

Use of Machine Learning models to differentiate neurodevelopment conditions through digitally collected questionnaires: a retrospective study

Silvia Grazioli, Alessandro Crippa, Noemi Buo, Silvia Busti Ceccarelli, Massimo Molteni, Maria Nobile, Antonio Salandi, Sara Trabattoni, Gabriele Caselli, Paola Colombo

Submitted to: JMIR Formative Research
on: November 15, 2023

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 28

 Figures 29

 Figure 1..... 30

 Figure 2..... 31

 Figure 3..... 32

 Figure 4..... 33

Use of Machine Learning models to differentiate neurodevelopment conditions through digitally collected questionnaires: a retrospective study

Silvia Grazioli^{1,2,3}; Alessandro Crippa¹; Noemi Buo¹; Silvia Busti Ceccarelli¹; Massimo Molteni¹; Maria Nobile¹; Antonio Salandi¹; Sara Trabattoni¹; Gabriele Caselli^{2,3}; Paola Colombo¹

¹Child Psychopathology Unit, Scientific Institute IRCCS Eugenio Medea Bosisio Parini, Lecco IT

²Department of Psychology, Sigmund Freud University Milan IT

³Studi Cognitivi, Cognitive Psychotherapy School and Research Centre Milan IT

Corresponding Author:

Alessandro Crippa

Child Psychopathology Unit, Scientific Institute IRCCS Eugenio Medea

Via Don Luigi Monza, 20

Bosisio Parini, Lecco

IT

Abstract

Background: Child and adolescent psychopathology diagnosis involves a multifaceted approach, which integrates clinical observations, behavioral assessments, medical history, cognitive testing, and familial context information. Digital technologies, including online platforms for administering caregiver-rated questionnaires, are increasingly utilized in this field, especially during the first screening phase. With the rise of digital platforms for data collection, advanced psychopathology classification methods such as supervised machine learning (ML) have gained prominence in both research and clinical settings. This shift, recently referred to as psycho-informatics, has been facilitated by the gradual incorporation of computational devices into clinical workflows. However, an integrated telemedicine and ML approach is yet to be developed.

Objective: This study aims to develop ML models for the classification of Attention-Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) diagnoses using online parent-reported socio-anamnestic data, with a primary focus on accuracy over interpretability. The objective is therefore to address the need for an innovative computational psychometrics framework that harnesses the potential of digital platforms and ML to improve the assessment of neurodevelopmental conditions.

Methods: In this retrospective, single-center observational study, socio-anamnestic data were collected from 1688 children and adolescents referred for suspected neurodevelopmental conditions. The data included information on sociodemographic, clinical, environmental, and developmental factors, collected remotely through the first Italian online screening tool for neurodevelopmental disorders, the Medea Information and Clinical Assessment On-Line (MedicalBIT). Random Forest (RF) and Logistic Regression (LR) models were developed and evaluated using classification accuracy, sensitivity, specificity, and importance of independent variables.

Results: RF models demonstrated robust accuracy, achieving 84% (95% CI: 82%-85%, $P < .001$) for ADHD and 86% (95% CI: 84%-87%, $P < .001$) for ASD classifications, with high sensitivities of 93% for ADHD and of 95% for ASD. In contrast, the LR models exhibited both lower accuracy (61%, 95% CI: 57%-64%, $P < .001$ for ADHD, and 63%, 95% CI: 60%-67%, $P < .001$ for ASD) and sensitivities (62% for ADHD and 68% for ASD). Intriguingly, the independent variables considered for classification differed in importance between the two models, reflecting the distinct approaches and capabilities of RF and LR.

Conclusions: The present study provides evidence for the potential of an ML approach, especially when built on RF models, in enhancing the diagnostic process of child and adolescent psychopathology. While interpretability remains crucial, the developed approach might therefore provide valuable screening tools for clinicians, highlighting the significance of embedding computational techniques in the diagnostic process. This paves the way for the integration of AI-driven tools in clinical practice to pursue the long-term goal of reducing the workload on clinicians.

(JMIR Preprints 15/11/2023:54577)

DOI: <https://doi.org/10.2196/preprints.54577>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>



Original Manuscript

Use of Machine Learning models to differentiate neurodevelopment conditions through digitally collected questionnaires: a retrospective study

Authors: Silvia Grazioli^{1,2}, Alessandro Crippa^{1,*}, Noemi Buo¹, Silvia Busti Ceccarelli¹, Massimo Molteni¹, Maria Nobile¹, Antonio Salandi¹, Sara Trabattoni¹, Gabriele Caselli², Paola Colombo¹

¹ Child Psychopathology Unit, Scientific Institute, IRCCS Eugenio Medea, Lecco, Italy

² Department of Psychology, Sigmund Freud University, Milan, Italy

* Corresponding author: alessandro.crippa@lanostrafamiglia.it

Abstract

Background: Child and adolescent psychopathology diagnosis involves a multifaceted approach, integrating clinical observations, behavioral assessments, medical history, cognitive testing, and familial context information. Digital technologies, especially online platforms for administering caregiver-rated questionnaires, are increasingly utilized in this field, particularly during the screening phase. The ascent of digital platforms for data collection has propelled advanced psychopathology classification methods like supervised machine learning (ML) into the forefront of both research and clinical environments. This shift, recently called psycho-informatics, has been facilitated by gradually incorporating computational devices into clinical workflows. However, an actual integration between telemedicine and the ML approach has yet to be fulfilled.

Objective: Under these premises, exploring the potential of ML applications for analyzing digitally collected data may have significant implications for supporting the clinical practice of diagnosing early psychopathology. The purpose of the present study was, therefore, to exploit ML models for the classification of Attention-Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) using online parent-reported socio-anamnestic data, aiming at obtaining accurate predictive models for new help-seeking families.

Methods: In this retrospective, single-center observational study, socio-anamnestic data were collected from 1688 children and adolescents referred for suspected neurodevelopmental conditions. The data included sociodemographic, clinical, environmental, and developmental factors, collected remotely through the first Italian online screening tool for neurodevelopmental disorders, the Medea Information and Clinical Assessment On-Line (MedicalBIT). Random Forest (RF), Decision Tree (DT), and Logistic Regression (LR) models were developed and evaluated using classification accuracy, sensitivity, specificity, and importance of independent variables.

Results: The RF model demonstrated robust accuracy, achieving 84% (95% CI: 82%-85%, $P < .001$) for ADHD and 86% (95% CI: 84%-87%, $P < .001$) for ASD classifications. Sensitivities were also high, with 93% for ADHD and 95% for ASD. In contrast, the DT and LR models exhibited lower accuracy (DT: 74%, 95% CI: 71%-77%, $P < .001$ for ADHD, and 79%, 95% CI: 77%-82%, $P < .001$ for ASD. LR: 61%, 95% CI: 57%-64%, $P < .001$ for ADHD, and 63%, 95% CI: 60%-67%, $P < .001$ for ASD) and sensitivities (DT: 82% for ADHD and 88% for ASD. LR: 62% for ADHD and 68% for ASD). The independent variables considered for classification differed in importance between the two models, reflecting the distinct characteristics of the three ML approaches.

Conclusions: This study highlights the potential of ML models, particularly RF, in enhancing the diagnostic process of child and adolescent psychopathology. Altogether, the present findings underscore the significance of leveraging digital platforms and computational techniques in the diagnostic process. While interpretability remains crucial, the

developed approach might provide valuable screening tools for clinicians, highlighting the significance of embedding computational techniques in the diagnostic process.

Keywords: Digital-aided clinical assessment; Machine learning; Random Forest; Logistic regression; Computational psychometrics; Telemedicine; Neurodevelopmental conditions; Parent-report questionnaires; Attention-deficit/hyperactivity disorder; Autism spectrum disorder

Introduction

Child and adolescent psychopathology entail a complex diagnostic process, which includes clinical observations, behavioral assessments, medical history evaluations, cognitive testing, and familial contexts. This comprehensive approach aligns with the recommendations outlined in the National Institute for Health and Care Excellence Guidelines [1], emphasizing the significance of a multifaceted understanding of these conditions.

In contemporary diagnostic practice, the adoption of digital technologies has gained remarkable prominence, particularly in the digital administration of caregiver-rated questionnaires focused on capturing neurodevelopmental and behavioral symptoms [2,3]. This shift toward digitalization (often referred to as psycho-informatics) is fueled by the gradual incorporation of computational devices into clinical workflows. These tools enable assessments encompassing various aspects, including psychological testing and psychometry [4]. Furthermore, the onset of the COVID-19 pandemic has further supported adopting digital tools for telehealth support, particularly within the mental health sector [5]. Digitization also includes aspects that cut across diagnoses, such as socio-anamnestic information relevant to neurodevelopmental conditions [6].

Over the past two decades, there has been a significant rise in the application of advanced classification methods, such as supervised machine learning (ML), to enhance diagnostic research in the behavioral sciences. [1,3,7–16]. Most of these studies have applied ML-based models to different types of data (e.g., home videos, child/adult diagnostic testing), reaching excellent classification accuracies [12–17]. Supervised ML involves the development of algorithms that learn from previous experiences to simulate human cognitive processes. ML techniques have been also used to analyze data collected through digital platforms.

In this context, recent work from our group has effectively identified children who received a clinical diagnosis of ADHD with an accuracy of up to 82% based on parent- and teacher-reported child behavior filled online, using a DT model simple supervised ML approach, the decision tree [3]. Ben-Sasson and Yom-Tov [15] also adopted DT to address a possible early identification of ASD through online queries posed by parents suspecting that their children could have that diagnosis; the classifier achieved an area under the curve of 0.82, indicating good predictive accuracy for identifying ASD risk based on parental narratives. Duda and colleagues [16] recruited through crowdsourcing a large sample of parents of children with only ASD or only ADHD to test whether a supervised ML algorithm could differentiate the two diagnoses. Using Social Responsiveness Scale items, the linear discriminant analysis and elastic net classifiers achieved an area under the curve of 0.81.

Nevertheless, previous research conducted on adults with ADHD identified accurate classification models based on questionnaire scores. Trognon and Richard developed a psychometric screening scale for the identification of adult ADHD based on DSM-5 diagnostic criteria. They tested an XGBoost classifier to obtain a predictive model for subjects with ADHD compared to controls [10].

The questionnaire scores-based classifier reached an accuracy of .98, based on a training set of 154 subjects. Lastly, Christiansen and colleagues tested supervised models on the Conners' Adult ADHD Rating Scale for differentiating between subjects with ADHD, obesity, problematic gambling, and a control group [11]. The models reached a predictive accuracy of .82 (support vector machine), 0.80 (LightGBM), and 0.79 (logistic regression).

When testing ML classification models, a relevant focus should be placed on interpretability and accuracy [17]. The first deals with the need for researchers and clinicians to fully understand the relationship between inputs and outputs, which provides an automated decision-making process performed by an ML model. In this context, a model is interpretable when it provides all the critical information about between-variable relationships gained through the learning process. Conversely, accuracy describes how well the ML model performs in providing correct class membership predictions, and it increases with model complexity (hence, with less interpretability) [17]. One or the other must be favored when setting a trade-off between classification accuracy and model interpretability [17]. Traditionally, psychology has primarily focused on explanatory (interpretable) modeling, seeking to understand the causal underpinnings of behavior. However, this emphasis on explanation has often led to models that lack meaningful predictive capacity, raising questions about the robustness and generalizability of psychological research [18].

Within this framework, traditional statistical models like logistic regression (LR) have been widely used for clinical classification purposes (i.e., for predicting the probability that an observation belongs to one of two possible classes) because of their open interpretability [17]. Nonetheless, the adoption of less transparent ML methods, such as decision trees (DT) and random forests (RF), has garnered significant attention due to their ability to capture more complex patterns within data. DT and RF often outperform LR, especially when dealing with categorical predictors [19]. Moreover, recent technological advancements enable the enhancement of interpretability in "black box" models using the Shapley additive explanations (SHAP) framework [12]. The SHAP analysis evaluates the influence of classification features in augmenting the likelihood of accurate predictions, allowing researchers to gain more insight into ranking factors that make a diagnosis probable [12].

Given these premises, the hypothesis examined in the present study is whether an innovative computational psychometrics framework could exploit the potential of the ML approach to digitally collected data to support the clinical assessment of neurodevelopmental conditions such as attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorders (ASD). To this end, we developed classification ML models to identify either children with ADHD or with ASD using parent-report socio-anamnestic questionnaires collected through the first Italian online, web-based, comprehensive screening tool for neurodevelopmental disorders and emotional and behavioral problems, the Medea Information and Clinical Assessment On-Line (MedicalBIT) [6].

Methods

Recruitment

In the present retrospective, single-center observational study, we considered socio-anamnestic data from a sample of children and adolescents referred for suspected neurodevelopmental conditions at the Scientific Institute "IRCCS Eugenio Medea"—Associazione La Nostra Famiglia in Bosisio Parini (Lecco, Italy)—between October 2018 and May 2022. A workflow of the diagnostic procedure is shown in Figure 1.

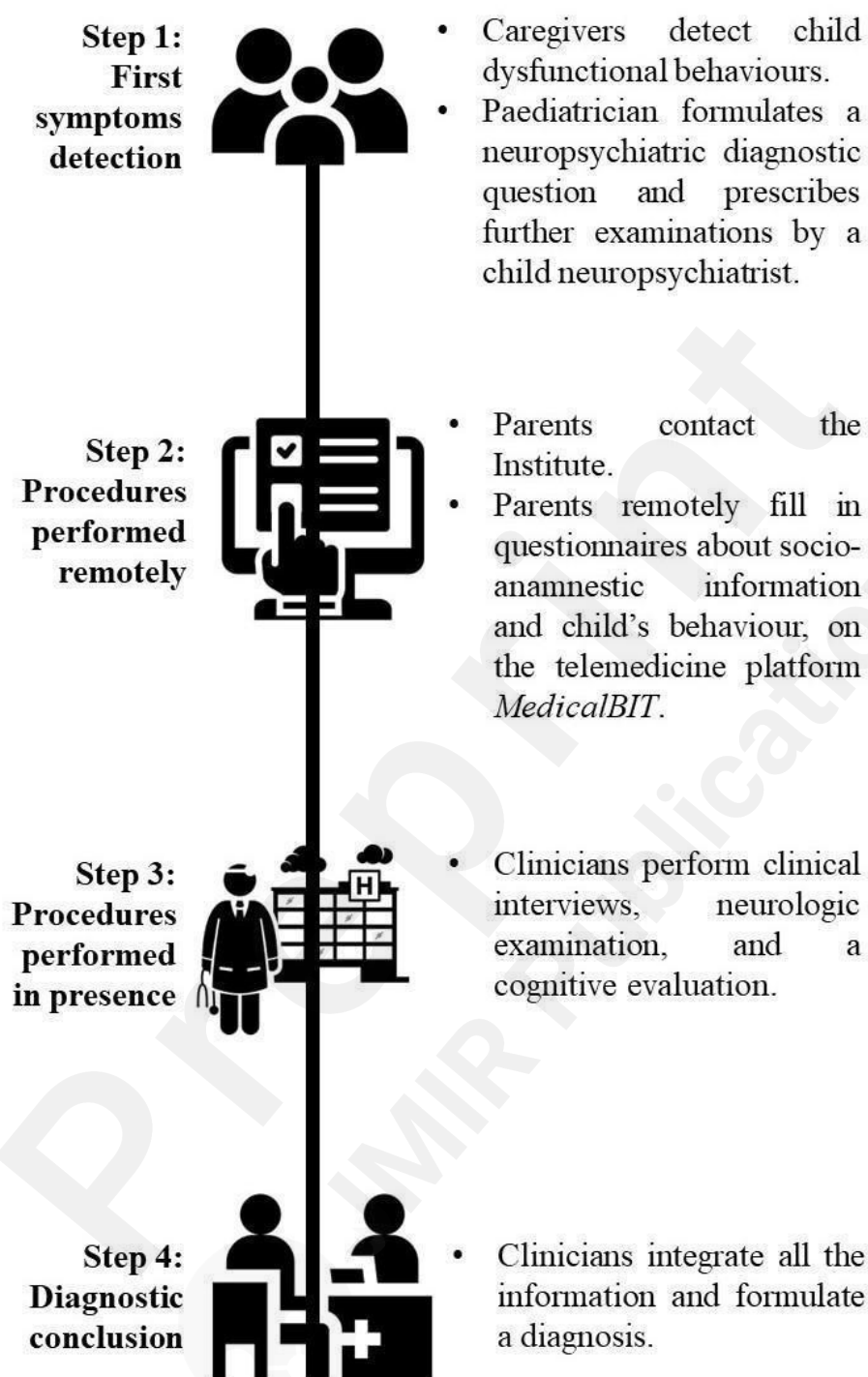


Figure 1. Diagnostic process scheme: the four steps from the first symptoms detection to the diagnostic conclusion.

Sample description

Participants included 1688 children and adolescents (35% females) living in Northern Italy, aged 2 to 16 years (mean age: 8 ± 3 SD), and who were referred for suspected neurodevelopmental and psychopathological diagnosis. All participant-related data was obtained remotely by parents using the MedicalBIT platform, except for the attending clinician's categorical diagnosis input into the

platform. At the access to our Institute, participants underwent a complete neuropsychiatric evaluation and received one or more clinical diagnoses following the DSM-5 criteria [20]. In some cases, symptom presentation fell below the diagnostic threshold, and the children did not receive a categorical diagnosis. Possible diagnoses comprised ADHD, ASD, anxiety disorders, communication disorders, developmental coordination disorders, emotional and behavioral disorders, intellectual disability, mood disorders, obsessive-compulsive disorder, sleep disorders, and specific learning disorders. Each diagnostic label could represent a single condition or a comorbid presentation (Figure 2). The prevalence of ADHD and ASD diagnoses, regardless of single or comorbid presentations, was equal, accounting for 16% of the sample each.

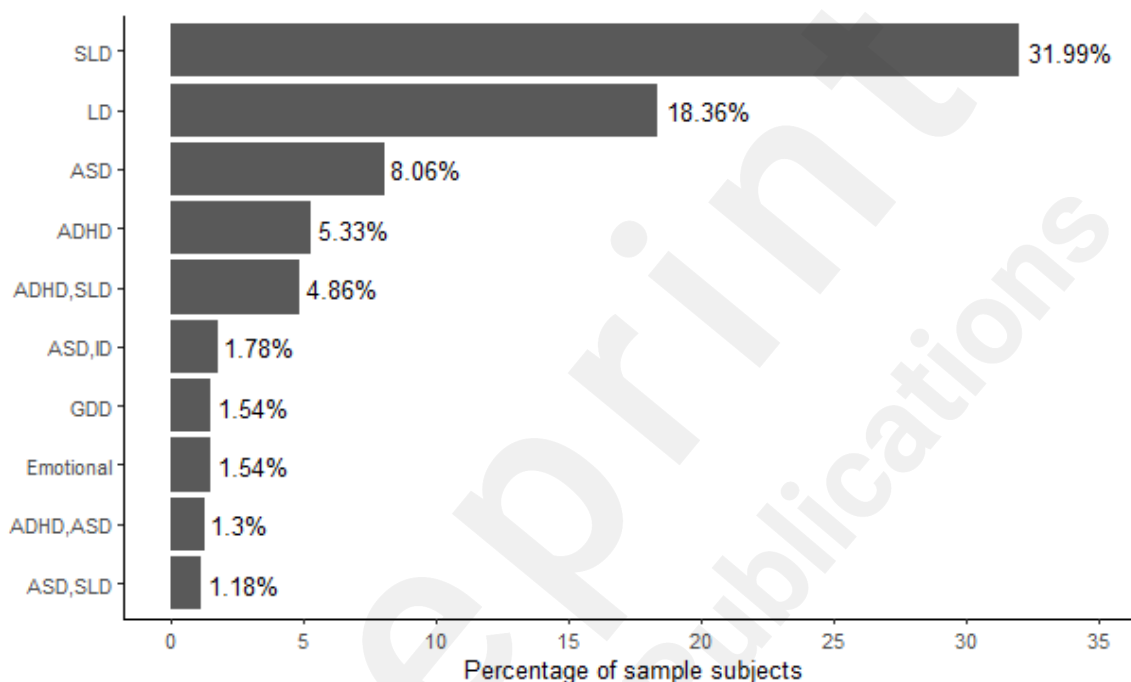


Figure 2. Most frequent configurations of diagnoses across the sample. The figure depicts the configuration of diagnostic categories and comorbidities presented by at least 1% of the participants. Note: ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; GDD: global developmental delay; ID: intellectual disability; LD: language disorder; SLD: specific learning disorder.

Independent variables

The independent variables were derived from a socio-anamnestic questionnaire tapping clinical, biological, environmental, and developmental information. These variables were selected from standard clinical practice and were completed by the participants' parents before accessing the Institute using the MedicalBIT online platform [6]. It is important to note that the questionnaire items were coded such that higher values corresponded to higher risk levels. Consequently, the item values were aggregated to obtain variables that quantified risk levels in *positive familiarity*, *pre- and perinatal risk*, *developmental concerns*, *early health problems*, and *stressful life events* (Table 1).

Table 1. List of independent and dependent variables. The two dependent variables were considered in separate statistical models.

Area	Variable name	Variable type (values or range)
<i>Socio demographic</i>	- Sex	Dichotomous (male; female)
<i>Positive familiarity</i>	- Neurologic or psychopathologic anamnesis in the family - Presence of family late talkers	Dichotomous (1 = presence; 0 = absence)
	- Preterm birth - Pregnancy problems - Fertility treatment - Smoke during pregnancy - mother - Smoke during pregnancy - father - Medications during the pregnancy - Risk of abortion - Birth problems - Problems after birth	Dichotomous (1 = presence; 0 = absence)
<i>Pre- and perinatal risk</i>	- Birth type	Categorical (1=natural; 2=caesarian; 3=induced; 4=problematic)
	- Birth weight	Categorical (0=normal; 1=moderately high or low; 2=extremely low)
	- APGAR at 1 minute - APGAR at 5 minutes	Categorical (1=9 or 10; 2=7 or 8; 3= 5 or 6; 4=1 to 4)
	- Breastfeeding	Dichotomous (0 = breastfeeding; 1 = formula)
	- Eating habits	Dichotomous (0=normal; 1=selective eating)
	- Sleeping problems	Dichotomous (0=no; 1=yes)
<i>Developmental concerns</i>	- Age of crawling	Categorical (1=6-9 months; 2=9-12 months, 3=>1 year; 4=never)
	- Walking	Categorical (1= <12 months; 2=12-18 months, 3=18-24 months; 4= >2 years)
	- First words at 18	Dichotomous (0=yes; 1=no)

		months - Word combinations at 36 months	
		- Any surgery - Any audiometry screening - Side effects from vaccines	Dichotomous (0=no; 1=yes)
<i>Early problems</i>	<i>health</i>	- Severe early health problems - Early hospitalization - Number of early ear infections	Categorical (0=no; 1=rarely; 2=sometimes; 3=often) Categorical (0=never; 1=1-2 days, 2=3-4 days, 3=1 week or more)
		- Any new family member - House moving - Hospitalization of any family member - Death of any family member - A parent lost work - Severe financial problems - Parents separation	Categorical (1=0-2; 2=3-5; 3=6-8; 4=9 or more)
		- ADHD diagnosis - ASD diagnosis	

Outcome variables

Clinicians' diagnostic conclusion regarding the presence or absence of ADHD or ASD was considered as a dichotomous outcome in separate classification models: i) the 'ADHD' group comprised participants with a clinical diagnosis of ADHD and possible comorbid conditions; the 'non-ADHD' group comprised participants without a clinical diagnosis of ADHD, i.e., subjects who received other neuropsychiatric diagnoses or no categorical diagnosis; ii) the 'ASD' group comprised participants with a clinical diagnosis of ASD, and possible comorbid conditions; the 'non-ASD' group comprised participants without a clinical diagnosis of ASD, i.e., subjects who received other neuropsychiatric diagnoses or no categorical diagnosis (Figure 3).

Statistical Analysis

Preliminary data handling

Data handling and statistical analyses were done through R software version 4.1.2 [21]. Data cleaning procedures were performed as described in De Jonge & Van Der Loo [22]. Missing data were imputed using the 10 nearest neighbors averaging [23].

Classification models

Separate classification models were obtained, addressing the following clinical and research questions: i) “Should a new help-seeking child be diagnosed with ADHD, considering the parent-report socio-anamnestic information?”; ii) “Should a new help-seeking child be diagnosed with ASD, considering the parent-report socio-anamnestic information?”. The three models are depicted in Figure 3. The latter highlights that heightened interpretability in models (specifically, decision tree and logistic regression) is associated with diminished predictive accuracy, which is higher for random forest models.

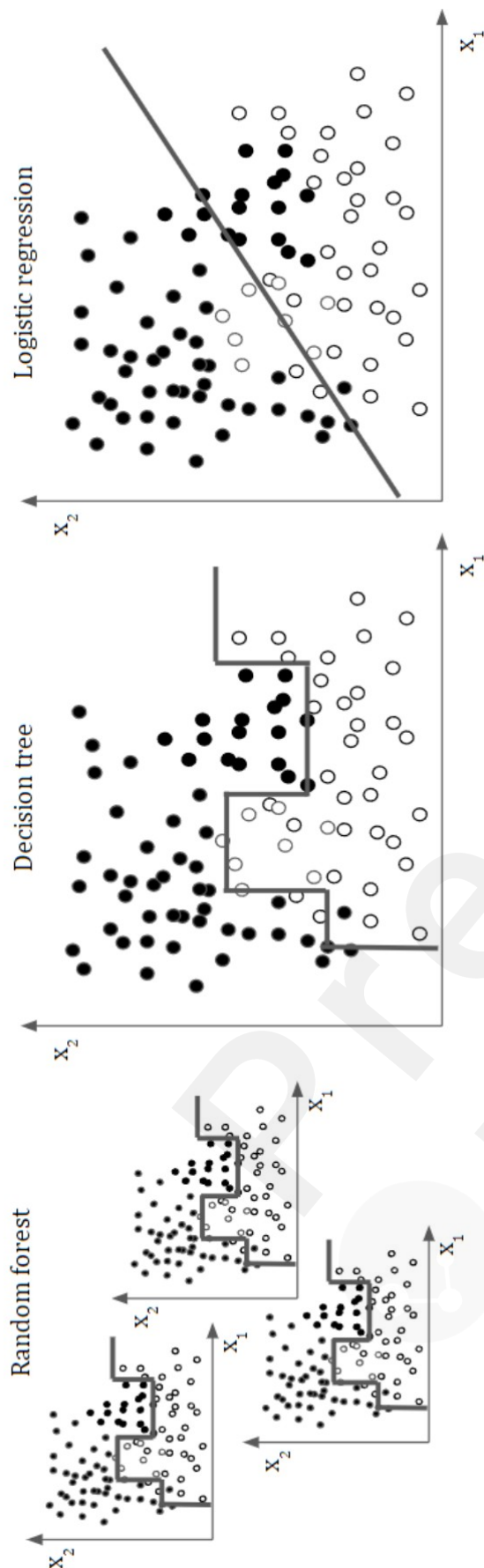


Figure 3. Decision boundaries of random forest (RF), decision tree (DT), and logistic regression (LR) in a binary classification example. The decision boundaries represent the separation between the two classes in a feature space. The decision boundary of an RF model is a nonlinear combination

of the individual decision boundaries from each decision tree. A DT decision boundary is a nonlinear combination of axis-parallel lines. The linear equation of a hyperplane in the feature space determines an LR decision boundary.

1) Random forest models. RF models were performed through the ‘randomForest’ toolbox [24], as previously done [3]. RF is an ensemble learning technique that generates many decision trees (DTs) and aggregates the results. To prevent overfitting, two layers of randomness are added in the procedure through bagging: 1) a bootstrap sample of the dataset is considered in each tree (the data that are not considered in the bootstrap sample are called out of bag, OOB); 2) a subset of m independent variables are selected at each tree node. New data categories are predicted by aggregating all predictions performed by the trees, i.e., choosing the majority of the voted categories [24]. In the tuning phase of the model selection, a leave-one-out cross-validation (LOOCV) approach was applied [25]. Lastly, a SHAP analysis was performed to gain insights into the interpretability of the model [26]. SHAP values are computed by comparing the model’s predictions with and without a particular feature, and this process is repeated iteratively for each feature and sample in the dataset. The magnitude of these values reflects the strength of the effect [26].

2) Decision trees. After conducting RF analyses, DT models were computed. The DT, characterized by a flowchart-like structure, is constructed by considering the entire dataset positioned at the top of a “root” node. At each decision point, observations meeting the specified splitting condition are allocated to the left branch, while those not meeting the condition are directed to the right branch [27]. Information gain is a node impurity measure for selecting attributes and dividing each node, continuing until the terminal node, referred to as the “leaf,” is reached [27]. Lastly, the algorithm assigns the most frequently observed class in each leaf as the classification prediction [27].

3) Logistic regression models. LR models were employed in addition to the DT and RF models. LR is a traditional statistical method widely used for binary classification tasks. It models the probability of a binary outcome (presence/absence of the considered diagnosis) based on one or more predictor variables. In our study, LR was applied using the ‘glm’ function in R.

Testing the classification accuracy

1) Fixed training and test set. To test the classification accuracy of the previously described models, we used 70% of the whole dataset as a training set and the remaining 30% as a test set; the two subsamples did not present overlapping subjects. The classification performances of the selected models were evaluated considering the following information on the test set:

- a) Classification accuracy: percentage of correctly performed classification concerning the total number of instances: $Acc = \frac{TP+TN}{TP+TN+FP+FN}$;
- b) The no information rate (NIR) represents the largest proportion of the observed classes, indicating the accuracy achievable by always predicting the majority class label.
- c) The P-value of $Acc > NIR$: a hypothesis test result to evaluate whether the classification accuracy performed by the algorithm is greater than the rate of the largest class (NIR).
- d) Specificity: percentage of correctly performed negative classification (non-ADHD or non-ASD) concerning the number of subjects without the actual diagnosis: $Sp = \frac{TN}{TN+FP}$;
- e) Sensitivity: percentage of correctly performed positive classification (ADHD or ASD) for the number of subjects with the actual diagnosis: $S = \frac{TP}{TP+FN}$.

2) 5-fold cross-validation. An additional cross-validation step was performed to test the results' robustness. The whole dataset was randomly split into five folds, and the three classification models were performed on each independent fold. The classification performances were calculated on each test set, and the mean performance values were estimated.

Ethical considerations

The study was approved by the Institute's Ethical Review Board (Prot. N. 7/23 - "Comitato Etico IRCCS E. Medea—Sezione Scientifica Associazione La Nostra Famiglia"). The research was conducted following guidelines and regulations depicted in the Declaration of Helsinki. The study data are de-identified; no identification of individual participants in any images of the manuscript is possible. All the participant's parents or legal guardians were informed of the aim of the study. Each subject was free to participate voluntarily and gave their written informed consent to the minor's participation. No monetary compensation was provided for participating in the study.

Results

Preliminary data handling

The maximum percentage of missing data per subject was 36% (3 subjects). Table 2 depicts the sample's demographic characteristics, considering the total sample and stratification by ADHD and ASD diagnosis.

Table 2. Descriptive statistics of demographic variables, considering the whole sample, and stratified by ADHD and ASD diagnosis. The considered descriptive statistics for age are the median (\pm standard deviation). Note: ADHD = Attention-deficit/hyperactivity disorder; ASD = Autism spectrum disorders; f = females; m = males; n = sample size.

Variable	Total sample (n=1688)	ADHD stratification		ASD stratification	
		ADHD (n=269)	non - ADHD (n=1419)	ASD (n=270)	non- ASD (n=1418)
Age	8 (\pm 3)	9 (\pm 3)	8 (\pm 3)	6 (\pm 4)	9 (\pm 3)
Sex	m = 65% f = 35%	m = 80% f = 20%	m = 63% f = 37%	m = 84% f = 16%	m = 62% f = 38%

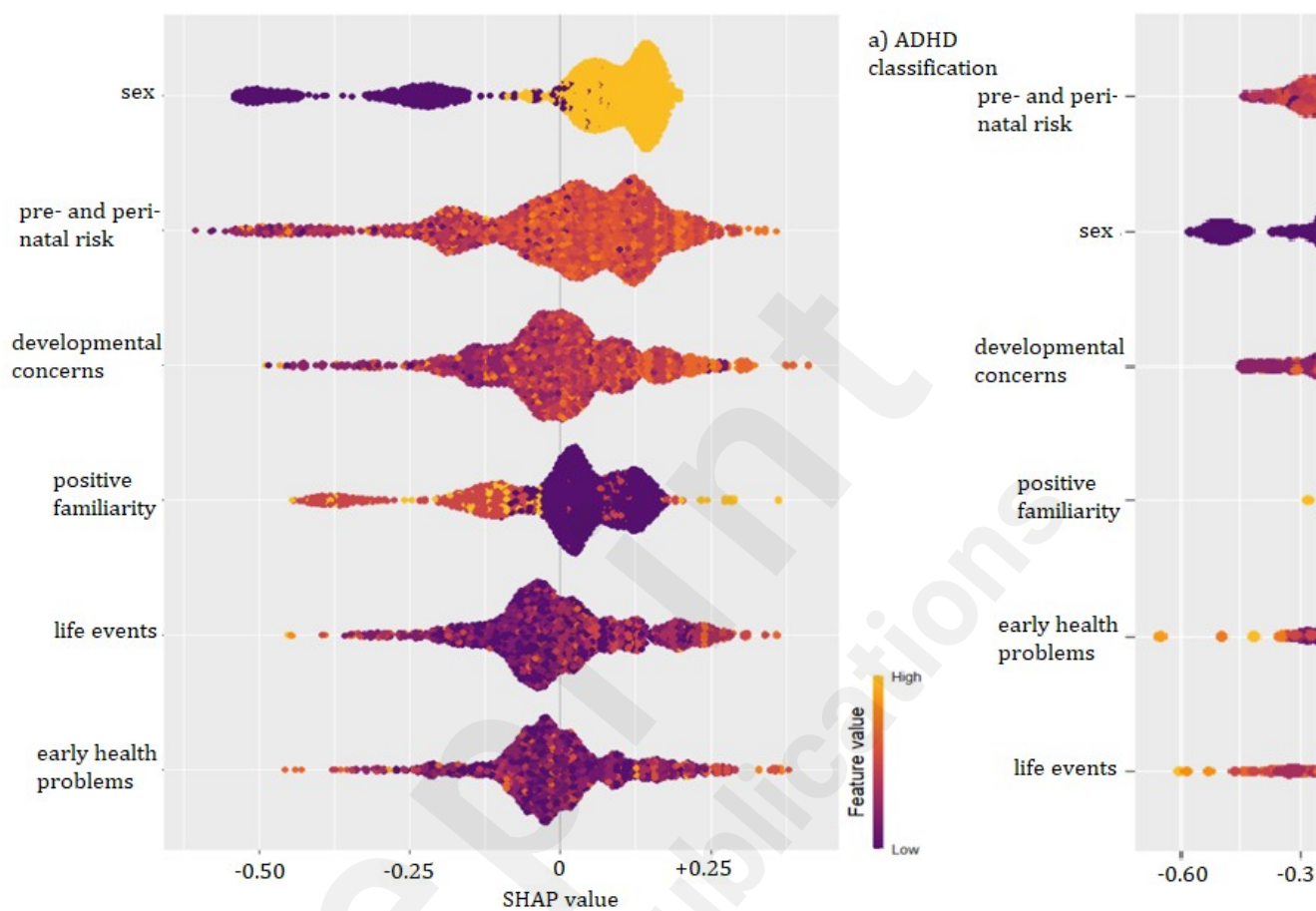
Random forest

Table 3 shows the RF classification models' performances. Figure 4 shows the SHAP values (i.e., the most important independent variables identified by the RF in accurately classifying the diagnoses).

Table 3. Performances of the RF models for classifying ADHD or ASD diagnoses. Note: Acc = accuracy; ADHD = Attention-deficit/hyperactivity disorder; ASD = Autism spectrum disorders; NIR = no information rate; sd = standard deviation; Se = sensitivity; Sp = specificity.

Classification model	Performance on the fixed training and test set	Mean SHAP values	Average performance on the 5-fold cross-validation sets (\pm sd)
ADHD vs. non-ADHD	<ul style="list-style-type: none"> ● Acc: 84% (95% CI: 82%-85%) ● NIR: 50% ● $P < .001$ ● Se: 93% ● Sp: 75% 	<ul style="list-style-type: none"> - Sex: 0.07 - Pre- and perinatal risk: 0.03 - Developmental concerns: -0.02 - Positive familiarity: 0.02 - Stressful life events: -0.01 - Early health problems: -0.01 	<ul style="list-style-type: none"> ● Acc: 68% (\pm 2.5) ● Se: 73% (\pm 2.5) ● Sp: 63% (\pm 2.5)
ASD vs. non-ASD	<ul style="list-style-type: none"> ● Acc: 86% (95% CI: 84%-87%) ● NIR: 50% ● $P < .001$ ● Se: 95% ● Sp: 77% 	<ul style="list-style-type: none"> - Pre- and perinatal risk: 0.08 - Sex: 0.05 - Developmental concerns: -0.04 - Positive familiarity: 0.03 - Early health problems: -0.02 - Stressful life events: -0.01 	<ul style="list-style-type: none"> ● Acc: 69% (\pm 2.2) ● Se: 70% (\pm 2.2) ● Sp: 68% (\pm 2.2)

Figure 4. SHAP values for a) ADHD classification through RF; b) ASD RF classification through RF. SHAP values are computed by comparing the model's predictions with and without a particular feature, and this process is repeated iteratively for each feature and sample in the dataset. The magnitude of these values reflects the strength of the effect.



Decision tree

Table 4 shows the DT model results and performances on the test sets.

Table 4. Performances of the DT models in the test set for classifying ADHD or ASD diagnoses. No Acc = accuracy; ADHD = Attention-deficit/hyperactivity disorder; ASD = Autism spectrum disorders; N = no information rate; sd = standard deviation; Se = sensitivity; Sp = specificity.			
Classificati on model	Performance on the fixed training and test set	Attribute importance to the training set	Average performance on the 5-fold cross- validation sets (± sd)
ADHD vs. non-ADHD	<ul style="list-style-type: none">● Acc: 74% (95% CI: 71%-77%)● NIR: 50%● P < .001● Se: 82%● Sp: 65%	<ul style="list-style-type: none">- Sex: 0.019- Early health problems: 0.008- Stressful life events: 0.007- Pre- and perinatal risk: 0.004- Developmental concerns: 0.001- Positive familiarity: 0.001	<ul style="list-style-type: none">● Acc: 59% (± 3.1)● Se: 57% (± 3.1)● Sp: 61% (± 3.1)
ASD vs. non-ASD	<ul style="list-style-type: none">● Acc: 79% (95% CI: 77%-82%)	<ul style="list-style-type: none">- Sex: 0.036- Pre- and perinatal risk: 0.016	<ul style="list-style-type: none">● Acc: 64% (± 3.1)● Se: 67% (± 3.4)

	<ul style="list-style-type: none">• NIR: 50%• $P < .001$• Se: 88%• Sp: 71%	<ul style="list-style-type: none">- Positive familiarity: 0.013- Early health problems: 0.010- Developmental concerns: 0.001- Stressful life events: 0.001	<ul style="list-style-type: none">• Sp: 61% (± 3.4)
--	---	---	--

Logistic regression

Table 4 shows the LR model results and performances.

Table 4. Performances of the LR models for classifying ADHD or ASD diagnoses. Note: Acc = accuracy; ADHD = Attention-deficit/hyperactivity disorder; ASD = Autism spectrum disorders; NIR = no information rate; sd = standard deviation; Se = sensitivity; Sp = specificity.			
Classification model	Performance on the fixed training and test set	OR coefficients in the training set (P)	Average performance on the 5-fold cross validation sets (\pm sd)
ADHD vs. non-ADHD	<ul style="list-style-type: none"> • Acc: 61% (95% CI: 57%-64%) • NIR: 50% • $P < .001$ • Se: 62% • Sp: 59% 	<ul style="list-style-type: none"> - Sex: 2.14 ($P < .001$) - Stressful life events: 1.19 ($P < .001$) - Early health problems: 1.09 ($P = .003$) - Pre- and perinatal risk: 1.08 ($P < .001$) - Developmental concerns: 1.08 ($P = .016$) - Positive familiarity: 1.03 ($P = .676$) 	<ul style="list-style-type: none"> • Acc: 57% (± 1.9) • Se: 61% (± 1.9) • Sp: 53% (± 1.9)
ASD vs. non-ASD	<ul style="list-style-type: none"> • Acc: 63% (95% CI: 60%-67%) • NIR: 50% • $P < .001$ • Se: 68% • Sp: 59% 	<ul style="list-style-type: none"> - Sex: 3.88 ($P < .001$) - Positive familiarity: 1.93 ($P < .001$) - Developmental concerns: 1.17 ($P < .001$) - Pre- and perinatal risk: 1.11 ($P < .001$) - Early health problems: 1.05 ($P = .134$) - Stressful life events: 0.87 ($P < .001$) 	<ul style="list-style-type: none"> • Acc: 62% (± 2.1) • Se: 67% (± 2.1) • Sp: 58% (± 2.1)

Discussion

Principal results

The primary objective of our study was to develop accurate classification models for the diagnosis of ADHD and ASD within a sample referred for clinical evaluation. To this

end, we employed an ML approach to analyze online parent-report socio-anamnestic questionnaires.

Our ML models reached overall reasonable classifications in the test sets for both ADHD and ASD. The RF models exhibited classification accuracies of 84% for ADHD and 86% for ASD, respectively, with high sensitivities (93% for ADHD and 95% for ASD). On the other hand, the DT and LR models reached lower accuracy rates, with 74% and 61% accuracy for ADHD and 79% and 63% for ASD, respectively. The DT and LR models also demonstrated lower sensitivities (82% and 62% for ADHD and 88% and 68% for ASD).

In the 5-fold experiment, all models showed a decline in predictive accuracy, as could be expected due to smaller sample sizes. Nevertheless, the RF model continued to exhibit greater accuracy than other models. Concerning the different levels of accuracy reached by our three ML models, it is crucial to acknowledge both the advantages and disadvantages of RF, DT, and LR. One of the distinctive features of RF models is that they can effectively capture complex relationships within the data that may elude human interpretation [17]. For this reason, RF models can occasionally be considered difficult to interpret. This characteristic needs adequate consideration in the clinical context because the primary aim is to provide clinicians with an accurate ‘first glance’ tool that supports them in forming initial diagnostic impressions.

Notwithstanding their eventual interpretability, RF models are remarkably effective in distinguishing different classes, thus representing an asset in psychopathology diagnosis. Conversely, as mentioned above, the DT and LR models are also readily interpretable for clinicians less familiar with ML techniques [17]. Therefore, the choice of approach depends on the decisional context and the desired degree of interpretability. In the present study, we preferred greater levels of classification accuracy over the readiness of the classification process. However, a noteworthy option to mitigate the interpretability concern associated with RF models is provided by Shapely additive explanation (SHAP) analysis. By assigning an important value to each feature in the classification model, SHAP analysis directly compares RF and other models regarding their interpretability.

Although slightly different in the achieved performance, the three models identified sex as the strongest predictor for both ADHD (all three models) and ASD (DT and LR models). It is well documented that males are more likely to be diagnosed with both ADHD [28] and ASD [29] than females. Interestingly, SHAP analysis indicated a relatively consistent ranking of features for RF models across the two clinical diagnoses. After sex, which showed by far the highest discriminative ability among the cases, the presence of pre- and perinatal risk and other developmental concerns featured as influential predictors of both ADHD and ASD classes. Not surprisingly, given the significant heritability of the two conditions- having a family member with reported difficulties was also a relevant predictor of the classification.

On the other hand, DT and LR models identified feature rankings that were, except for

sex, significantly different for ADHD and ASD classification. This discrepancy could be due to the underlying assumptions of the different ML methodologies. Whereas LR models assume linear relationships between predictors and outcomes, DT and RF models could exploit non-linear relationships and interactions within the data [19]. Consequently, some degree of variation in predictor ranking is expected, further highlighting the diverse nature of insights gained from different analytical methodologies. Lastly, it should be remembered that it is impossible to conclude the causality and direction of the interrelations among predictors in the ML model.

Comparison with prior work

Our RF model's accuracy was in line with previous ML classification approaches to questionnaire data [10,11,15,16,30] and other data sources [31–36,36–38].

Nevertheless, these classification models outperformed a recent work from our group, where we identified children with ADHD with an accuracy of up to 82% using a decision tree-based supervised machine learning approach [3]. Despite some methodological differences, the higher level of accuracy obtained in the present work underscores the potential of RF models in increasing the precision of computer-aided diagnosis. Altogether, this pattern of findings suggests that the RF model outperformed both the DT and LR models in effectively categorizing neurodevelopmental conditions based on parent-reported socio-anamnestic information, as highlighted by previous studies [28,29,39].

Implications for clinical practice and future research

In the domain of child and adolescent neuropsychiatry, the diagnostic process includes an initial stage where anamnestic, sociodemographic, and behavioral data need to be collected. This data gathering can be remotely performed via online parent reports, as evidenced by previous studies [5,6,40–42]. With this regard, the MedicalBIT platform currently represents the first Italian online screening instrument for child and adolescent neuropsychiatric conditions [6]. As the data are compiled in databases within MedicalBIT, the exploitation of ML models can prompt the classification of the probable diagnostic risk associated with new subjects seeking assistance. The significant predictive value of the models developed in this study might be valuable to support the clinical practice of diagnosing neurodevelopmental conditions.

Limitations

Despite the encouraging findings, this study is not free of limitations. First, our ML models exclusively rely on parent-reported data. Existing literature [43] has previously indicated that the reliability of these reports could be negatively influenced by factors such as the possibility of accessing digital tools, intrinsic comprehension difficulties, or

general parental educational attainment. Second, our sample exclusively included children and adolescents from a geographically restricted region (Northern Italy). The present findings' generalizability to populations from different areas needs cautious consideration. Third, the relatively low occurrence of ADHD-ASD comorbidity in our cohort prevents us from developing classification models tailored for more nuanced diagnostic presentations, such as either ADHD- or ASD-only versus ADHD-ASD comorbid presentation. Therefore, future extensions of this study should consider including broader cohorts of participants to consider this possibility.

Conclusions and future advancements

Within the rapidly evolving context of 'psycho-informatics,' we feel that the present work represents a novel pioneering effort in the realm of computational psychometrics [29]. Through an exploration of remotely collected parent-reported socio-anamnestic data, the present research has revealed promising avenues for enhancing the diagnostic process of neurodevelopmental and psychopathological conditions. Integrating digital platforms for data collection and ML could offer clinicians a dynamic tool supporting their diagnostic decisions. Within the healthcare systems, clinical teams confront a scarcity of personnel, with high emotional and cognitive demands for the actual staff [44]. In this context, the present research represents a preliminary effort to mitigate the clinicians' workload by automating specific tasks (such as data collection and analysis). If proven effective, this approach could leave more time for clinicians to nurture the essential patient-clinician bond, a facet that remains irreplaceable by AI technologies.

Acknowledgments

SG performed statistical analyses and wrote the first draft. SBC, NB, MN, AS, ST, MM, and PC contributed with clinical knowledge regarding neurodevelopmental and psychopathological conditions and telemedicine. AC, GC, and PC contributed with scientific knowledge on health care technology. SBC, NB, and PC contributed to the development of the telemedicine platform through clinical consultation for variable selection. All authors revised and approved the final manuscript.

Data Availability

The dataset analyzed in the present study is available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ASD: autism spectrum disorder
DTs: decision trees
GDD: global developmental delay
ID: intellectual disability
LD: language disorder

LOOCV: leave-one-out cross-validation

LR: logistic regression

ML: machine learning

NIR: no information rate

OOB: out of the bag

RF: random forest

SLD: specific learning disorder

References

1. National Institute for Health and Clinical Excellence (Great Britain). *Attention Deficit Hyperactivity Disorder: Diagnosis and Management of ADHD in Children, Young People and Adults*. National Institute for Health and Clinical Excellence; 2009.
2. Alavi N, Yang M, Stephenson C, et al. Using the online psychotherapy tool to address mental health problems in the context of the COVID-19 pandemic: protocol for an electronically delivered cognitive behavioral therapy program. *JMIR Res Protoc*. 2020;9(12):e24913. doi:<https://doi.org/10.2196/24913>
3. Grazioli S, Crippa A, Rosi E, et al. Exploring telediagnostic procedures in child neuropsychiatry: addressing ADHD diagnosis and autism symptoms through supervised machine learning. *Eur Child Adolesc Psychiatry*. 2023;1-11. doi:<https://doi.org/10.1007/s00787-023-02145-4>
4. Markowetz A, Błaszczewicz K, Montag C, Switala C, Schlaepfer TE. Psychoinformatics: big data shaping modern psychometrics. *Med Hypotheses*. 2014;82(4):405-411. doi:<https://doi.org/10.1016/j.mehy.2013.11.030>
5. Sherwood AR, MacDonald B. A teleneuropsychology consultation service model for children with neurodevelopmental and acquired disorders residing in rural state regions. *Arch Clin Neuropsychol*. 2020;35(8):1196-1203. doi:<https://doi.org/10.1093/arclin/acia099>
6. Colombo P, Ceccarelli SB, Pacchiarini S, Cribellati S, Molteni M. A Web Platform for Standardized Data Acquisition, Processing, and Export in the Child Psychopathology Clinical Routine (MedicalBIT): Design and Implementation Study. *JMIR Form Res*. 2022;6(7):e36757. doi:<https://doi.org/10.2196/36757>
7. Briganti G, Le Moine O. Artificial intelligence in medicine: today and tomorrow. *Front Med*. 2020;7:27. doi:<https://doi.org/10.3389/fmed.2020.00027>
8. Crippa A, Salvatore C, Perego P, et al. Use of machine learning to identify children with autism and their motor abnormalities. *J Autism Dev Disord*. 2015;45:2146-2156.
9. Crippa A, Salvatore C, Molteni E, et al. The utility of a computerized algorithm based on a multi-domain profile of measures for the diagnosis of attention deficit/hyperactivity disorder. *Front Psychiatry*. 2017;8:189.
10. Trognon A, Richard M. Questionnaire-based computational screening of adult ADHD. *BMC Psychiatry*. 2022;22(1):401.
11. Christiansen H, Chavanon ML, Hirsch O, et al. Use of machine learning to classify adult ADHD and other conditions based on the Conners' Adult ADHD Rating Scales. *Sci Rep*. 2020;10(1):18871.

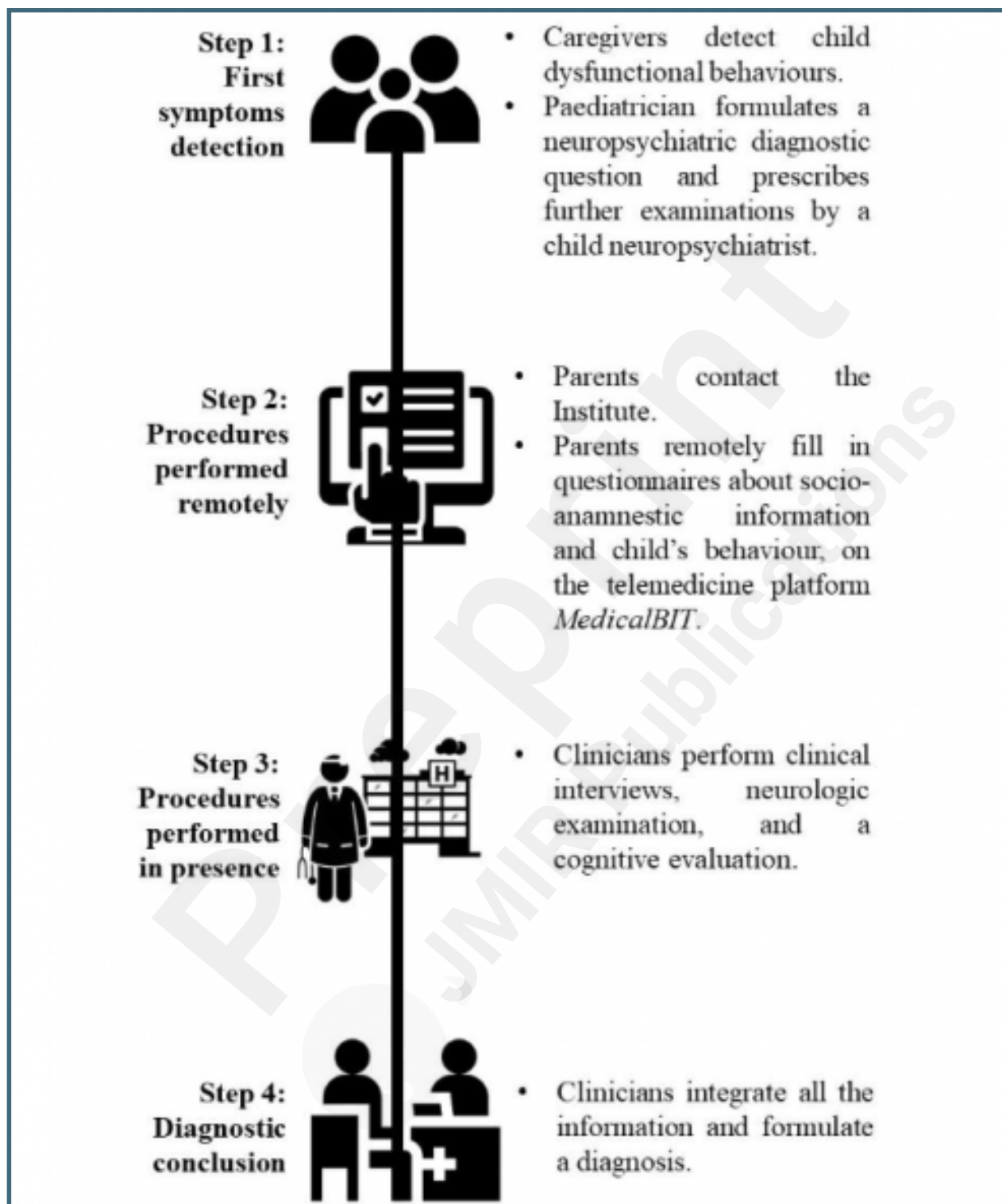
12. Dziugaite GK, Ben-David S, Roy DM. Enforcing interpretability and its statistical impacts: Trade-offs between accuracy and interpretability. *ArXiv Prepr ArXiv201013764*. 2020. doi:<https://doi.org/10.48550/arXiv.2010.13764>
13. Yarkoni T, Westfall J. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect Psychol Sci*. 2017;12(6):1100-1122.
14. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32. doi:<https://doi.org/10.1023/A:1010933404324>
15. Caly H, Rabiei H, Coste-Mazeau P, et al. Machine learning analysis of pregnancy data enables early identification of a subpopulation of newborns with ASD. *Sci Rep*. 2021;11(1):6877.
16. Akter T, Khan MI, Ali MH, Satu MS, Uddin MJ, Moni MA. Improved machine learning based classification model for early autism detection. In: IEEE; 2021:742-747.
17. Young SR, Maddocks DL, Caemmerer JM. Computer-enhanced practice: The benefits of computer-assisted assessment in applied clinical practice. *Prof Psychol Res Pract*. 2022.
18. Cao X, Cao J. Commentary: Machine learning for autism spectrum disorder diagnosis—challenges and opportunities—a commentary on Schulte-Rüther et al. (2022). *J Child Psychol Psychiatry*. 2023. doi:<https://doi.org/10.1111/jcpp.13764>
19. Ben-Sasson A, Yom-Tov E. Online concerns of parents suspecting autism spectrum disorder in their child: content analysis of signs and automated prediction of risk. *J Med Internet Res*. 2016;18(11):e300.
20. Duda M, Haber N, Daniels J, Wall D. Crowdsourced validation of a machine-learning classification system for autism and ADHD. *Transl Psychiatry*. 2017;7(5):e1133-e1133.
21. APA APA. Diagnostic and statistical manual of mental disorders. *Am Psychiatr Assoc*. 2013.
22. Team RDC. R: A language and environment for statistical computing. (No Title). 2010. <https://www.R-project.org/>.
23. de Jonge E, van der Loo M. An introduction to data cleaning with R. *Stat Neth*. 2013.
24. Franzin A, Sambo F, Di Camillo B. bnstruct: an R package for Bayesian Network structure learning in the presence of missing data. *Bioinformatics*. 2017;33(8):1250-1252. doi:<https://doi.org/10.1093/bioinformatics/btw807>
25. Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2(3):18-22.
26. Refaeilzadeh P, Tang L, Liu H, Liu L, Özsu M. Encyclopedia of database systems. *Cross-Valid*. 2009;5:532-538.
27. Perochon S, Di Martino JM, Carpenter KL, et al. Early detection of autism using digital behavioral phenotyping. *Nat Med*. 2023;29(10):2489-2497.
28. Patel HH, Prajapati P. Study and analysis of decision tree based classification algorithms. *Int J Comput Sci Eng*. 2018;6(10):74-78.
29. Devika Varshini G, Chinnaiyan R. Optimized machine learning classification approaches for prediction of autism spectrum disorder. *Ann Autism Dev Disord* 2020 1 1. 2020;1001.
30. Albahri AS, Hamid RA, Zaidan A, Albahri OS. Early automated prediction model for the diagnosis and detection of children with autism spectrum disorders based on

- effective sociodemographic and family characteristic features. *Neural Comput Appl.* 2023;35(1):921-947.
31. van der Meer D, Hoekstra P, Van Donkelaar M, et al. Predicting attention-deficit/hyperactivity disorder severity from psychosocial stress and stress-response genes: a random forest regression approach. *Transl Psychiatry.* 2017;7(6):e1145-e1145.
 32. Zhang-James Y, Chen Q, Kuja-Halkola R, Lichtenstein P, Larsson H, Faraone SV. Machine-Learning prediction of comorbid substance use disorders in ADHD youth using Swedish registry data. *J Child Psychol Psychiatry.* 2020;61(12):1370-1379.
 33. Faraone SV, Asherson P, Banaschewski T, et al. Attention-deficit/hyperactivity disorder. *Nat Rev Dis Primer.* 2015;1(1):1-23.
 34. Elsabbagh M, Divan G, Koh Y, et al. Global prevalence of autism and other pervasive developmental disorders. *Autism Res.* 2012;5(3):160-179.
 35. Hewitt KC, Rodgin S, Loring DW, Pritchard AE, Jacobson LA. Transitioning to telehealth neuropsychology service: Considerations across adult and pediatric care settings. *Clin Neuropsychol.* 2020;34(7-8):1335-1351. doi:https://doi.org/10.1080/13854046.2020.1811891
 36. Pasca L, Zanaboni MP, Grumi S, et al. Impact of COVID-19 pandemic in pediatric patients with epilepsy with neuropsychiatric comorbidities: A telemedicine evaluation. *Epilepsy Behav.* 2021;115:107519. doi:https://doi.org/10.1016/j.yebeh.2020.107519
 37. Pritchard AE, Sweeney K, Salorio CF, Jacobson LA. Pediatric neuropsychological evaluation via telehealth: Novel models of care. *Clin Neuropsychol.* 2020;34(7-8):1367-1379. doi:https://doi.org/10.1080/13854046.2020.1806359
 38. Tahillioğlu A, Bilaç Ö, Uysal T, Ercan ES. Who predict ADHD with better diagnostic accuracy?: Parents or teachers? *Nord J Psychiatry.* 2021;75(3):214-223. doi:https://doi.org/10.1080/08039488.2020.1867634
 39. O'Connor K, Neff DM, Pitman S. Burnout in mental health professionals: A systematic review and meta-analysis of prevalence and determinants. *Eur Psychiatry.* 2018;53:74-99. doi:https://doi.org/10.1016/j.eurpsy.2018.06.003

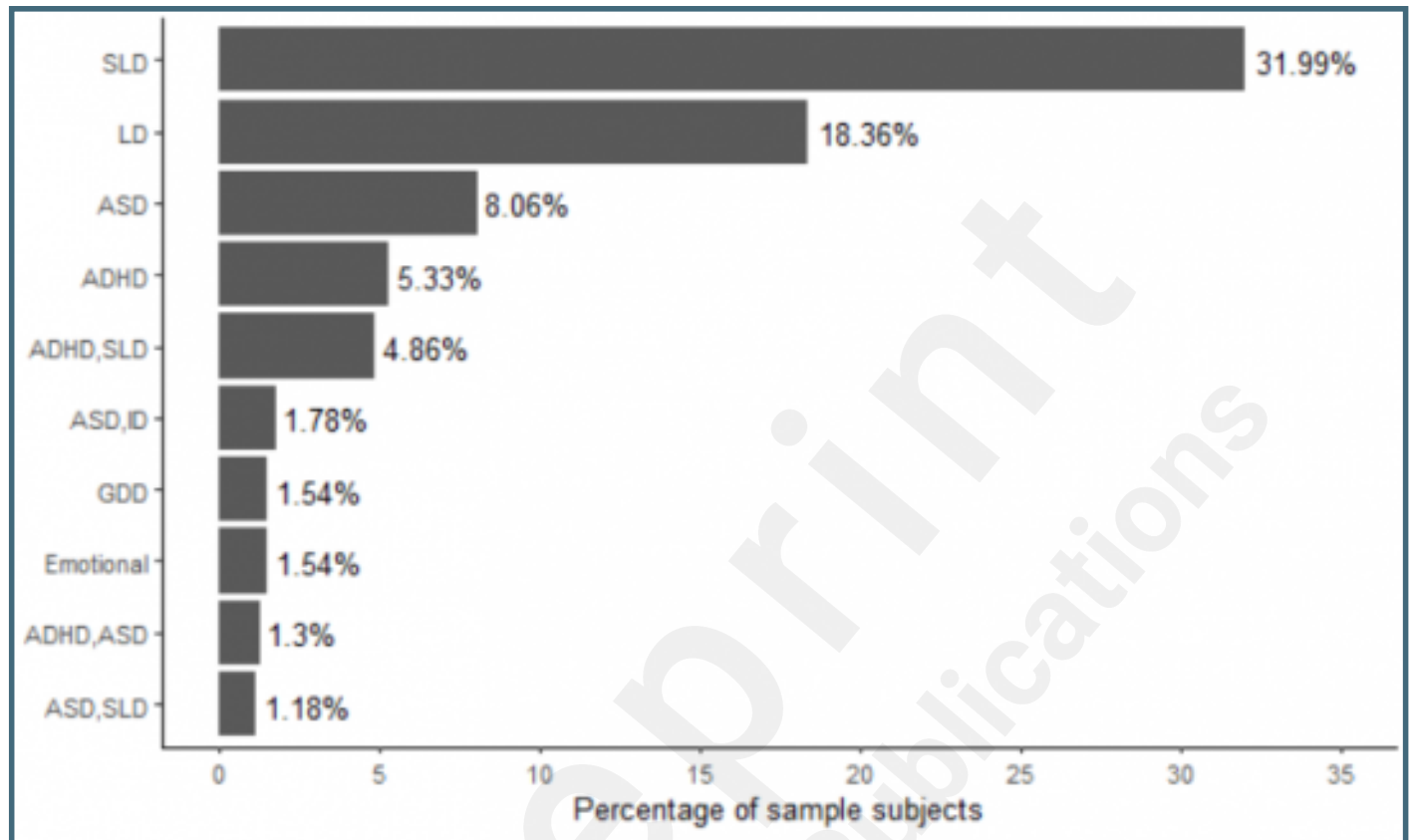
Supplementary Files

Figures

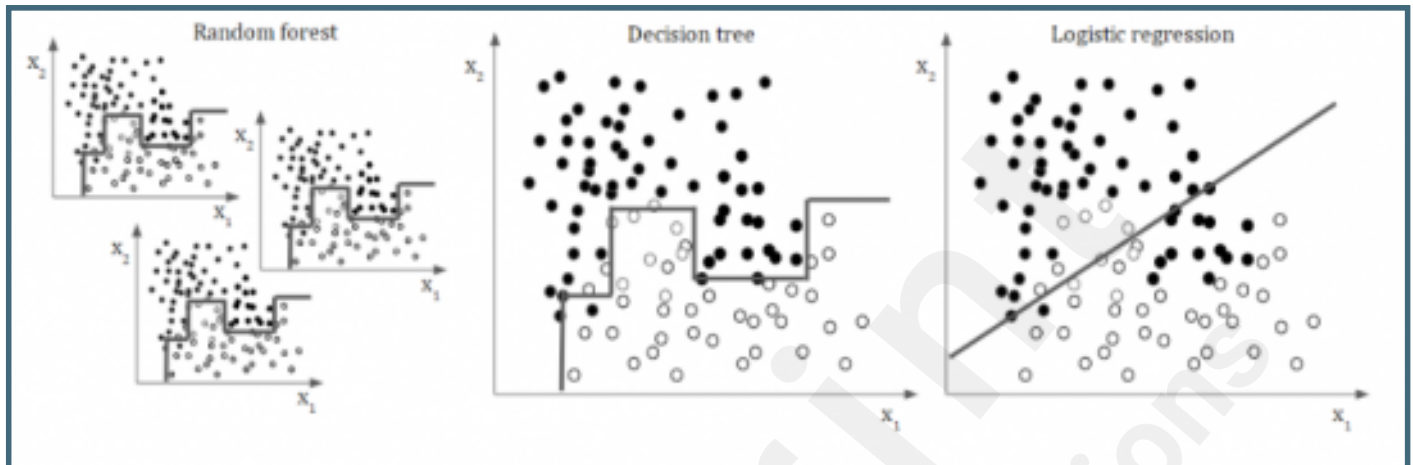
Diagnostic process scheme.



Most frequent configurations of diagnoses across the sample. The figure depicts the configuration of diagnostic categories and comorbidities presented by at least 1% of the participants. Note: ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; GDD: global developmental delay; ID: intellectual disability; LD: language disorder; SLD: specific learning disorder.



Decision boundaries of random forest (RF), decision tree (DT), and logistic regression (LR) in a binary classification example. The decision boundaries represent the separation between the two classes in a feature space. The decision boundary of an RF model is a nonlinear combination of the individual decision boundaries from each decision tree. A DT decision boundary is a nonlinear combination of axis-parallel lines. The linear equation of a hyperplane in the feature space determines an LR decision boundary.



SHAP values for a) ADHD classification through RF; b) ASD RF classification through RF.

