99.99% Precision = 99.97% Recall = 100% Specificity = 99.98%	The Enhanced Deep Learning technique surpasses previous CNNs for facial emotion recognition in autistic children.	The study's limitation is the reliance on a small dataset.
[73]	Visualization of Eye-Tracking Scan Paths	Deep visual features	LSTM	Chaotic Butterfly Optimization	Accuracy = 99% Precision = 98.22% Sensitivity = 98.25% Specificity = 98.85%	The U-Net segmentation, Inception v3 embeddings, Chaotic butterfly optimization, and LSTM gaze patterns to capture eye movement cues.	The study lack a clear link between eye-movement patterns and ASD behaviors, as high variability complicates interpretation.
[70]	GeoPref (Geometric Preference) Test dynamic stimulus eye-tracking	Gaze/eye-movement	CNN-LSTM	Fusion	Accuracy = 98.59% AUC = 96% Sensitivity = 100% Specificity = 76.47%	An ACLNet + SVM (Support Vector Machine) pipeline on dynamic GeoPref stimuli effectively distinguishes ASD from TD and predicts ASD symptom severity with top performance.	Preschoolers should be tested across ages, real-world environments, and various eye-trackers in controlled lab settings.
[61]	In-house	EEG features	CNN	Reinforcement Learning	Accuracy = 85.75% AUC = 94% Sensitivity = 88.43% Specificity = 85.71%	Joint multi-task reinforcement learning significantly outperforms single-task and manually crafted deep models for EEG-based ASD comorbidity discrimination	User-friendly interfaces and multicenter studies are still needed to assess deployment feasibility.
[101]	In-house	EEG features	Auto-Encoder	Transfer Learning	Accuracy = 98.88% F1-score = 99.19% Sensitivity = 100% Specificity = 96.40%	A novel sparse-coding signal-to-image mapping combined with a fine-tuned ResNet automatically distinguishes ASD from neurotypical EEG with reliability	Real-time or low-cost implementations and validation on portable EEG systems are needed.
[18]	Multimodal ASD	Questionnaire-derived features	CNN–LSTM	Federated Learning	Accuracy = 93% Sensitivity = 95%	Automated ASD detection with minimal computing overhead by integrating multimodal datasets	Exploring attention-based or graph fusion could enhance interpretability and performance
[75]	ADHD-200 consortium dataset	Neuroimaging features	CNN-LSTM	Fusion	Accuracy = 98.12% F1-score = 97.72% Sensitivity = 97.50% Specificity = 97.72%	A hybrid CNN–LSTM model outperforms a standalone 2D CNN in discriminating ADHD from controls using resting-state fMRI images.	Using only 2D slices with 3D context or functional connectivity metrics could enhance performance.
[84]	ADHD-200 consortium dataset	Neuroimaging features	SVM	Ensembling	Accuracy = 86.43%	The three-objective SVM scheme effectively addresses dataset imbalance and outperforms standard SVM, Random Forest, and ELM across all ADHD-200 sites.	Incomplete metrics, precision, F1-score, and AUC were not reported, limiting a complete comparison with other methods.

(*continued on next page*)

Table 2 *(continued)*

Article	Dataset	Feature Space	Model	Learning Strategies	Performance	Findings	Research Gap
[98]	Self-Stimulatory Behavior Dataset (SSBD)	Visual features	CNN	Ensembling	Accuracy = 98.3%	By combining unsupervised feature extraction with TCDN (Temporal Coherency Deep Network) and supervised classification with SVM, the classification of self-stimulatory behaviors in children with autism	The model needs more testing in uncontrolled environments with dynamic data for real-time detection.
[57]	ADHD-200 consortium	fMRI time-series features	DCNN	Ensembling	Accuracy (NYU) = 73.1% Accuracy (Peking) = 67.9% Accuracy (NI) = 67.9%	A deep learning for classifying ADHD from fMRI data, highlighting the importance of functional connectivity as a key biomarker for ADHD classification	The main challenge is data heterogeneity across imaging sites (NYU, Peking, and NI), impacting the model's generalizability.
[33]	ASD vs TD eye-tracking	Spatial visual gaze features	DCNN	Fusion	Accuracy = 99.8% F1-score = 99.4% Sensitivity = 99% Specificity = 100% F1 = score = 99.4%	Combining classical feature-engineering with FFNNs yields near-perfect ASD vs TD discrimination	Although CNNs (GoogLeNet, ResNet-18) can extract deep gaze patterns directly, they demand heavy computing and long training.
[72]	ADHD-200 consortium	Visual features	CNN	Ensembling	Sensitivity = 93.28% Specificity = 91.38% AUC = 97.0% F1-score = 93.86%	A CNN-DNN applied to eye-tracking scan-path images can effectively differentiate ASD from TD children, achieving high AUC	The small dataset (59 participants) lacks diversity, necessitating larger, varied samples across age groups
[68]	ASD	Spatial features	CNN	Transfer Learning	Accuracy = 88.89%	A novel Joint Energy Image (JEI) that integrates 3D joint positions with temporal and depth features outperforms traditional kinematic models	The dataset would benefit from greater diversity and size. Real-world applicability can improve with controlled environments and better noise-handling techniques, while further generalization can be achieved through varied testing conditions
[63]	EEG recordings	EEG features	BiLSTM	Ensembling	Accuracy = 89.7%	The architecture classified ASD from controls using EEG waveform data, achieving better performance with fewer parameters than Transformer-based models	The diversity of the ASD population, including factors like age, sex, and co-occurring disorders, needs more focus
[112]	In-house	Radiographic features	Mask-RCNN (Mask Region-based Convolutional Neural Network)	Generative Models	Accuracy = 91.5% Sensitivity = 84%	The model demonstrated high accuracy in automatically measuring key sagittal balance parameters in ASD	Future work should test the algorithm on diverse radiographic settings and spinal deformities beyond ASD, and incorporate more spinopelvic parameters to enhance clinical application.
[1]	ADHD-200 consortium dataset	Topological features	CNN	Transfer Learning	Accuracy = 81.6% Sensitivity = 78.3% AUC = 81.0%	The model effectively improves the prediction of neurodevelopmental disorders by using multi-filter convolutional layers	The model should be validated with larger datasets and diverse neurological conditions
[97]	Continuous Performance Test (CPT)	Custom features	Hidden Markov Model	Ensembling	Accuracy = 93.68% Sensitivity = 90.66% Specificity = 87.72% F1-score = 87.78	The proposed model successfully predicts ADHD with improved accuracy and reduced batch effects	The model applicability to diverse populations and settings should be explored, and integrating fuzzy inference systems with behavioral assessments could enhance its predictive power.
[60]	ADHD-200 dataset	Visual features	GradCAM (Gradient-weighted Class Activation Mapping)	Fusion	Accuracy = 72.41%	The model outperforms state-of-the-art methods in classifying and interpreting brain disorders	There is a need for further generalization of the MDCN (Multivariate Distance-based Connectome Network) approach to larger, more diverse clinical populations
[42]	In-house	EEG Features	CNN	Transfer Learning	Accuracy = 97.35%	The study presents a novel framework using EEG-based SPWVD-TFD (Smoothed Pseudo Wigner-Ville Distribution-based Time-Frequency Distribution) images with the ASD-Net model, achieving high accuracy for ASD detection	Real-time clinical applications and reducing computational complexity are key future directions.

(*continued on next page*)

Table 2 *(continued)*

Article	Dataset	Feature Space	Model	Learning Strategies	Performance	Findings	Research Gap
[74]	Autistic Children Video Dataset (ACVD)	Visual features	LSTM	Generative Models	Accuracy = 98.8% Sensitivity = 91.1% Specificity = 96.7%	The method provides an effective way to estimate gaze direction from raw video, outperforming existing methods for ASD diagnosis based on gaze patterns	The model is susceptible to issues like illumination changes and background clutter, which need more robust preprocessing and gaze estimation
[2]	ADHD-200 dataset	Visual features	GAN	Generative Models	Accuracy = 75.1%	The model identifies multiple mental disorders from multimodal MRI data using a multioutput conditional GAN (Generative Adversarial Network) for data augmentation and a gating fusion model for classification.	Using multi-atlas combinations for more generalized results can improve feature selection based on the empirical choice of ROIs and atlases.
[100]	ADHD-200 dataset	fMRI features	VAE (Variational Autoencoder)-GAN	Generative Models	Accuracy = 75.1%	The model effectively identifies ASD and ADHD by leveraging multimodal MRI data and attention-based fusion methods for improved diagnostic performance	Future work could explore more advanced feature selection techniques, especially for handling large datasets and improving robustness.
[92]	In-house	EEG Features	LSTM	Ensembling	Accuracy = 92.15% Sensitivity = 90.95% Specificity = 93.43%	The study suggests that spectral features and specific EEG recording statuses significantly improve ADHD classification accuracy	Future work may include optimization methods such as arithmetic, grasshopper, and sine cosine algorithms for improved feature selection.
[93]	Eye-tracking	Eye-tracking features	DCNN (Deep Convolutional Neural Network)	Transfer Learning	Accuracy = 100%	Eye-tracking technology shows promise for diagnosing ASD by capturing distinct visual attention patterns in affected children.	Future research could further optimize the model architectures and investigate the integration of additional eye-tracking features to improve ASD detection accuracy.

extraction protocols [18], explainable fusion mechanisms for clinical trust [87], and containerized preprocessing modules for EHR integration, transforming feature extraction from ad-hoc signal processing into a reproducible, equitable pillar of AI-augmented NDDs screening.

5.2. Deep learning architectures in end-to-end diagnostic informatics pipelines (RQ2)

Deep learning architectures have transformed biomedical informatics by embedding computer vision for automated, scalable conversion of multimodal pediatric data into standardized diagnostic biomarkers [51,55,88–90]. Unlike traditional machine learning which requires manual feature engineering [72], modern architectures (CNNs, RNNs, transformers) learn hierarchical representations directly from raw data, capturing essential spatiotemporal dynamics for neurobehavioral analysis [29,59]. This addresses the critical informatics challenge of converting heterogeneous behavioral inputs into ontology-aligned, EHR-integrable features [51,56,91].

These architectures succeed through modality-specific optimization and cross-modal alignment [55], demonstrated via three key paradigms: feature extraction as phenotype mapping, where 3D CNNs map behavioral videos to clinical phenotypes (e.g., stereotyped movements) for standardized biomarker registries [51,68]; multimodal fusion via latent space harmonization, where autoencoders project EEG and eye-tracking into shared embeddings, improving ASD detection by 12%–18% over unimodal approaches [36,55,61,92–94]; and attention mechanisms for interpretable weighting, where vision transformers highlight diagnostically relevant regions (e.g., reduced face fixation) to produce clinician-readable rationales [73,87,95].

The informatics pipeline ensures reproducibility through systematic governance: preprocessing for interoperability via standardized normalization enabling federated multi-site training [18,33]; hierarchical representation learning, where CNNs map facial dynamics to clinical severity scores for EHR integration [69]; and temporal modeling using bidirectional LSTMs to capture gaze evolution as prognostic markers of joint attention deficits [59,74,96,97].

Advanced strategies address translational bottlenecks: federated learning preserves privacy while enabling multi-site generalization [18,23,51]; explainable AI (Grad-CAM) visualizes gaze clusters driving ASD risk scores to bridge AI opacity with clinical intuition [56,70,87,98]; and multiscale connectivity analysis combines CNNs and LSTMs to reveal local activation and network-level disruptions for precision subtyping [59,75,99].

Deep learning thus positions computer vision as a core clinical informatics engine, standardizing data ingestion while ensuring reproducibility, privacy, and interpretability for scalable NDDs diagnosis [23,51,55,56].

5.3. Optimizing performance with reproducible AI strategies (RQ3)

Transfer learning, multimodal fusion, and ensemble methods dominate NDDs diagnosis by addressing core informatics challenges: data scarcity, modality heterogeneity, and model variance [51,55,56].

1. Transfer Learning: Pre-trained CNNs (VGG, ResNet) overcome pediatric data limitations (<500 samples) by transferring visual hierarchies to NDDs domains, achieving 79.09%–99.29% accuracy while reducing overfitting in resource-constrained settings [18,51,67,73].
2. Multimodal Fusion: Integrating spatial (vision), temporal (EEG/fNIRS), and connectivity (fMRI) features captures NDDs heterogeneity through modality synergy, achieving 85–99.8% accuracy by combining behavioral phenotyping with neural dynamics [17,33,36,55,71,100,101].
3. Ensemble Learning: Aggregating diverse learners enhances robustness through variance reduction, achieving 0.928 Dice scores and 72.41% classification accuracy, though dependent on model diversity and computationally intensive [54,60,102–104].

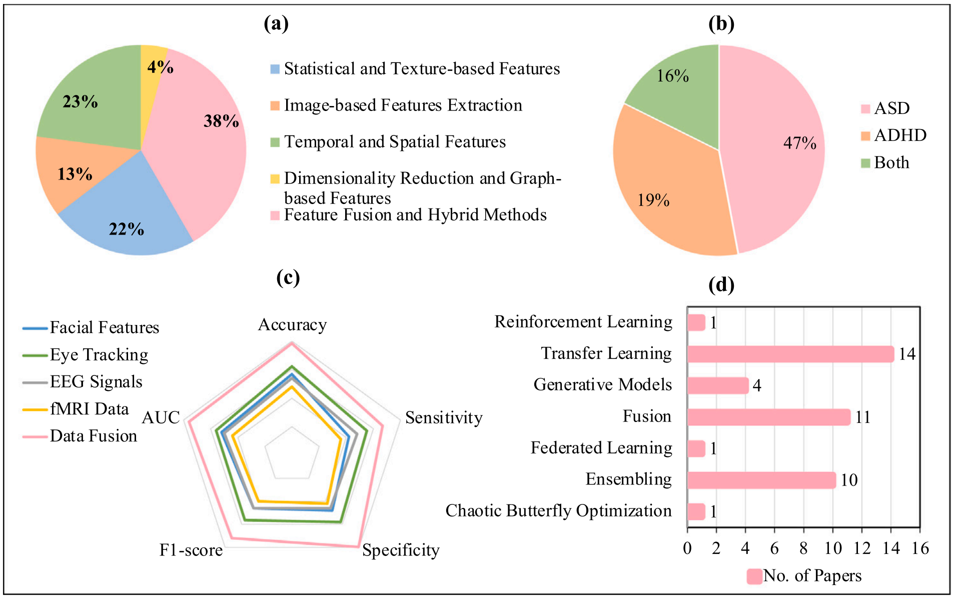

Fig. 3. Quantitative synthesis of AI methodologies for NDDs diagnosis across 43 studies. (a) Feature development techniques: hybrid and fusion methods dominate (38%), reflecting the field's emphasis on multimodal integration. (b) Diagnostic focus: ASD is disproportionately studied (47%) compared to ADHD (19%) or combined conditions (16%), revealing a gap in unified diagnostic frameworks. (c) Performance across modalities: data fusion achieves higher accuracy and AUC than unimodal approaches by integrating complementary information. (d) Learning strategies: transfer learning (14 studies) is most common, while federated learning, reinforcement learning, and XAI remain critically underutilized—limiting privacy preservation, adaptability, and clinical interpretability.

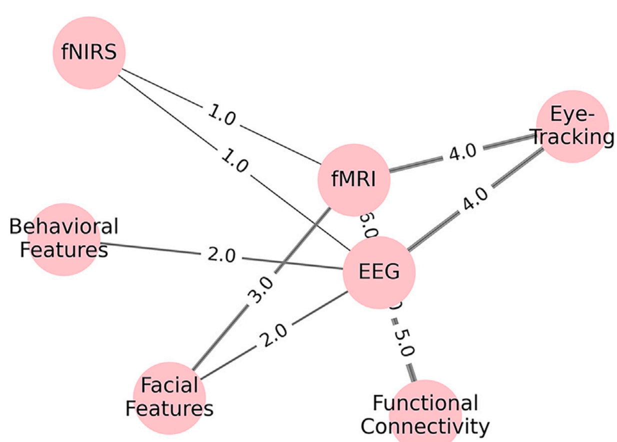

Fig. 4. Feature co-occurrence network (directed, weighted) of different features. Edge thickness = co-occurrence frequency. EEG emerges as the central integration hub, EEG-eye-tracking (weight 4.0) showing the strongest synergy for multimodal NDDs diagnosis.

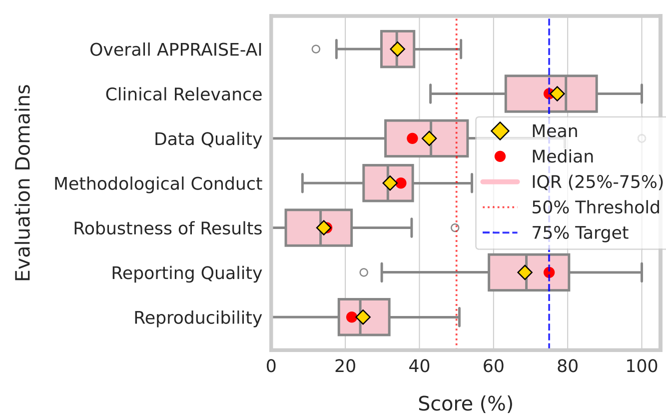

Fig. 5. Methodological quality assessment across 43 studies using APPRAISE-AI. Clinical relevance scores highest (72.8%) while reproducibility (41.0%) and robustness (45.1%) score lowest, identifying these as primary barriers to clinical deployment.

Despite strong performance, critical gaps remain: transfer learning risks demographic bias, fusion requires impractical synchronized data, and ensembles increase computational costs. Underutilization of federated learning [18] and explainable AI limits privacy preservation and clinical trust [23,56,87], reflected in poor robustness scores (45.1%) due to missing reproducibility materials [56].

Future frameworks must integrate federated transfer learning, XAI-enhanced ensembles, and containerized registries for deployable pediatric diagnostic pipelines [18,51,55].

5.4. Overcoming informatics barriers in multimodal integration (RQ4)

Multimodal integration for NDDs screening faces fundamental informatics barriers—data heterogeneity, privacy constraints, computational demands, and interpretability gaps—that hinder clinical translation [51,55,56]. These stem from systemic issues in data governance rather than algorithmic limitations [18,23]. Data heterogeneity arises from inconsistent acquisition protocols and annotation schemas that fragment multimodal datasets, with only 41.0% of studies meeting robustness criteria due to missing standardization [54,56,57,105]. Data scarcity and privacy restrictions limit pediatric datasets to <100 cases, preventing centralized training and creating institutional silos with biased models [18,106]. Computational bottlenecks from high-dimensional data and temporal misalignment increase hardware demands, making real-time screening impractical in low-resource settings [33,61,107]. Interpretability and bias gaps leave black-box models lacking clinical rationale, performing poorly on underrepresented populations beyond WEIRD cohorts [60,87,108].

Table 3
Comparative analysis of bias risks across modalities in deep learning-driven NDDs diagnosis.

Modality	Studies References	Potential Biases	Evidence from APPRAISE-AI	Suggested Mitigation
Facial Imaging	[17,33]	High cultural/demographic bias (facial expressions vary by ethnicity, culture, age); sensitive to lighting, pose, occlusion; overfits to WEIRD (Western, Educated, Industrialized, Rich, and Democratic) datasets.	Multiple countries: 0%, Low/middle income: 0%, Subgroup analysis: 0%, Error analysis: 0%, Model explanations: 0%	Diverse global datasets (e.g., Child Affective Facial Expression (CAFE) [44]); XAI (SHapley Additive exPlanations (SHAP) or Gradient-weighted Class Activation Mapping (Grad-CAM)); federated learning across cultures; bias audits per ISO/TR 23265.
Eye-Tracking	[33,36,61,66–71]	Cultural norms in gaze (e.g., eye contact avoidance in some cultures); sensor variability; underrepresents comorbidities (e.g., ADHD + ASD); calibration bias in young children.	Multiple countries: 11%, Low/middle income: 0%, Subgroup analysis: 11%, Error analysis: 0%, Model explanations: 0%	BIDS-EEGLAB standardization; include low-resource cohorts; error analysis by cultural subgroup; transfer learning from adult gaze models.
EEG	[36,54,61,66,67,72–75]	Less cultural but artifact-prone (movement, sweat in children); demographic bias in cohorts; high computational load; missing channels in real-world settings.	Multiple countries: 44%, Low/middle income: 0%, Subgroup analysis: 0%, Error analysis: 11%, Model explanations: 33%	Independent Component Analysis (ICA)-based artifact removal; federated multi-site training; XAI (e.g., Deep Learning Important FeaTures (DeepLIFT) on time-frequency maps); low-cost wearable EEG (e.g., Muse).
fMRI	[54,57,67,72–75]	Access bias (expensive, unavailable in low-resource settings); motion intolerance in NDDs children; bias toward high-SES, high-IQ cohorts; spatial resolution limits behavioral inference.	Multiple countries: 71%, Low/middle income: 0%, Subgroup analysis: 14%, Error analysis: 29%, Model explanations: 43%	resting-state fMRI for low-cost alternatives; Fast Healthcare Interoperability Resources (FHIR)-compliant data lakes; external validation in community clinics; decision curve analysis for clinical utility.
Multimodal Fusion	[36,61,66,67,76]	Compounded biases from all modalities; synchronization errors; black-box integration hides bias sources; high data requirements.	Multiple countries: 20%, Low/middle income: 0%, Subgroup analysis: 0%, Error analysis: 60%, Model explanations: 20%	Ontology-aligned fusion (e.g., NeuroLex); XAI for modality contribution (e.g., attention weights); APPRAISE-AI-guided reporting; containerized pipelines (Docker).

Full study lists in Supplementary File S2. Metrics = % of studies in each modality scoring “Y” on APPRAISE-AI items. Metrics computed by grouping 43 studies by primary modality (Fig. 3(c) analysis) and averaging binary APPRAISE-AI scores.

Fig. 6. Translational framework for multimodal ASD/ADHD diagnosis: mapping methodological challenges to AI solutions and human collaboration roles for quality-aware integration. The framework maps primary barriers paired with corresponding emerging technological solutions in deep learning-driven multimodal data integration.

To address these challenges, emerging informatics solutions leverage AI and edge computing advances (Fig. 6). Federated learning enables multi-site collaboration without data sharing, achieving 97.5% accuracy while preserving privacy and reducing bias [18,23]. Self-supervised learning (Simple Contrastive Learning of Representations (SimCLR), Generative Adversarial Networks (GANs)) generates synthetic data, reducing annotation needs by 80% while improving generalization [69,109]. XAI integration with Grad-CAM and SHAP provides clinical rationales, while interactive dashboards support clinician validation [70,87,95]. Edge computing with quantized models enables <2s inference on wearables, facilitating deployment in low-resource clinics [61,110]. Standardized ontologies through BIDS-compliant repositories with Human Phenotype Ontology (HPO) / Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) mapping ensure interoperability across healthcare systems [60,65]. These solutions form a unified framework that transforms multimodal integration from a technical challenge into a scalable, equitable component of precision pediatric care [18,51,55,56].

6. Discussion

This systematic review identifies multimodal fusion as a prevalent informatics paradigm in NDDs diagnosis (38% of studies; Fig. 3a), driven by complementary modality resolution—EEG for temporal fidelity, fMRI for network topology, and computer vision for behavioral quantification [36,51,55,58]. However, a notable translational paradox exists, with APPRAISE-AI revealing critical gaps in reproducibility and robustness despite strong clinical relevance [56], primarily due to absent open-source code, data dictionaries, and external validation. Recent work [111] emphasizes that reproducibility requires empirical verification—including replication attempts, environment capture, dependency pinning, seed control, and documented compute environments—beyond merely sharing digital artifacts. This reframes reproducibility as an auditable continuum rather than a binary property. Furthermore, environmental efficiency aligns with our discussion of computational bottlenecks, suggesting that reproducible pipelines must also be computationally sustainable for real-world clinical deployment.

Federated learning—employed in only one study [18]—emerges as a promising solution, enabling privacy-preserving, multi-site model aggregation while mitigating data silos and demographic bias [23,51]. Similarly, underutilized XAI methods can generate clinician-auditable rationales through SHAP, Grad-CAM, or attention heatmaps [70,87,95], transforming black-box outputs into FHIR-compliant decision support artifacts essential for regulatory approval [56,112].

The feature co-occurrence network (Fig. 4) provides evidence-based modality selection guidance, indicating EEG-inclusive triplets maximize diagnostic yield. Combined with BIDS-compliant synchronization and HPO-mapped feature registries [60,65], this enables automated pipeline generation across multiple sites. Edge-optimized models further facilitate real-time screening in low-resource settings with minimal accuracy loss [61,110].

Emerging systems demonstrate potential for translation: vision-language models align imaging with clinical reports for differential diagnosis [74,113], while agentic AI reduces workflow time through optimized triage [114,115]. These operate within federated, EHR-integrated ecosystems [18,23,51,65], leveraging FHIR-compliant outputs and privacy-preserving orchestration [56].

Based on these findings, the proposed unified framework (Fig. 6)—integrating federated pre-training, standardized fusion, XAI validation, and edge deployment—directly addresses core informatics challenges in multimodal NDDs screening [18,23,51,55,56,87,116]. It operationalizes key priorities, including data provenance, interoperability, clinical utility, and bias reduction, while enabling equitable deployment through privacy-preserving pipelines.

Although the methods in Table 4 show clear methodological advantages and varying degrees of clinical relevance across modalities, none have been deployed in routine clinical practice or reported for prospective patient diagnosis. Reported patient numbers reflect retrospective validation cohorts rather than real-world use. Federated learning [18], edge-optimized lightweight architectures [61,110], and explainable AI [87,95] represent the most promising methodological pathways for clinical deployment, as they directly address the privacy, computational, and interpretability barriers identified in our APPRAISE-AI analysis. This highlights a persistent deployment gap between high algorithmic performance and clinical adoption.

6.1. Limitations of our review

This review was limited to Q1/Q2 articles from Web of Science and Scopus (2020–2024), excluding other databases and gray literature. This selection prioritized methodologically rigorous, high-impact studies with robust peer review, aligning with our focus on reproducible AI pipelines for clinical translation. While ensuring methodological rigor, this approach may have introduced selection bias by underrepresenting innovative pipelines from clinical or open-access venues [56] and omitting real-world deployment insights from technical reports and federated learning initiatives [18,51]. Additionally, restricting to Q1/Q2 journals in WoS and Scopus—including the Scopus Open Access-only filter—may have excluded relevant studies from subscription-based venues, lower-tier journals, and conferences, potentially overlooking emerging methodologies. This restriction may also underrepresent real-world deployment studies and certain modalities (e.g., behavioral sensing) that are more frequently published in clinical or specialized venues outside our search scope. This review was not prospectively registered (e.g., PROSPERO), which is acknowledged as a limitation. Despite the availability of deployment-oriented methodologies, such as federated learning and lightweight architectures, none of the reviewed frameworks have been integrated into routine clinical practice (Table 5).

6.2. Future directions

Future research must prioritize clinically viable informatics systems that extend beyond accuracy metrics. Critical directions include

Table 4
Comparative analysis of methodological diversity and performance heterogeneity across diagnostic modalities and their clinical relevance.

Modality	# of Studies	Accuracy Range	Mean Accuracy	Heterogeneity
Neuroimaging (fMRI, structural Magnetic Resonance Imaging (sMRI), T1)	12	72.4% – 99.8%	86.50%	High: Variation driven by multi-site (e.g., ADHD-200) versus single-site data and architectural depth (2D vs. 3D CNN).
Electrophysiological (EEG, fNIRS)	8	85.7% – 98.8%	92.70%	Low: Consistent performance across studies, benefiting from transfer learning and frequency-domain features.
Behavioral & Visual (Eye-tracking, Face, Video)	17	81.0% – 100%	94.60%	Moderate: High ceiling effects but sensitivity to environmental factors (e.g., lighting, calibration) and sample size.
Hybrid / Multi-modality	6	68.0% – 93.6%	83.10%	High: Performance depends on clinical instrument selection and fusion strategy.

Table 5
Summary of AI methodologies with clinical deployment Characteristics and validation Status.

Study	Learning Strategy	Deployment Advantage	Status	Cohort Size	Key Finding
[18]	CNN-LSTM (Federated)	Collaborative training across hospitals without sharing patient data	Pre-clinical prototype	$N > 1{,}000$	98% accuracy (simulated); preserves privacy in multimodal decentralized training
[12]	Gaze tracking (CV)	Uses camera instead of lab eyetrackers for home screening	Clinical validation	$N = 104$	Detected reduced social attention in ASD toddlers using consumer devices
[13]	XGBoost (ML)	Routine retinal imaging as objective biomarker, bypassing subjective surveys	Retrospective validation	$N = 323$	96.9% AUC for ADHD; stratified symptom severity via retinal vasculature
[42]	CNN (Time-Frequency Distribution (TFD))	Integrates time-frequency distribution for automated EEG artifact removal	Experimental	–	97.35% accuracy; outperformed ResNet/DenseNet in ASD detection
[26]	ViT (Video)	Captures long-range temporal correlations for subtle behavioral markers	Experimental	$N = 146$	High accuracy on 146 children (ages 2–8); learned discriminative video representations

advancing multimodal integration, interpretable learning, federated benchmarking, and longitudinal clinical tracking to overcome current deployment barriers [18,51]. Key priorities also include prospective trials and fairness-aware modeling to ensure equitable performance across diverse populations. Focusing on open-source pipelines and clinician-in-the-loop validation will accelerate global deployment of AI-augmented NDDs screening, reducing diagnostic delays while ensuring equitable intervention access.

7. Conclusion

Deep learning-driven computer vision is reshaping biomedical informatics for NDDs diagnosis, shifting from subjective assessment to objective, multimodal decision support. This PRISMA-guided [117] review of 43 studies confirms that feature fusion and hybrid pipelines (38%; Fig. 3(c)) outperform unimodal approaches by integrating complementary modalities—EEG for temporal precision, fMRI for network mapping, and eye-tracking for behavioral quantification. However, reproducibility (41%) and robustness (45.1%) remain critical barriers to translation. The underuse of federated learning and explainable AI limits privacy-compliant, clinician-trustworthy deployment. By embedding standardized data harmonization, XAI-driven interpretability, and EHR-integrated validation, these frameworks can accelerate scalable screening—transforming research prototypes into deployable clinical tools.

CRediT authorship contribution statement

Abdur Rasool: Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Faizan Ahmad:** Writing – review & editing, Visualization, Validation, Software, Investigation, Formal analysis. **Chayut Bunterngchit:** Writing – review & editing, Validation, Resources, Methodology, Data curation, Conceptualization. **Saba Aslam:** Writing – review & editing, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.ijmedinf.2026.106417.

References

[1] M. Chen, H. Li, H. Fan, J.R. Dillman, H. Wang, M. Altaye, B. Zhang, N.A. Parikh, L. He, Conceptcnn: a novel multi-filter convolutional neural network for the prediction of neurodevelopmental disorders using brain connectome, Med. Phys. 49 (5) (2022) 3171–3184.
[2] R. Liu, Z.A. Huang, Y. Hu, Z. Zhu, K.C. Wong, K.C. Tan, Attention-like multimodality fusion with data augmentation for diagnosis of mental disorders using MRI, IEEE Trans. Neural Netw. Learn. Syst. 35 (6) (2022) 7627–7641.
[3] B.C. Loftness, J. Halvorson-Phelan, A. O'Leary, C. Bradshaw, S. Prytherch, I. Berman, J. Torous, W.E. Copeland, N. Cheney, R.S. McGinnis, E.W. McGinnis, The champ app: a scalable mhealth technology for detecting digital phenotypes of early childhood mental health, IEEE J. Biomed. Health Inform. 28 (4) (2023) 2304–2313.
[4] Y. Tang, J. Sun, C. Wang, Y. Zhong, A. Jiang, G. Liu, X. Liu, ADHD classification using auto-encoding neural network and binary hypothesis testing, Artif. Intell. Med. 123 (2022) 102209.
[5] E.B. Varghese, M. Qaraqe, D. Al-Thani, Attention level evaluation in children with autism: leveraging head pose and gaze parameters from videos for educational intervention, IEEE Trans. Learn. Technol. 17 (2024) 1737–1753.
[6] S.S. Rajagopalan, K. Tammimies, Predicting neurodevelopmental disorders using machine learning models and electronic health records – status of the field, J. Neurodev. Disord. 16 (1) (2024) 63.
[7] C. Ko, J.-H. Lim, J. Hong, S.-B. Hong, Y.R. Park, Development and validation of a joint attention-based deep learning system for detection and symptom severity assessment of autism spectrum disorder, JAMA Netw. Open 6 (5) (2023) e2315174.
[8] S. Bandyopadhyay, M. Sarma, D. Samanta, AI-enabled computational intelligence approach to neurodevelopmental disorders detection using rs-fmri data, Comput. Electr. Eng. 123 (2025) 110117.
[9] K.A. Shaw, S. Williams, M.E. Patrick, E. al., Prevalence and early identification of autism spectrum disorder among children aged 4 and 8 years – autism and developmental disabilities monitoring network, 16 sites, united states, 2022, MMWR Surveill Summ 74 (SS-2) (2025) 1–22.
[10] S.L. Hyman, S.E. Levy, S.M. Myers, D.Z. Kuo, S. Apkon, L.F. Davidson, K.A. Ellerbeck, J.E. Foster, G.H. Noritz, M.O.C. Leppert, B.S. Saunders, Identification, evaluation, and management of children with autism spectrum disorder, Pediatrics 145 (1) (2020).
[11] S. Cheekaty, G. Muneeswari, Advancing autism prediction through visual-based AI approaches: integrating advanced eye movement analysis and shape recognition with kalman filtering, Vis. Comput. 41 (4) (2025) 2251–2270.
[12] M. Bovery, G. Dawson, J. Hashemi, G. Sapiro, A scalable off-the-shelf framework for measuring patterns of attention in young children and its application in autism spectrum disorder, IEEE Trans. Affect. Comput. 12 (3) (2021) 722–731.
[13] H. Choi, J. Hong, H.G. Kang, M.-H. Park, S. Ha, J. Lee, E. al., Retinal fundus imaging as biomarker for ADHD using machine learning for screening and visual attention stratification, NPJ Digit. Med. 8 (1) (2025) 164.
[14] P. Amado-Caballero, P. Casaseca-de-la Higuera, S. Alberola-López, J.M. Andrés-de Llano, J.A. López-Villalobos, C. Alberola-López, Insight into ADHD diagnosis with deep learning on actimetry: quantitative interpretation of occlusion maps in age and gender subgroups, Artif. Intell. Med. 143 (102630) (2023) 102630.
[15] A. Wiebe, B. Selaskowski, M. Paskin, L. Asché, J. Pakos, B. Aslan, E. al., Virtual reality-assisted prediction of adult ADHD based on eye tracking, EEG, actigraphy and behavioral indices: a machine learning analysis of independent training and test samples, Transl. Psychiatry 14 (1) (2024) 508.
[16] R.A.J. de Belen, T. Bednarz, A. Sowmya, D. Del Favero, Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019, Transl. Psychiatry 10 (1) (2020) 333.
[17] B. Awaji, E.M. Senan, F. Olayah, E.A. Alshari, M. Alsulami, H.A. Abosaq, J. Alqahtani, P. Janrao, Hybrid techniques of facial feature image analysis for early detection of autism spectrum disorder based on combined CNN features, Diagnostics 13 (18) (2023) 2948.

[18] A. Lakhan, M.A. Mohammed, K.H. Abdulkareem, H. Hamouda, S. Alyahya, Autism spectrum disorder detection framework for children based on federated learning integrated cnn-lstm, Comput. Biol. Med. 166 (2023) 107539.
[19] Y. Fan, H. Xiong, G. Sun, Deepasdpred: a cnn-lstm-based deep learning method for autism spectrum disorders risk RNA identification, BMC Bioinform. 24 (1) (2023) 261.
[20] F. Ramzan, M.U.G. Khan, A. Rehmat, S. Iqbal, T. Saba, A. Rehman, Z. Mehmood, A deep learning approach for automated diagnosis and multi-class classification of alzheimer's disease stages using resting-state FMRI and residual neural networks, J. Med. Syst. 44 (2) (Dec 2019).
[21] Y. Xu, Z. Yu, Y. Li, Y. Liu, Y. Li, Y. Wang, Autism spectrum disorder diagnosis with EEG signals using time series maps of brain functional connectivity and a combined cnn-lstm model, Comput. Methods Programs Biomed. 250 (2024) 108196.
[22] K. Khan, R. Katarya, Mcbert: a multi-modal framework for the diagnosis of autism spectrum disorder, Biol. Psychol. 194 (2025) 108976.
[23] H. Wang, H. Jing, J. Yang, C. Liu, L. Hu, G. Tao, E. al., Identifying autism spectrum disorder from multi-modal data with privacy-preserving, Npj Ment. Health Res. 3 (1) (2024) 15.
[24] S. Vidivelli, P. Padmakumari, P. Shanthi, Multimodal autism detection: deep hybrid model with improved feature level fusion, Comput. Methods Programs Biomed. 260 (2025) 108492.
[25] T. Mahmood, A. Rehman, T. Saba, Y. Wang, F.S. Alamri, Alzheimer's disease unveiled: cutting-edge multi-modal neuroimaging and computational methods for enhanced diagnosis, Biomed. Signal Process. Control 97 (Nov 2024) 106721.
[26] C. Xia, H. Chen, J. Han, D. Zhang, K. Li, Identifying children with autism spectrum disorder via transformer-based representation learning from dynamic facial cues, IEEE Trans. Affect. Comput. 16 (1) (2025) 83–97.
[27] M. Mujahid, A. Rehman, T. Alam, F.S. Alamri, S.M. Fati, T. Saba, An efficient ensemble approach for alzheimer's disease detection using an adaptive synthetic technique and deep learning, Diagnostics (Basel) 13 (15) (Jul 2023) 2489.
[28] P. Chinnasamy, V. Nirmala, B.R.S. Reddy, R.V.V.N.B. Rao, A. Udhayakumar, D.P. Degala, Early alzheimer's disease stage identification and categorization through deep learning, in: 2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA), 2025, pp. 1146–1152.
[29] H. Alkahtani, Z.A. Ahmed, T.H. Aldhyani, M.E. Jadhav, A.A. Alqarni, Deep learning algorithms for behavioral analysis in diagnosing neurodevelopmental disorders, Mathematics 11 (19) (2023) 4208.
[30] Y. Wang, H. Long, T. Bo, J. Zheng, Residual graph transformer for autism spectrum disorder prediction, Comput. Methods Programs Biomed. 247 (2024) 108065.
[31] J. Smith-Young, R. Chafe, R. Audas, "Managing the wait": parents' experiences in accessing diagnostic and treatment services for children and adolescents diagnosed with autism spectrum disorder, Health serv. insights 13 (2020) 1178632920902141.
[32] S. Zhang, AI-assisted early screening, diagnosis, and intervention for autism in young children, Front. Psychiatry 16 (2025) 1513809.
[33] I.A. Ahmed, E.M. Senan, T.H. Rassem, M.A. Ali, H.S.A. Shatnawi, S.M. Alwazer, M. Alshahrani, Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques, Electronics 11 (4) (2022) 530.
[34] M.U.G. Khan, Z. Naz, J. Khan, T. Saba, I. Abunadi, A. Rehman, U. Tariq, Cognitive skill enhancement system using neuro-feedback for ADHD patients, Comput. Mater. Contin. 68 (2) (2021) 2363–2376.
[35] J. Chen, C. Chen, R. Xu, L. Liu, Autism identification based on the intelligent analysis of facial behaviors: an approach combining coarse-and fine-grained analysis, Children 11 (11) (2024) 1306.
[36] J. Han, G. Jiang, G. Ouyang, X. Li, A multimodal approach for identifying autism spectrum disorders in children, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 2003–2011.
[37] M. Cordova, K. Shada, D.V. Demeter, O. Doyle, O. Miranda-Dominguez, A. Perrone, E. al., Heterogeneity of executive function revealed by a functional random forest approach across ADHD and ASD, NeuroImage. Clinical 26 (2020) 102245.
[38] J. Hashemi, G. Dawson, K.L.H. Carpenter, K. Campbell, Q. Qiu, S. Espinosa, G. Sapiro, E. al., Computer vision analysis for quantification of autism risk behaviors, IEEE Trans. Affect. Comput. 12 (1) (2021) 215–226.
[39] R. Knott, B.P. Johnson, J. Tiego, O. Mellahn, A. Finlay, K. Kallady, E. al., The monash autism-adhd Genetics and neurodevelopment (magnet) project design and methodologies: a dimensional approach to understanding neurobiological and genetic aetiology, Mol. Autism 12 (1) (2021) 55.
[40] D.K. Moorthy, P. Chinnasamy, P. Nagaraj, Optimization enabled ResNet features with transfer learning for alzheimer's disease detection, Comput. Biol. Chem. 119 (2025) 108613.
[41] A. Hendry, E.J.H. Jones, R. Bedford, L. Andersson Konke, J. Begum Ali, S. Bolte, E. al., Atypical development of attentional control associates with later adaptive functioning, autism and ADHD traits, J. Autism Dev. Disord. 50 (11) (2020) 4085–4105.
[42] R.S. Lalawat, V. Bajaj, An automatic framework for detecting autism spectrum disorder from EEG signals using tfd, IEEE Sens. J. 24 (7) (2024) 10 632–10 639.
[43] N. Wolff, G. Kohls, J.T. Mack, A. Vahid, E.M. Elster, S. Stroth, E. al., A data driven machine learning approach to differentiate between autism spectrum disorder and attention-deficit/hyperactivity disorder based on the best-practice diagnostic instruments for autism, Sci. Rep. 12 (2022) 18744.
[44] D.E. Astle, J. Holmes, R. Kievit, S.E. Gathercole, Annual research review: the transdiagnostic revolution in neurodevelopmental disorders, J. Child Psychol. Psychiatry Allied Discip. 63 (4) (2022) 397–417.
[45] S. Baker, Y. Kandasamy, Machine learning for understanding and predicting neurodevelopmental outcomes in premature infants: a systematic review, Pediatr. Res. 93 (2) (2023) 293–299.
[46] D.Y. Song, C.C. Topriceanu, D.C. Ilie-Ablachim, M. Kinali, S. Bisdas, Machine learning with neuroimaging data to identify autism spectrum disorder: a systematic review and meta-analysis, Neuroradiology 63 (12) (2021) 2057–2072.
[47] M.O. Ribas, M. Micai, A. Caruso, F. Fulceri, M. Fazio, M.L. Scattoni, Technologies to support the diagnosis and/or treatment of neurodevelopmental disorders: a systematic review, Neurosci. Biobehav. Rev. 145 (2023) 105021.
[48] F.R.D. de Barros, C.N.F. da Silva, G. de Castro Michelassi, H. Brentani, F.L. Nunes, A. Machado-Lima, Computer aided diagnosis of neurodevelopmental disorders and genetic syndromes based on facial images–a systematic literature review, Heliyon 9 (10) (2023).
[49] S.D. Dlima, S. Shevade, S.R. Menezes, A. Ganju, Digital phenotyping in health using machine learning approaches: scoping review, JMIR Bioinform. Biotechnol. 3 (1) (2022) e39618.
[50] J.P. Mendes, I.R. Moura, P. Van de Ven, D. Viana, F.J. Silva, L.R. Coutinho, S. Teixeira, J.J. Rodrigues, A.S. Teles, Sensing APPS and public data sets for digital phenotyping of mental health: systematic review, J. Med. Internet Res. 24 (2) (2022) e28735.
[51] I. Moura, A. Teles, D. Viana, J. Marques, L. Coutinho, F. Silva, Digital phenotyping of mental health using multimodal sensing of multiple situations of interest: a systematic literature review, J. Biomed. Inform. 138 (2023) 104278.
[52] A.S. Cohen, C.R. Cox, M.D. Masucci, T.P. Le, T. Cowan, L.M. Coghill, T.B. Holmlund, B. Elvevåg, Digital phenotyping using multimodal data, Curr. behav. neurosci. rep. 7 (4) (2020) 212–220.
[53] K. Khowaja, S.S. Salim, A. Asemi, S. Ghulamani, A. Shah, A systematic review of modalities in computer-based interventions (cbis) for language comprehension and decoding skills of children with autism spectrum disorder (ASD), Univers. Access Inf. Soc. 19 (2) (2020) 213–243.
[54] M. Svanera, M. Savardi, A. Signoroni, S. Benini, L. Muckli, Fighting the scanner effect in brain MRI segmentation with a progressive level-of-detail network trained on multi-site data, Med. Image Anal. 93 (2024) 103090.
[55] Y. Li, M. El Habib Daho, P.-H. Conze, R. Zeghlache, H. Le Boité, R. Tadayoni, E. al., A review of deep learning-based information fusion techniques for multimodal medical image classification, Comput. Biol. Med. 177 (2024) 108635.
[56] J.C.C. Kwong, A. Khondker, K. Lajkosz, M.B.A. McDermott, X.B. Frigola, M.D. McCradden, M. Mamdani, G.S. Kulkarni, A.E.W. Johnson, Appraise-AI tool for quantitative evaluation of AI studies for clinical decision support, JAMA Netw Open 6 (9) (2023) e2335377.
[57] A. Riaz, M. Asad, E. Alonso, G. Slabaugh, Deepfmri: end-to-end deep learning for functional connectivity and classification of ADHD using FMRI, J. Neurosci. Methods 335 (2020) 108506.
[58] T. Tang, C. Li, S. Zhang, Z. Chen, L. Yang, Y. Mu, J. Chen, P. Xu, D. Gao, F. Li, B. He, A hybrid graph network model for ASD diagnosis based on resting-state EEG signals, Brain Res. Bull. 206 (2024) 110826.
[59] J. Li, X. Kong, L. Sun, X. Chen, G. Ouyang, X. Li, S. Chen, Identification of autism spectrum disorder based on electroencephalography: a systematic review, Comput. Biol. Med. 170 (2024) 108075.
[60] Y. Yang, C. Ye, T. Ma, A deep connectome learning network using graph convolution for connectome-disease association study, Neural Netw. 164 (2023) 91–104.
[61] H. Dong, D. Chen, Y. Chen, Y. Tang, D. Yin, X. Li, A multi-task learning model with reinforcement optimization for ASD comorbidity discrimination, Comput. Methods Programs Biomed. 243 (2024) 107865.
[62] W. Zheng, T. Eilam-Stock, T. Wu, A. Spagna, C. Chen, B. Hu, J. Fan, Multi-feature based network revealing the structural abnormalities in autism spectrum disorder, IEEE Trans. Affect. Comput. 12 (3) (2021) 732–742.
[63] M. Kulyabin, P.A. Constable, A. Zhdanov, I.O. Lee, D.A. Thompson, A. Maier, Attention to the electroretinogram: gated multilayer perceptron for ASD classification, IEEE Access 12 (2024) 52 352–52 362.
[64] M. Ruan, P.J. Webster, X. Li, S. Wang, Deep neural network reveals the world of autism from a first-person perspective, Autism Res. 14 (2) (2021) 333–342.
[65] R.J. Schachar, A. Dupuis, P.D. Arnold, E. Anagnostou, E. Kelley, S. Georgiades, J. Crosbie, Autism spectrum disorder and attention-deficit/hyperactivity disorder: shared or unique neurocognitive profiles? Res. Child Adolesc. Psychopathol. 51 (1) (2023) 17–31.
[66] J. Shin, S. Konnai, M. Maniruzzaman, M.A.M. Hasan, K. Hirooka, A. Megumi, A. Yasumura, Identifying ADHD for children with coexisting ASD from fnirs signals using deep learning approach, IEEE Access 11 (2023) 82 794–82 801.
[67] A. Ashraf, Z. Qingjie, W.H.K. Bangyal, M. Iqbal, Analysis of brain imaging data for the detection of early age autism spectrum disorder using transfer learning approaches for internet of things, IEEE Trans. Consum. Electron. 70 (1) (2023) 4478–4489.
[68] B. Henderson, P. Yogarajah, B. Gardiner, T.M. McGinnity, Encoding kinematic and temporal gait data in an appearance-based feature for the automatic classification of autism spectrum disorder, IEEE Access 11 (2023) 134 100–134 117.
[69] F.M. Talaat, Real-time facial emotion recognition system among children with autism based on deep learning and IOT, Neural Comput. Appl. 35 (17) (2023) 12 717–12 728.
[70] R.A.J. de Belen, V. Eapen, T. Bednarz, A. Sowmya, Using visual attention estimation on videos for automated prediction of autism spectrum disorder and symptom severity in preschool children, PLOS ONE 19 (2) (2024) e0282818.
[71] A. Atyabi, F. Shic, J. Jiang, C.E. Foster, E. Barney, M. Kim, B. Li, P. Ventola, C.H. Chen, Stratification of children with autism spectrum disorder through fusion of

temporal information in eye-gaze scan-paths, ACM Trans. Knowl. Discov. Data 17 (2) (2023) 1–20.
[72] M.R. Kanhirakadavath, M.S.M. Chandran, Investigation of eye-tracking scan path as a biomarker for autism screening using machine learning algorithms, Diagnostics 12 (2) (2022) 518.
[73] T. Thanarajan, Y. Alotaibi, S. Rajendran, K. Nagappan, Eye-tracking based autism spectrum disorder diagnosis using chaotic butterfly optimization with deep learning model, Comput. Mater. Contin. 76 (2) (2023).
[74] J. Li, Z. Chen, Y. Zhong, H.K. Lam, J. Han, G. Ouyang, X. Li, H. Liu, Appearance-based gaze estimation for ASD diagnosis, IEEE Trans. Cybern. 52 (7) (2022) 6504–6517.
[75] V. Khullar, K. Salgotra, H.P. Singh, D.P. Sharma, Deep learning-based binary classification of ADHD using resting state mr images, Augment. Hum. Res. 6 (1) (2021) 5.
[76] H. Shahamat, M.S. Abadeh, Brain MRI analysis using a deep learning based evolutionary approach, Neural Netw. 126 (2020) 218–234.
[77] D.Y. Isaev, M. Sabatos-DeVito, J.M. Di Martino, K. Carpenter, R. Aiello, E.A. Compton, Computer vision analysis of caregiver-child interactions in children with neurodevelopmental disorders: a preliminary report, J. Autism Dev. Disord. 54 (6) (2024) 2286–2297.
[78] B.C. Dos Santos Melício, K. Karaköse, A. Fodor, L. Xiang, V. Varga, E.A. Soorya, Multimodal framework for automatic behavior analysis of children with autism during ados-2, Cogn. Comput. 17 (4) (2025).
[79] I. Ahmad, J. Rashid, M. Faheem, A. Akram, N.A. Khan, R.U. Amin, Autism spectrum disorder detection using facial images: a performance comparison of pretrained convolutional neural networks, Healthc. Technol. Lett. 11 (4) (2024) 227–239.
[80] X. Chen, S. Wang, X. Yang, C. Yu, F. Ni, J. Yang, Y. Tian, J. Ye, H. Liu, R. Luo, Utilizing artificial intelligence-based eye tracking technology for screening ADHD symptoms in children, Front. Psychiatry 14 (2023) 1260031.
[81] N. Stuart, A. Whitehouse, R. Palermo, E. Bothe, N. Badcock, Eye gaze in autism spectrum disorder: a review of neural evidence for the eye avoidance hypothesis, J. Autism Dev. Disord. 53 (5) (2023) 1884–1905.
[82] T. Su, M. Yang, W. Zhou, Y. Tao, M. Ni, W. Zheng, B. Hu, Joint probabilistic modeling analysis of head and eye behaviors in autism spectrum disorder children based on a social interaction paradigm, Biomed. Signal Process. Control 106 (2025) 107669.
[83] N. Zhang, M. Ruan, S. Wang, L. Paul, X. Li, Discriminative few shot learning of facial dynamics in interview videos for autism trait classification, IEEE Trans. Affect. Comput. 14 (2) (2022) 1110–1124.
[84] L. Shao, Y. You, H. Du, D. Fu, Classification of ADHD with FMRI data and multi-objective optimization, Comput. Methods Programs Biomed. 196 (2020) 105676.
[85] V.G. Prakash, M. Kohli, S. Kohli, A.P. Prathosh, T. Wadhera, D. Das, D. Panigrahi, J.V.S. Kommu, Computer vision-based assessment of autistic children: analyzing interactions, emotions, human pose, and life skills, IEEE Access 11 (2023) 47 907– 47 929.
[86] G. Marsicano, C. Bertini, L. Ronconi, Decoding cognition in neurodevelopmental, psychiatric and neurological conditions with multivariate pattern analysis of EEG data, Neurosci. Biobehav. Rev. 164 (2024) 105795.
[87] C. Bunterngchit, L.H. Baniata, A. Rasool, Neuroxai: explainable deep learning for eeg-based detection of alzheimer's and parkinson's diseases, pp. 612–619, 2025.
[88] W. Wei, Z. Liu, L. Huang, Z. Wang, W. Chen, T. Zhang, E. al., Identify autism spectrum disorder via dynamic filter and deep spatiotemporal feature extraction, Signal Process., Image Commun. 94 (2021) 116195.
[89] Z. Xu, C.L.P. Chen, T. Zhang, Tfagl: a novel agent graph learning method using time-frequency EEG for major depressive disorder detection, IEEE Trans. Affect. Comput. 16 (3) (2025) 1592–1605.
[90] M. Bakhtyari, S. Mirzaei, ADHD detection using dynamic connectivity patterns of EEG data and convlstm with attention framework, Biomed. Signal Process. Control 76 (2022) 103708.
[91] K. Gao, Z. Fan, J. Su, L.L. Zeng, H. Shen, J. Zhu, D. Hu, Deep transfer learning for cerebral cortex using area-preserving geometry mapping, Cereb. Cortex 32 (14) (2022) 2972–2984.
[92] M. Tosun, Effects of spectral features of EEG signals recorded with different channels and recording statuses on ADHD classification with deep learning, Phys Eng Sci Med 44 (2021) 693–702.
[93] N. Alsharif, M.H. Al-Adhaileh, M. Al-Yaari, N. Farhah, Z.I. Khan, Utilizing deep learning models in an intelligent eye-tracking system for autism spectrum disorder diagnosis, Front. Med. 11 (2024) 1436646.
[94] J. Li, Y. Zhong, J. Han, G. Ouyang, X. Li, H. Liu, Classifying ASD children with LSTM based on raw videos, Neurocomputing 390 (2020) 226–238.
[95] E. Kerz, S. Zanwar, Y. Qiao, D. Wiechmann, Toward explainable AI (XAI) for mental health detection based on language behavior, Front. Psychiatry 14 (2023) 1219479.
[96] M. Varma, P. Washington, B. Chrisman, A. Kline, E. Leblanc, K. Paskov, N. Stockham, J.Y. Jung, M.W. Sun, D.P. Wall, Identification of social engagement indicators associated with autism spectrum disorder using a game-based mobile app: comparative study of gaze fixation and visual scanning methods, J. Med. Internet Res. 24 (2) (2022) e31830.
[97] T.R. Mahesh, T. Goswami, S. Sriramulu, N. Sharma, A. Kumari, G. Khekare, Cognitive based attention deficit hyperactivity disorder detection with ability assessment using auto encoder based hidden markov model, Int. J. Commun. Netw. Inf. Secur. 14 (2) (2022) 53–65.
[98] S. Liang, A.Q.M. Sabri, F. Alnajjar, C.K. Loo, Autism spectrum self-stimulatory behaviors classification using explainable temporal coherency deep features and SVM classifier, IEEE Access 9 (2021) 34 264– 34 275.
[99] C. Gupta, P. Chandrashekar, T. Jin, C. He, S. Khullar, Q. Chang, D. Wang, Bringing machine learning to research on intellectual and developmental disabilities: taking inspiration from neurological diseases, J. Neurodev. Disord. 14 (1) (2022) 28.
[100] N. Qiang, J. Gao, Q. Dong, H. Yue, H. Liang, L. Liu, J. Yu, J. Hu, S. Zhang, B. Ge, Y. Sun, Functional brain network identification and FMRI augmentation using a vae-gan framework, Comput. Biol. Med. 165 (2023) 107395.
[101] B. Ari, N. Sobahi, O. Alçin, A. Sengur, U.R. Acharya, Accurate detection of autism using douglas-peucker algorithm, sparse coding based feature mapping and convolutional neural network techniques with EEG signals, Comput. Biol. Med. 143 (2022) 105311.
[102] Y. Zhang-James, J.K. Buitelaar, E.-A.W. Group, D. van Rooij, S.V. Faraone, Ensemble classification of autism spectrum disorder using structural magnetic resonance imaging features, JCPP Advances 1 (3) (2021) e12042.
[103] F. Hajjej, S. Ayouni, M.A. Alohali, M. Maddeh, Novel framework for autism spectrum disorder identification and tailored education with effective data mining and ensemble learning techniques, IEEE Access 12 (2024) 35 448– 35 461.
[104] A. Rasool, S. Aslam, Y. Wang, Y. Pan, Deep neurocomputational fusion for ASD diagnosis using multi-domain EEG analysis, Neurocomputing 641 (2025) 130353.
[105] A.B. Arnett, B.P. Flaherty, A framework for characterizing heterogeneity in neurodevelopmental data using latent profile analysis in a sample of children with ADHD, J. Neurodev. Disord. 14 (1) (2022) 45.
[106] J. Shin, S. Konnai, M. Maniruzzaman, Y. Tomioka, Y.S. Hwang, A. Megumi, A. Yasumura, A multimodal approach for ADHD with coexisting ASD detection for children, Sci. Rep. 15 (1) (2025) 21182.
[107] A.G. Papaioannou, E. Kalantzi, C.C. Papageorgiou, K. Korombili, A. Bokou, A. Pehlivanidis, E. al., Complexity analysis of the brain activity in autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) due to cognitive loads/demands induced by aristotle's type of syllogism/reasoning. A power spectral density and multiscale entropy (MSE) analysis, Heliyon 7 (9) (2021) e07984.
[108] A. Federico, A. Zgodic, K. Flory, R.M. Hantman, J.M. Eberth, A.C. Mclain, J. Bradshaw, Predictors of autism spectrum disorder and ADHD: results from the national survey of children's health, Disabil. Health J. 17 (1) (2024) 101512.
[109] S. Liaqat, C. Wu, P.R. Duggirala, S.-C.S. Cheung, C.-N. Chuah, S. Ozonoff, G. Young, Predicting ASD diagnosis in children with synthetic and image-based eye gaze data, Signal Process., Image Commun. 94 (2021) 116198.
[110] Deepika, M. Sharma, S. Arora, Redefining parameter-efficiency in ADHD diagnosis: a lightweight attention-driven kolmogorov-arnold network with reduced parameter complexity and a novel activation function, Psychiatry Res. Neuroimaging 351 (2025) 112016.
[111] C. Jean-Quartier, et al., Reproducibility and environmental efficiency of metabolomics cancer modeling, Appl. Sci. 16 (2) (2026) 588.
[112] J. Löchel, M. Putzier, M. Dreischarf, P. Grover, K. Urinbayev, F. Abbas, K. Labbus, R. Zahn, Deep learning algorithm for fully automated measurement of sagittal balance in adult spinal deformity, Eur. Spine J. 33 (11) (2024) 4119–4124.
[113] G. Wang, G. Yang, Z. Du, L. Fan, X. Li, Clinicalgpt: large language models fine-tuned with diverse medical data and comprehensive evaluation, arXiv preprint arXiv:2306.09968, 2023.
[114] D. Baradari, N. Kosmyna, O. Petrov, R. Kaplun, P. Maes, Neurochat: A neuroadaptive ai chatbot for customizing learning experiences, arXiv preprint arXiv:2503.07599, 2025.
[115] A.P. Kiraly, S. Baur, K. Philbrick, F. Mahvar, L. Yatziv, T. Chen, B. Sterling, N. George, F. Jamil, J. Tang, K. Bailey, Health ai developer foundations, arXiv preprint arXiv:2411.15128, 2024.
[116] S. Aslam, M. Hussain, X. Li, H. Li, H. Wu, Hamd-net: a biologically informed hybrid attention model for EEG classification of multiple neurological disorders, Biomed. Signal Process. Control 120 (Part A) (2026) 110049.
[117] M.J. Page, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ 372 (2021) n71.

Author biography

Abdur Rasool is a Postdoctoral Fellow at the University of Hawaii, U.S.A. He received a Ph.D. in computer applied technology (in 2024) from Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (SIAT-CAS), Shenzhen, China. He has authored over 35 publications in international journals and academic conferences in the field of data mining, with a focus on DNA data storage, bioinformatics, security, and text mining. He was an excellent international graduate at the University of Chinese Academy of Sciences (UCAS) and a winner of various research contests. His interests lie in multidisciplinary computer science and technology, with a focus on AI, Bioinformatics, Autism, DNA data storage, and LLMs for research.

Faizan Ahmad received the B.S. degree in electrical engineering from the National University of Computer and Emerging Sciences, Pakistan, in 2013, the M.S. degree in communication and information systems from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2020, and the Ph.D. degree in pattern recognition and intelligent systems from the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, in 2025. He is currently a Postdoctoral Fellow at Shenzhen Technology University, Shenzhen, China. His research interests include medical imaging, deep learning, augmented reality, and medical robotics.

Chayut Bunterngchit received a Bachelor of Engineering in Industrial Engineering from Kasetsart University in 2013, a Master of Engineering Management from the University of Wollongong in 2016, and a Ph.D. in Control Theory and Control Engineering from the Institute of Automation, Chinese Academy of Sciences in 2025. He is currently a lecturer at the Faculty of Engineering and Technology, King Mongkut's University of Technology North Bangkok, Rayong Campus, Thailand. His research interests include brain-computer interfaces, biomedical signal processing, and deep learning.

Saba Aslam is pursuing a Ph.D. from Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (SIAT-CAS), Shenzhen, China. She received a Master's degree from Xiamen University, China, and a BS (Hons.) degree in Software Engineering from Government College University, Faisalabad, Pakistan, in 2015. She has also served in many educational institutions in Pakistan, including at the National University of Computer and Emerging Sciences – Chiniot-Faisalabad as an Artificial Intelligence instructor. During her student life, she was an active member of the IEEE CS Society and organized many national and international conferences, workshops, and seminars. Her research interests include AI in Bioinformatics, Continual Learning, Neurological Disorders Diagnosis, and Time Series Forecasting with Machine Learning and Deep Learning.