Mediterranean Journal of Clinical Psychology

ISSN 2282-1619

MJCP

Volume 14, n 2, 2026

Articles

Machine Learning Predictive Model for ADHD Diagnosis: Presentations and Comorbidities

*Catalina Quintero-López [1] *, Víctor Daniel Gil Vera [2], María José Ruiz-Quintero [3], Andrés Gallego [4]*

Abstract

Background: Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental condition with heterogeneity and clinical variability. It has three diagnostic presentations (inattentive, hyperactive/impulsive, combined). It is common for it to occur concomitantly with other psychopathologies (neurodevelopmental disorder or behavioral disorder), which increases the complexity of its detection.

Methods: This work developed and evaluated a multiclass machine learning (ML) model to identify ADHD and comorbidities in a pediatric population. The research approach was quantitative and cross-sectional, with a non-experimental design and non-probability sampling. The sample has 892 children aged [6-12] years from Medellín, Colombia (ADHD=780, typically developing controls=112). Cognitive and behavioral assessments were used as predictor variables. Python 3.12.13 was used in the programming algorithms; a computational pipeline was implemented with preprocessing techniques, class balancing using Synthetic Minority Over-sampling Technique (SMOTE) and stratified cross-validation.

Results: Multiple classification algorithms were evaluated, Random Forest (RF) was identified as the optimal model (accuracy=.9778, precision=.9798, recall=.9778, F1-score=.9777; weighted averages for multiclass metrics). The results indicate an adequate capacity of the model to differentiate complex clinical profiles with a performance superior to that of traditional binary classification methods.

Conclusions: Use of accessible and clinically interpretable psychometric variables enhances the applicability of the tool in real-world healthcare settings with limited access to resources and specialized professionals. ML is a promising complementary tool for ADHD detection. External validation is needed to confirm its generalizability across different populations and clinical contexts.

[1] Faculty of Social Sciences, Health and Well-being, Catholic University Luis Amigó, Medellín, Colombia

[2] Faculty of Engineering, Catholic University Luis Amigó, Medellín, Colombia

[3] Faculty of Sciences, National University of Colombia, Medellín, Colombia

[4] Faculty of Engineering, National University of Colombia, Bogotá, Colombia

E-mail corresponding author: catalina.quinterolo@amigo.edu.co

Keywords:

ADHD, Clinical psychology, Comorbidity, Machine learning, Multiclass classification, Psychometric testing.

Received: 5 May 2026

Accepted: 3 August 2026

Published: 15 August 2026

Citation: Quintero-López, C., Gil Vera, D. V., Ruiz-Quintero, M. J., & Gallego, A. (2026). Machine learning predictive model for ADHD diagnosis: Presentations and comorbidities. *Mediterranean Journal of Clinical Psychology, 14*(2). https://doi.org/10.13129/2282-1619/mjcp-5422

1. Introduction

ADHD is the most common neurodevelopmental disorder in children, with an estimated global prevalence of around 5.3% (Johnson et al., 2026; Polanczyk et al., 2014; Salari et al., 2023). It is a highly relevant reason for seeking professional help due to its academic, familial and social repercussions (Karteczka-Świętek et al., 2022; Schmengler et al., 2023). Although the diagnostic definition is based on dimensions of inattention, hyperactivity and impulsivity, the clinical presentation is far from uniform (Ging-Jehli & Pine, 2026). Far from constituting a unitary entity, ADHD comprises a heterogeneous set of manifestations that change in intensity, symptomatic combination, evolutionary trajectory and level of functional impairment; this complicates the nosological delimitation and recognition in clinical practice (American Psychiatric Association, 2022; Yue et al., 2022).

This diversity does not depend solely on the variability between presentations (inattentive, hyperactive-impulsive, or combined); in clinical practice, the condition often coexists with other neurodevelopmental or behavioral conditions (autism, learning difficulties, oppositional defiant disorder, among others), which modifies the behavioral profile, broadens the phenotypic spectrum and makes a linear or dichotomous diagnostic approach less likely (Bonti et al., 2024; Nigg, 2026). The clinical challenge lies in establishing the presence or absence of the disorder and distinguishing complex clinical presentations, which are frequently complicated by comorbidities (Nazmin et al., 2024). The conventional diagnostic process continues to rely on a combination of clinical interviews, developmental history, behavioral observation, parent and teacher questionnaires and psychometric tests (Musullulu, 2025). Although this approach constitutes the clinical standard, its application depends on the availability of trained professionals, requires lengthy evaluation times and is not always easily accessible in settings with limited healthcare resources or financial constraints (Willig et al., 2024). This situation takes on particular importance in settings where access to a comprehensive neuropsychological assessment is limited, so the need for complementary, objective and scalable tools has taken center stage in recent research (da Costa et al., 2026; Rodríguez-Prieto et al., 2024).

In recent years, artificial intelligence and in particular ML and deep learning methods have been incorporated into the study of ADHD with the aim of improving diagnostic accuracy (ACC) and advancing toward clinical support models (Zaheer & Akhtar, 2025). The literature reports approaches built from medical data, questionnaires, cognitive tests, demographic variables, electroencephalographic signals and structural and functional magnetic resonance imaging (Lohani & Rana, 2023; Quintero López et al., 2024).

Deep learning-based approaches have played a particularly prominent role in the analysis of electroencephalograms and neuroimaging, due to their ability to model complex nonlinear relationships and extract high-dimensional patterns from biological signals (Agarwal et al., 2024; Sanchis et al., 2024). The growing interest in these developments does not eliminate their limitations (Yen et al., 2023).

Some studies have been conducted with small sample sizes, in highly controlled environments and under methodological conditions that make it difficult to generalize the findings to real-world clinical populations (Ghasemi et al., 2022; Karabiber Cura et al., 2023; Maniruzzaman et al., 2022). Related works have used public repositories whose value for training and comparing models varies significantly in acquisition protocols, processing pipelines and sample characteristics, introducing variability and limiting comparability between studies (Begum & Shaik, 2025; Chen et al., 2025; Guo et al., 2024). The data acquisition cost, dependence on specialized technological infrastructure and the need for highly trained personnel for data capture and interpretation are factors that restrict transferability to everyday healthcare settings, especially in health systems with access gaps (Hu et al., 2024; Huynh et al., 2024). Advanced neuroimaging approaches may further increase accessibility barriers, while greater technological complexity does not necessarily translate into a proportional clinical advantage (Knyazhansky & Shrot, 2025; Miller et al., 2021).

In children, behavioral measures and functional reports on cognitive processes have shown predictive value for ADHD classification, whereas the incremental contribution of structural brain imaging measures has been more modest than expected in terms of diagnostic classification (Bledsoe et al., 2020; Li et al., 2025). Psychometric tests are important in the construction of clinically useful predictive models. They are standardized instruments that are more accessible and widely used in clinical practice with a direct relationship to the functional manifestations of the disorder (Goh et al., 2023; Öztekin et al., 2021). Recent ML applications based on clinical assessment data have also demonstrated the feasibility of classifying ADHD and its subtypes in children using standardized behavioral and cognitive measures (Qin et al., 2025). The combination of behavioral and cognitive measures has shown high classification performance when integrated with supervised classification algorithms (Caselles-Pina et al., 2024). Most models have been oriented toward binary classification problems, generally ADHD versus typical development or have focused the analysis on partial subgroups, without adequately representing the clinical diversity of the disorder (Dai & Hsu, 2025).

Studies that incorporate the three diagnostic presentations are less frequent and even more so those that explicitly consider the burden of comorbidity, even though this constitutes one of the main sources of complexity in everyday clinical assessment (Alqaysi et al., 2022; Lin et al., 2023; Qureshi et al., 2017). While an important portion of methodological development has prioritized algorithmic optimization, it remains necessary to move toward models that more faithfully reflect the real structure of the diagnostic problem (Narzisi et al., 2026).

This study aimed to develop and evaluate a predictive model for ADHD based on standardized psychometric tests, to distinguish the three subtypes of the disorder and clinically relevant comorbidities in a pediatric sample. This approach did not seek to replace comprehensive neuropsychological assessment; rather, it is a complementary diagnostic support tool, useful in contexts with limited access to specialized psychometric testing or in settings with social disparities in healthcare. Thirteen categories were established in a multi-class classification scheme, derived from the three presentations: 1) inattentive ADHD, inattentive ADHD with oppositional defiant disorder (ODD), inattentive ADHD with autism (ASD) and inattentive ADHD with specific learning disorder (SLD). 2) hyperactive/impulsive ADHD, hyperactive/impulsive ADHD with ODD, hyperactive/impulsive ADHD with ASD and hyperactive/impulsive ADHD with SLD. 3) combined ADHD, combined ADHD with ODD, combined ADHD with ASD and combined ADHD with SLD; and 4) typically developing group.

2. Methodology

2.1 Approach and Design

The study employed a quantitative, predictive-correlational approach with a non-experimental, cross-sectional design. It focused on the development, evaluation and validation of ML models for the classification of ADHD.

2.2 Participants

The sample size was estimated using an ACC-based approach, considering publications that employed ML to diagnose ADHD using standardized psychometric tests (Table 1). A structured literature search was conducted in Web of Science and Scopus for studies published between January 2014 and June 2023. Search terms were defined using the APA PsycINFO and IEEE thesauri and combined "predictive modelling" OR "machine learning" with "ADHD". Eligible publications were empirical studies conducted in pediatric populations that applied machine-learning methods to ADHD prediction, used training and validation procedures and reported model-performance metrics.

Studies focused on adult populations and non-empirical publications were excluded from the selection. For the purpose of the sample-size calculation, the studies considered in Table 1 were those that used standardized psychometric, cognitive, or behavioral assessments as the data source for model development. Nine studies met this methodological criterion. These studies examined measures related to ADHD symptoms, attention, executive functioning, behavior and other clinically relevant domains, making their reported classification ACC directly pertinent to the type of variables used in the present study.

Their performance values were therefore used to estimate the expected ACC underlying the sample-size calculation. The selection criterion was defined by the type of data used for model development, rather than by the magnitude of the reported classification performance.

Based on these studies, an average precision of p=.88 was obtained and a desired precision of .90133 was established, with a significance level of α=.05 and $Z\alpha/2$=1.96 (Eq. 1), where d denotes the difference between the desired and expected precision.

$$n = \left(\frac{Z_{\alpha/2} \cdot \sqrt{p \cdot (1-p)}}{d} \right)^2 = 891.02 \approx 892 \qquad (1)$$

Non-probability sampling was used. The sample consisted of 892 Colombian children with and without ADHD. The following inclusion criteria were established: a) children aged 6 to 12 years; b) currently enrolled in the education system; c) IQ of 70 or higher (IQ≥70); d) for the clinical group, a diagnosis of ADHD in any of its subtypes (inattentive, hyperactive/impulsive, or combined), made by a specialized professional; e) children with ADHD with and without comorbidities (ODD, ASD, SLD); f) for the control group, typically developing children with no prior neuropsychiatric diagnoses. Exclusion a) in the clinical group, presence of psychopathologies other than ADHD or comorbid diagnoses not corresponding to those defined in the study; b) history of neurological or genetic disorders; c) chronic medical conditions affecting cognitive or behavioral performance.

Diagnostic classifications were derived from comprehensive neuropsychological evaluations conducted by specialized professionals. ADHD presentations and comorbid diagnoses of ASD, ODD, and SLD were established according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR) diagnostic criteria (American Psychiatric Association, 2022), supported by the clinical and psychometric information obtained during the assessment process. Diagnosis-specific standardized instruments were used when clinically indicated as part of the comprehensive evaluation. The diagnostic labels used as the target variable were established independently of the subset of clinical, behavioral and cognitive measures subsequently selected as predictors for the ML models.

Table 1

Foundational studies for estimating precision in sample calculations

Authors	Sample	Test	Model	Accuracy	Validation
(Bledsoe et al., 2020)	23 ADHD 12 Control	Conners BASC-2 d2 Attention	SVM DT	ACC=97.1%	LOOCV
(Duda et al., 2017)	174 ADHD 248 ASD	SRS	LDA	ACC=83.0%	3-fold cross-validation
(Goh et al., 2023)	222 ADHD 177 Control	K-SADS ADHD-RS	RF	ACC=92.0%	LOOCV
(Kim et al., 2023)	79 ADHD 1011 Control	K-SADS	LGBM	ACC=97.3%	Training 70% Test 10% Retention 20%
(Lin et al., 2023)	328 ADHD inattentive 239 ADHD combined	SNAP-IV Conners	ANN	ACC=86.7%	k-fold cross-validation
(Mikolas et al., 2022)	153 ADHD 139 Control	Conners-3 CBCL TRF SDQ	SVM	ACC=66.1%	30-fold cross-validation
(Öztekin et al., 2021)	87 ADHD 75 Control	BRIEF	SVM	ACC=92.6%	5-fold cross-validation
(Slobodin et al., 2020)	213 ADHD 245 Control	CPT MOXO	RF	ACC=87.0%	100-fold cross-validation
(Yasumura et al., 2020)	108 ADHD 108 Control	STROOP	SVM	ACC=86.3%	NR

Note. ACC=accuracy, ASD=Autism Spectrum Disorder, BRIEF=Behavioral Assessment of Executive Function, SNAP-IV = Scale for the Detection of ADHD, Conners=ADHD Behavior Questionnaire, CPT=Continuous Performance Test, MOXO=Sustained Performance Test, STROOP=Color-Word Test, SRS=Social Responsiveness Scale, K-SADS= Children's Scale for the Assessment of Depression, ADHD-R= ADHD Rating Scale, CBCL=Child Behavior Checklist, TRF=Teacher Report Form, SDQ=Strengths and Difficulties Questionnaire, BASC-2=Behavior Assessment System for Children and Adolescents, d2 Attention=Attention Test, SVM=Support Vector Machines, ANN=Artificial Neural Networks, RF= Random Forests, LDA=Linear Discriminant Analysis, LGBM=Light Gradient Boosting Machine, DT= Decision Trees, NR= Not Reported, LOOCV= Leave-One-Out Cross-Validation.

Table 1 summarizes previous studies that used ML models to classify ADHD based on psychometric tests, which incorporate behavioral instruments and cognitive tests. Variability is observed in sample sizes, instruments and algorithmic approaches. They reported ACC values ranging from [66.1%–97.3%]. These results served as the empirical basis for estimating the expected precision parameter in sample size calculation. Table 2 characterizes the sample according to diagnosis, distribution by sex, age and educational level.

Table 2

Demographic profile

Diagnosis	Total	Sex		Age							Education level	
		M	F	6	7	8	9	10	11	12	Elementary (1°-5°)	Secondary (6°-7°)
ADHD combined	88	69	19	20	18	12	9	5	5	19	67	21
ADHD combined with ASD	25	24	1	5	6	1	3	0	3	7	17	8
ADHD combined with SLD	85	67	18	11	12	25	13	10	7	7	76	9
ADHD combined with ODD	91	76	15	19	13	8	8	6	11	26	59	32
ADHD hyperactive/impulsive	43	30	13	14	6	5	8	4	3	3	41	2
ADHD hyperactive/impulsive with ASD	38	27	11	7	6	2	4	8	0	11	29	9
ADHD hyperactive/impulsive with SLD	51	24	27	9	8	7	9	7	4	7	46	5
ADHD hyperactive/impulsive with ODD	56	42	14	9	8	2	9	5	3	20	35	21
ADHD inattentive	107	74	33	15	10	11	9	9	13	40	58	49
ADHD inattentive with SLD	98	60	38	8	12	13	19	10	10	26	70	28
ADHD inattentive with ODD	51	41	10	4	8	5	7	3	7	17	29	22
ADHD inattentive with ASD	47	37	10	5	5	3	6	10	6	12	33	14
Typical development	112	75	37	18	14	17	13	21	15	14	91	21
Total	892	646	246	144	126	111	117	98	87	209	651	241

Table 2 presents the sample (N=892). There is a predominance of males (n=646) over females (n=246). The age distribution falls within the range of [6–12] years. Most participants are in elementary school. The sample was distributed across 13 diagnostic categories: combined ADHD (n=88), combined ADHD with ASD (n=25), combined ADHD with SLD (n=85), combined ADHD with ODD (n=91); hyperactive/impulsive ADHD (n=43), hyperactive/impulsive ADHD with ASD (n=38), hyperactive/impulsive ADHD with SLD (n=51), hyperactive/impulsive ADHD with ODD (n=56); inattentive ADHD (n=107), inattentive ADHD with SLD (n=98), inattentive ADHD with ODD (n=51), inattentive ADHD with ASD (n=47) and typical development (n=112). This distribution demonstrates the inclusion of multiple clinical combinations of the disorder, which allows for capturing heterogeneity and supports the development of clinically valid multiclass predictive models.

2.3 Instruments and variables

Standardized psychometric tests administered by professionals specializing in neuropsychology at neurological centers in the city of Medellín, Colombia, were used as predictor variables.

Conners Revised Clinician Rating Scale (CPRS-R; Conners, 1998). This is a standard instrument for assessing ADHD symptoms in children and adolescents. It assesses the dimensions of inattention, hyperactivity and impulsivity using Likert-scale items. The total ADHD index, inattention index and hyperactivity-impulsivity index were used. Recent psychometric studies of the Conners scales have reported adequate to high internal consistency (Cronbach's α=.77-.94), with evidence of test–retest reliability and convergent validity (Mehrzad et al., 2026; Nguyen et al., 2024).

Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V; Wechsler, 2015). A standardized test for assessing general cognitive functioning in children. In this study, the full-scale IQ (FSIQ) was used as an overall indicator of cognitive performance. The FSIQ corresponds to a composite score with a mean ($\bar{X}$=100) and standard deviation (σ=15), derived from multiple cognitive subtests. The WISC-V exhibits high reliability, with coefficients ranging from [.88-.95] and its validity is supported by confirmatory factor analyses and correlations with other measures of intelligence (Watkins et al., 2022). The FSIQ showed adequate temporal stability in clinical samples (r=.86) and adequate internal consistency (ω=.904), supporting its psychometric reliability for clinical interpretation (Watkins & Canivez, 2021).

Behavior Assessment System for Children and Adolescents (BASC-3; Kamphaus & Reynolds, 2015). An instrument that assesses adaptive and maladaptive behaviors in clinical and educational settings. In this study, the clinical scales for aggression, hyperactivity, conduct problems, attention problems and atypicality were used. The scores correspond to standardized typical scores derived from normative scales. Psychometric evidence supports the reliability and validity of the BASC-3. The clinical scales have shown adequate to high internal consistency, with Cronbach's alpha coefficients ranging from .82 to .97 in parent- and teacher-report forms (Akrami et al., 2021). Evidence from other BASC-3 components also supports test–retest and inter-rater reliability, as well as convergent, divergent, predictive and discriminant validity (Schmidt et al., 2023).

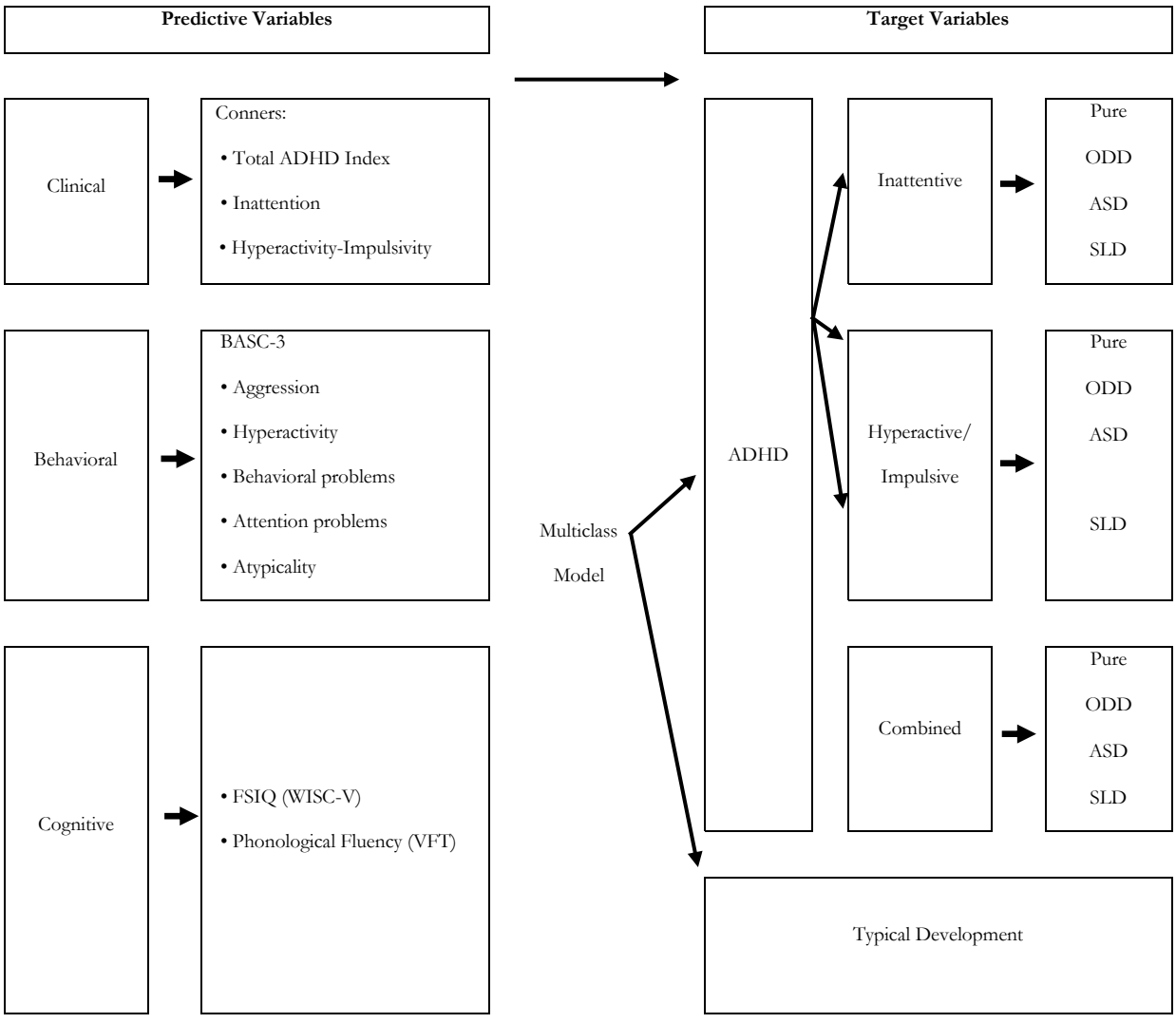
Verbal Fluency Test (VFT; Portellano Pérez & Martínez Arias, 2020). An individually administered neuropsychological test designed for the rapid assessment of expressive language and executive functions. In this study, phonological verbal fluency was used, which consists of generating words beginning with a specific letter within a limited time frame. It demonstrates adequate validity as a measure of language and executive functions, specifically in processes of lexical access, working memory and attentional control. Recent evidence in Spanish-speaking children supports the convergent validity and test–retest reliability of verbal fluency measures, with coefficients above .80 for total word productivity and high stability when performance was modeled as a unified construct (ϕ=.94) (Arán Filippetti et al., 2024).

Predictor selection was guided by the clinical relevance of the measures to the symptomatic, behavioral, and cognitive domains involved in ADHD and by evidence from previous ML studies using standardized clinical and psychometric measures (Bledsoe et al., 2020; Goh et al., 2023). The predictor set comprised the three Conners indices, five BASC-3 clinical scales, FSIQ, and phonological verbal fluency. Similar symptom, behavioral, and cognitive measures have been incorporated in previous ADHD classification studies (Lin et al., 2023; Mikolas et al., 2022). These variables were selected before model development to represent complementary clinically interpretable domains. No automated feature-selection procedure was applied.

Figure 1 shows the variables considered in the construction of the multiclass classification models. Clinical, behavioral and cognitive tests were included as predictors, 12 diagnostic categories of ADHD with and without comorbidities were used as the target variable and 1 category of typical development was included.

Figure 1

Variables for the multiclass classification of ADHD



2.4 Computational pipeline for data preprocessing and model development

A unified computational pipeline was implemented in Python 3.12.13 (Python Software Foundation, 2026) to integrate data preprocessing, model training, and evaluation. Data manipulation and structuring were performed using the pandas and numpy libraries. Numerical variables were screened for consistency with their expected measurement ranges, and no automated statistical outlier-removal procedure was applied. The original sample was first divided using stratified sampling into training (n=713; 79.9%), validation (n=89; 10.0%), and test (n=90; 10.1%) sets before any preprocessing procedure was fitted. Preprocessing was fitted exclusively on the training data and included median imputation for missing values followed by MinMaxScaler normalization to the [0,1] range (Appendix 1 - Normalized dataset). MinMaxScaler was selected to place predictors measured on different psychometric scales within a common bounded range while preserving their relative ordering. Scaling preceded SMOTE because synthetic observations are generated from nearest-neighbor relationships and are therefore sensitive to differences in predictor scale (Chawla et al., 2002).

Class imbalance was addressed using SMOTE exclusively within the training data, consistent with recommendations for resampling imbalanced medical datasets (Khushi et al., 2021). SMOTE was configured with sampling_strategy="auto", random_state=42, and k_neighbors=3. Under this configuration, each minority category was resampled to the size of the majority category within the corresponding training data. The validation and test sets remained composed exclusively of original observations. Model training and comparison incorporated 5-fold stratified cross-validation (Hao & Ho, 2019). Within each training fold, median imputation, MinMax scaling, and SMOTE were fitted sequentially before model estimation. Imputation and scaling parameters were estimated only from observations in the corresponding training fold and synthetic observations were generated exclusively within that fold. The held-out fold remained composed of original observations. This procedure prevented information from evaluation observations from entering model training, as applying oversampling before cross-validation can produce optimistically biased performance estimates (Demircioğlu, 2024).

Seven supervised classification algorithms were evaluated: Logistic Regression, SVM, Gaussian Naive Bayes, Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), Decision Tree and RF. Models were compared on the independent validation set using ACC, weighted precision, weighted recall, and weighted F1-score. The optimal model was selected primarily according to weighted F1-score, with the remaining metrics used as secondary criteria. After model selection, the selected algorithm was refitted using the combined original training and validation data through the same preprocessing and SMOTE pipeline and evaluated once on the untouched

test set of 90 real participants. The Python implementation is presented in a Google Colab notebook (Appendix 2- Python code).

2.5 Data Analysis

An exploratory Pearson correlation analysis was conducted among the 10 numerical predictors using the complete original sample (N=892), prior to SMOTE and model development. The analysis was descriptive and was not used for predictor selection or feature elimination. The correlation matrix was calculated in Python and visualized as a heatmap. Seven supervised classification algorithms were evaluated: Logistic Regression, SVM, Gaussian Naive Bayes, MLP, KNN, Decision Tree and RF. Logistic Regression was configured with max_iter=1000, random_state=42, and class_weight="balanced"; SVM with random_state=42, class_weight="balanced", and probability=True; MLP with hidden_layer_sizes=(100,50), max_iter=1000, and random_state=42; KNN with n_neighbors=5; Decision Tree with random_state=42 and class_weight="balanced"; and RF with n_estimators=100, random_state=42, and class_weight="balanced". Gaussian Naive Bayes was implemented using its default configuration. No automated hyperparameter-search procedure was performed, and the complete predictor set defined in Section 2.3 was retained across models.

Model development incorporated 5-fold stratified cross-validation (Hao & Ho, 2019) using the preprocessing sequence described in Section 2.4. Models were compared on the independent validation set using ACC, weighted precision, weighted recall, and weighted F1-score. The optimal model was selected primarily according to weighted F1-score, with the remaining metrics used as complementary criteria. RF was selected as the final model. It was subsequently refitted using the combined original training and validation sets (n=802) through the same preprocessing and SMOTE pipeline and evaluated once on the untouched test set (n=90). Final performance was summarized using ACC, weighted precision, weighted recall, weighted F1-score, and the multiclass confusion matrix.

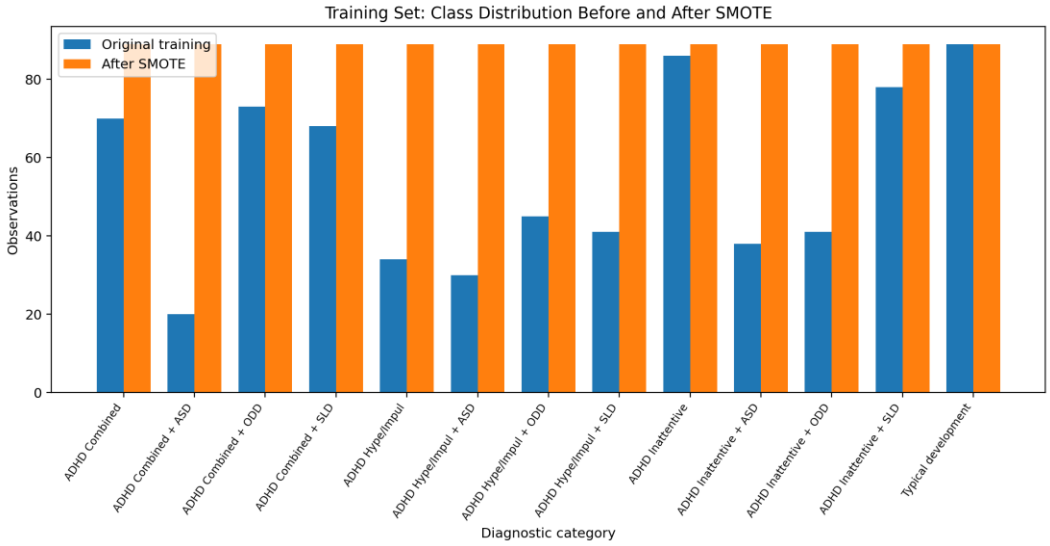
3. Results

3.1 Class distribution and SMOTE resampling

The original training set (n=713) showed an unequal distribution across the 13 diagnostic categories. SMOTE was applied exclusively to the training data, as described in Section 2.4. Using sampling_strategy="auto", each minority category was resampled to the size of the largest category in the training set (n=89). The validation and test sets were not resampled and contained only original observations. Figure 2 presents the class distribution in the training set before and after SMOTE.

Figure 2

Class distribution in the training set before and after SMOTE



After SMOTE, each training category contained 89 observations. The number and proportion of synthetic observations depended on the original size of each category. No synthetic observations were generated for typical development, whereas ADHD combined with ASD required the largest synthetic contribution (n=69; 77.53%). Synthetic observations were restricted to model training and were not included in validation or test performance estimates. Table 3 details the original and synthetic composition of each training category.

Table 3

Original and synthetic observations in the training set after SMOTE

Diagnostic category	Original n	After SMOTE n	Synthetic n	Original %	Synthetic %
ADHD Combined	70	89	19	78.65	21.35
ADHD Combined + ASD	20	89	69	22.47	77.53
ADHD Combined + ODD	73	89	16	82.02	17.98
ADHD Combined + SLD	68	89	21	76.40	23.60
ADHD Hype/Impul	34	89	55	38.20	61.80
ADHD Hype/Impul + ASD	30	89	59	33.71	66.29
ADHD Hype/Impul + ODD	45	89	44	50.56	49.44
ADHD Hype/Impul + SLD	41	89	48	46.07	53.93
ADHD Inattentive	86	89	3	96.63	3.37
ADHD Inattentive + ASD	38	89	51	42.70	57.30
ADHD Inattentive + ODD	41	89	48	46.07	53.93
ADHD Inattentive + SLD	78	89	11	87.64	12.36
Typical development	89	89	0	100.00	.00

3.2 Correlation analysis

The Pearson correlation matrix was calculated using the complete original sample (N=892), before SMOTE and model development, as specified in Section 2.5. Figure 3 shows the associations among the 10 predictors.

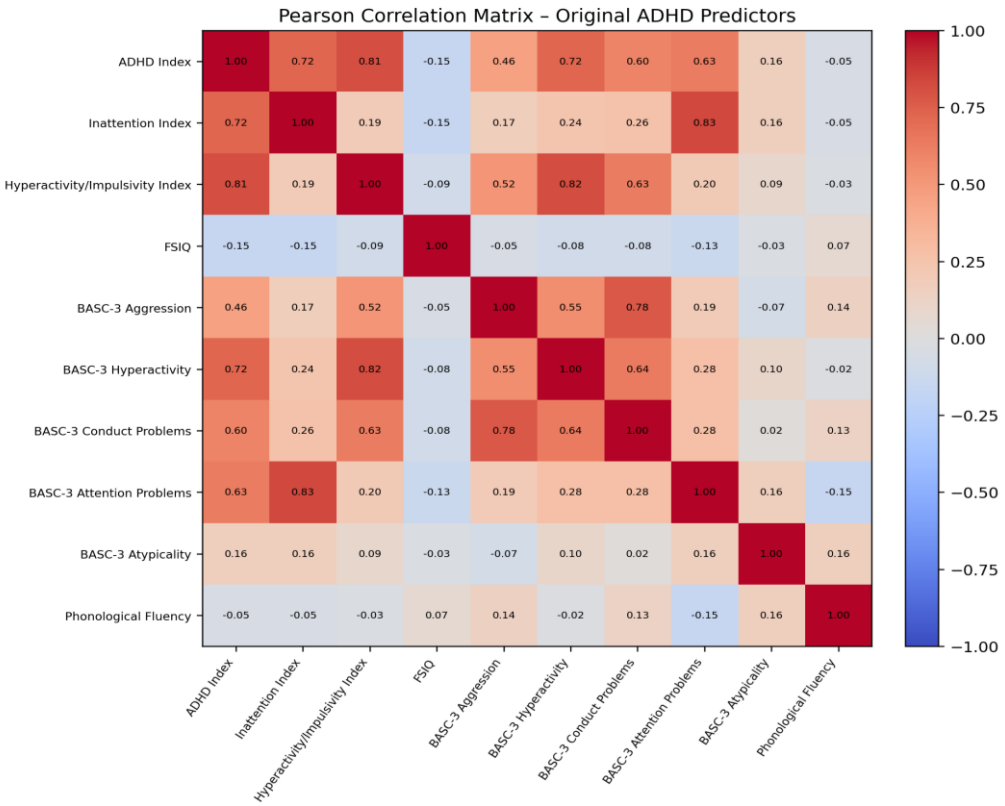
The overall ADHD index showed strong associations with the impulsivity/hyperactivity (r=.81) and inattention (r=.72) indices, reflecting convergence among the core dimensions (Wu et al., 2022).

Strong correlations were identified between BASC-3 hyperactivity and the impulsivity/hyperactivity index (r=.82), as well as between aggression and conduct problems (r=.78), suggesting a direct relationship between behavioral dysregulation and externalizing symptoms (Rosenthal et al., 2024). BASC-3 attention problems were strongly correlated with the inattention index (r=.83).

FSIQ showed weak negative correlations with the ADHD index (r=-.15), inattention index (r=-.15), and BASC-3 attention problems (r=-.13). Phonological fluency also showed weak associations with the remaining predictors, with coefficients ranging from r=-.15 to r=.16. BASC-3 atypicality was weakly associated with the other predictors (|r|<=.16). The correlation analysis was descriptive and was not used for predictor selection or feature elimination.

Figure 3

Pearson correlation matrix of the original predictors



3.3 Model comparison and validation performance

Across 5-fold stratified cross-validation, RF showed the highest mean performance, with ACC=.9762 (SD=.0094) and weighted F1-score=.9761 (SD=.0095). The cross-validation variability estimates for all models are reported in Table 4.

On the held-out validation set (n=89), RF, SVM, and Gaussian Naive Bayes each achieved ACC=.9888. RF obtained the highest weighted F1-score (.9892) and weighted precision (.9916), and was therefore selected as the final model according to the prespecified selection criterion. Figure 4 summarizes validation performance across the seven algorithms.

Figure 4

Performance comparison of classification models on the validation set

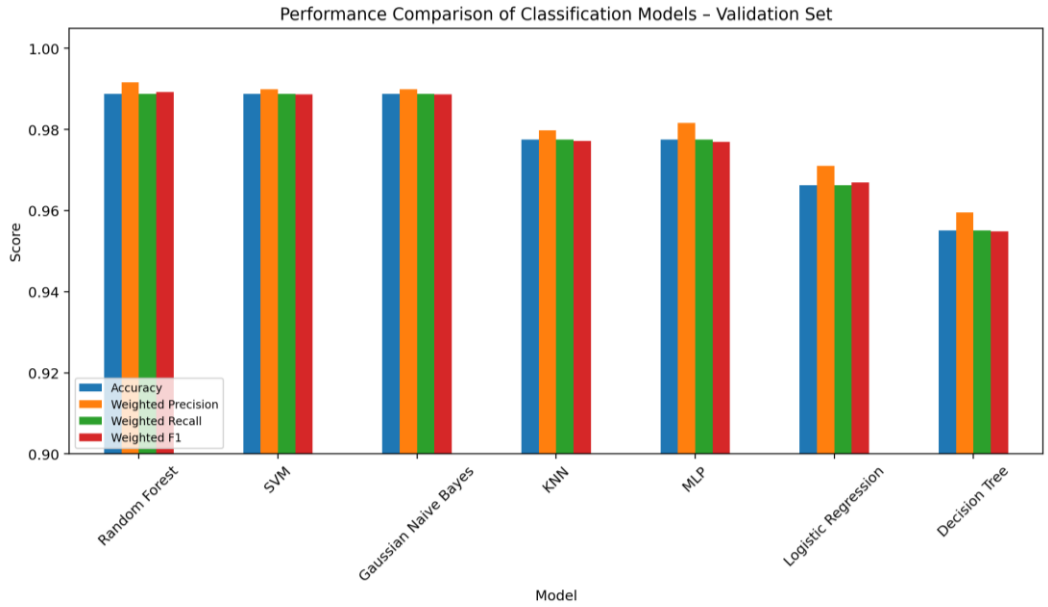


Table 4

Cross-validation variability and held-out validation performance

Model	CV ACC mean (SD)	CV weighted F1 mean (SD)	Validation ACC	Weighted precision	Weighted recall	Weighted F1
Random Forest	.9762 (.0094)	.9761 (.0095)	.9888	.9916	.9888	.9892
SVM	.9607 (.0128)	.9607 (.0133)	.9888	.9899	.9888	.9887
Gaussian Naive Bayes	.9593 (.0174)	.9587 (.0184)	.9888	.9899	.9888	.9887
MLP	.9523 (.0181)	.9521 (.0183)	.9775	.9816	.9775	.9770
Decision Tree	.9383 (.0075)	.9383 (.0078)	.9551	.9596	.9551	.9549
Logistic Regression	.9369 (.0180)	.9360 (.0193)	.9663	.9710	.9663	.9669
KNN	.9215 (.0207)	.9209 (.0216)	.9775	.9798	.9775	.9772

Note. CV values are mean (SD) across the five stratified folds. Precision, recall, and F1-score are weighted averages for the multiclass problem.

3.4 Final RF performance on the held-out test set

After model selection, RF was refitted using the combined original training and validation sets (n=802) through the same preprocessing and SMOTE pipeline and was evaluated once on the untouched held-out test set (n=90). The final model achieved ACC=.9778, weighted precision=.9798, weighted recall=.9778, and weighted F1-score=.9777.

At the class level, all test observations were correctly classified in 11 of the 13 true diagnostic categories. One ADHD combined + SLD observation was classified as ADHD combined, and one ADHD inattentive + SLD observation was classified as ADHD inattentive. These two errors produced recall values of .8889 and .9000, respectively, for the corresponding SLD comorbidity categories. Table 5 presents the complete classification report, and Figure 5 shows the final confusion matrix.

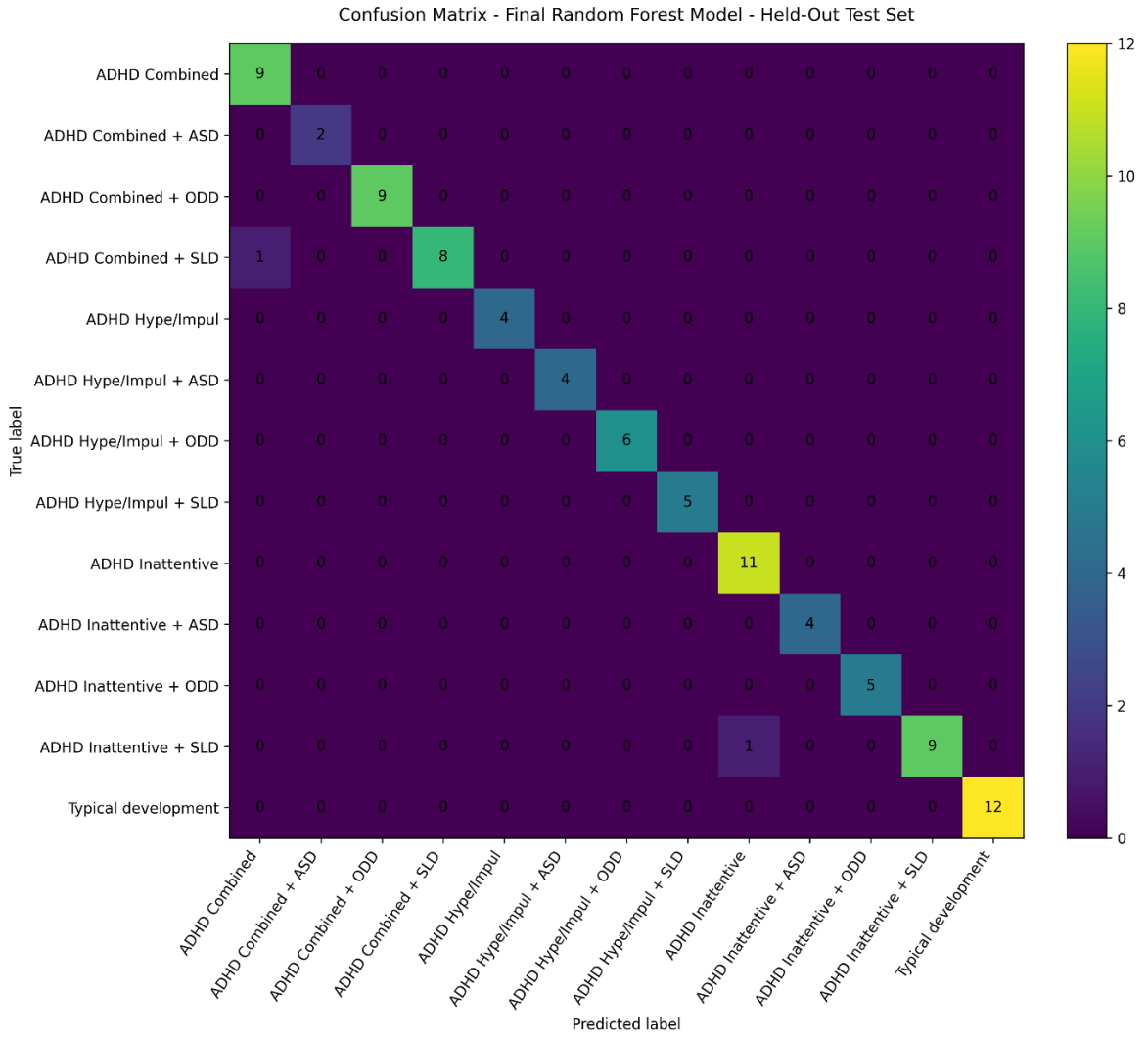
Table 5

Classification report for the final RF model on the held-out test set

Diagnostic category	Precision	Recall	F1-score	Support
ADHD combined	.9000	1.0000	.9474	9
ADHD combined + ASD	1.0000	1.0000	1.0000	2
ADHD combined + ODD	1.0000	1.0000	1.0000	9
ADHD combined + SLD	1.0000	.8889	.9412	9
ADHD hyperactive/impulsive	1.0000	1.0000	1.0000	4
ADHD hyperactive/impulsive + ASD	1.0000	1.0000	1.0000	4
ADHD hyperactive/impulsive + ODD	1.0000	1.0000	1.0000	6
ADHD hyperactive/impulsive + SLD	1.0000	1.0000	1.0000	5
ADHD inattentive	.9167	1.0000	.9565	11
ADHD inattentive + ASD	1.0000	1.0000	1.0000	4
ADHD inattentive + ODD	1.0000	1.0000	1.0000	5
ADHD inattentive + SLD	1.0000	.9000	.9474	10
Typical development	1.0000	1.0000	1.0000	12
Macro average	.9859	.9838	.9840	90
Weighted average	.9798	.9778	.9777	90

Figure 5

Confusion matrix of the final RF model on the held-out test set



Because the held-out test set contained relatively few observations in some diagnostic categories (e.g., n=2-4 for several ASD and hyperactive/impulsive categories), class-specific estimates should be interpreted in relation to their support. The final test set contained only original observations and was not used during preprocessing, model training, model selection, or SMOTE generation.

4. Discussion

This work aimed to develop and evaluate a predictive model for ADHD based on standardized psychometric tests, oriented towards the multiclass classification of the different presentations of the disorder and comorbidities in a pediatric population. The results showed high internal classification performance across the 13 diagnostic categories, while the interpretation of this performance is restricted to the internal validation framework used in the present study. This finding extends previous literature that has frequently focused on binary classification schemes

(Madububambachu et al., 2024; Zaheer & Akhtar, 2025). This research conceptualized ADHD as a heterogeneous and multidimensional entity (Morris et al., 2023; Niina et al., 2022; Sonuga-Barke et al., 2023). Unlike most studies that address the problem as a dichotomy (ADHD vs. typical development), the proposed model integrated 13 diagnostic categories that combine the 3 clinical presentations with relevant comorbidities (Alsharif et al., 2024; Sethurao et al., 2021). Previous ML studies have also moved beyond simple ADHD-control comparisons by examining ADHD presentations and comorbid neurocognitive profiles; however, these approaches have generally included fewer classification targets or a more restricted set of clinical groupings (Lin et al., 2023; Quintero-López et al., 2023). The present framework therefore extends this line of work by simultaneously representing the three ADHD presentations, selected comorbidities, and typical development within a single multiclass problem.

This approach is consistent with real-world clinical practice, where ADHD is characterized by variability in symptom profiles, severity and comorbidity with other neurodevelopmental disorders (Ging-Jehli & Pine, 2026). Comorbidity may also modify the behavioral and functional expression of ADHD, as observed in children and adolescents with co-occurring ODD (D'Aiello et al., 2024). Within the revised leakage-safe pipeline, RF showed the highest mean cross-validation performance, with ACC=.9762 (SD=.0094) and weighted F1-score=.9761 (SD=.0095), and retained high performance on the held-out test set (ACC=.9778; weighted F1-score=.9777). The relatively large sample (N=892) and the inclusion of clinically relevant psychometric variables remain strengths of the study (Eng et al., 2024; Rajput et al., 2023). The predefined preprocessing sequence, with imputation, scaling, and SMOTE restricted to training data and training folds, reduced the risk of information leakage from evaluation observations and strengthened the internal validity of the analysis (Demircioğlu, 2024). The consistency detected in the relationships between variables suggests that the included predictors adequately capture key dimensions of the behavioral and cognitive functioning involved in ADHD. These findings are relevant compared to previous literature. A high proportion of studies were conducted with small samples, highly specialized data, or experimental conditions that limit their applicability in real clinical contexts (Ashraf et al., 2024; López et al., 2025; Tian et al., 2024). The use of standardized psychometric tests as the basis of the model is advantageous, given that they are accessible tools, widely used in clinical practice and directly related to the functional expression of the disorder (Bigler et al., 2024; da Costa et al., 2026). Evidence that cognitive processes such as working memory are also related to broader functional and social outcomes in children with ADHD supports the value of integrating cognitive and behavioral information rather than relying on a single domain (Quintero-López et al., 2024). This strengthens the model viability as a clinical decision-support tool in settings with limited access to advanced

technologies or specialized assessments (Gizaw et al., 2022). Unlike approaches based on neuroimaging or electrophysiological recordings, which involve high costs, complex technological infrastructure and the availability of highly specialized personnel, psychometric tests can be applied in a wide variety of healthcare contexts (Hua et al., 2026; Song et al., 2022). The proposed model does not substitute the comprehensive neuropsychological evaluation carried out by a specialized professional; rather, it may serve as a complementary decision-support approach based on information obtained from standardized clinical, behavioral, and cognitive assessments. This characteristic expands the potential for implementation in health systems with gaps in care (Beauchamp et al., 2022).

One clinically relevant aspect of this model is that its capacity to discriminate among several diagnostic groupings reflects the heterogeneity and complexity of ADHD presentations and comorbid profiles (Ebadi et al., 2025). Within the broader landscape of AI-assisted ADHD care, this model differs from systems based on unstructured clinical text, large language models, or natural language processing because it relies on a predefined set of standardized psychometric variables and an explicitly specified computational pipeline. This structured design may favor transparency and reproducibility, although clinical implementation still requires attention to usability, data governance, privacy, equity, and integration with professional decision-making, which are central considerations for AI-augmented clinical decision support systems (Dahò & Caci, 2025).

5. Clinical implications, limitations, and future directions

Practical implementation would require availability of the same predictor information used in model development: the three Conners indices, five BASC-3 clinical scales, FSIQ, and phonological verbal fluency. Once these scores are available, the computational requirements are modest because prediction can be performed through the preprocessing and RF pipeline described in Sections 2.4 and 2.5. Accordingly, the model would not eliminate the need for standardized assessment or professional interpretation; its potential role would be to organize already collected information and provide an additional classification signal for clinical review. Potential implementation barriers include access to licensed assessment instruments, staff training, interoperability with clinical information systems, local data-governance requirements, and the need to evaluate usability and acceptance among clinicians (Dahò & Caci, 2025).

The cross-sectional nature of the study precludes claims on the causal direction between the variables included in this study and the non-probability sampling may introduce biases into which we cannot ascertain whether our respondents are representative of a broader population (Mahjoob et al., 2024; Stratton, 2023). In addition, the current results represent internal

validation only. Training, validation, and test observations were obtained from the same Colombian clinical context and diagnostic framework, and some diagnostic categories had limited support in the held-out test set (e.g., n=2-4). Consequently, class-specific estimates for these less frequent categories should be interpreted cautiously and cannot yet establish performance in independent populations. The use of SMOTE also requires specific consideration. Although synthetic observations were generated only within the training data and within training folds during cross-validation, the proportion of synthetic observations varied substantially across categories and reached 77.53% for ADHD combined + ASD. SMOTE observations are mathematical interpolations of existing cases rather than independently assessed patients; therefore, the clinical plausibility of individual synthetic profiles was not directly validated (Gholampour, 2024). This creates the possibility that part of the learned decision structure may reflect properties of the oversampling process, particularly in categories with fewer original observations. For this reason, the strong internal performance should be interpreted together with the original class sizes and the synthetic proportions reported in Table 3.

The model performed well, but more external validation steps are required to test its robustness and transferability to other populations and clinical settings (Santos & Amorim-Lopes, 2025). The importance of culturally informed assessment is particularly relevant when considering the transfer of ADHD and SLD diagnostic tools across populations and clinical contexts (McWhirter & Walters, 2024). Future studies should evaluate the model in independent cohorts from different hospitals, healthcare services, educational settings and sociocultural contexts; increase the number of real participants in the least frequent diagnostic categories; and prospectively examine its contribution to clinical workflow, screening prioritization and referral decisions. Comparative analyses using alternative imbalance-management strategies may also help determine whether the observed multiclass performance is maintained without substantial dependence on synthetic resampling.

6. Conclusions

The results show that standardized psychometric variables can support the identification of ADHD heterogeneity in pediatric populations, including clinical presentations as well as comorbid profiles. By moving beyond the traditional ADHD versus typical development distinction, the proposed multiclass approach offers a closer representation of real clinical assessment, where symptom overlap, behavioral variability, together with comorbidity, shape diagnostic decision-making. The performance of the RF model suggests that behavioral, cognitive, and clinically interpretable measures may contribute to complementary decision-support approaches for ADHD classification. This is particularly relevant in healthcare settings

lacking access to specialized neuropsychological assessment, advanced neuroimaging, or highly trained professionals. Previous research has also highlighted the value of ML as a complementary strategy for early ADHD detection, especially when based on psychometric, clinical or behavioral data rather than costly technological procedures. The proposed model should be understood as a support tool within a comprehensive clinical process rather than a replacement for professional diagnosis. Its applicability is linked to the characteristics of the sample, composed of Colombian children aged 6 to 12 years. External validation in independent cohorts, broader sociocultural contexts, as well as different healthcare environments is required. Future research should examine model performance in prospective designs while also assessing its usefulness for clinical workflow, screening prioritization and early referral pathways.

Abbreviations

Attention-Deficit/Hyperactivity Disorder (ADHD); Machine Learning (ML); Synthetic Minority Over-sampling Technique (SMOTE); Random Forest (RF); Accuracy (ACC); Oppositional Defiant Disorder (ODD); Autism Spectrum Disorder (ASD); Specific Learning Disorder (SLD); Conners Revised Clinician Rating Scale (CPRS-R); Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V); Intellectual Quotient (IQ); Full Scale Intellectual Quotient (FSIQ); Behavior Assessment System for Children and Adolescents (BASC-3); Verbal Fluency Test (VFT); Support Vector Machine (SVM); Multi-Layer Perceptron (MLP); K-Nearest Neighbors (KNN); Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR).

Ethical approval

It respected the confidentiality and identity of the participants. Data protection was guaranteed. The study was approved by the Bioethics Committee of two Colombian higher education institutions: University of San Buenaventura (001-2024) and Luis Amigó Catholic University (719905-2025). The research was classified as minimal risk in accordance with Article No. 11 of 008430 Resolution of 1993 of the Colombian Ministry of Health.

Informed Consent Statement

Informed consent was obtained from the subjects, parents or legal custodians before participation.

Data Availability Statement

Appendix 1 - Normalized dataset (Excel File)

Appendix 2: Python code

https://colab.research.google.com/drive/1BZkBLFX7SZSO9c5lQiwgFkHRFi_tCISn?usp=sharing

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any potential conflict of interest.

Funding

This research was funded by Luis Amigo Catholic University Cost Center 05020299196.

Authors' Contribution

QLC: Writing – original draft, Investigation, Data curation, Conceptualization. DV: Investigation, Data curation, Review & editing, Statistical analysis. RQMJ: Data recollection, Investigation and Writing – original draft. GA: Data evaluation.

References

1. Agarwal, S., Raj, A., Chowdhury, A., Aich, G., Chatterjee, R., & Ghosh, K. (2024). Investigating the impact of standard brain atlases and connectivity measures on the accuracy of ADHD detection from fMRI data using deep learning. *Multimedia Tools and Applications*, *83*(25), 67023–67057. https://doi.org/10.1007/s11042-023-17962-7
2. Akrami, L., Malekpour, M., & Abedi, A. (2021). Developing and accessing psychometric properties of the Persian version of Behavior Assessment System for Children (3rd Edition) in children with mild intellectual disabilities and normal children. *Iranian Journal of Psychiatry and Clinical Psychology, 27*(3), 388–405. https://doi.org/10.32598/ijpcp.27.4.3462.1
3. Alqaysi, M. E., Albahri, A. S., & Hamid, R. A. (2022). Diagnosis-based hybridization of multimedical tests and sociodemographic characteristics of autism spectrum disorder using artificial intelligence and machine learning techniques: a systematic review. *International Journal of Telemedicine and Applications, 2022*(1), 3551528. https://doi.org/10.1155/2022/3551528
4. Alsharif, N., Al-Adhaileh, M. H., Alsubari, S. N., & Al-Yaari, M. (2024). ADHD diagnosis using text features and predictive machine learning and deep learning algorithms. *Journal of Disability Research*, *3*(7), 20240082. https://doi.org/10.57197/JDR-2024-0082
5. American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.; DSM-5-TR). American Psychiatric Publishing.
6. Arán Filippetti, V., Gutierrez, M., & Krumm, G. (2024). Norms, convergent validity, test–retest reliability, and practice effects for verbal fluency overall performance, clustering, and switching in Spanish-speaking children. *The Clinical Neuropsychologist, 38*(7), 1726–1755. https://doi.org/10.1080/13854046.2024.2315729
7. Ashraf, I., Jung, S., Hur, S., & Park, Y. (2024). A systematic literature review of neuroimaging coupled with machine learning approaches for diagnosis of attention deficit hyperactivity disorder. *Journal of Big Data*, *11*(1), 140. https://doi.org/10.1186/s40537-024-00998-3
8. Beauchamp, M. L., Amorim, K., Wunderlich, S. N., Lai, J., Scorah, J., & Elsabbagh, M. (2022). Barriers to access and utilization of healthcare services for minority-language speakers with neurodevelopmental disorders: A scoping review. *Frontiers in Psychiatry*, *13*, 915999. https://doi.org/10.3389/fpsyt.2022.915999
9. Begum, Z., & Shaik, K. (2025). Optimized temporal denoised convolutional autoencoder for enhanced ADHD classification using fMRI data. *IEEE Access, 13*, 29031–29045. https://doi.org/10.1109/ACCESS.2025.3539706
10. Bigler, E. D., Allder, S., & Victoroff, J. (2024). What traditional neuropsychological assessment got wrong about mild traumatic brain injury. II: limitations in test development, research design, statistical and psychometric issues. *Brain Injury, 38*(13), 1053–1074. https://doi.org/10.1080/02699052.2024.2376261
11. Bledsoe, J. C., Xiao, C., Chaovalitwongse, A., Mehta, S., Grabowski, T. J., Semrud-Clikeman, M., Pliszka, S., & Breiger, D. (2020). Diagnostic Classification of ADHD Versus Control: Support Vector Machine Classification Using Brief Neuropsychological Assessment. *Journal of Attention Disorders, 24*(11), 1547–1556. https://doi.org/10.1177/1087054716649666

12. Bonti, E., Zerva, I. K., Koundourou, C., & Sofologi, M. (2024). The high rates of comorbidity among neurodevelopmental disorders: reconsidering the clinical utility of distinct diagnostic categories. *Journal of Personalized Medicine*, *14*(3), 300. https://doi.org/10.3390/jpm14030300
13. Caselles-Pina, L., Quesada-López, A., Sújar, A., Hernandez, E. M. G., & Delgado-Gómez, D. (2024). A systematic review on the application of machine learning models in psychometric questionnaires for the diagnosis of attention deficit hyperactivity disorder. *European Journal of Neuroscience, 60*(3), 4115–4127. https://doi.org/10.1111/ejn.16288
14. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953
15. Chen, L., Yang, Y., Yu, A., Guo, S., Ren, K., Liu, Q., & Qiao, C. (2025). An explainable spatio-temporal graph convolutional network for the biomarkers identification of ADHD. *Biomedical Signal Processing and Control, 99*, 106913. https://doi.org/10.1016/j.bspc.2024.106913
16. Conners, C. K. (1998). *Conners' Rating Scales–Revised: Technical manual.* Multi-Health Systems.
17. da Costa, D. I., de Souza, R. V., & Fiamoncini, T. C. (2026). The role of neuropsychological assessment in the investigation of neurodevelopmental disorders. *Jornal de Pediatria*, *102*, 101470. https://doi.org/10.1016/j.jped.2025.101470
18. Dahò, M., & Caci, B. (2025). Clinical Decision Support Systems and Artificial Intelligence in ADHD Assessment and Rehabilitation: Opportunities and Challenges for Technology-Assisted Care. *Healthcare, 13*(23), 3171. https://doi.org/10.3390/healthcare13233171
19. Dai, Y. W., & Hsu, C. F. (2025). Machine Learning for ADHD Diagnosis: Feature Selection from Parent Reports, Self-Reports and Neuropsychological Measures. *Children, 12*(11), 1448. https://doi.org/10.3390/children12111448
20. D'Aiello, B., Di Vara, S., De Rossi, P., Vicari, S., & Menghini, D. (2024). The effect of a single dose of methylphenidate on attention in children and adolescents with ADHD and comorbid oppositional defiant disorder. *PLoS One*, *19*(8), e0299449. https://doi.org/10.1371/journal.pone.0299449
21. Demircioğlu, A. (2024). Applying oversampling before cross-validation will lead to high bias in radiomics. *Scientific Reports, 14,* 11563. https://doi.org/10.1038/s41598-024-62585-z
22. Duda, M., Haber, N., Daniels, J., & Wall, D. P. (2017). Crowdsourced validation of a machine-learning classification system for autism and ADHD. *Translational Psychiatry, 7*, e1133. https://doi.org/10.1038/tp.2017.86
23. Ebadi, A., Allouch, S., Mheich, A., Tabbal, J., Kabbara, A., Robert, G., Lefebvre, A., Iftimovici, A., Rodríguez-Herreros, B., Chabane, N., & Hassan, M. (2025). Beyond homogeneity: Charting the landscape of heterogeneity in neurodevelopmental and psychiatric electroencephalography. *Translational Psychiatry, 15*(1), 223. https://doi.org/10.1038/s41398-025-03441-0
24. Eng, A. G., Bansal, P. S., Goh, P. K., Nirjar, U., Petersen, M. K., & Martel, M. M. (2024). Evidence-based assessment for attention-deficit/hyperactivity disorder. *Assessment, 31*(1), 42–52. https://doi.org/10.1177/10731911221149957

25. Ghasemi, E., Ebrahimi, M., & Ebrahimie, E. (2022). Machine learning models effectively distinguish attention-deficit/hyperactivity disorder using event-related potentials. *Cognitive Neurodynamics, 16*(6), 1335–1349. https://doi.org/10.1007/s11571-021-09746-2
26. Gholampour, S. (2024). Impact of nature of medical data on machine and deep learning for imbalanced datasets: clinical validity of SMOTE is questionable. *Machine Learning and Knowledge Extraction, 6*(2), 827–841. https://doi.org/10.3390/make6020039
27. Ging-Jehli, N. R., & Pine, D. S. (2026). From symptom-based heterogeneity to mechanism-based profiling in youth ADHD: the promise of computational psychiatry. *Neuropsychopharmacology, 51*(1), 234–243. https://doi.org/10.1038/s41386-025-02254-5
28. Gizaw, Z., Astale, T., & Kassie, G. M. (2022). What improves access to primary healthcare services in rural communities? A systematic review. *BMC Primary Care, 23*(1), 313. https://doi.org/10.1186/s12875-022-01919-0
29. Goh, P. K., Elkins, A. R., Bansal, P. S., Eng, A. G., & Martel, M. M. (2023). Data-Driven Methods for Predicting ADHD Diagnosis and Related Impairment: The Potential of a Machine Learning Approach. *Research on Child and Adolescent Psychopathology, 51*, 679–691. https://doi.org/10.1007/s10802-023-01022-7
30. Guo, J., Wang, M., & Guo, X. (2024). Hierarchical hyperedge graph transformer toward dynamic interactions of brain networks for neurodevelopmental disease diagnosis. *IEEE Access, 12*, 162145–162156. https://doi.org/10.1109/ACCESS.2024.3468439
31. Hao, J., & Ho, T. K. (2019). Machine learning made easy: a review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics, 44*(3), 348-361. https://doi.org/10.3102/1076998619832248
32. Hu, Y., Ran, J., Qiao, R., Xu, J., Tan, C., Hu, L., & Tian, Y. (2024). Identifying ADHD-related abnormal functional connectivity with a graph convolutional neural network. *Neural Plasticity, 2024*(1), 8862647. https://doi.org/10.1155/2024/8862647
33. Hua, C., Wang, X. L., & Sheng, H. (2026). Neurodevelopmental disorders in children: the role of MRI in early detection and intervention planning. *Frontiers in Neuroscience*, *20*, 1758568. https://doi.org/10.3389/fnins.2026.1758568
34. Huynh, N., Yan, D., Ma, Y., Wu, S., Long, C., Sami, M. T., Almudaifer, A., Jiang, Z., Chen, H., Dretsch, M. N., Denney, T. S., Deshpande, R., & Deshpande, G. (2024). The use of generative adversarial network and graph convolution network for neuroimaging-based diagnostic classification. *Brain Sciences, 14*(5), 456. https://doi.org/10.3390/brainsci14050456
35. Johnson, S., Lim, E., Jacoby, P., Faraone, S. V., Su, B. M., Solmi, M., Forrest, B., Furfaro, B., von Klier, K., Downs, J., & Chen, W. (2026). Prevalence of attention deficit hyperactivity disorder/hyperkinetic disorder of pediatric and adult populations in clinical settings: A systematic review, meta-analysis and meta-regression. *Molecular Psychiatry, 31*(1), 576–586. https://doi.org/10.1038/s41380-025-03178-8
36. Kamphaus, R. W., & Reynolds, C. R. (2015). *Behavior Assessment System for Children (3rd ed.)*. Pearson.

37. Karabiber Cura, O., Kocaaslan Atli, S., & Akan, A. (2023). Attention deficit hyperactivity disorder recognition based on intrinsic time-scale decomposition of EEG signals. *Biomedical Signal Processing and Control, 81*, 1–11. https://doi.org/10.1016/j.bspc.2022.104512
38. Karteczka-Świętek, K., Opozda-Suder, S., & Strojny, A. (2022). Psychosocial functioning of adolescents with ADHD in the family, school and peer group: A scoping review protocol. *Plos One, 17*(6), e0269495. https://doi.org/10.1371/journal.pone.0269495
39. Khushi, M., Shaukat, K., Alam, T. M., Hameed, I. A., Uddin, S., Luo, S., Yang, X., & Reyes, M. C. (2021). A comparative performance analysis of data resampling methods on imbalance medical data. *IEEE Acess, 9*, 109960–109975. https://doi.org/10.1109/ACCESS.2021.3102399
40. Kim, W. P., Kim, H. J., Pack, S. P., Lim, J. H., Cho, C. H., & Lee, H. J. (2023). Machine Learning-Based Prediction of Attention-Deficit/Hyperactivity Disorder and Sleep Problems with Wearable Data in Children. *JAMA Network Open, 6*(3), e233502. https://doi.org/10.1001/jamanetworkopen.2023.3502
41. Knyazhansky, M., & Shrot, T. (2025). ADHD diagnostic tools across ages: traditional and digital approaches. *Frontiers in Psychiatry*, *16*, 1668070. https://doi.org/10.3389/fpsyt.2025.1668070
42. Li, Y., Zeng, W., Dong, W., Cai, L., Wang, L., Chen, H., Yan, H., Bian, L., & Wang, N. (2025). MHNet: Multi-view High-Order Network for Diagnosing Neurodevelopmental Disorders Using Resting-State fMRI. *Journal of Imaging Informatics in Medicine*, *38*, 2994–3014. https://doi.org/10.1007/s10278-025-01399-5
43. Lin, I. C., Chang, S. C., Huang, Y. J., Kuo, T. B. J., & Chiu, H. W. (2023). Distinguishing different types of attention deficit hyperactivity disorder in children using artificial neural network with clinical intelligent test. *Frontiers in Psychology, 13*, 1067771. https://doi.org/10.3389/fpsyg.2022.1067771
44. Lohani, D. C., & Rana, B. (2023). ADHD diagnosis using structural brain MRI and personal characteristic data with machine learning framework. *Psychiatry Research: Neuroimaging*, *334*, 111689. https://doi.org/10.1016/j.pscychresns.2023.111689
45. López, C. Q., Vera, V. D., & Quintero, M. J. (2025). Diagnosis of ADHD in Children with EEG and Machine Learning: Systematic Review and Meta-analysis. *Clinical and Health, 36*(2), 109–121. https://doi.org/10.5093/clh2025a16
46. Madububambachu, U., Ukpebor, A., & Ihezue, U. (2024). Machine learning techniques to predict mental health diagnoses: A systematic literature review. *Clinical Practice and Epidemiology in Mental Health: CP & EMH*, *20*, e17450179315688. https://doi.org/10.2174/0117450179315688240607052117
47. Mahjoob, M., Cardy, R., Penner, M., Anagnostou, E., Andrade, B. F., Crosbie, J., Kelley, E., Ayub, M., Brian, J., Iaboni, A., Schachar, R., Georgiades, S., Nicolson, R., Jones, J., & Kushki, A. (2024). *Predictors of health-related quality of life for children with neurodevelopmental conditions*. *Scientific Reports, 14*(1), 6377. https://doi.org/10.1038/s41598-024-56821-9
48. Maniruzzaman, M., Shin, J., Hasan, M. A. M., & Yasumura, A. (2022). Efficient Feature Selection and Machine Learning Based ADHD Detection Using EEG Signal. *Computers, Materials & Continua, 72*(3), 5179–5195. https://doi.org/10.32604/cmc.2022.028339

49. McWhirter, A. C., & Walters, K. P. (2024). The Development of Ecologically Integrated and Culturally Informed ADHD and SLD Diagnostic Tools for Doctoral Assessment Training. *Disabilities*, *4*(3), 583–615. https://doi.org/10.3390/disabilities4030037
50. Mehrzad, F., Tehrani-Doost, M., Fallahinia, N., Shahrivar, Z., & Mahmoudi Gharaei, J. (2026). Psychometric properties of the Conners' 3-parent rating scale in a clinical sample of Iranian children. *BMC Psychiatry, 26*, 468. https://doi.org/10.1186/s12888-026-07928-y
51. Mikolas, P., Vahid, A., Bernardoni, F., Süß, M., Martini, J., Beste, C., & Bluschke, A. (2022). Training a machine learning classifier to identify ADHD based on real-world clinical data from medical records. *Scientific Reports, 12*, 12934. https://doi.org/10.1038/s41598-022-17126-x
52. Miller, L. E., Kaseda, E. T., Koop, J. I., Mau, K. A., & Heffelfinger, A. K. (2021). Differential access to neuropsychological evaluation in children with perinatal complications or autism spectrum disorder: Impact of sociodemographic factors. *The Clinical Neuropsychologist*, *35*(5), 988–1008. https://doi.org/10.1080/13854046.2020.1837247
53. Morris, S. S., Timmons, A., & Musser, E. D. (2023). An individualized, Data-Driven biological approach to Attention-Deficit/Hyperactivity disorder (ADHD) heterogeneity. *Research on Child and Adolescent Psychopathology*, *51*(11), 1565–1579. https://doi.org/10.1007/s10802-023-01104-6
54. Musullulu, H. (2025). Evaluating attention deficit and hyperactivity disorder (ADHD): a review of current methods and issues. *Frontiers in Psychology*, *16*, 1466088. https://doi.org/10.3389/fpsyg.2025.1466088
55. Narzisi, A., Barbetti, F., Fabbri-Destro, M., Berloffa, S., Fantozzi, P., Viglione, V., Muccio, R., Valente, E., Accorinti, I., Foti, E., Milone, A., Cardillo, R., & Masi, G. (2026). Cognitive and emotional profiles in children with ASD, ADHD and comorbid presentations: evidence for a distinct clinical phenotype. *Frontiers in Psychiatry, 17*, 1765698. https://doi.org/10.3389/fpsyt.2026.1765698
56. Nazmin, F., Roy, A., Bushra, T., Retina, I. J., Arnab, K. H., & Ferdaus, F. (2024). Exploring the Prevalence and Social Determinants of ADHD and Comorbidities Among Urban School Aged Children in Bangladesh. *Asia Pacific Journal of Medical Innovations*, *1*(2), 61–67. https://doi.org/10.70818/apjmi.2024.v01i02.010
57. Nguyen, V., Montout, C., Mura, T., Purper-Ouakil, D., & Lopez-Castroman, J. (2024). Concordance and validity between versions of the ADHD Conners scale for parents. *L'Encéphale, 50*(4), 373–379. https://doi.org/10.1016/j.encep.2023.07.006
58. Nigg, J. T. (2026). The evolving ADHD phenotype in the externalizing context. *American Journal of Psychiatry*, *183*(1), 18–30. https://doi.org/10.1176/appi.ajp.20251165
59. Niina, A., Eyre, O., Wootton, R., Stergiakouli, E., Thapar, A., & Riglin, L. (2022). Exploring ADHD Symptoms and Associated Impairment across Development. *Journal of Attention Disorders, 26*(6), 822–830. https://doi.org/10.1177/10870547211025612
60. Öztekin, I., Finlayson, M. A., Graziano, P. A., & Dick, A. S. (2021). Is there any incremental benefit to conducting neuroimaging and neurocognitive assessments in the diagnosis of ADHD in young children? A machine learning investigation. *Developmental Cognitive Neuroscience, 49*, 100966. https://doi.org/10.1016/j.dcn.2021.100966

61. Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., & Rohde, L. A. (2014). ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *International Journal of Epidemiology, 43*(2), 434–442. https://doi.org/10.1093/ije/dyt261
62. Portellano Pérez, J. A., & Martínez Arias, R. (2020). *TFV: Test de Fluidez Verbal: Para la evaluación del lenguaje y de las funciones ejecutivas*. TEA Ediciones.
63. Python Software Foundation. (2026). *Python (Version 3.12.13)* [Software]. https://www.python.org/
64. Qin, H., Zhang, L., Wang, J., Yan, W., Wang, X., Qu, X., Peng, N., & Wang, L. (2025). Interpretable machine learning approaches for children's ADHD detection using clinical assessment data: An online web application deployment. *BMC Psychiatry, 25*, 139. https://doi.org/10.1186/s12888-025-06573-1
65. Quintero López, C., Gil Vera, V. D., Cerpa Bernal, R. M., & Herrera Martínez, M. (2024). Prediction of ADHD with machine learning: A systematic literature review. *Revista Argentina de Ciencias del Comportamiento, 16*(3), 14–32. https://doi.org/10.32348/1852.4206.v16.n3.42221
66. Quintero-López, C., Gil-Vera, V. D., Landinez-Martínez, D. A., Vargas-Gaviria, J. P., & Gómez-Muñoz, N. (2023). Predictive Neurocognitive Model of Attention Deficit Hyperactivity Disorder Diagnosis. *Mediterranean Journal of Clinical Psychology, 11*(1). https://doi.org/10.13129/2282-1619/mjcp-3606
67. Quintero-López, C., Gil-Vera, V. D., Valencia Ochoa, J., & Herrera Velez, A. (2024). The Effect of Working Memory on Social Skills in Male Children with ADHD: A Structural Equation Model. *Mediterranean Journal of Clinical Psychology,* 12(2). https://doi.org/10.13129/2282-1619/mjcp-4130
68. Qureshi, M. N. I., Oh, J., Min, B., Jo, H. J., & Lee, B. (2017). Multi-modal, multi-measure and multi-class discrimination of ADHD with hierarchical feature extraction and extreme learning machine using structural and functional brain MRI. *Frontiers in Human Neuroscience, 11*, 157. https://doi.org/10.3389/fnhum.2017.00157
69. Rajput, D., Wang, W. J., & Chen, C. C. (2023). Evaluation of a decided sample size in machine learning applications. *BMC bioinformatics*, *24*(1), 48. https://doi.org/10.1186/s12859-023-05156-9
70. Rodríguez-Prieto, P., Giral-Oliveros, N. A., Simpson, I. C., & Ibáñez-Alfonso, J. A. (2024). Cognitive stimulation in socioeconomically disadvantaged children with neurodevelopmental disorders: a case series. *Frontiers in Psychology*, *15*, 1365697. https://doi.org/10.3389/fpsyg.2024.1365697
71. Rosenthal, E. A., Broos, H. C., Timpano, K. R., & Johnson, S. L. (2024). Does emotion-related impulsivity relate to specific ADHD symptom dimensions and do the effects generalize across comorbid internalizing and externalizing syndromes? *Journal of Attention Disorders, 28*(2), 178–188. https://doi.org/10.1177/10870547231210283
72. Salari, N., Ghasemi, H., Abdoli, N., Rahmani, A., Shiri, M. H., Hashemian, A. H., Akbari, H., & Mohammadi, M. (2023). The global prevalence of ADHD in children and adolescents: A systematic review and meta-analysis. *Italian Journal of Pediatrics, 49*(1), 48. https://doi.org/10.1186/s13052-023-01456-1
73. Sanchis, J., García-Ponsoda, S., Teruel, M. A., Trujillo, J., & Song, I. Y. (2024). A novel approach to identify the brain regions that best classify ADHD by means of EEG and deep learning. *Heliyon*, *10*(4), e26028. https://doi.org/10.1016/j.heliyon.2024.e26028

74. Santos, C. S., & Amorim-Lopes, M. (2025). Externally validated and clinically useful machine learning algorithms to support patient-related decision-making in oncology: a scoping review. *BMC Medical Research Methodology*, *25*(1), 45. https://doi.org/10.1186/s12874-025-02463-y
75. Schmengler, H., Peeters, M., Stevens, G. W., Hartman, C. A., Oldehinkel, A. J., & Vollebergh, W. A. (2023). ADHD symptoms and educational level in adolescents: The role of the family, teachers and peers. *Research on Child and Adolescent Psychopathology*, *51*(7), 1051–1066. https://doi.org/10.1007/s10802-023-01047-y
76. Schmidt, E. M., Rothenberg, W. A., Davidson, B. C., Barnett, M., Jent, J., Cadenas, H., Fernandez, C., & Davis, E. (2023). Psychometric properties of the behavior assessment system for children student observation system (BASC-3 SOS) with young children in special education. *Journal of Behavioral Education, 32*(3), 543–564. https://doi.org/10.1007/s10864-021-09458-x
77. Sethurao, A., Amit, H. S., Vaishnavi, M. V., Balachandra, A., & Manoj Kumar, M. V. (2021). *Classification of attention deficit hyperactivity disorder using machine learning*. 2021 International Conference on Forensics, Analytics, Big Data and Security (FABS). https://doi.org/10.1109/FABS52071.2021.9702614
78. Slobodin, O., Yahav, I., & Berger, I. (2020). A Machine-Based Prediction Model of ADHD Using CPT Data. *Frontiers in Human Neuroscience, 14*, 560021. https://doi.org/10.3389/fnhum.2020.560021
79. Song, C., Jiang, Z. Q., Liu, D., & Wu, L. L. (2022). Application and research progress of machine learning in the diagnosis and treatment of neurodevelopmental disorders in children. *Frontiers in Psychiatry*, *13*, 960672. https://doi.org/10.3389/fpsyt.2022.960672
80. Sonuga-Barke, E. J. S., Becker, S. P., Bölte, S., Castellanos, F. X., Franke, B., Newcorn, J. H., Nigg, J. T., Rohde, L. A., & Simonoff, E. (2023). Annual research review: Perspectives on progress in ADHD science from characterization to cause. *Journal of Child Psychology and Psychiatry, 64*(4), 506–532. https://doi.org/10.1111/jcpp.13696
81. Stratton, S. J. (2023). Population sampling: Probability and non-probability techniques. *Prehospital and Disaster Medicine*, *38*(2), 147-148. https://doi.org/10.1017/S1049023X23000304
82. Tian, L., Zheng, H., Zhang, K., Qiu, J., Song, X., Li, S., Zeng, Z., Ran, B., Deng, X., & Cai, J. (2024). Structural or/and functional MRI-based machine learning techniques for attention-deficit/hyperactivity disorder diagnosis: A systematic review and meta-analysis. *Journal of Affective Disorders, 355*, 459–469. https://doi.org/10.1016/j.jad.2024.03.111
83. Watkins, M. W., & Canivez, G. L. (2021). Assessing the psychometric utility of IQ scores: A tutorial using the Wechsler Intelligence Scale for Children–Fifth Edition. *School Psychology Review, 50*(4), 577–596. https://doi.org/10.1080/2372966X.2020.1816804
84. Watkins, M. W., Canivez, G. L., Dombrowski, S. C., McGill, R. J., Pritchard, A. E., Holingue, C. B., & Jacobson, L. A. (2022). Long-term stability of Wechsler Intelligence Scale for Children–fifth edition scores in a clinical sample. Applied Neuropsychology: *Child, 11*(3), 422–428. https://doi.org/10.1080/21622965.2021.1875827
85. Wechsler, D. (2015). *Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V)*. Pearson.

86. Willig, T. N., Dajon, M., Assathiany, R., Brun, L., Fourneret, P., Massé, M., Monge, A., Piollet, A., Thiollier, M. C., Cortese, S., & Purper-Ouakil, D. (2024). Healthcare pathways and practitioners' knowledge about ADHD in children. *L'Encéphale, 50*(4), 363–372. https://doi.org/10.1016/j.encep.2023.07.005

87. Wu, Z. M., Wang, P., Liu, L., Liu, J., Cao, X. L., Sun, L., Cao, Q. J., Yang, L., Wang, Y. F., & Yang, B.-R. (2022). ADHD-inattentive versus ADHD-combined subtypes: A severity continuum or two distinct entities? A comprehensive analysis of clinical, cognitive and neuroimaging data. *Journal of Psychiatric Research, 149*, 28–36. https://doi.org/10.1016/j.jpsychires.2022.02.012

88. Yasumura, A., Omori, M., Fukuda, A., Takahashi, J., Yasumura, Y., Nakagawa, E., Koike, T., Yamashita, Y., Miyajima, T., Koeda, T., Aihara, M., Tachimori, H., & Inagaki, M. (2020). Applied Machine Learning Method to Predict Children with ADHD Using Prefrontal Cortex Activity: A Multicenter Study in Japan. *Journal of Attention Disorders, 24*(14), 2012–2020. https://doi.org/10.1177/1087054717740632

89. Yen, C., Lin, C. L., & Chiang, M. C. (2023). Exploring the frontiers of neuroimaging: a review of recent advances in understanding brain functioning and disorders. *Life*, *13*(7), 1472. https://doi.org/10.3390/life13071472

90. Yue, X., Liu, L., Chen, W., Preece, D. A., Liu, Q., Li, H., Wang, Y., & Qian, Q. (2022). Affective-cognitive-behavioral heterogeneity of Attention-Deficit/Hyperactivity Disorder (ADHD): Emotional dysregulation as a sentinel symptom differentiating "ADHD-simplex" and "ADHD-complex" syndromes? *Journal of Affective Disorders, 307,* 133–141. https://doi.org/10.1016/j.jad.2022.03.065

91. Zaheer, A., & Akhtar, A. (2025). Artificial intelligence as a support to diagnose ADHD: an insight of unorthodox approaches: a scoping review. *Child Neuropsychology*, *31*(8), 1324–1358. https://doi.org/10.1080/09297049.2025.2468411

DOI: 10.13129/2282-1619/mjcp-5422