Review | Open Access

Journal of Environmental Exposure Assessment

Zheng *et al. J. Environ. Expo. Assess.* **2026**, *5*, 26 **DOI:**10.20517/jeea.2026.25

Artificial intelligence in environmental etiology of autism spectrum disorder: progress, opportunities, and challenges

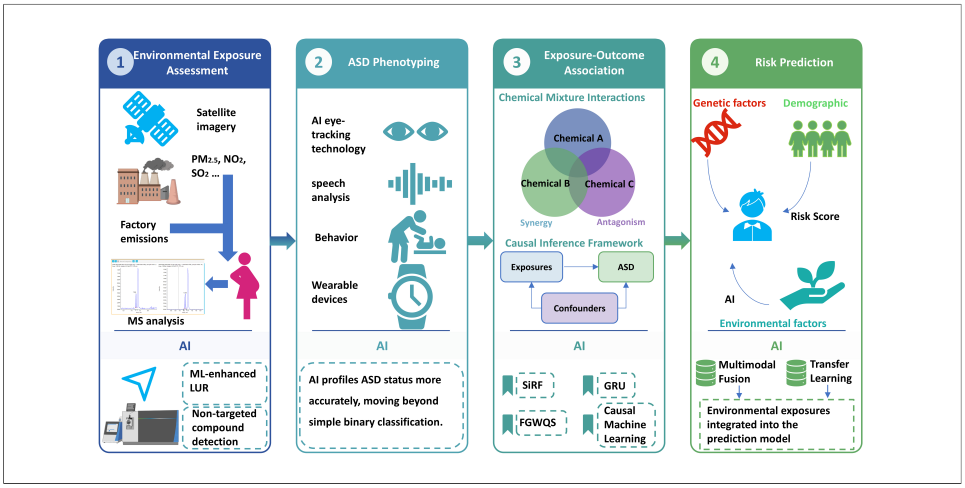
Ruizhi Zheng[1], Lin Wen[2], Yun Huang[1], Ziying Gan[1], Hongyan Yin[3], Wei Xia[1], Hongxiu Liu[1,4]

Keywords:
Autism spectrum disorder, environmental exposure, AI, machine learning, risk prediction, causal inference

Citation: Zheng, R.; Wen, L.; Huang, Y.; Gan, Z.; Yin, H.; Xia, W.; Liu, H. Artificial intelligence in environmental etiology of autism spectrum disorder: progress, opportunities, and challenges. *J. Environ. Expo. Assess.* **2026**, *5*, 26. https://dx.doi.org/10.20517/jeea.2026.25

Received: 16 May 2026
First Decision: 15 Jun 2026
Revised: 16 Jul 2026
Accepted: 17 Jul 2026
Published: 12 Aug 2026

Academic Editor:
Stuart Harrad
Copy Editor:
Tong Wang
Production Editor:
Tong Wang



Abstract

Traditional epidemiological analyses may be limited when addressing examining the environmental etiology of autism spectrum disorder (ASD), including assumptions of linearity, difficulties in handling high-dimensional exposure data, and challenges in detecting gene-environment or exposure-exposure interactions. The emergence of artificial intelligence (AI) - ranging from conventional machine learning to deep learning and large language models (LLMs) - has provided new tools for modeling the complex associations between environmental exposures and health outcomes. Machine learning models have shown strong predictive performance in individual-level air pollution exposure assessment and has enabled the detection of untargeted chemical substances from large-scale mass spectrometry data. In this review, we examine AI applications in four key areas of ASD environmental etiology research: environmental exposure assessment, ASD phenotyping,

[1]Key Laboratory of Environment and Health (HUST), Ministry of Education & Ministry of Environmental Protection, State Key Laboratory of Environmental Health (Incubation), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei, China.
[2]Department of Nutrition and Food Hygiene, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei, China.
[3]School of Tropical Agriculture and Forestry, Hainan University, Haikou 570228, Hainan, China.
[4]Department of Environmental Health, School of Environmental Science and Engineering, Hainan University, Haikou 570228, Hainan, China.

Correspondence to: Dr. Hongxiu Liu, Department of Environmental Health, School of Environmental Science and Engineering, Hainan University, Haikou 570228, Hainan, China. E-mail: lhx@hainanu.edu.cn

exposure-outcome association analysis, and risk prediction. Our findings suggest that although LLMs are emerging in environmental toxicology research, ASD environmental etiology studies remain predominantly based on conventional machine-learning methods, with deep learning and LLMs being less applied in this field. Further progress will require large-scale prospective cohort studies, adoption of advanced AI methods, unified data infrastructure, and interation across disciplines.

INTRODUCTION

With the rapid development of artificial intelligence (AI), this technology has been rapidly integrated into environmental etiology research[1,2]. Machine learning (ML) and deep learning (DL) offer strengths in data processing, pattern recognition, prediction, and modeling, and have been used to improve exposure assessment, disease diagnosis, and the identification of environmental risk factors[3,4]. The reported prevalence of autism spectrum disorder (ASD) has increased in recent decades and substantially affect children's development and daily functioning[5]. Growing evidence also points to environmental exposures as contributors to neurodevelopmental disorders[5,6]. The intersection of AI and ASD environmental research has accordingly drawn increasing attention.

In real-world settings, people are exposed to multiple pollutants, and the combined effects of pollutants and genes may lead to adverse health outcomes[7]. However, the traditional biomonitoring approaches are primarily based on targeted detection, only measuring preset chemicals, making it difficult to evaluate complicated exposure environments systematically[8]. Detection of thousands of chemical features by nontarget analysis is possible, but compound identification remains a major bottleneck[9]. Furthermore, the traditional land use regression (LUR) model assumes a linear relationship, and it is difficult to capture the spatial and temporal dynamics of pollutants[10]. Questionnaire-based exposure assessment is susceptible to recall bias[11], and single biological samples may not adequately capture long-term exposure for pollutants with short half-lives[12]. These limitations have long been recognized in exposure assessment. Recent advances in AI offer promising opportunities to address some of these challenges. The evolution of AI, ranging from conventional ML to DL, large language models (LLMs), and causal ML, provides new possibilities to solve these bottlenecks[13]. In exposure assessment, ML improves the accuracy of air pollution prediction through enhanced LUR[14]; LLMs can predict the persistence, bioaccumulation, and toxicity of chemicals from natural language descriptions (e.g., physical appearance, melting point, industrial use, *etc.*)[15]; ML can also assist in processing non-targeted analysis data from mass spectrometry (MS) to prioritize unknown chemical features for further identification[16]. In the statistical analysis, AI-based approaches may complement traditional statistical methods in analyzing complex exposure-response relationships and prioritizing the screening of neurodevelopmental toxicants from high-dimensional data. To overcome these bottlenecks, ML methods have been proposed to automatically screen potential risk substances from high-dimensional data and identify complex mixture effects[17]. In addition, causal ML can estimate heterogeneous exposure effects from observational data and identify susceptible subgroups[18,19]. Moreover, traditional models typically estimate odds ratios at the group level in the context of risk prediction[20]. The traditional forecasting model is highly dependent on genetic and demographic factors and seldom incorporates environmental exposure variables[21]. In statistical modeling, causal ML can be used to estimate heterogeneous exposure effects, whereas DL models can generate individual-level risk predictions[18,22].

Currently, validated biological markers for ASD phenotypes are still lacking[23]. Gold-standard diagnostic tools such as the autism diagnostic interview-Revised (ADI-R) require 1-3 h of interview time and trained clinicians, making them impractical for large-scale studies[24]. Changes in diagnostic criteria over time have also made it difficult to compare across different studies[25]. AI may have the potential to assist diagnostic accuracy, quantify behavioral traits, shorten assessment time, reduce reliance on experts, and provide

standardized phenotyping for large epidemiological studies. Technologies such as facial image recognition[26], eye movement tracking technology[27], and voice analysis[28] have been explored as potential tools to support ASD diagnosis, though most remain at the research stage. Moreover, AI can also identify potential ASD subgroups from brain imaging data, offering a complementary approach to traditional phenotype-based classifications[29].

We searched PubMed on April 14, 2026, using a combination of MeSH terms and free-text keywords across three domains: ASD, environmental exposures, and AI/computational methods. After screening, 13 studies met the inclusion criteria. The full search strategy, screening process, and preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram are provided in the Supplementary Materials [Supplementary Text 1 and Supplementary Figure 1]. This review summarizes AI applications in ASD environmental etiology research - covering exposure assessment, phenotyping, association analysis, and risk prediction. Figure 1 provides an overview of these applications[30]. It also identifies key gaps in current studies and discusses why advanced AI methods remain less commonly used in research on the environmental etiology of ASD. Finally, it outlines a framework for integrating emerging AI tools into future studies. To this end, the review aims to inform future interdisciplinary research on environmental risk factors for ASD.

AI-ENABLED ENVIRONMENTAL EXPOSURE ASSESSMENT

AI approaches in environmental exposure assessment

The exposome concept - the sum of environmental exposures from conception onward - has shifted environmental health research from single-pollutant to multi-exposure approaches[31]. This shift matters for ASD, where complex exposures during critical developmental windows may contribute to disease risk. But putting the exposome into practice is not easy. Exposure data are high-dimensional, correlated, and dynamic. They require integrating biomonitoring, geospatial, and sensor data[32]. ML and DL are increasingly used to handle the high dimensionality, correlation, and nonlinearity in exposome data[32,33]. Most studies use supervised or unsupervised learning to examine multiple exposure factors in chronic disease, with DL emerging more recently[34]. For ASD research, the convergence of exposome and AI offers a way to move beyond single-pollutant studies toward a more complete picture of environmental risk.

AI is increasingly being applied within the exposome framework for dynamic air pollution monitoring and for detecting pollutants in biological samples. AI has advanced environmental exposure assessment by integrating information from multiple sources, including satellite images, land-use data, meteorological data, and mobile sensor networks[35]. This capability has improved the modeling of complex nonlinear relationships that may be difficult to capture using conventional parametric models[36]. Commonly used AI techniques in environmental etiology research include ensemble learning methods such as random forest and eXtreme Gradient Boosting (XGBoost), and deep learning architectures, including convolutional neural networks and GRUs[37-39]. These tools can predict pollutant levels[40], reconstruct historical exposure trends[40], and identify compounds based on MS[41,42], providing modeled estimates at a higher spatial resolution than traditional monitoring systems AI-based models can now generate exposure estimates at kilometer-level spatial resolution with fine temporal granularity, offering a new way to study environmental exposures[43]. AI-driven wearables can support continuous monitoring[44]. Real-time tracking of physiological and environmental data through wearables is a promising direction for exposome research. The socio-exposome framework has expanded the concept of exposure to include social and structural determinants - such as neighborhood socioeconomic status, education, and inequality[45]. Ren *et al.* showed that random forest and XGBoost can capture nonlinear associations between these factors and health outcomes that traditional geostatistical models missed[45]. AI is also becoming better at integrating diverse exposome data - MS, transcriptomics, geospatial information - into unified exposure profiles[46,47]. At a broader scale, whole-body exposome models offer a valuable reference for integrating multi-scale data, although their application in ASD research remains to be explored[48].

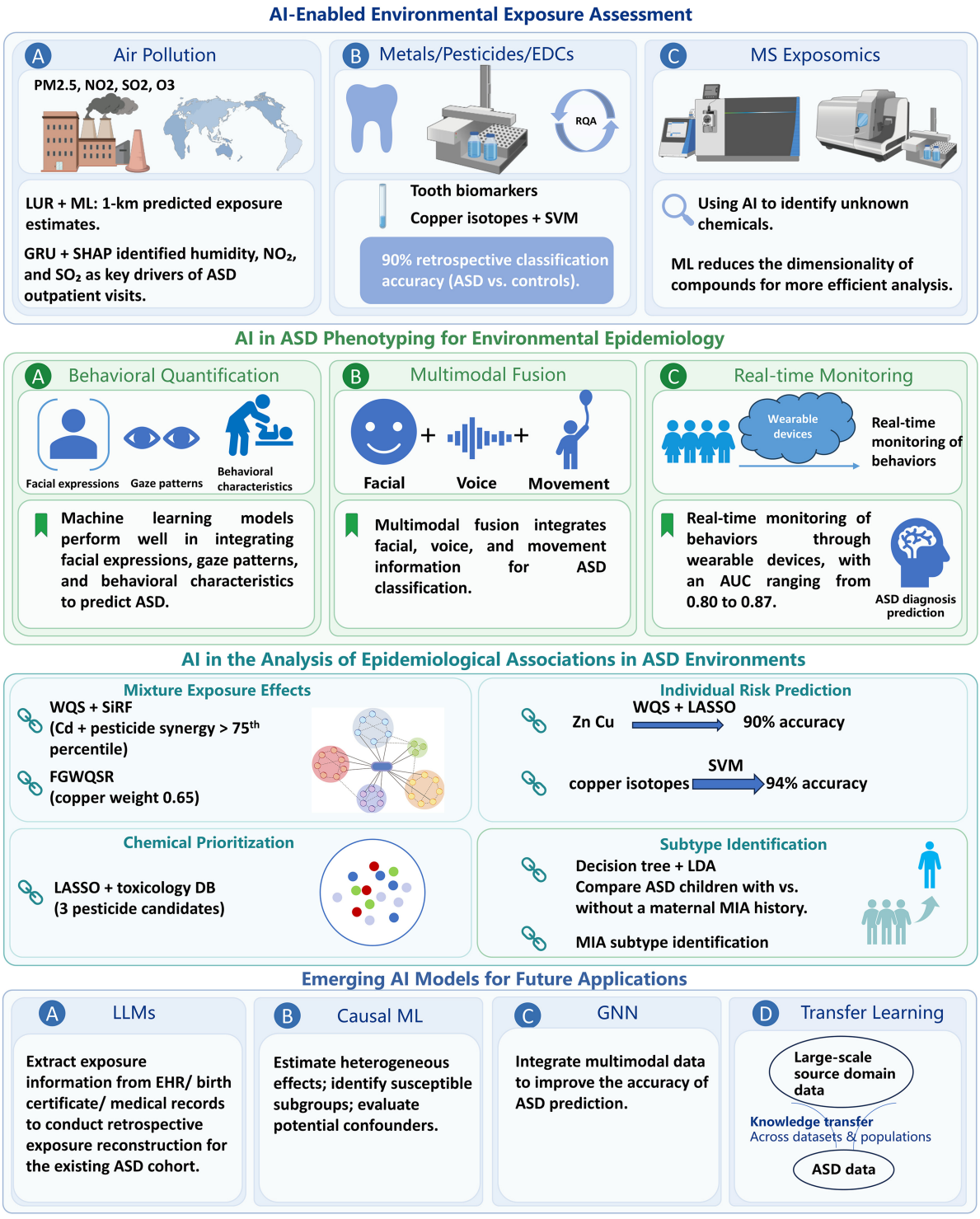


Figure 1. AI applications in environmental etiology research on ASD. Icons in this figure were created using BioGDP (https://BioGDP.com)[30] and the built-in library in Microsoft PowerPoint. LUR: Land use regression; ML: machine learning; GRU: gated recurrent unit; SHAP: Shapley Additive explanations; ROA: recurrence quantification analysis; SVM: support vector machine; WQS: weighted quantile sum; SiRF: signed iterative random forests; FGWQSR: frequentist grouped weighted quantile sum regression; MIA: maternal immune activation; ASD: autism spectrum disorder; MS: mass spectrometry; LASSO: least absolute shrinkage and selection operator; LLMs: large language models; GNN: graph neural network; EDCs: endocrine disrupting chemicals.

Table 1 summarizes the six studies in this review that applied AI to environmental exposure assessment in ASD research. As shown in Figure 1, among the studies screened in this review, the application of ML in ASD-related exposure research remains predominantly based on traditional methods. Some studies have

Table 1. AI applications in environmental exposure assessment for ASD-related research

Authors (Year)	Study type	Exposures	AI method	Population	Sample size	Objectives	Key findings
O'Sharkey *et al.* (2025)[14]	Cohort study	$PM_{2.5}$, NO_2, O_3	D/S/A algorithm; ML-enhanced LUR	California (2013-2018)	2,371,379 (44,173 ASD cases)	To assess the association between prenatal air pollution exposure and social demographic modifying factors and the risk of ASD	Adjusted R^2: 83.6% ($PM_{2.5}$), 65.2% (NO_2), 93.1% (O_3) at 1 km × 1 km; effect modification by race/ethnicity and SES
O'Sharkey *et al.* (2024)[49]	Cohort study	$PM_{2.5}$ components (brake/tire wear markers)	LUR; co-kriging	Southern California (2016-2019)	444,651 (11,466 ASD cases)	Identify traffic-related $PM_{2.5}$ components associated with ASD	Brake wear (Ba) and tire wear (Zn) particles linked to ASD incidence
O'Sharkey *et al.* (2025)[51]	Cohort study	$PM_{2.5}$, NO_2, O_3, benzene, 1,3-butadiene, Cr, Pb, Ni, Zn	D/S/A algorithm; ML-enhanced LUR	California (1990-2019)	527,710 (9585 ASD cases)	Develop high-resolution exposure models (1989-2018) for California births and assess prenatal exposure associations with ASD risk	R^2 at 100 m: 84% (NO_2), 65% ($PM_{2.5}$), 92% (O_3); ASD ORs: $PM_{2.5}$ = 1.10, NO_2 = 1.25, benzene = 1.55, Ni = 1.32 per IQR
Wang *et al.* (2023)[52]	Time-series study	Air pollutants, meteorological factors	GRU; SHAP	Nanjing, China (2015-2019)	~1.47 million visits	Identify key drivers of ASD outpatient visits	Humidity, NO_2, SO_2 identified as key drivers; synergistic and antagonistic pollutant interactions distinguished
Curtin *et al.* (2018)[53]	Case-control	Fetal/postnatal zinc-copper metabolic cycles	RQA; WQS/LASSO	Sweden, UK, USA (4 cohorts)	193 (80 ASD cases)	Predict ASD from fetal metal exposure dynamics	90% accuracy across 4 cohorts; cyclical features distinguished ASD
Ling *et al.* (2023)[54]	Case-control	Serum copper isotopic signature	SVM	Guangxi, China	60 (30 ASD, 30 controls)	Classify ASD based on serum copper isotopes	94.4% accuracy (30 ASD, 30 controls)

Each row corresponds to a cited study; full references are listed in the reference list. AI: Artificial intelligence; ASD: autism spectrum disorder; D/S/A: deletion-substitution-addition; DEET: N,N-diethyl-meta-toluamide; FGWQSR: Frequentist Grouped Weighted Quantile Sum Regression; GRU: Gated Recurrent Unit; LASSO: Least Absolute Shrinkage and Selection Operator; LDA: linear discriminant analysis; LUR: land use regression; MIA: maternal immune activation; ML: machine learning; PCA: principal component analysis; RQA: recurrence quantification analysis; SHAP: SHapley Additive exPlanations; SiRF: Signed Iterative Random Forest; SVM: support vector machine; WQS: weighted quantile sum; SES: .

used ML-enhanced LUR models to address exposure assessment challenges[14,49,50]. O'Sharkey *et al.* conducted a California population-based cohort study of 44,173 ASD cases among 2,371,379 children, finding that prenatal $PM_{2.5}$ exposure was associated with an increased risk of ASD in all models, with the postnatal NO_2 effect strongest among Black and Hispanic children[14]. The same team also developed a high-resolution LUR model for California spanning 1989 to 2018 using a deletion-substitution-addition algorithm with collaborative Kriging[51]. At 100-meter resolution, the model achieved adjusted R^2 values of 84% for NO_2, 65% for $PM_{2.5}$, and 92% for O_3, demonstrating that ML-enhanced LUR showed good predictive performance for prenatal and early life periods relevant to ASD studies[50]. These models were later used to link specific $PM_{2.5}$ components (including black carbon from exhaust emissions, barium from brake wear, and zinc from tire wear) to the risk of ASD in Southern California[49]. A gated recurrent unit network combined with SHapley Additive exPlanations (SHAP) analysis identified humidity, NO_2, and SO_2 as key drivers of ASD outpatient visits and identified potential nonlinear joint patterns[52].

For pesticide and metal exposure, traditional methods relying on single-time biological samples struggle to capture the dynamic exposure changes[11]. Curtin *et al.* used tooth biomarkers, which grow gradually like tree rings, to reconstruct continuous fetal and postnatal zinc-copper exposure histories[53]. Applying recurrence quantification analysis (RQA), a nonlinear dynamics method, they extracted cyclical features, including cycle

duration, regularity, and complexity. These features distinguished ASD cases from controls with 90% accuracy across four independent cohorts, whereas raw metal concentrations showed no group differences, suggesting that exposure dynamics might provide additional discriminatory information[53]. In a small sample (30 ASD, 30 controls), support vector machine classification of serum copper isotope composition achieved 94.4% accuracy in distinguishing ASD cases from controls[54]. However, accuracy estimates from small samples may be overestimated, and this finding requires external validation.

In addition to targeted metal analysis, both targeted and non-targeted MS methods have been applied. The targeted method can accurately quantify preset chemicals but cannot detect unexpected exposures. Non-targeted high-resolution MS can detect thousands of chemical features, but compound identification remains a bottleneck because most spectral signals do not match existing databases[55]. Recent AI advances have begun to address these challenges. ML strategies have enabled concentration prediction in the absence of reference standards - achieving "stratified semi-quantification" - and have improved metabolite annotation through observation-data-based retention time prediction[42,56]. However, despite these advances in general exposomics, one ASD-related study has applied AI-assisted MS to microbial markers: a gut microbiota study combining gas chromatography-mass spectrometry (GC-MS) with an ML classifier distinguished ASD from controls with 89% accuracy based on volatile organic compound profiles[57]. The application of AI-driven MS identification in ASD exposure assessment remains largely unexplored.

Challenges and future directions

The above examples show the potential of AI in promoting research on ASD-related environmental exposure, but several obstacles continue to limit its broad application[50,52,53]. Beyond the lag in translating advanced AI methods into ASD research, the acquisition of exposome data in ASD etiological studies also faces inherent obstacles. One major challenge is the uncertainty of susceptibility windows[52,58]. Although AI methods such as gated recurrent unit networks have shown promise in exploring short-term lagged associations, only one ASD-related study has employed this approach to date[52]. Residential mobility adds further complexity. Pregnant women frequently change address during pregnancy, yet most studies only assign exposure based on their birth address, which may lead to greater bias[59]. While AI-based geospatial models capable of incorporating multi-address spatiotemporal information are technically feasible, they have not been systematically applied in ASD research. Addressing these technical challenges will require greater integration of longitudinal mobility data and advanced time-series AI methods into ASD exposure assessment workflows.

AI IN ASD PHENOTYPING FOR ENVIRONMENTAL EPIDEMIOLOGY

The diagnosis of ASD relies on behavioral assessment tools such as the Autism Diagnostic Observation Schedule and the ADI-R. However, these instruments are costly, time-consuming, and subject to clinician-dependent variability, which limits their feasibility for large-scale epidemiological studies[60-62]. The transition from diagnostic and statistical manual of mental disorders, fourth edition (DSM-IV) to diagnostic and statistical manual of mental disorders, fifth edition (DSM-5) reduced the ASD diagnosis rate by an estimated 20.8%, and the diagnosis is often delayed until 4 to 5 years of age, complicate the retrospective assessment of early-life environmental exposures in epidemiological studies[63,64].

Figure 2 summarizes the key advantages and technical framework of AI in ASD phenotyping. ASD is a complex disease[65], with considerable variation in clinical manifestations, neuroimaging findings, and developmental trajectories across individuals. Grouping all ASD cases into a single diagnostic category may obscure exposure effects that are specific to a particular subtype. Based on AI methods[65-68], this problem has begun to be addressed by identifying clinical characteristics or neuroanatomical subtypes in ASD. For example, one study identified four reproducible ASD subtypes based on resting-state fMRI features using a

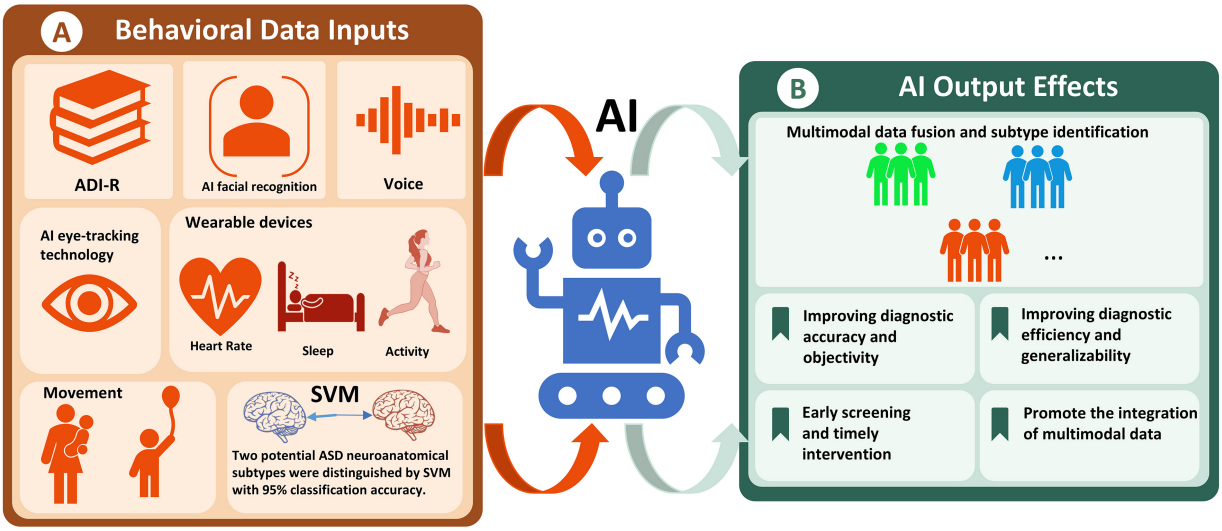


Figure 2. Advantages and technical framework of AI in ASD phenotyping. (A) illustrates the various behavioral input data for AI-assisted ASD diagnosis, including ADI-R, AI facial recognition, voice analysis, wearable devices, AI eye-tracking technology, and physiological/behavioral data such as heart rate, sleep, and activity; (B) summarizes the core values of AI-assisted diagnosis, which include multimodal data fusion and subtype identification, improved diagnostic accuracy and objectivity, improved diagnostic efficiency and scalability, early screening and timely intervention, and privacy protection and non-intrusiveness. Icons in this figure were created using BioGDP (https://BioGDP.com) [30]. AI: Artificial intelligence; ASD: autism spectrum disorder; SVM: support vector machine; ADI-R: autism diagnostic interview-revised.

ML approach, and subsequently examined clinical differences across these subtypes[67]. Fan *et al.* identified two neuroanatomical subtypes of ASD, which were distinguished by an support vector machine model with 95% classification accuracy[68]. Wang *et al.* used semi-supervised clustering to identify two reproducible ASD subtypes - hyper-connected and hypo-connected - with distinct connectivity patterns across large-scale brain networks[65]. Li *et al.* further developed a dual-autoencoder framework based on population graphs, integrating resting-state functional magnetic resonance imaging (MRI) and clinical non-imaging data, which revealed two ASD subtypes with distinct functional connectivity patterns and behavioral correlates[66]. These AI-driven subtypes may help characterize outcome heterogeneity, allowing researchers to test whether environmental exposures are more strongly associated with specific clinical phenotypes than with a broad autism diagnosis.

AI has also shown considerable promise in quantifying behavioral characteristics through multiple data modalities[69]. Traditional ML models perform well in analyzing facial expressions and gaze patterns[69]. A script-based behavioral understanding framework using multiple LLMs in a collaborative manner achieved an F1 score of 95.24% on ASD classification, outperforming both behavioral signal processing baselines and human raters while preserving interpretability[26].A multimodal strategy that integrated behavioral, genetic, and structural MRI data achieved a diagnostic accuracy of 98.7% on a held-out test set, with model selection based on validation performance[70]. Eye-tracking technology has shown utility in ASD diagnosis, and the EarliPoint System—an Food and Drug Administration (FDA)-cleared device—reported a sensitivity of 71% in identifying children with ASD in a clinical study[71]. AI-enabled wearable devices can support continuous behavioral monitoring and can predict behavioral escalations 1 to 3 min in advance[72].

From the perspective of environmental etiology, the core value of AI-assisted phenotyping lies in shifting the diagnostic framework from binary classification toward continuous, dimensional outcome measures. The behavioral characteristics of AI quantification (such as social gaze duration and eye movement patterns) can

be incorporated as continuous variables, which not only improve statistical efficiency but also help identify environmental risk factors related to specific behavioral dimensions rather than a general ASD diagnosis. In addition, AI-driven real-time monitoring enables the capture of dynamic profiles of both environmental exposures and individual symptom fluctuations. Rather than relying on point-in-time comparisons, this approach allows researchers to examine how changes in exposure trajectories relate to changes in outcome trajectories over time.

AI FOR ASSOCIATION ANALYSIS IN ASD ENVIRONMENTAL ETIOLOGY RESEARCH

Methodological landscape: from variable screening to causal inference

ML methods have been widely used in variable screening and exposure-reaction modeling. Least absolute shrinkage and selection operator (Lasso), random forest, and XGBoost can help screen or prioritize key predictors from high-dimensional related exposure data[73]. Gradient boosting can flexibly fit nonlinear dose-response relationships without requiring prespecified functional forms[74]. However, these models often lack interpretability, which may limit their usefulness for environmental etiology research. Explainable AI methods such as SHAP and local interpretable model-agnostic explanations (LIME) explain how individual features contribute to model predictions, but they do not estimate epidemiological or causal effect[75]. In mixed exposure analysis, weighted quantile sum (WQS) can estimate the overall mixture effect and identify the contribution of individual components, whereas Bayesian kernel machine regression (BKMR) can further explore nonlinear exposure–response relationships and potential interactions among component[73]. Causal forest, targeted maximum likelihood estimation (TMLE), and double/debiased ML can estimate how effects vary across different subgroups from observational data, rather than just giving a single average effect for the whole population[76].

In recent years, DL and LLMs have further expanded the ability of mixture analysis. Traditional regression approaches are generally designed to assess associations with individual pollutants or prespecified combinations, which may limit their ability to capture joint effects from unmeasured or emerging chemicals detected through nontargeted analysis. One study used a fine-tuned LLM to extract toxicity-related information from nearly 10,000 publications, generating over 20,000 toxicity records that were then combined with ML to prioritize previously uncharacterized chemicals for risk assessment[77]; another study developed an LLM-based system for functional annotation and initial screening of unknown substances detected through non-targeted analysis, reporting approximately 85% accuracy[78]. In addition, the NeuralPLSI model[79] combines the flexibility of a neural network with an interpretable semi-parametric model, allowing researchers to identify key pollutants while maintaining predictive performance.

Current applications in ASD environmental association studies

Table 2 summarizes the AI methods applied to exposure-ASD association analysis, mixture discovery, and risk prediction in the reviewed studies. Mixed-exposure analysis remains one of the most critical challenges in ASD environmental etiology research. In real-world settings, individuals are exposed to complex pollutant mixtures, and the traditional single-pollutant model is difficult to handle the problem of high collinearity among mixed-exposure pollutants. A study combining WQS regression with symbolic iterative random forests found two synergistic effects that only occur when the exposure level exceeds the 75th percentile: the pairing of cadmium with organophosphorus pesticide metabolites and chlorophenol paired with the same metabolite, both of which are ignored in a single chemical analysis[80]. Rud *et al.* analyzed more than 300,000 mother-child pairs from Southern California using frequentist grouped weighted quantile sum (FGWQS) regression[81]. They found that trace metals, particularly copper, carried the greatest weight in the association with ASD risk, suggesting that metal composition rather than total $PM_{2.5}$ mass may be more relevant to neurodevelopment[81]. A key advantage of this method is that it can simultaneously estimate group mixture effects and individual chemical weights without data splitting, and its computational efficiency on large-scale data is substantially higher than that of Bayesian approaches[81].

Table 2. AI and advanced statistical methods for exposure-ASD association analysis, mixture discovery, and risk prediction

Authors (Year)	Study type	Analytical task	AI/ML method	Population	Sample size	Objectives	Key findings
Association analysis							
Midya *et al.* (2023)[80]	Case-control	Estimate mixture effects and component weights	WQS; SiRF	CHARGE Study, USA (2006-2017)	479 (231 ASD cases, 248 controls)	Detecting potential synergistic effects in chemical mixtures	Cadmium paired with an organophosphate pesticide metabolite, and chlorophenol paired with the same metabolite, showed potential synergistic effects only when exposure levels exceeded the 75th percentile
Rud *et al.* (2025)[81]	Cohort study	Estimate mixture effects and component weights	FGWQSR	Southern California (2001-2014)	317,767 (4,522 ASD cases)	Estimate joint mixture effects o PM2.5 components and identify dominant components in $PM_{2.5}$ mixtures	Copper (weight 0.65) showed the strongest association with ASD (OR = 1.13 per quintile)
Wang *et al.* (2023)[52]	Time-series study	Air pollution-meteorology interaction	GRU; SHAP	Nanjing, China (2015-2019)	~1.47 million visits	Identify key drivers of ASD outpatient visits	Humidity, NO_2, SO_2 as key drivers; distinguished synergistic *vs.* antagonistic interactions
Qi *et al.* (2025)[82]	Bioinformatics	Chemical prioritization	Lasso	Public transcriptomic datasets (GEO)	Not specified (bioinformatics analysis)	Reverse screen for neurodevelopmental toxicants	Identified 3 pesticides (epoxiconazole, flusilazole, DEET) as candidates for further validation
Ellul *et al.* (2023)[83]	Case-control	Compare ASD subgroups defined by MIA exposure	Decision tree; LDA	France (Paris + 8 ASD Expert Centers)	295 (209 Paris, 86 multicenter)	Distinguish ASD subgroups by environmental exposure history	MIA-positive subgroup showed more severe socialization difficulties
Risk prediction							
Curtin *et al.* (2018)[53]	Case-control	Risk prediction	WQS ensemble; Lasso	Sweden, UK, USA (4 cohorts)	193 (80 ASD cases)	Predict ASD from fetal zinc-copper exposure cycles	90% accuracy across 4 independent cohorts; cyclical features (not raw concentrations) distinguished cases
Ling *et al.* (2023)[54]	Case-control	Risk prediction	SVM	Guangxi, China	60 (30 ASD, 30 controls)	Classify ASD based on serum copper isotopes	94.4% accuracy (30 ASD, 30 controls)
Doi *et al.* (2024)[91]	Birth cohort	Risk prediction	PCA; Lasso	the Chiba Study of Mother and Child Health (C-MACH), Japan	115 mother-infant pairs	Combining PCB exposure with infant movement patterns for ASD risk prediction	Improved predictive performance *vs.* single modality

Each row corresponds to a cited study; full references are listed in the reference list. AI: Artificial intelligence; ASD: autism spectrum disorder; CI: confidence interval; DEET: N,N-diethyl-meta-toluamide; FGWQSR: frequentist grouped weighted quantile sum regression; GRU: gated recurrent unit; Lasso: least absolute shrinkage and selection operator; LDA: linear discriminant analysis; MIA: maternal immune activation; ML: machine learning; OR: odds ratio; PCA: principal component analysis; SiRF: signed iterative random forest; SVM: support vector machine; WQS: weighted quantile sum; SHAP: SHapley Additive exPlanations; GEO: gene expression omnibus; PCB: pol chlorinated biphenls.

A key challenge in environmental etiology research is identifying the most hazardous chemicals from the vast exposome data, allowing subsequent analyses to focus on the highest-priority candidates. To address this problem, one study implemented a reverse screening strategy. The researchers used Lasso to regress the ASD-related transcriptome data to identify the central hub genes, then compared these genes with the toxicological genomics database, thereby prioritizing three pesticides, epoxiconazole, flusilazole, and N,N-diethyl-meta-toluamide (DEET) , for further analysis[82]. Furthermore, a time-series study used gated recurrent unit networks combined with SHAP analysis to identify humidity, NO_2, and SO_2 as the key drivers of ASD outpatient visits and to distinguish synergistic and antagonistic effects among the pollutants[52]. Previous models usually assume that all cases of ASD are homogeneous, which makes it difficult to identify the effect of subtype specificity. AI methods can automatically identify subtypes from data, helping researchers examine whether exposure factors are associated with specific clinical phenotypes. Ellul *et al.* applied classification decision trees and linear discriminant analysis to distinguish ASD children based on maternal immune activation during pregnancy, and found that the maternal immune activation (MIA)-positive subgroup showed more severe social adaptation difficulties, independent of overall ASD severity[83].

AI methods for causal inference

Predictive models and causal inference are different things. Most AI methods reviewed here, including random forest, support vector machines, and gradient boosting - are predictive. They find associations and generate hypotheses, but they do not estimate causal effects. This distinction is often overlooked in environmental etiology research, yet it matters for interpretation and policy.

Exposures are often correlated with numerous demographic, behavioral, and socioeconomic factors, making traditional adjustment for confounders impractical. Causal machine learning methods can flexibly model measured confounders without requiring prespecified parametric forms, which may help reduce residual confounding due to model misspecification[84]. For instance, Zhou *et al.* developed a spatial causal tensor completion framework that jointly models multiple per- and polyfluoroalkyl substances (PFAS) exposures [including perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS)] and 13 chronic disease outcomes, while adjusting for latent spatial confounders using graph Laplacian eigenvectors. Applied to national PFAS monitoring data, the approach yielded more conservative causal estimates than conventional methods[85]. Some methods also address unmeasured confounding - a longstanding challenge in observational environmental studies. The Spatial Deconfounder, for instance, reconstructs a substitute confounder from local treatment vectors using a conditional variational autoencoder and estimates causal effects via a flexible outcome model—an approach that may help identify direct and spillover effects under specific model assumptions, though these remain to be empirically validated[86]. Similarly, a diffusion-based causal model has been proposed to apply backdoor adjustments during sampling to reduce bias from unmeasured confounders with spatial or temporal patterns[87]. Some AI-based frameworks go beyond prediction. Causal forests, for example, extend random forests to estimate how effects vary across subgroups. Naito *et al.* used this approach in the UK Biobank and BioBank Japan and found that the obesity-diabetes link varied with polygenic risk. People in the top 10% of genetic risk had a 2.3-fold higher risk than those in the bottom 10%[19]. TMLE combines ML with a bias-correction step and can be used to estimate causal effects when the relevant assumptions are met. Herrera *et al.* used TMLE to estimate the health impact of living near mines in Chile[88]. They estimated that moving children away from mining sites could cut respiratory disease by about four percentage points[88]. Gao *et al.* recently applied causal machine learning - including refutation tests and exposure-response curves - to e-waste exposures and kidney injury markers[89]. About one-third of the pollutant-biomarker associations they tested withstood causal refutation, with primary aromatic amines showing the strongest potential effects on kidney injury[89]. Double/debiased ML has also been used to assess $PM_{2.5}$ components and cognitive decline while accounting for measurement error. After correction, bromine, manganese, lead, and silicon showed the strongest links to faster decline[90].

None of the studies we reviewed applied TMLE, causal forests, or double/debiased ML to causal questions in this field.

AI IN INDIVIDUAL RISK PREDICTION FOR ASD

Most applications in Table 2 focus on chemical mixtures and substance prioritization, while the number of risk prediction studies remains limited. Unlike association analyses, prediction models estimate an individual's probability of ASD using a set of predictors. Risk prediction has received more attention in ASD research, but environmental exposure is rarely included. Most existing models rely on sociodemographic, familial, and developmental factors - such as birth parameters, growth trajectories, parental characteristics, and early language or motor milestones. For example, a retrospective cohort study of 604,835 children used a gradient boosting model based on routine wellness records from birth to 24 months to predict ASD likelihood, achieving an AUC of 0.81 (SD = 0.004) in threefold cross-validation[18,22].

Some studies with a proof-of-concept character suggest that measuring early-life exposure might improve prediction. A retrospective case-control study used cyclical features reconstructed from shed deciduous teeth to distinguish ASD cases from controls, achieving 90% classification accuracy across four cohorts[53]. Another study used ratios with a support vector machine and achieved 94.4% accuracy in distinguishing ASD cases from controls - but only 30 participants per group[54]. A prospective cohort study of prenatal polychlorinated biphenyls (PCB) exposure suggested a potential association with elevated ASD risk based on modified checklist for autism in toddlers screening at 18 months, although the sample size was small and the outcome was based on a screening tool rather than clinical diagnosi[91].

At present, many challenges exist in incorporating environmental exposure factors into risk prediction of ASD due to small sample size and the lack of routine exposure measurements in the clinical environment. Future studies could pretrain multi-modal representations using large birth cohorts with harmonized exposure and phenotypic data, followed by fine-tuning, recalibration, and external validation in ASD-specific cohorts. However, these methods need to be carefully verified before they can be applied to clinical or public health practice.

CHALLENGES AND FUTURE DIRECTIONS

Despite rapid advances in AI across many scientific fields, its application to research on the environmental etiology of ASD has been limited. Most ASD studies to date have relied on conventional ML methods, such as random forests, support vector machines, and gradient boosting, whereas DL methods have been used less frequently. This is partly due to small sample sizes and the complexity of ASD etiology. An additional challenge is that existing cohort infrastructure may not fully meet the data requirements of advanced AI methods: large, harmonized, well-annotated datasets are lacking for conventional ML, and domain-specific text corpora are additionally needed for LLM applications. Data are not standardized, and interdisciplinary collaboration remains insufficient.

Cohorts differ in exposure measurements, outcome definitions, and data formats, making harmonization difficult. The Environmental Influences on Child Health Outcomes (ECHO) program, which includes 69 cohorts and over 57,000 children, required substantial time and resources just to align data across sites[92]. The ECHO experience illustrates the substantial resources required for harmonizing heterogeneous cohort data, a challenge that is also relevant to ASD research.

The limitations of general-purpose LLMs in specialized environmental tasks illustrate the gap between promise and practice. In an air pollution control evaluation, 11 LLMs showed clear limitations in accuracy and hallucination rates, suggesting that general models need task-specific adaptation[93]. In public health

policy identification, agreement between generative AI and experts ranged from 67% to 78%, with the lowest consistency in policy interpretation[94]. These numbers are a reminder that LLMs are not ready for high-stakes environmental health applications without careful validation.

But there are reasons to be optimistic. Data standardization is achievable, and domain-specific fine-tuning is already improving LLM performance in specialized tasks[95]. Federated learning may support collaborative model training without centralizing raw data, although additional privacy safeguards and evaluation under between-site heterogeneity remain necessary[96,97]. This could enable the kind of large-scale, collaborative research that ASD environmental studies have long needed - without compromising participant privacy.

More broadly, the potential of AI in ASD environmental etiology research extends beyond current applications. As exposome data accumulate, AI will be able to integrate multiple environmental exposures, multi-omics data, and clinical phenotypes into more complete individual exposure profiles. For phenotype prediction, multimodal models combining imaging, behavioral, and genetic data could improve the resolution of ASD subtype identification, making it possible to link environmental exposures to specific subtypes more reliably. Transfer learning can leverage patterns learned from large general-population cohorts, helping to overcome the small-sample constraints that have long limited ASD research. Graph neural networks may help characterize complex relationships among exposures, genes, and health outcomes, but causal interpretation requires explicit causal assumptions, appropriate study designs, and dedicated causal inference methods. These methods are already being used in other areas of environmental health. These methods may be adaptable to ASD environmental etiology research, but their validity and utility require evaluation in ASD-specific datasets. Integrating them into ASD environmental studies may not yield immediate breakthroughs, but it points in a clear direction: away from single-dataset, single-model approaches, and toward more systematic, mechanism-oriented research that can meaningfully address the exposure-etiology questions that matter.

CONCLUSION

AI has begun to reshape research on the environmental etiology of ASD, but progress is uneven. In exposure assessment, AI has improved model resolution; in phenotyping, it has enabled more objective measurement; in association analysis, it has helped prioritize key toxicants; and in risk prediction, early-life markers have distinguished ASD cases with considerable accuracy.

Yet most studies still rely on ML. DL, causal inference, and LLMs are rarely used. The obstacles are not just technical - small cohorts, unstandardized data, and weak interdisciplinary collaboration - but also infrastructural. Progress will require better data, clearer causal questions, and closer collaboration.

DECLARATIONS

Acknowledgments

We thank the BioGDP platform (https://BioGDP.com) for providing the factory and mass spectrometer icons used in the Graphical Abstract, as well as several icons used in the figures within the main text.

Authors' contributions

Conceptualization and study design: Liu, H.; Xia, W.
Literature search and data extraction: Zheng, R.; Wen, L.; Huang, Y.; Gan, Z.
Analysis and interpretation: Zheng, R.; Yin, H.
Drafting of the manuscript: Zheng, R.
Critical revision for important intellectual content: Liu, H.; Xia, W.; Yin, H.
Supervision and funding acquisition: Liu, H.
All authors reviewed and approved the final manuscript.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool DeepSeek (version 3.2, released 2026-04-06) was used solely for language polishing and expression refinement. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

This work was supported by the National Natural Science Foundation of China (No. 42477463; No. 22236001) and the Open Project Funding of the Hubei Key Laboratory of Pollution Damage Assessment and Environmental Health Risk Prevention and Control (HAES-HJJK202402).

Conflicts of interest

Xia, W. and Liu, H. are the Guest Editors of the Special Issue "AI in Environmental Exposure and Health" of the journal *Journal of Environmental Exposure Assessment*. Liu, H. is also a Junior Editorial Board Member of the journal *Journal of Environmental Exposure Assessment*. They were not involved in any steps of editorial processing, notably including reviewers' selection, manuscript handling, and decision-making. The other authors declare no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Supplementary Materials

Supplementary Materials

REFERENCES

1. Wilk-Jakubowski, J. L.; Kuchcinski, A.; Wilk-Jakubowski, G. K.; Palej, A.; Pawlik, L. Digital technologies and machine learning in environmental hazard monitoring: a synthesis of evidence for floods, air pollution, earthquakes, and fires. *Sensors* **2026**, *26*, 893. DOI PubMed PMC
2. Lin, M.; Xu, Y.; Peng, J.; et al. From passive prevention to proactive intervention: an AI-assisted integrated governance system for algal bloom monitoring and control. *Environ. Res.* **2026**, *289*, 123379. DOI PubMed
3. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44-56. DOI PubMed
4. Esteva, A.; Robicquet, A.; Ramsundar, B.; et al. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24-9. DOI PubMed
5. Shaw, K.A.; Williams, S.; Patrick, M. E.; et 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.* **2025**, *74*, 1-22. DOI PubMed PMC
6. Bhattacharya, B.; Toor, D.; Chatterjee, M. Connecting the dots: environmental pollution and Autism Spectrum Disorder. *Rev. Environ. Health.* **2025**, *40*, 602-15. DOI PubMed
7. Ma, Y.; Li, D.; Cui, F.; et al. Exposure to air pollutants and myocardial infarction incidence: a UK Biobank study exploring gene-environment interaction. *Environ. Health. Perspect.* **2024**, *132*, 107002. DOI PubMed PMC
8. Lee, T. H. Y.; Li, C.; Dos Santos, M. M.; et al. Assessment of emerging and persistent contaminants in an anthropogenic-impacted watershed: application using targeted, non-targeted, and in vitro bioassay techniques. *Chemosphere* **2024**, *364*, 143067. DOI PubMed
9. Vitale, G. A.; Xia, S. N.; Dührkop, K.; et al. Enhancing tandem mass spectrometry-based metabolite annotation with online chemical labeling. *Nat. Commun.* **2025**, *16*, 6911. DOI PubMed PMC
10. Basu, B.; Alam, M. S.; Ghosh, B.; Gill, L.; Mcnabola, A. Mcnabola, A. Augmenting limited background monitoring data for improved performance in land use regression modelling: using support vector regression and mobile monitoring. *Atmos. Environ.* **2019**, *201*, 310-22. DOI

11. Ryu, H.; Choi, Y. H.; Kim, E.; et al. Misclassification and characterization of exposure to humidifier disinfectants using a questionnaire. *BMC. Public. Health.* **2021**, *21*, 1458. DOI PubMed PMC
12. Bradman, A.; Kogut, K.; Eisen, E. A.; et al. Variability of organophosphorous pesticide metabolite levels in spot and 24-hr urine samples collected from young children during 1 week. *Environ. Health. Perspect.* **2013**, *121*, 118-24. DOI PubMed PMC
13. Pati, A. K.; Tripathy, A. R.; Subudhi, S. Intelligence frameworks for environmental pollution assessment: a review on air and water quality monitoring systems. *Measurement* **2026**, *258*, 119122. DOI
14. O'Sharkey, K.; Mitra, S.; Chow, T.; et al. Air pollution and autism spectrum disorder: unveiling multipollutant risks and sociodemographic influences in California. *Environ. Health. Perspect.* **2025**, *133*, 067010. DOI PubMed PMC
15. Swaminathan, S.; Park, N.; Soares, E. A.; Brazil, E. A. V. From descriptions to chemical hazards: predicting persistence, bioaccumulation, and toxicity from natural language using LLMs. In *Book of Abstracts*, ML4PS Workshop at NeurIPS 2025, San Diego, CA, December 2025; Paper No. 225. https://ml4physicalsciences.github.io/2025/files/NeurIPS_ML4PS_2025_225.pdf (accessed 2026-08-10).
16. Qiang, H.; Wang, F.; Lu, W.; et al. Language model-guided anticipation and discovery of mammalian metabolites. *Nature* **2026**, *651*, 211-20. DOI PubMed PMC
17. Esu, C. O.; Pyo, J.; Cho, K. Machine learning-derived dose-response relationships considering interactions in mixtures: Applications to the oxidative potential of particulate matter. *J. Hazard. Mater.* **2024**, *475*, 134864. DOI PubMed
18. Galatro, D.; Di Nardo, A.; Pai, V.; et al. Considerations for using tree-based machine learning to assess causation between demographic and environmental risk factors and health outcomes. *Environ. Sci. Pollut. Res. Int.* **2024**, *31*, 60927-35. DOI PubMed
19. Naito, T.; Inoue, K.; Namba, S.; et al. ; BioBank Japan. Machine learning reveals heterogeneous associations between environmental factors and cardiometabolic diseases across polygenic risk scores. *Commun. Med.* **2024**, *4*, 181. DOI PubMed PMC
20. Chatterjee, A.; Woodruff, H.; Wu, G.; Lambin, P. Limitations of only reporting the odds ratio in the age of precision medicine: a deterministic simulation study. *Front. Med.* **2021**, *8*, 640854. DOI PubMed PMC
21. Zhang, L.; He, W.; Huang, P.; Zhong, L. A risk prediction model for autism spectrum disorder integrating biopsychosocial factors: a systematic review and meta-analysis with multicenter validation. *Transl. Pediatr.* **2025**, *14*, 3219-30. DOI PubMed PMC
22. Nam, J. W.; Choi, E. Y.; Ailshire, J. A.; Chiang, Y. Y. Unveiling population heterogeneity in health risks posed by environmental hazards using regression-guided neural network. *Sci. Rep.* **2026**, *16*, 24483. DOI PubMed
23. Cortese, S.; Bellato, A.; Gabellone, A.; et al. Latest clinical frontiers related to autism diagnostic strategies. *Cell. Rep. Med.* **2025**, *6*, 101916. DOI PubMed PMC
24. Brian, J. A.; Zwaigenbaum, L.; Ip, A. Standards of diagnostic assessment for autism spectrum disorder. *Paediatr. Child. Health.* **2019**, *24*, 444-60. DOI PubMed PMC
25. Grzadzinski, R.; Huerta, M.; Lord, C. DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes. *Mol. Autism.* **2013**, *4*, 12. DOI PubMed PMC
26. Liu, W.; Pan, Y.; Zhang, D.; Deng, H.; Zou, X.; Li, M. Detecting children with autism spectrum disorder based on script-centric behavior understanding with emotional enhancement. *Neurocomputing* **2026**, *670*, 132434. DOI
27. Kayış, H.; Çelik, M.; Gedizlioğlu, Ç.; et al. A new approach in autism diagnosis: evaluating natural interaction using point of view (POV) glasses. *Asian. J. Psychiatr.* **2026**, *116*, 104798. DOI PubMed
28. Hu, C.; Thrasher, J.; Li, W.; et al. Speech pattern disorders in verbally fluent individuals with autism spectrum disorder: a machine learning analysis. *Front. Neuroinform.* **2025**, *19*, 1647194. DOI PubMed PMC
29. Aarthi, D.; Kannimuthu, S. MACAFNet transformer-based multi-atlas fusion framework for autism spectrum disorder classification using functional connectivity. *Sci. Rep.* **2026**, *16*, 20075. DOI PubMed PMC
30. Jiang, S.; Li, H.; Zhang, L.; et al. Generic Diagramming Platform (GDP): a comprehensive database of high-quality biomedical graphics. *Nucleic. Acids. Res.* **2025**, *53*, D1670-6. DOI PubMed PMC
31. Wild, C. P. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer. Epidemiol. Biomarkers. Prev.* **2005**, *14*, 1847-50. DOI PubMed
32. Hu, H.; Liu, X.; Zheng, Y.; et al. Methodological challenges in spatial and contextual exposome-health studies. *Crit. Rev. Environ. Sci. Technol.* **2023**, *53*, 827-46. DOI PubMed PMC
33. Ponzano, M.; Rotem, R. S.; Bellavia, A. Complex methods for complex data: key considerations for interpretable and actionable results in exposome research. *Eur. J. Epidemiol.* **2025**, *40*, 1399-403. DOI PubMed
34. Isola, S.; Murdaca, G.; Brunetto, S.; Zumbo, E.; Tonacci, A.; Gangemi, S. The use of artificial intelligence to analyze the exposome in the development of chronic diseases: a review of the current literature. *Informatics* **2024**, *11*, 86. DOI
35. Rajesh, M.; Babu, R. G.; Moorthy, U.; Easwaramoorthy, S. V. Machine learning-driven framework for realtime air quality assessment and predictive environmental health risk mapping. *Sci. Rep.* **2025**, *15*, 28801. DOI PubMed PMC

36. Cox, L. A. Jr. Implications of nonlinearity, confounding, and interactions for estimating exposure concentration-response functions in quantitative risk analysis. *Environ. Res.* **2020**, *187*, 109638. DOI PubMed PMC
37. Gianquintieri, L.; Oxoli, D.; Caiani, E. G.; Brovelli, M. A. State-of-art in modelling particulate matter (PM) concentration: a scoping review of aims and methods. *Environ. Dev. Sustain.* **2025**, *27*, 25889-911. DOI
38. Guo, F.; Chang, Y.; Zhang, J.; et al. Predicting gestational hypertension risk in high-density cities: a time-weighted heat-humidity exposure and built environment analysis using XGBoost in Shanghai. *IEEE. Access.* **2025**, *13*, 194197-211. DOI
39. Rao, L.; Zhao, S.; Wang, Q.; et al. A GRU-based framework for air quality forecasting and spatiotemporal visual analytics. In *Proceedings of the 2025 6th International Conference on Computer Science and Management Technology (ICCSMT '25)*, ACM, 2026; pp 121-6. DOI
40. Kheder, A.; Toropainen, H.; Peng, W.; et al. TopoFlow: topography-aware pollutant Flow learning for high-resolution air quality prediction. *npj. Clim. Atmos. Sci.* **2026**, *9*, 143. DOI
41. Xu, Y.; Ma, Y.; Xu, W.; Yang, Z.; Ting, K. M. A large language model for deriving spectral embeddings for accurate compound identification in mass spectrometry. *Commun. Chem.* **2025**, *8*, 326. DOI PubMed PMC
42. Sillé, F. C. M.; Prasse, C.; Luechtefeld, T.; Hartung, T. AI redefines mass spectrometry chemicals identification: retention time prediction in metabolomics and for a Human Exposome Project. *Front. Public. Health.* **2025**, *13*, 1687056. DOI PubMed PMC
43. Iyer, H. S.; Karasaki, S.; Yi, L.; Hswen, Y.; James, P.; Vopham, T. Harnessing geospatial artificial intelligence (GeoAI) for environmental epidemiology: a narrative review. *Curr. Environ. Health. Rep.* **2025**, *12*, 34. DOI PubMed PMC
44. Konstantinou, G. N.; Anderson, W. C. 3rd; Bagkis, E.; et al. Artificial intelligence-driven wearable and connected technology for allergy: real-time monitoring and predictive management for personalized care. *J. Allergy. Clin. Immunol. Pract.* **2025**, *13*, 2903-13.e2. DOI PubMed PMC
45. Ren, X.; Mi, Z.; Georgopoulos, P. G. Socioexposomics of COVID-19 across New Jersey: a comparison of geostatistical and machine learning approaches. *J. Expo. Sci. Environ. Epidemiol.* **2024**, *34*, 197-207. DOI PubMed PMC
46. Johnson, T. DigitalExposome: A dataset for wellbeing classification using environmental air quality and human physiological data. *Data. Brief.* **2025**, *59*, 111442. DOI PubMed PMC
47. Johnson, T.; Kanjo, E.; Woodward, K. DigitalExposome: quantifying impact of urban environment on wellbeing using sensor fusion and deep learning. *Comput. Urban. Sci.* **2023**, *3*, 14. DOI PubMed PMC
48. Ibáñez, A.; Duran-Aniotz, C.; Migeot, J.; et al. Computational whole-body-exposome models for global precision brain health. *Nat. Commun.* **2025**, *16*, 11078. DOI PubMed PMC
49. O'Sharkey, K.; Meng, Q.; Mitra, S.; et al. Associations between brake and tire wear-related $PM_{2.5}$ metal components, particulate oxidative stress potential, and autism spectrum disorder in Southern California. *Environ. Int.* **2024**, *185*, 108573. DOI PubMed
50. O'sharkey, K.; Mitra, S.; Chow, T.; et al. Prenatal exposure to criteria air pollution and traffic-related air toxics and risk of autism spectrum disorder: a population-based cohort study of California births (1990-2018). *Environ. Int.* **2025**, *201*, 109562. DOI PubMed
51. O'Sharkey, K.; Chow, T.; Mitra, S.; et al. Exploring the link between grandmaternal air pollution exposure and Grandchild's ASD risk: A multigenerational population-based study in California. *Environ. Int.* **2025**, *200*, 109526. DOI PubMed
52. Wang, C.; Qi, Y.; Chen, Z. Explainable gated recurrent unit to explore the effect of co-exposure to multiple air pollutants and meteorological conditions on mental health outcomes. *Environ. Int.* **2023**, *171*, 107689. DOI PubMed
53. Curtin, P.; Austin, C.; Curtin, A.; et al. ; (for the Emergent Dynamical Systems Group). Dynamical features in fetal and postnatal zinc-copper metabolic cycles predict the emergence of autism spectrum disorder. *Sci. Adv.* **2018**, *4*, eaat1293. DOI PubMed PMC
54. Ling, W.; Zhao, G.; Wang, W.; et al. Metallomic profiling and natural copper isotopic signatures of childhood autism in serum and red blood cells. *Chemosphere* **2023**, *330*, 138700. DOI PubMed
55. Hupatz, H.; Rahu, I.; Wang, W. C.; Peets, P.; Palm, E. H.; Kruve, A. Critical review on in silico methods for structural annotation of chemicals detected with LC/HRMS non-targeted screening. *Anal. Bioanal. Chem.* **2025**, *417*, 473-93. DOI PubMed PMC
56. Stienstra, C. M. K.; Nazdrajić, E.; Hopkins, W. S. From reverse phase chromatography to HILIC: graph transformers power method-independent machine learning of retention times. *Anal. Chem.* **2025**, *97*, 4461-72. DOI PubMed
57. Vernocchi, P.; Marangelo, C.; Guerrera, S.; et al. Gut microbiota functional profiling in autism spectrum disorders: bacterial VOCs and related metabolic pathways acting as disease biomarkers and predictors. *Front. Microbiol.* **2023**, *14*, 1287350. DOI PubMed PMC
58. Cloutier, M.; Yu, C.; Talarico, R.; et al. Prenatal exposure to fine particulate matter components and autism risk in childhood. *JAMA. Netw. Open.* **2025**, *8*, e2538882. DOI PubMed PMC
59. Heo, S.; Afanasyeva, Y.; Trasande, L.; Bell, M. L. Ghassabian, A. Residential mobility in pregnancy and potential exposure misclassification of air pollution, temperature, and greenness. *Environ. Epidemiol.* **2023**, *7*, e273. DOI PubMed PMC
60. Gupta, N.; Srinivasan, S.; Gupta, M. Rethinking psychometric testing in autism: overcoming the challenges of comorbidity and diagnostic overshadowing. *CNS. Spectr.* **2025**, *30*, e19. DOI PubMed PMC

61. Lingham, S.; Perumal, T. L.; Lewis, R.; Rodrigues, S.; Mukherjee, M. 8406 Use of a local bespoke social communication difficulties (SCD) proforma instead of ADI-R in children less than 5 years to capture DSM-5 criteria from history. *Arch. Dis. Child.* **2025**, *110*, A380-2. DOI
62. Parish-Morris, J. Viewing early diagnosis through a developmental lens - "Autistic, yet still a girl". *JAMA. Netw. Open.* **2025**, *8*, e2525893. DOI PubMed
63. Kulage, K. M.; Goldberg, J.; Usseglio, J.; Romero, D.; Bain, J. M.; Smaldone, A. M. How has DSM-5 affected autism diagnosis? A 5-year follow-up systematic literature review and meta-analysis. *J. Autism. Dev. Disord.* **2020**, *50*, 2102-27. DOI PubMed
64. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1000-7. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators. Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2010. *MMWR Surveill. Summ.* **2014**, *63*, 1-21. Available online: https://www.cdc.gov/mmwr/preview/mmwrhtml/ss6302a1.htm (accessed on 10 August 2026).
65. Wang, Y.; Chen, Z.; Song, P.; et al. Diagnosis-informed neuro-subtyping reveals subgroups of autism spectrum disorder with reliable and distinct functional connectivity profiles. *Prog. Neuropsychopharmacol. Biol. Psychiatry.* **2025**, *141*, 111452. DOI PubMed
66. Li, X.; Xu, G.; Geng, G.; et al. Subtyping autism spectrum disorder with a population graph-based dual autoencoder: revealing two distinct biotypes. *CNS. Neurosci. Ther.* **2025**, *31*, e70675. DOI PubMed PMC
67. Li, J.; Zheng, W.; Fu, X.; et al. Individual deviation-based functional hypergraph for identifying subtypes of autism spectrum disorder. *Brain. Sci.* **2024**, *14*, 738. DOI PubMed PMC
68. Fan, X. R.; He, Y.; Wang, Y. S.; et al. Profiling brain morphology for autism spectrum disorder with two cross-culture large-scale consortia. *Commun. Biol.* **2025**, *8*, 1157. DOI PubMed PMC
69. Mohammadi, F.; Shahrokhi, H.; Asadzadeh, A.; Pirmoradi, S.; Moghtader, A.; Rezaei-Hachesu, P. Artificial intelligence in autism spectrum disorder diagnosis: a scoping review of face, voice, and text analysis methods. *Health. Sci. Rep.* **2025**, *8*, e71476. DOI PubMed PMC
70. Malik, W.; Fahiem, M. A.; Khan, J.; Jung, Y.; Alturise, F. An adaptive transfer learning framework for multimodal autism spectrum disorder diagnosis. *Life* **2025**, *15*, 1524. DOI PubMed PMC
71. Jones, W.; Klaiman, C.; Richardson, S.; et al. Eye-tracking-based measurement of social visual engagement compared with expert clinical diagnosis of autism. *JAMA* **2023**, *330*, 854-65. DOI PubMed PMC
72. Rad, A. B.; Villavicencio, T.; Kiarashi, Y.; et al. From motion to emotion: exploring challenging behaviors in autism spectrum disorder through analysis of wearable physiology and movement. *Physiol. Meas.* **2025**, *13*, 015004. DOI PubMed PMC
73. Yu, L.; Liu, W.; Wang, X.; et al. A review of practical statistical methods used in epidemiological studies to estimate the health effects of multi-pollutant mixture. *Environ. Pollut.* **2022**, *306*, 119356. DOI PubMed
74. Schwartz, J.; Feng, Y.; Castro, E.; Wei, Y. Causal concentration-response modeling with continuous curves and exposure error correction: PM2.5 and mortality in the medicare cohort. *Environ. Health. Perspect.* **2025**, *133*, 067007. DOI PubMed PMC
75. Tasioulis, T.; Bagkis, E.; Kassandros, T.; Karatzas, K. The quest for the best explanation: comparing models and XAI methods in air quality modeling tasks. *Appl. Sci.* **2025**, *15*, 7390. DOI
76. Arif, S. Estimating causal effects with machine learning: a guide for ecologists. *Methods. Ecol. Evol.* **2025**, *16*, 2771-83. DOI
77. Zhang, S.; Guo, J.; Qian, Y.; Mao, M.; An, D. AI-enabled mapping of structure-hazard relationships for emerging contaminants. *Environ. Health.* **2026**, *4*, 1251-62. DOI PubMed PMC
78. Cheng, F.; Li, Q.; He, L.; et al. Leveraging large language models for contextual prioritization of contaminants of emerging concern in chemical mixtures. *Environ. Sci. Technol.* **2026**, *60*, 11380-91. DOI PubMed
79. Do, H.; Wang, Y.; Liu, M.; Lee, M. Neural network-based partial-linear single-index models for environmental mixtures analysis. *arXiv* **2025**, arXiv:2512.11593. Available online: https://arxiv.org/abs/2512.11593 (accessed 21 July 2026).
80. Midya, V.; Alcala, C. S.; Rechtman, E.; et al. Machine learning assisted discovery of interactions between pesticides, phthalates, phenols, and trace elements in child neurodevelopment. *Environ. Sci. Technol.* **2023**, *57*, 18139-50. DOI PubMed PMC
81. Rud, D.; Mostafijur Rahman, M.; Xiang, A. H.; et al. Frequentist grouped weighted quantile sum regression for correlated chemical mixtures. *Stat. Med.* **2025**, *44*, e70078. DOI PubMed PMC
82. Qi, L.; Yang, J.; Niu, Q.; Li, J. Exploring pesticide risk in autism via integrative machine learning and network toxicology. *Ecotoxicol. Environ. Saf.* **2025**, *297*, 118233. DOI PubMed
83. Ellul, P.; Maruani, A.; Vantalon, V.; et al. Maternal immune activation during pregnancy is associated with more difficulties in socio-adaptive behaviors in autism spectrum disorder. *Sci. Rep.* **2023**, *13*, 17687. DOI PubMed PMC
84. Huang, Z.; Dong, K.; Lin, T.; Antonelli, J. Multivariate incremental effects for continuous treatments: studying the health effects of environmental mixtures. *arXiv* **2026**, arXiv:2604.23534v1. Available online: https://arxiv.org/html/2604.23534v1 (accessed 21 July 2026).

85. Zhou, X.; Reich, B. J.; Yang, S. Spatial causal tensor completion for multiple exposures and outcomes: an application to the health effects of PFAS pollution. *arXiv* **2026**, arXiv:2603.16854. Available online: https://arxiv.org/abs/2603.16854 (accessed 21 July 2026).
86. Khot, A.; Oprescu, M.; Schröder, M.; Kagawa, A.; Luo, X. Spatial Deconfounder: interference-aware deconfounding for spatial causal inference. *arXiv* **2026**, arXiv:2510.08762. Available online: https://arxiv.org/abs/2510.08762 (accessed 21 July 2026).
87. Liu, X.; Qian, L.; Chen, S. X.; Tang, N. Partially Functional Dynamic Backdoor Diffusion-Based Causal Model. *arXiv* **2026**, arXiv:2509.00472. Available online: https://arxiv.org/abs/2509.00472 (accessed 21 July 2026).
88. Herrera, R.; Berger, U.; Von Ehrenstein, O. S.; et al. Estimating the causal impact of proximity to gold and copper mines on respiratory diseases in chilean children: an application of targeted maximum likelihood estimation. *Int. J. Environ. Res. Public. Health.* **2017**, *15*, 39. DOI PubMed PMC
89. Yang, L.; Gao, C.; Qin, R.; Wu, Q.; Sun, H.; Zhang, T. Impact of e-waste pollutant exposure on renal injury and oxidative stress biomarkers: Evidence from causal machine learning. *J. Hazard. Mater.* **2025**, *495*, 138831. DOI PubMed
90. Zhang, B.; Hart, J. E.; Laden, F.; et al. The association between PM2.5 components and cognitive decline: the impact of measurement error correction. *Environ. Res.* **2026**, *301*, 124533. DOI PubMed PMC
91. Doi, H.; Furui, A.; Ueda, R.; et al. Risk of autism spectrum disorder at 18 months of age is associated with prenatal level of polychlorinated biphenyls exposure in a Japanese birth cohort. *Sci. Rep.* **2024**, *14*, 31872. DOI PubMed PMC
92. Jacobson, L. P.; Parker, C. B.; Cella, D.; et al. Approaches to protocol standardization and data harmonization in the ECHO-wide cohort study. *Pediatr. Res.* **2024**, *95*, 1726-33. DOI PubMed PMC
93. Dai, S.; Wu, Q.; Zhang, H.; Wang, Y.; Wang, S. Evaluating the performance of large language models for one atmosphere using automated extracted datasets. *Environ. Sci. Technol.* **2025**, *60*, 725-35. DOI PubMed
94. Wilson, R.; Weets, C. M.; Rosner, A.; Katz, R. Evaluating generative artificial intelligence's limitations in health policy identification and interpretation. *PLoS. One.* **2024**, *19*, e0312078. DOI PubMed PMC
95. Zhang, Y.; Lin, S.; Xiong, Y.; et al. Fine-tuning large language models for interdisciplinary environmental challenges. *Environ. Sci. Ecotechnol.* **2025**, *27*, 100608. DOI PubMed PMC
96. Vimalapriya, M. D.; Mythili, R.; Poonguzhali, A.; Manikandan, G.; Alabdulatif, A.; Yasmin, J. H. J. A hierarchical privacy-preserving federated learning framework with differential privacy for collaborative healthcare diagnostics. *SN. Comput. Sci.* **2026**, *7*, 398. DOI
97. Wang, H.; Zhang, X.; Ren, X.; et al. Collaborative and privacy-preserving cross-vendor united diagnostic imaging via server-rotating federated machine learning. *Commun. Eng.* **2025**, *4*, 148. DOI PubMed PMC