

Review

Machine Learning Approaches to Identify and Classify ADHD: A Narrative Review with Tabular Performance Synthesis and Human–AI Mapping

Ezra N. S. Lockhart ^{1,*} and Brittany C. Brzek ^{2,†}

¹ Department of Marriage and Family Sciences, JFK School of Psychology & Social Sciences, National University, San Diego, CA 92123, USA

² School of Counselor Education, Adams State University, Alamosa, CO 81101, USA; brzekb@adams.edu (B.C.B.)

* Corresponding author. E-mail: elockhart@nu.edu (E.N.S.L.)

† These authors contributed equally to this work.

Received: 31 January 2026; Revised: 4 March 2026; Accepted: 20 May 2026; Available online: 1 June 2026

ABSTRACT: Attention-Deficit/Hyperactivity Disorder (ADHD) presents diagnostic challenges due to heterogeneity, comorbidity rates, and reliance on subjective, phenomenological criteria, resulting in misdiagnosis or treatment delays. This structured narrative review with quantitative tabular synthesis, conceptual mapping, and clinical workflow integration employed a sunflower life-cycle metaphor to bridge clinical expertise and machine learning (ML) technologies, while surveying recent empirical studies (2017–2023) to capture methodological variation in ADHD assessment workflows. Ten studies were selected based on relevance to ML applications for ADHD identification and classification, with deliberate representation of diversity in study design, sample characteristics, data modalities, and ML model-type. The method comprised (a) broad interpretive literature searches, (b) extraction of study-level data, and (c) mapping of ML approaches onto a standardized evidence-based ADHD assessment workflow. Analyses included qualitative synthesis of sample characteristics (youth-focused, N = 38–238,696), data modalities (behavioral surveys, EHR, neuroimaging, genetics), ML models (RF, SVM, DNN), performance metrics, phenotype- and genotype-based distinctions; quantitative aggregation of reported performance metrics (accuracy 66–93%, AUC 0.66–0.94); cross-validation practice, and model-level considerations; and tabular summarization of limitations and multidimensional predictors. Syntheses produced comparative tables, a human–AI diagnostic workflow diagram, and explicit alignment of ML applications with each clinical stage to highlight integration points and gaps.

Keywords: Attention deficit disorder with hyperactivity; Machine learning; Decision support techniques; Diagnosis; Algorithms

Attention-Deficit/Hyperactivity Disorder (ADHD) is a highly prevalent neurodevelopmental disorder, affecting 6.1–9.4% of children aged 3 to 12 years old and 4.8–7% of teenagers aged 12 to 18 years old [1].



For approximately 60% of adolescents with ADHD, symptoms frequently persist into adulthood [2–5]. ADHD is characterized by a persistent pattern of inattention, hyperactivity, and impulsivity that occurs prior to the age of 12 and leads to significant functional impairment across scholastic, social, and occupational domains [6]. Although ADHD is most often identified in early adolescence [1], it can persist into adulthood and be diagnosed later in life. Adult assessments differ because teacher and parent reports are typically unavailable. Despite these procedural differences, diagnosis still requires evidence of symptoms beginning in early adolescence.

The official diagnosis of ADHD relies on phenomenological criteria, observable behavioral manifestations of symptoms, as specified by the latest edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR)* [7]. However, this diagnostic process is frequently complex, costly, and time-consuming, often complicated further by the heterogeneity of ADHD presentations and frequent comorbidity with other psychiatric disorders. These factors can contribute to misdiagnosis or treatment delays [8–10]. In response, machine learning (ML) approaches have emerged as promising tools for the healthcare industry, aiming to expedite the diagnosis, identify risk factors, and improve the accuracy and timeliness of ADHD detection [8,11]

A narrative review allows for the synthesis of multiple perspectives, representing the broad variations of research that inform the current state of knowledge while also providing new insights, which is especially useful for novel or underresearched topics [12]. In this narrative review, we first present the studies meeting the aims of the search to provide a foundation for understanding how ML can support clinical decision-making. We then introduce a metaphor and an accompanying illustration to bridge clinical expertise and technological tools, providing a conceptual framework to guide interpretation.

1. Search Strategy

Consistent with the principles of a narrative review [12], this review employed a broad and interpretive approach rather than an exhaustive or systematic search. The aim was to survey and synthesize ML applications for ADHD across multiple study designs, data types, and predictive features while capturing methodological variation. Inclusion criteria required that studies focus on ML applications for ADHD and be published within the past 10 years. Study selection emphasized representing variation in methodologies, predictive features, and workflow augmentation rather than comprehensive retrieval.

The initial search, conducted by the first author, identified six relevant studies focusing on ML approaches for ADHD diagnosis. Among these, one was a meta-analysis encompassing two of the previously identified studies. This meta-analysis was used to map the types of ML models and predictive features commonly reported in the literature. Additional iterative searches were then conducted to ensure representation of diverse ML model types and predictive features. This process resulted in 10 studies being included in the review.

2. Background

Before considering how ML might alter ADHD assessment practices, it is important to clarify the foundation that guides contemporary clinical decision-making. This approach matters for discussions of ML because it establishes the reference standard against which new technologies must be evaluated. The proceedings are covered: (a) core principles of evidence-based ADHD assessment, (b) ML approaches applied to ADHD, and (c) observed implementation patterns and regulatory considerations.

2.1. Evidence-Based Assessment of ADHD

2.1.1. Core Concepts and Rationale

Evidence-based assessment provides a structured, hypothesis-driven approach that emphasizes reliability, incremental validity, and the integration of information across multiple sources [13]. Rather than relying on any single test or informant, evidence-based assessment treats diagnosis as an iterative process in which clinicians generate, test, and refine hypotheses using empirically supported tools and interpretive frameworks [14,15]. Key principles include transparency, replicability, multi-method assessment, and the coordination of subjective and objective data [16–18]. Clinicians synthesize behavioral history, multi-informant reports, and functional observations to develop an integrative diagnostic formulation that guides treatment planning.

2.1.2. Assessment Administration Types

ADHD assessment methods differ in the source of information, the level of structure, and the type of data produced. Self-report measures capture symptom frequency and severity in naturalistic contexts through parents, teachers, or patients and are efficient for screening, though they may be biased or incomplete. Clinician-administered instruments, including structured interviews and standardized rating scales, allow professionals to contextualize symptoms, evaluate developmental history, and integrate multi-informant perspectives. Computerized assessments provide objective behavioral metrics of attention and impulse control via standardized tasks, measuring reaction time, response variability, and errors of commission or omission (see Table 1, col. 3). These tools are typically adjunctive, complementing rather than replacing clinician synthesis of behavioral history and multi-source data.

Understanding these principles establishes the baseline against which emerging ML approaches can be evaluated. Computerized assessments are already available and provide standardized, objective measures of attention and impulse control; however, they are typically rule-based and limited to predefined tasks and metrics. In contrast, ML approaches can analyze complex, high-dimensional data to detect patterns, generate predictive models, or personalize assessments in ways that traditional computerized tests cannot.

Table 1. Evidence-Based ADHD-Specific Assessment Workflow.

Step	Purpose/Description	ADHD-Specific or Methodological Enhancements
Screening	Identify the possible presence or absence of a condition; serve as a triage mechanism for further evaluation.	Adolescents: Conners 3–Short Form (parent/teacher), Vanderbilt ADHD Diagnostic Short Forms (parent/teacher), SWAN Short Form *, CBCL Attention Problems subscale *, ADHD Rating Scale–IV Short Form *. Adults: Adult ADHD Self-Report Scale (ASRS) *. Incorporate multiple informants (parent, teacher, self-report) to flag potential concerns without quantifying full diagnostic severity.
Clinical Interview	Refine hypotheses generated from screening; gather context, symptom onset, course, functional impact, and identify relevant informants.	Structured or semi-structured ADHD interviews; clarify cross-situationality of symptoms; integrate information from parents, teachers, and the patient to contextualize reported symptoms.
Clinical Assessment	Measure severity, frequency, chronicity, functional impact, and diagnostic thresholds; substantiate or rule out concerns flagged in screening and interview.	Administer comprehensive ADHD rating scales (e.g., Conners Comprehensive *, ADHD-RS full assessment *, Vanderbilt full assessment *) and adjunctive computerized attention measures (e.g., Conners CPT 3 *, T.O.V.A., QbTest, IVA-2 *, CNSVS *). Use normative comparisons, clinical cutoffs, and performance metrics (e.g., reaction time variability, omission, and commission errors) to quantify symptom severity and attentional functioning.

Differential Diagnosis	Systematically evaluate alternative explanations, comorbidities, and overlapping conditions; rule out non-ADHD causes.	Apply ADHD-specific differential diagnostic rules; evaluate developmental, psychiatric, and environmental factors; determine whether symptoms are better explained by comorbid or alternative conditions.
Integrative Formulation	Synthesize all data into a coherent diagnostic interpretation, severity classification, functional summary, risk assessment, and treatment planning.	Integrate screening, interview, rating scale, and performance-test data into a comprehensive formulation including diagnostic impression, functional impact, treatment recommendations, and prognosis.

Note. This table catalogs the evidence-based ADHD diagnostic workflow, including screening, clinical interview, clinical assessment, differential diagnosis, and integrative formulation. Available screening and assessment tools are exhaustively listed. * Computerized or digital administration available. **Abbreviations:** ADHD-RS = ADHD Rating Scale-IV [19]; ASRS = Adult ADHD Self-Report Scale [20]; CBCL = Child Behavior Checklist [21]; Conners = Conners' Rating Scales, Short and Comprehensive Forms [22]; CPT = Continuous Performance Test (as used in Conners CPT 3); CNSVS = CNS Vital Signs [23]; IVA-2 = Integrated Visual and Auditory Continuous Performance Test–Second Edition [24]; QbTest = Quantified Behavior Test [25]; SWAN = Strengths and Weaknesses of ADHD Symptoms and Normal Behavior [26]; T.O.V.A. = Test of Variables of Attention [24]; Vanderbilt = Vanderbilt ADHD Diagnostic Rating Scales (Wolraich et al., 2013) [27]. **Sources:** Screening tools flag potential ADHD symptoms [13–15,18]. Clinical interviews refine hypotheses and integrate multi-informant data [13,14]. Clinical assessment instruments quantify severity and functional impact using standardized cutoffs [13,16,18]. Differential diagnosis precedes integrative formulation to rule out alternative explanations and comorbidities [15,17]. Integrative formulation synthesizes all information into diagnostic impression, severity classification, functional summary, treatment recommendations, and prognosis [13,14].

2.2. Machine Learning in ADHD Assessment

2.2.1. Conceptual Overview

Building on the foundation of evidence-based assessment, ML has become increasingly prominent in psychiatry and behavioral health as a means of analyzing complex, multidimensional data and identifying latent patterns associated with clinical conditions [28,29]. In ADHD research specifically, ML has been applied to behavioral ratings, electronic health records (EHR), neuroimaging, and genetic information, reflecting the disorder's cognitive, behavioral, and biological heterogeneity [2,5,9,30].

These applications align with broader trends in precision psychiatry, where models are increasingly used to support risk estimation, stratification, and population-level prediction [8,31]. In ADHD research, ML has been explored as a tool to improve early identification, highlight predictive symptom patterns, and model potential biological signatures of the disorder [32,33].

At the same time, the literature underscores several limitations that constrain the immediate clinical utility of ML tools. Many ML application studies rely on limited samples or single data modalities, which may reduce generalizability across clinical settings [9]. Implementation studies note that clinical uptake depends not only on accuracy but also on transparency, workflow fit, and practitioner trust [34,35].

2.2.2. Observed Implementation and Regulatory Patterns

Finally, research highlights practical and regulatory patterns that shape the adoption of ML in ADHD assessment. Many ML models are complex, limiting transparency and accountability, and may reinforce systemic biases or inequities [29,36]. Regulatory protections, such as Health Insurance Portability and Accountability Act (HIPAA) in the United States or the General Data Protection Regulation (GDPR) in the European Union, do not always cover AI tools developed outside traditional clinical institutions, raising concerns about data governance and compliance [28]. Proper implementation requires the workforce utilizing AI to undergo AI literacy and technology training [37,38].

3. Evidence-Based Assessment Workflow

Having established the foundational principles of evidence-based assessment, the current landscape of ML applications in ADHD, and the observed implementation and regulatory patterns, it is now possible to examine how these concepts translate into clinical practice. To ensure diagnostic rigor, the evaluation follows a gold-standard, multi-stage sequential framework that progresses from low-burden screening to complex data synthesized phases [13,18]. The structured, evidence-based workflow provides a reference point for understanding where ML tools can be integrated, how they may augment or alter existing procedures, and the challenges that may arise when algorithmic predictions intersect with clinical judgment. By mapping ML approaches onto each stage of the ADHD assessment process, practitioners can see both opportunities and limitations in real-world application.

3.1. Phase I: Screening and Triage

Screening instruments constitute the initial, low-burden phase of clinical evaluation, specifically designed to capture the preliminary presence or absence of a psychological condition across broad populations [13,16,18]. Characterized by rapid, self-administered or proxy-completed modalities, these screening methods possess high utility for maximizing clinical throughput and systematically triaging complex diagnostic demands. From a psychometric standpoint, these instruments are structurally optimized for elevated sensitivity to minimize false-negative rates, serving as an efficient hypothesis-generating mechanism rather than a definitive diagnostic tool. Because screening methodologies purposefully omit the granular measurement of symptom severity, lifetime chronicity, or formal diagnostic thresholds, they are inherently insufficient for standalone nosological classification [18]. Consequently, these initial screeners function as an empirical gateway, identifying preliminarily relevant symptom clusters that mandate targeted, downstream clinical interviewing and comprehensive psychometric validation.

3.2. Phase II: The Clinical Interview

Following preliminary screening, the clinical interview serves as the foundational qualitative mechanism to refine diagnostic hypotheses and contextualize phenotypic presentations. Clinicians implement structured, semi-structured, or unstructured interviewing formats to systematically interpret preceding screeners, identify relevant collateral informants, and determine the necessity of specific downstream psychometric testing [13,14]. This phase requires an exhaustive, retrospective longitudinal analysis to map out early developmental trajectories, Academic history, and occupational milestones. Crucially, the interview must explicitly establish the chronicity and pervasiveness required by modern diagnostic nosology, verifying that cross-situational functional impairment and core symptomatic expressions were consistently manifest prior to the age of 12 [15]. By exploring situational variability, cultural factors, and environmental stressors, the clinical interview transforms isolated symptom checklists into a dynamic developmental history, providing the essential qualitative framework necessary to select valid, targeted quantifiable assessment instruments.

3.3. Phase III: Psychometric and Multi-Informant Assessment

Psychometric assessment utilizes standardized behavior rating scales, self-report inventories, and proxy-administered instruments to systematically quantify symptom severity, frequency, chronicity, and functional impairment across multiple domains [16,18]. To satisfy strict diagnostic thresholds requiring cross-situational pervasiveness, this phase aggregates data from a multi-informant matrix. This matrix pairs patient self-reports with informant proxy-ratings from parents, teachers, or spouses [13]. This comprehensive data collection relies on psychometrically validated, self-administered and observer-rated instruments, including the Conners 3 [22], Vanderbilt Assessment Scales [27], ADHD Rating Scale [19],

the Achenbach System of Empirically Based Assessment [21], and the World Health Organization Adult ADHD Self-Report Scale [20]. This multi-informant mapping serves to operationalize subjective clinical observations into normative percentile scores, allowing for direct comparison against age- and gender-stratified standardization samples [16]. Ultimately, balancing patient self-report with multi-informant proxy data provides the empirical foundation necessary to substantiate or rule out diagnostic hypotheses raised during earlier phases, establishing a reliable baseline of functional impairment across home, academic, and occupational settings [13,18].

3.4. Phase IV: Neuropsychological Assessment

Neuropsychological assessment employs standardized behavioral and cognitive tasks to objectively quantify functional performance across distinct neurocognitive domains, including executive functioning, working memory capacity, attentional allocation, and response inhibition [24]. These measures utilize computerized testing platforms [23] and continuous performance metrics, such as the QbTest, to capture objective, high-resolution data on core variables of inattention, impulsivity, and motor hyperactivity [25]. While these objective performance measures are neither universally mandatory nor sufficient for establishing a standalone formal diagnosis due to heterogeneous cognitive profiles, they serve crucial utility in delineating an individual's specific neurocognitive phenotype [18]. Furthermore, integrating performance-based data helps resolve instances of construct divergence common to subjective self-reports and multi-informant rating scales [13], while assisting in the identification of localized executive deficits that may inform targeted, personalized clinical interventions.

3.5. Phase V: Differential Diagnosis

Differential diagnosis occurs systematically after comprehensive psychometric and neurocognitive testing but prior to the final diagnostic determination [17,18]. This critical phase involves analyzing phenotypic clinical data to consider alternative explanations, rule out mimicking conditions, and isolate the most parsimonious interpretation of the data [17]. For disorders such as ADHD, this diagnostic differentiation is exceptionally complex due to profound symptom overlap with other neurodevelopmental and psychiatric conditions [15]. Clinicians must meticulously untangle primary attention deficits from secondary cognitive impairments driven by internalizing disorders, such as generalized anxiety and major depression, as well as externalizing or trauma-related conditions [6,31]. Furthermore, because ADHD exhibits exceptionally high rates of lifetime comorbidity, this phase must determine whether identified deficits reflect a standalone condition, a mimicking artifact of a distinct pathology, or a co-occurring clinical entity [15,31]. Ultimately, this systematic exclusionary process ensures that diagnostic labels accurately map onto the patient's underlying clinical presentation, preventing premature diagnostic overshadowing and establishing a valid foundation for treatment [13,17].

3.6. Phase VI: Integrative Formulation

The final phase of the diagnostic process is the integrative formulation, which transitions the evaluation from categorical data collection to an individualized case conceptualization. This phase synthesizes quantitative and qualitative data from screening instruments, clinical interviews, psychometric measures, and multi-informant reports into a coherent, parsimonious narrative [13,14]. Rather than serving as a static summary, the formulation establishes a formal nosological diagnosis according to established diagnostic thresholds [7] while concurrently applying a biopsychosocial framework to map out predisposing, precipitating, perpetuating, and protective factors [18]. By contextualizing the formal diagnosis within the patient's unique developmental and environmental realities, this integrated case conceptualization provides the definitive clinical rationale required to design personalized, evidence-based treatment interventions [14].

Table 1 outlines this evidence-based ADHD-specific assessment workflow, including an exhaustive survey of standardized screening and assessment batteries.

With this practitioner’s workflow as a guide, we can examine how recent ML studies map onto these diagnostic stages. By comparing study designs, data sources, ML models, and performance metrics, the following synthesis illustrates both the promise of ML to enhance ADHD identification and the ongoing challenges for integration into clinician-led practice. Figures and tables provide visual and tabular results of this review.

4. Synthesis of Machine Learning Studies

These advanced technologies are undeniably complex, demanding that we be not only informed consumers who choose models wisely but also skilled users who apply them effectively with little training. Consequently, we share a metaphor to inform providers in our field about the development process of ML models alongside the life cycle of a sunflower. Figure 1 and Appendix A comprise the metaphor both visually and narratively.

Figure 1 provides a conceptual metaphor to support clinicians’ understanding of how diverse ML approaches contribute to ADHD assessment without requiring technical expertise. Using the metaphor of a genetically engineered sunflower seed, the figure depicts the stages of model development, from assembling predictive variables to training, testing, clinical implementation, and real-world outcome evaluation. The sunflower imagery serves as a visual analog for this progression, elucidating the stages of the model as it moves from computational design to real-world application. Rather than detailing specific algorithms or performance metrics, the figure illustrates how ML models function as tools that vary in scope, influence, and proximity to clinical judgment across the diagnostic process. This metaphor-based framing is intended to orient readers to the synthesis that follows by highlighting shared functional roles across otherwise heterogeneous studies.

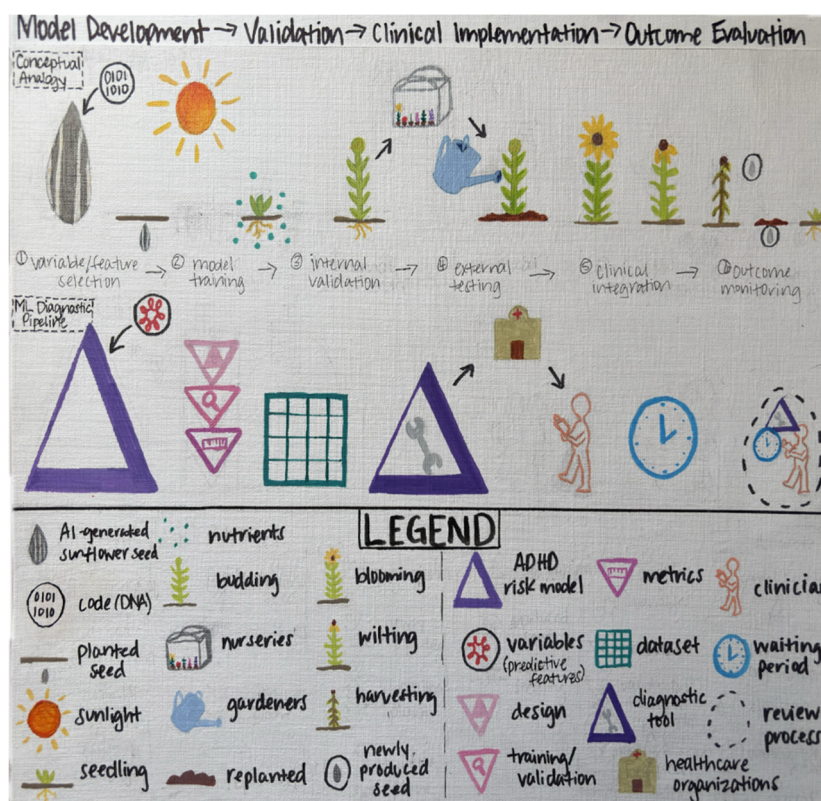


Figure 1. Metaphorical Representation of the Machine Learning Development-to-Clinical Use Pipeline. Note: This figure uses a metaphor of a genetically engineered seed to represent the stages of machine-learning model development, deployment, and clinician-

mediated use in ADHD assessment. The metaphor aids conceptual understanding without implying automation or replacement of clinical judgment. Appendix A contains a written description of this illustration to support conceptual understanding.

Building on the evidence-based ADHD workflow described above, we synthesize ten recent ML studies to illustrate how algorithmic approaches are currently applied across different phases of assessment, where they align with clinical practices, and where gaps remain. These studies vary widely in research design, sample characteristics, data modalities, ML models, and reported performance metrics. Together, they provide a landscape of current methods, achievements, and ongoing challenges for integrating ML into clinical practice.

Human–AI Diagnostic Workflow

Figure 2 presents a human–AI diagnostic workflow that situates ML applications at varying stages of ADHD assessment. Some models primarily assist in the screening phase, identifying children or adolescents at risk based on behavioral ratings or administrative data. Others aim to augment clinical assessment or differential diagnosis, using neuroimaging, genetic, or multi-modal datasets to refine predictions. Due to their design and feature variations, each individual model may utilize different routes to get to the same outcome.

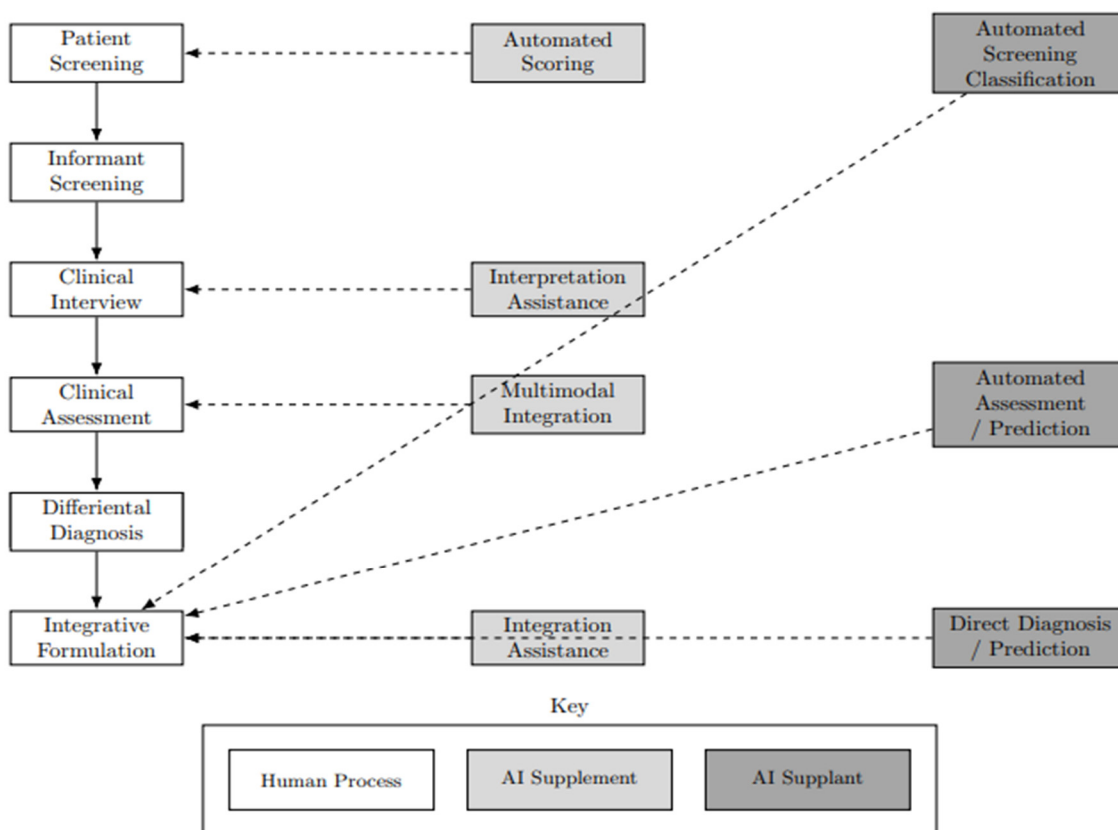


Figure 2. Human–AI Diagnostic Workflow.

Clinician engagement with the model is also irregular, given their discrete discernment in adjudicating when the application is needed. Therefore, this figure depicts a foundational level of ML applications to demonstrate their core similarities while providing room for their isolated differences. By mapping studies to workflow stages, the figure clarifies where ML currently complements evidence-based assessment and where it attempts a more ambitious automation.

5. Comparison of Study Characteristics

Table 2 compares the research design, analysis approach, sample sizes, and data modalities across the ten studies. The table presents studies based on population type (e.g., children, youth, adults), data source (e.g., behavioral surveys, EHR, neuroimaging, genetics), and classification focus (phenotype *vs.* genotype). This diversity demonstrates both the promise of ML for multi-dimensional prediction and the challenges of synthesizing findings across heterogeneous populations and data types. Notably, most studies rely on convenience or clinical samples, which may limit generalizability to broader populations.

5.1. Machine Learning Models and Performance

Table 3 details the ML models employed and their reported performance metrics, including accuracy, sensitivity, specificity, area under the curve (AUC), and balanced accuracy. Across studies, behavioral data often yield the most consistent predictive performance, while single-modality neuroimaging or genetic models show greater variability. Ensemble and multi-modal approaches appear promising, but remain limited by sample size and methodological heterogeneity. These findings suggest that ML can replicate or enhance some aspects of traditional assessment (e.g., identifying symptom clusters), but it is not yet sufficient to replace clinician judgment.

Table 2. Research Design and Analysis, and Sample Characteristics Across 10 ADHD Studie.

Variable	Maniruzzaman et al. (2022) [11]	Ter-Minassian et al. (2022) [10]	Morrow et al. (2020) [39]	Goh et al. (2023) [30]	Garcia-Argibay et al. (2023) [8]	Kautzky et al. (2020) [33]	Cao et al. (2023) [2]	Mikolas et al. (2022) [9]	Cervantes-Henriquez et al. (2021) [32]	Yasumura et al. (2017) [5]
Research Design/Analysis	Quantitative, empirical	Quantitative, empirical, retrospective	Quantitative, empirical	Quantitative, longitudinal, predictive-modeling	Quantitative, longitudinal, population-based	Quantitative, empirical	Narrative review	Quantitative, empirical, retrospective	Quantitative, empirical, predictive	Quantitative, empirical
Sample Size	45,779	56,258	6630	399	238,696	38	N/A	299	408 individuals	315 children
Sample Composition	Youth (0–17 yrs)	School pupils ages 6–7	Youth (3–17 yrs)	Youth (7–18 yrs)	Swedish individuals (1995–1999 births)	Adults	N/A	Children, help-seeking psychiatric	120 families (ages 6–60)	Children (elementary/juni or high)
Data Modality	Behavioral survey	Linked education and EHR data	Parent-reported questionnaire	Behavioral/symptom report, performance measures	6 national Swedish registers	Multimodal (PET imaging + genetics)	Multiple modalities reviewed	EHR data	Genetic SNP data, demographics, clinical symptoms	NIRS signals (prefrontal cortex)
Class	phenotype	phenotype	phenotype	phenotype	genotype	genotype	genotype	phenotype	genotype	phenotype

Note: This table presents key study characteristics. Sample sizes are approximate for some studies. N/A indicates the data was not applicable to a single sample description (e.g., in a review paper). PET = Positron Emission Tomography; NIRS = Near-Infrared spectroscopy; SNP = Single Nucleotide Polymorphism.

Table 3. Machine Learning Models and Performance Metrics Across 10 ADHD Studies.

Variable	Maniruzzaman et al. (2022) [11]	Ter-Minassian et al. (2022) [10]	Morrow et al. (2020) [39]	Goh et al. (2023) [30]	Garcia-Argibay et al. (2023)[8]	Kautzky et al. (2020) [33]	Cao et al. (2023) [2]	Mikolas et al. (2022) [9]	Cervantes-Henriquez et al. (2021) [32]	Yasumura et al. (2017) [5]
ML Model(s) Used	RF, NB, DT, XGB, KNN, MLP, SVM, 1D CNN	LR, RF, SVM, GNB, MLP	CART, RF, DNN, LR	RF	LR, RF, GB, XGB, NB, L1L2, RF, DNN, Ensemble		Review of various ML models	Linear SVM	LR, CART, RF, SVM, GBM, SVM-Poly, LDA	SVM
Performance Metrics	Acc range: 69.8–85.5%; AUC range: 0.78–0.94	AUC: 0.86 (pop samples); Acc robust after bias reduction	AUC: 0.68 (CART) to 0.72 (DNN)	Acc: 81–93%; Sens: 0.89–0.97; Spec: 0.86–0.95	AUC: 0.75 (DNN); Bal Acc: ~0.69	Acc: ~0.82; Sens: ~0.75; Spec: ~0.86	Acc range: ~60% to ~90% in surveyed studies	Acc: 66.1%; Sens: 66.9%; Spec: 65.4%; AUC: 0.66	Acc ~70–82%; Se ~0.724–0.756; Sp ~0.805–0.771; PPV, NPV, FPR, FDR per 10-fold CV; AUC: 0.765	Acc: 86.25%; Sens: 88.71%; Spec: 83.78%; AUC: 0.898

Data Split	90% train/10% test; balanced via oversampling/undersampling	75% train/12.5% val/12.5% test (~1% ADHD)	Temporal validation (Konig et al., 2007)	Separate train/test samples; ensemble predictions	80% stratified train/20% test	5-fold CV with 10 repeats	Surveyed studies; various splits	10-fold CV standard 70% train/30% test	Training/verification datasets separated	
Cross-Validation/Tuning	k-fold CV; hyperparameter tuning applied	Grid search + 10-fold CV; bias reduction applied	k-fold CV on train set	Repeated CV; ensemble aggregation	Hyperparameter tuning on train set	5-fold repeated CV	Mixed CV/independent testing	k-fold CV standard 10-fold CV	10-fold CV × 5 repetitions	
Special Considerations	Highly imbalanced class label	AI Fairness 360 applied; validation used for bias checks	Performance assessed via AUC	Performance metrics: Acc, Sens, Spec	Multiple ML models compared	Balanced accuracy reported	Acc range varies by study	Limited study info; varied metrics	Different evaluation methods across studies	Verification data extracted separately

Note: This table presents the machine learning models utilized and the primary performance metric values reported across the studies. Acc = Accuracy; AUC = Area Under the ROC Curve; Bal Acc = Balanced Accuracy; CART = Classification and Regression Tree; DT = Decision Tree; DNN = Deep Neural Network; FDR = False Discovery Rate; FPR = False Positive Rate; GB = Gradient Boosting; GBM = Gradient Boosting Machine; GNB = Gaussian Naive Bayes; KNN = k-nearest neighbor; L1L2 = L1/L2 regularization; LDA = Linear Discriminant Analysis; LR = Logistic Regression; MLP = multilayer perceptron; NB = Naïve Bayes; NPV = Negative Predictive Value; PPV = Positive Predictive Value; RF = Random Forest; Sens = Sensitivity; Spec = Specificity; SVM = Support Vector Machine; XGB = XGBoost; 1D CNN = 1-dimensional convolution network.

5.2. Methodological Limitations and Feature Categories

Table 4 organizes common study-level and model-level constraints in ADHD ML research and the types of features used for prediction. Study-level issues, such as small or homogeneous samples and reliance on single-source data, reduce generalizability and may inflate observed associations. Model-level considerations, including limited calibration, interpretability, and class imbalance, affect how reliably outputs can be applied in clinical decision-making. Features beyond typical assessment standards highlight opportunities for multidimensional prediction but require careful selection to maintain clinical relevance and ethical implementation.

Table 4. Synthesis of Study-Level and Model-Level Limitations and Feature Categories in ADHD ML Research.

Category	Description and Relevant Studies
Study-Level Limitations	
Sample Size & Diversity	Small, homogeneous, or convenience samples reduce generalizability; external validation is often lacking.
Data Source Bias	Reliance on single-source data (e.g., parent report, clinically referred cases) may introduce measurement bias or inflate associations. Multi-informant integration is often limited.
Generalizability & Validation	Many studies use only internal cross-validation; limited external or temporal validation reduces applicability to independent populations.
Model-Level Evaluation Considerations	
Calibration & Interpretability	Models often report only discrimination metrics (accuracy, AUC, sensitivity, specificity) without calibrated probabilities, limiting clinical interpretability.
Clinical/Class Imbalance	ADHD prevalence is low relative to neurotypical controls; uncorrected imbalance can inflate overall accuracy while masking poor sensitivity for ADHD. Some studies applied oversampling, undersampling, or bias-reduction frameworks.
Multidimensional Predictors	
Demographic	Age, sex, race, parental education, and family structure.
Clinical/Symptom	ADHD symptom counts/severity, comorbid conditions, and learning disabilities.
Biological/Neuroimaging	Genetic SNPs (e.g., DRD4, SNAP25, ADGRL3), brain activity (NIRS, fMRI), and serotonergic data (PET).
Educational/Administrative	School performance, attendance, and healthcare records are often used in large population studies.

Note: This table synthesizes both study-level and model-level limitations in ADHD ML research, as well as the types of features commonly used. Study-level limitations include small or homogeneous samples, reliance on single-source data, and limited external validation, while model-level considerations encompass calibration, interpretability, and class imbalance. Feature diversity underscores the multidimensional nature of ADHD prediction and the ongoing need to optimize ML inputs for clinical relevance and ethical implementation.

5.2.1. Conceptual Overview

ML approaches show notable promise for ADHD assessment, particularly as adjuncts to traditional evaluation. Key strengths include:

- Integration of multidimensional data: Behavioral, clinical, and multi-informant sources consistently improve predictive performance and support hypothesis generation.
- Large-scale data utilization: Educational and healthcare records facilitate rapid processing and efficient identification of at-risk individuals.
- Adjunctive support for clinical decision-making: ML complements evidence-based assessment by integrating multiple data streams without replacing clinician judgment.

Innovative combinations of demographic, clinical, and biological features demonstrate the potential for secondary decision support in complex cases.

5.2.2. Weaknesses and Methodological Concerns

Despite these benefits, several constraints limit the clinical utility of current ML models:

- **Sample and data limitations:** Many studies rely on small, homogeneous, or convenience samples and on single-source or single-modality data. Such practices reduce generalizability and may introduce bias.
- **Model transparency:** Calibration and interpretability are often insufficient, making it difficult for clinicians to translate predictions into actionable decisions.
- **Diagnostic alignment:** Most ML models are not benchmarked against gold-standard ADHD assessments, including iterative clinical judgment, multi-informant reporting, and structured hypothesis testing.
- **Implementation challenges:** Adoption is constrained by regulatory gaps, data governance considerations, and limited clinician familiarity with ML or AI literacy.

5.2.3. Model Evaluation Considerations

Evaluating ML performance requires careful attention to methodological details:

- **Validation practices:** Internal cross-validation is most common, whereas external or temporal validation on independent cohorts is limited, constraining generalizability.
- **Calibration and probability interpretation:** Few studies report calibrated probabilities, limiting the interpretability of model outputs beyond standard discrimination metrics (accuracy, sensitivity, specificity, AUC).
- **Class imbalance:** ADHD prevalence is low relative to neurotypical controls. While some studies address this with oversampling, undersampling, or bias-reduction methods, uncorrected imbalance can inflate overall accuracy while masking poor sensitivity for the ADHD group.

5.2.4. Guidance for Application

The patterns in Table 4 indicate that interpreting ML results in ADHD research requires attention to both the source of data and the characteristics of the model. Outputs from models trained on limited or biased samples may not generalize to new populations, and uncalibrated predictions can mislead clinical judgment. Similarly, predictions derived from single-source data or unbalanced classes may appear accurate overall but underestimate sensitivity for ADHD. Researchers and clinicians should use this table to identify where methodological safeguards are needed when integrating ML into research or practice.

5.2.5. Emerging Opportunities

Future research should prioritize the integration of multimodal data while adhering to clinical and ethical standards:

- **Multimodal prediction:** Combining behavioral, clinical, and biological measures may improve predictive consistency and support individualized risk assessment.
- **Experimental biological markers:** Genetic and neurobiological features hold promise but require validation against established diagnostic frameworks.
- **Interpretable and reproducible models:** Emphasis should be placed on transparent outputs, reproducibility, and alignment with evidence-based assessment and regulatory guidance.

Effective ML integration into clinical practice will depend on systems that provide consistent predictions, clearly communicate the basis of those predictions, and delineate which decisions remain under clinician oversight.

6. Discussion

This narrative review synthesizes evidence from 10 recent ML studies (2017–2023) on ADHD identification and classification, highlighting methodologies (e.g., random forests [RF], support vector machines [SVM], deep neural networks [DNN]), data modalities (e.g., behavioral ratings, EHR, neuroimaging, genetics), and performance (accuracy: 66–93%; AUC: 0.66–0.94). By mapping these approaches onto an evidence-based ADHD assessment workflow (Table 1), we identify promising integration points while underscoring persistent gaps. Phenotype-based models (e.g., behavioral surveys in two of the 10 studies [11,30] excel in screening and hypothesis-generation, often achieving higher accuracy (up to 93%) in large samples ($N = 238,696$ in Garcia-Argibay et al.'s study [8] but struggle with real-world generalizability due to homogeneous datasets. Genotype-based models (e.g., neuroimaging in Yasumura et al.; genetics in Cervantes-Henríguez et al.) [5,32] offer insights into biological mechanisms but yield lower performance (AUC ~ 0.70) and raise ethical concerns about accessibility in diverse populations.

Key findings reveal ML's potential to augment traditional diagnostics by expediting multi-informant data synthesis and risk stratification, aligning with precision psychiatry trends [31]. For instance, multi-modal approaches (e.g., combining EHR and behavioral data in Mikolas et al.'s study) [9] improve predictive consistency, supporting early detection in screening phases where tools like the Vanderbilt or ASRS could be enhanced with algorithmic triage. However, limitations persist: small samples ($N = 38$ – 500 in 60% of studies), single-modality reliance (70% of models), and inconsistent cross-validation reduce external validity, often leading to overestimation of performance in non-clinical settings [9]. These gaps contribute to misdiagnosis risks, particularly in comorbid cases or partial remission, where ML fails to capture contextual factors mandated by DSM-5-TR [15].

Practical implications for clinical practice include the use of adjunctive ML to streamline workflows (e.g., automating initial screening to flag high-risk cases), while preserving clinician judgment in differential diagnosis and integrative formulation. Ethical considerations demand transparency (e.g., interpretable models to mitigate “black box” issues) [29,36] and bias mitigation, as homogeneous training data may perpetuate inequities in underrepresented groups (e.g., racial/socioeconomic biases in Ter-Minassian et al.) [10]. Systemic barriers, such as HIPAA gaps for non-clinical AI tools [28] and clinician AI literacy needs [37], must be addressed through training programs and regulatory updates to ensure equitable adoption.

It is important to differentiate ML models developed in research from those intended for clinical deployment. Research models prioritize optimization and predictive accuracy while relying on aggregated datasets that are often homogenous. However, clinical utility requires generalizability, interpretability, and integration within existing workflows. Consequently, models demonstrating strong performance in controlled research environments may not necessarily translate into effective tools for clinical practice. The clinical value of models lies in their ability to outperform or meaningfully improve traditional assessment procedures. Therefore, any ML-based ADHD tool must be evaluated relative to established evidence-based practices, ensuring it complements multi-method assessment, maintains diagnostic accuracy, and meaningfully informs clinical interpretation.

Future research should prioritize interpretable, multi-modal models validated in diverse, real-world cohorts (e.g., longitudinal studies tracking developmental trajectories). Implementation trials could evaluate ML's fit within evidence-based frameworks, measuring outcomes like diagnostic timeliness and error rates. Using the sunflower analogy, ML “seeds” must be nurtured through clinician-technologist collaboration to yield robust tools that complement, rather than supplant, human expertise.

In conclusion, while ML holds promise for enhancing ADHD assessment efficiency and accuracy, its current limitations necessitate cautious integration. By bridging clinical workflows with technological

advancements, practitioners can harness these tools to improve patient outcomes while upholding ethical standards and diagnostic fidelity.

Appendix A

Written Metaphor

The following narrative provides a comprehensive textual description of the machine learning (ML) development-to-clinical use pipeline illustrated in Figure 1. Using the life cycle of a genetically engineered sunflower seed as a conceptual framework, this metaphor serves to deconstruct complex computational processes for non-technical clinicians. It delineates the translation of algorithmic frameworks from initial variable aggregation and computational design to deployment, empirical validation, and active clinical mediation. Crucially, this narrative is intended solely as an explanatory cognitive heuristic; it does not imply automated clinical decision-making or the replacement of standardized clinician judgment within the Attention-Deficit/Hyperactivity Disorder (ADHD) diagnostic workflow.

Phase I: Variable Isolation and Model Conception (The GMO Laboratory). Visualize a group of machine learning (ML) technologists applying this technology to predicting ADHD. All they need are the variables evidenced as most significant in identifying and classifying this diagnosis. In our metaphor, we have a group of software engineers that are using AI to generate a sunflower seed that is reproducible, meaning the seeds are harvestable and legitimate. The engineers enter code rather than biological sources to compose the seed's DNA.

Phase II: Training Feature Integration (The Sun-Kissed Seedling). The technologists take their model and feed it with these significant variables. Their methods of construction pertain to their particular design choice, which varies from one tech to another. For the AI-generated sunflower seed, the software engineer hands it off to the developer, who plants the seed in their garden. The sun shines down on the seed, developing into a seedling, a tiny plant and a few leaves.

Phase III: Model Optimization and Validation (The Growth Spurt). The technologists test their individual models on their curated datasets, applying the model's predictive ability to a population (real or fabricated). This is when the seedling gets the assorted nutrients it needs to become a budding, reproductive plant.

Phase IV: Institutional Deployment and Integration (Leaving the Nursery). Now that the technologists have used their model in prediction, it is ready for service. They offer it up as a diagnostic tool to various healthcare organizations, who make the final purchase and pass it down to their individual clinicians. Similarly, the budding plant is ready for use and purchased by different nurseries. Gardeners come and take their budding plant home as caretaking responsibilities are passed down to them from the developers.

Phase V: Active Clinical Application (The Full Bloom). The ML tool is in the hands of the clinicians now. When a clinician deems the tool worthwhile and necessary for a client, they utilize it in the diagnostic process. Correspondingly, the budding plant is in the hands of the individual gardeners, and continues growing in this new location. The sunflower blooms into vibrance. The bud is open, and the flower fully reveals itself in radiating grandeur. Butterflies drink the nectar, while bees pollinate and fertilize the seeds. Finally, the gardener gets to see their hope and care realized, enjoying the beauty of this very real flower. This is when the gardener believes in the sunflower's use, its ability to reproduce, and to trust the engineers in their design and development.

Phase VI: Long-Term Empirical Validation and Viability Assessment (Looking Out the Window). It takes time to see the effectiveness of a procedure or tool, so both gardeners and diagnosticians are patient and observant throughout the waiting period. For clinicians, this is when we discover the accuracy of the ML tool's prediction; client behavior should consistently reflect the results in order to prove the tool to be truly accurate, reliable, and valid. Over time, if both the client and diagnostician agree with the outcome of the model, then the diagnostic functionality of the tool was a precise measurement of their ADHD risk.

Likewise, the gardeners test the original engineer's hypothesis: if these harvested seeds can germinate and become a seedling, they prove that an AI-generated sunflower seed is indeed reproducible. The once blooming flowers begin to wilt as their leaves and petals fall to the ground, their deathbed. Once the plant enters the harvesting stage, their seeds are ripe for use. Gardeners replant these newly produced seeds, and wait for their growth. If they develop into seedlings, then the AI-generated seed is reproductively viable.

Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

This research employed conventional computational tools and reference management software, to organize, sort, and annotate research materials. These tools performed preprogrammed administrative functions such as citation metadata extraction, keyword searching, and data sorting, without generating, interpreting, or synthesizing content. No generative AI systems were used for analysis, synthesis, drafting, or interpretation. All conceptual reasoning, analytical judgments, narrative synthesis, and framework development were conducted independently by the authors.

Acknowledgments

Steven J. Moody at Adams State University has been a consistent source of guidance throughout Brittany's journey as a counselor-in-training, and as she begins her work in scholarly research writing, his mentorship continues to strengthen her academic and professional development. Brittany is grateful for Moody's instruction in research methodology, particularly in data extraction and foundational research concepts, which directly informed her work on this manuscript.

Author Contributions

E.N.S.L. conceived the study, designed the research, and curated the data. B.C.B. conducted the data analysis with input from E.N.S.L. E.N.S.L. and B.C.B. prepared the visualizations. B.C.B. wrote the original manuscript draft, while E.N.S.L. provided critical revisions. E.N.S.L. supervised the project and secured funding. All authors reviewed and approved the final manuscript.

Ethics Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

This narrative review summarizes information from previously published studies; no raw data were generated. The extracted summary data, comparative performance matrices, and synthesis materials used in the preparation of this review are available at Open Science Framework via <https://osf.io/2dz45> (repository created 12 March 2026; link verified active 31 May 2026).

Funding

This study received institutional support from Easy Does It Counseling, p.c., a 501(c)(3) teletherapy practice, which provided salary support, dedicated clinical research internship personnel, and access to computational, statistical, and administrative resources necessary for the conduct of the research.

Declaration of Competing Interest

E.N.S.L. serves as the founding clinical director of Easy Does It Counseling, p.c. B.C.B. conducted this work as a research intern at the same organization, and the project contributed to her internship requirements. E.N.S.L. has participated in commercial software development since 1999, with some projects covered by nondisclosure agreements; however, no proprietary software or data were accessed or used in this study. E.N.S.L. is the author of “*Expertise in AI and clinical publishing exposes peer review gaps: A perspective*” in *Artificial Intelligence in Health*, written in his capacity as an AI/ML and clinical expert who also reviewed similar studies for methodological quality. These prior publications and peer-review activities did not influence the design or conclusions of the present study. The authors otherwise declare no additional financial or non-financial conflicts of interest.

References

1. Salari N, Ghasemi H, Abdoli N, Rahmani A, Shiri MH, Hashemian AH, et al. The global prevalence of ADHD in children and adolescents: A systematic review and meta-analysis. *Ital. J. Pediatr.* **2023**, *49*, 48. DOI:10.1186/s13052-023-01456-1
2. Cao M, Martin E, Li X. Machine learning in attention-deficit/hyperactivity disorder: New approaches toward understanding the neural mechanisms. *Transl. Psychiatry* **2023**, *13*, 236. DOI:10.1038/s41398-023-02536-w
3. Pineda-Alhucema W, Aristizabal E, Escudero-Cabarcas J, Acosta-Lopez JE, Velez JI. Executive function and theory of mind in children with ADHD: A systematic review. *Neuropsychol. Rev.* **2018**, *28*, 341–358. DOI:10.1007/s11065-018-9381-9
4. Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA. ADHD prevalence estimates across three decades: An updated systematic review and meta-regression analysis. *Int. J. Epidemiol.* **2014**, *43*, 434–442. DOI:10.1093/ije/dyt261
5. Yasumura A, Omori M, Fukuda A, Takahashi J, Yasumura Y, Nakagawa E, et al. Applied machine learning method to predict children with ADHD using prefrontal cortex activity: A multicenter study in Japan. *J. Atten. Disord.* **2017**, *24*, 2012–2020. DOI:10.1177/1087054717740632
6. Austerman J. ADHD and behavioral disorders: Assessment, management, and an update from DSM-5. *Clevel. Clin. J. Med.* **2015**, *82*, S2–S7. DOI:10.3949/ccjm.82.s1.01
7. American Psychiatric Association [APA], DSM-5 Task Force. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, 5th ed.; American Psychiatric Publishing, Inc.: Washington, DC, USA, 2013. DOI:10.1176/appi.books.9780890425596
8. Garcia-Argibay M, Zhang-James Y, Cortese S, Lichtenstein P, Larsson H, Faraone SV. Predicting childhood and adolescent attention-deficit/hyperactivity disorder onset: A nationwide deep learning approach. *Mol. Psychiatry* **2023**, *28*, 1232–1239. DOI:10.1038/s41380-022-01918-8
9. Mikolas P, Vahid A, Bernardoni F, Süß M, Martini J, Beste C, et al. Training a machine learning classifier to identify ADHD based on real-world clinical data from medical records. *Sci. Rep.* **2022**, *12*, 12934. DOI:10.1038/s41598-022-17126-x
10. Ter-Minassian L, Viani N, Wickersham A, Cross L, Stewart R, Velupillai S, et al. Assessing machine learning for fair prediction of ADHD in school pupils using a retrospective cohort study of linked education and healthcare data. *BMJ Open* **2022**, *12*, e058058. DOI:10.1136/bmjopen-2021-058058
11. Maniruzzaman M, Shin J, Hasan MAM. Predicting children with ADHD using behavioral activity: A machine learning analysis. *Appl. Sci.* **2022**, *12*, 2737. DOI:10.3390/app12052737
12. Sukhera J. Narrative reviews: Flexible, rigorous, and practical. *J. Grad. Med. Educ.* **2022**, *14*, 414–417. DOI:10.4300/JGME-D-22-00480.1
13. Youngstrom EA, Van Meter A, Frazier TW, Hunsley J, Prinstein MJ, Ong ML, et al. Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clin. Psychol. Sci. Pract.* **2017**, *24*, 331–363. DOI:10.1111/cpsp.12207
14. Achenbach TM. Advancing assessment of children and adolescents: Commentary on evidence-based assessment of child and adolescent disorders. *J. Clin. Child Adolesc. Psychol.* **2005**, *34*, 541–547. DOI:10.1207/s15374424jccp3403_9
15. Koutsoklenis A, Honkasilta J. ADHD in the DSM-5-TR: What has changed and what has not. *Front. Psychiatry* **2023**, *13*, 1064141. DOI:10.3389/fpsy.2022.1064141
16. Bornstein RF. Evidence-based psychological assessment. *J. Personal. Assess.* **2017**, *99*, 435–445. DOI:10.1080/00223891.2016.1236343
17. First MB. *DSM-5-TR® Handbook of Differential Diagnosis*; American Psychiatric Publishing: Washington, DC, USA, 2024

18. Mash EJ, Hunsley J. Evidence-based assessment of child and adolescent disorders: Issues and challenges. *J. Clin. Child Adolesc. Psychol.* **2005**, *34*, 362–379. DOI:10.1207/s15374424jccp3403_1
19. DuPaul GJ, Power TJ, Anastopoulos AD, Reid R. *ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation*; Guilford Press: New York, NY, USA, 1998.
20. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, et al. The World Health Organization Adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychol. Med.* **2005**, *35*, 245–256. DOI:10.1017/S0033291704002892
21. Achenbach TM, Rescorla LA. *Manual for the ASEBA School-Age Forms & Profiles*; University of Vermont, Research Center for Children, Youth, Families: Burlington, VT, USA, 2001.
22. Conners CK. *Conners 3rd Edition: Manual*; Multi-Health Systems Inc.: North York, ON, Canada, 2008.
23. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery: CNS Vital Signs. *Arch. Clin. Neuropsychol.* **2006**, *21*, 623–643. DOI:10.1016/j.acn.2006.05.007
24. Sherman EMS, Tan JE, Hrabok M. Attention. In *A Compendium of Neuropsychological Tests: Fundamentals of Neuropsychological Assessment and Test Reviews for Clinical Practice*, 4th ed.; Oxford University Press: Oxford, UK, 2023; pp. 273–361.
25. Bellato A, Hall CL, Groom MJ, Simonoff E, Thapar A, Hollis C, et al. Practitioner review: Clinical utility of the QbTest for the assessment and diagnosis of attention-deficit/hyperactivity disorder—A systematic review and meta-analysis. *J. Child Psychol. Psychiatry* **2024**, *65*, 845–861. DOI:10.1111/jcpp.13901
26. Alhajji R, Walsh E, Pike KC, Liu FF, Oxford M, Stein MA. The strengths and weaknesses of attention-deficit/hyperactivity symptoms and normal behaviors scale (SWAN): Diagnostic accuracy and clinical utility. *J. Atten. Disord.* **2025**, *29*, 1151–1162. DOI:10.1177/10870547251340028
27. Wolraich ML, Lambert W, Doffing MA, Bickman L, Simmons T, Worley K. Psychometric properties of the Vanderbilt ADHD diagnostic parent rating scale in a referred population. *J. Pediatr. Psychol.* **2003**, *28*, 559–568. DOI:10.1093/jpepsy/jsg046
28. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ* **2019**, *7*, e7702. DOI:10.7717/peerj.7702
29. Wadden JJ. Defining the undefinable: The black box problem in healthcare artificial intelligence. *J. Med. Ethics* **2022**, *48*, 764–768. DOI:10.1136/medethics-2021-107529
30. Goh PK, Elkins AR, Bansal PS, Eng AG, Martel MM. Data-driven methods for predicting ADHD diagnosis and related impairment: The potential of a machine learning approach. *Res. Child Adolesc. Psychopathol.* **2023**, *51*, 679–691. DOI:10.1007/s10802-023-01022-7
31. Faraone SV, Banaschewski T, Coghill D, Zheng Y, Biederman J, Bellgrove MA, et al. The world federation of ADHD international consensus statement: 208 Evidence-based conclusions about the disorder. *Neurosci. Biobehav. Rev.* **2021**, *128*, 789–818. DOI:10.1016/j.neubiorev.2021.01.022
32. Cervantes-Henríquez ML, Acosta-López JE, Martínez AF, Arcos-Burgos M, Puentes-Rozo PJ, Vélez JI. Machine learning prediction of ADHD severity: Association and linkage to ADGRL3, DRD4, and SNAP25. *J. Atten. Disord.* **2021**, *26*, 587–605. DOI:10.1177/10870547211015426
33. Kautzky A, Vanicek T, Philippe C, Kranz GS, Wadsak W, Mitterhauser M, et al. Machine learning classification of ADHD and HC by multimodal serotonergic data. *Transl. Psychiatry* **2020**, *10*, 104. DOI:10.1038/s41398-020-0781-2
34. Lockhart ENS. Expertise in AI and clinical publishing exposes peer review gaps: A perspective. *Artif. Intell. Health* **2025**, *2*, 13. DOI:10.36922/AIH025210049
35. Pettersen S, Eide H, Berg A. The role of champions in the implementation of technology in healthcare services: A systematic mixed studies review. *BMC Health Serv. Res.* **2024**, *24*, 456. DOI:10.1186/s12913-024-10867-7
36. Zednik C. Solving the black box problem: A normative framework for explainable artificial intelligence. *Philos. Technol.* **2021**, *34*, 265–288. DOI:10.1007/s13347-019-00382-7
37. Fahim YA, Hasani IW, Kabba S, Ragab WM. Artificial intelligence in healthcare and medicine: Clinical applications, therapeutic advances, and future perspectives. *Eur. J. Med. Res.* **2025**, *30*, 848. DOI:10.1186/s40001-025-03196-w
38. Rizvi YS, Zaheer S. Training healthcare professionals in artificial intelligence augmented services. In *Industry 4.0 and Intelligent Business Analytics for Healthcare*; Nova Science Publishers: New York, NY, USA, 2022; pp. 117–134.
39. Morrow AS, Campos Vega AD, Zhao X, Liriano MM. Leveraging machine learning to identify predictors of receiving psychosocial treatment for attention deficit/hyperactivity disorder. *Adm. Policy Ment. Health Ment. Health Serv. Res.* **2020**, *47*, 680–692. DOI:10.1007/s10488-020-01045-y