A Machine-Based Prediction Model of ADHD Using CPT Data

Ortal Slobodin[1]*, Inbal Yahav[2] and Itai Berger[3,4]

[1]Department of Education, Ben-Gurion University, Beer-Sheva, Israel, [2]Coller School of Management, Tel Aviv University, Tel Aviv, Israel, [3]Pediatric Neurology, Assuta Ashdod University Hospital, Ashdod, Israel, [4]Faculty of Health Sciences, Ben-Gurion University, Beer-Sheva, Israel

Despite the popularity of the continuous performance test (CPT) in the diagnosis of attention-deficit/hyperactivity disorder (ADHD), its specificity, sensitivity, and ecological validity are still debated. To address some of the known shortcomings of traditional analysis and interpretation of CPT data, the present study applied a machine learning-based model (ML) using CPT indices for the Prediction of ADHD.Using a retrospective factorial fitting, followed by a bootstrap technique, we trained, cross-validated, and tested learning models on CPT performance data of 458 children aged 6–12 years (213 children with ADHD and 245 typically developed children). We used the MOXO-CPT version that included visual and auditory stimuli distractors. Results showed that the ML proposed model performed better and had a higher accuracy than the benchmark approach that used clinical data only. Using the CPT total score (that included all four indices: Attention, Timeliness, Hyperactivity, and Impulsiveness), as well as four control variables [age, gender, day of the week (DoW), time of day (ToD)], provided the most salient information for discriminating children with ADHD from their typically developed peers. This model had an accuracy rate of 87%, a sensitivity rate of 89%, and a specificity rate of 84%. This performance was 34% higher than the best-achieved accuracy of the benchmark model. The ML detection model could classify children with ADHD with high accuracy based on CPT performance. ML model of ADHD holds the promise of enhancing, perhaps complementing, behavioral assessment and may be used as a supportive measure in the evaluation of ADHD.

Keywords: attention-deficit/hyperactivity disorder, continuous performance test, machine learning, prediction, children

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorders (Barkley, 2015), with an estimated prevalence of 9.4% in USA children (Centers for Disease Control and Prevention, 2018). The rates of ADHD diagnoses have been rising in recent decades. In 2003, 7.8% of the USA children were diagnosed with ADHD, compared to 9.5% in 2007 and 11% in 2011–2012 (Danielson et al., 2018). ADHD is characterized by symptoms of inattention and/or impulsivity and hyperactivity, which can adversely impact the behavioral, emotional, and social aspects of life. In approximately 80% of children with ADHD, symptoms persist into adolescence and may continue into adulthood (Faraone et al., 2003).

Because early behavioral and developmental interventions for ADHD could improve outcomes (Sonuga-Barke and Halperin, 2010; Halperin et al., 2012), there is a need for reliable diagnostic ADHD markers that can be identified early in life.

ADHD diagnosis is based on the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM), which are clinically judged and therefore are subjective to clinician and reporter's bias (Rousseau et al., 2008; Berger, 2011). Diagnostic criteria bias may reflect socio-cultural influences on symptom manifestation and diagnostic procedures (for review see Slobodin and Masalha, 2020) as well as the substantial overlap between ADHD symptoms and other psychiatric, developmental and neurological conditions (e.g., learning disabilities, depression, anxiety; Nikolas et al., 2019). Thus, predicting ADHD impairment using objective, easy-to-collect variables by noninvasive methods might be useful as a supportive measure in the evaluation of ADHD and other neurological/psychiatric disorders (Na, 2019).

Using CPT in the Diagnosis of ADHD

The continuous performance test (CPT) is one of the most popular objective measures of ADHD-related inattention and impulsivity (Edwards et al., 2007). CPTs usually include a serial presentation of visual or auditory target and non-target stimuli (numbers, letters, number/letter sequences, or geometric figures). Failing to respond to a target stimulus ("omission error") is assumed to measure inattention. A response to a non-target stimulus ("commission error") is considered to measure impulsivity. Other standard measures of CPT responses include the number of correct responses, the response time (RT), and the variability in RT.

Several studies have supported the utility of the CPT in the diagnostic process of ADHD (for review, Hall et al., 2016). For example, a recent study in the U.K. found that the QbTest (a computerized CPT combined with an infra-red camera to detect motor activity; Qbtech Limited) increased the speed and efficiency of ADHD clinical decision making without compromising diagnostic accuracy. Furthermore, the economic analysis revealed that the QbTest could increase patient throughput and reduce waiting times without significant increases in overall healthcare system costs (Hollis et al., 2018).

Despite its popularity, the utility of the CPT in the diagnostic process of ADHD has been long debated, due to its limited specificity, sensitivity, and ecological validity (Nigg et al., 2005; Toplak et al., 2013). Most of the methods used to discriminate between children with ADHD and typically developed children were based on standard statistical techniques, such as analysis of variance that were run on the data obtained from CPT measurements. These methods have led to inconsistent results between the researchers in the studies of ADHD children and adolescents (Hall et al., 2016). For example, an analysis of eight CPT studies revealed a wide variety in measures of sensitivity (9–88%) and specificity (23–100%) to ADHD (Pan et al., 2007). Similarly, a meta-analysis of 47 studies of CPT performance in children with ADHD found that the large effect sizes identified in previous research were significantly attenuated by unidentified true moderators or uncorrected artifacts, such as sampling error and measurement unreliability (Huang-Pollock et al., 2012). Traditional data analytic methods of CPT considerably limit the number of variables that can be used in a given analysis and, especially, the analysis of interactions. These methods of analysis also have limited ability to shed light on causality when the data are not based on randomized experimental designs (Deshpande et al., 2013). Most importantly, although standard approaches to CPT may distinguish clinical and non-clinical populations, they do not guarantee predictive optimality or parsimony in a data analysis-independent manner (Saxe et al., 2017). Taken together, the above findings emphasize the need to develop reliable validation techniques for both the clinical and research implications of CPT.

Machine learning (ML) is a rapidly emerging field that has allowed the exploitation of large datasets to generate predictive models. In "supervised learning," machines develop ways of linking a target outcome from a set of predictors ("features") in existing data. Such models may generalize to novel predictor data. In contrast to traditional statistical approaches, ML focuses on prediction rather than explanation (Hatton et al., 2019).

Availability and affordability of data collecting devices have opened doors for the use of ML to predict the likelihood of individuals developing a set of mental disorders such as depression, anxiety, autism, dementia, brain tumors, schizophrenia, psychosis, et cetera (Sen et al., 2018; Sakai and Yamada, 2019; Vieira et al., 2019). A growing number of supervised ML studies have been carried out on discriminating ADHD from control groups using the data obtained from electroencephalogram (EEG; Tenev et al., 2014), brain structural magnetic resonance imaging (MRI; Peng et al., 2013), MRI and functional magnetic resonance imaging (fMRI; Sen et al., 2018), Near-infrared spectroscopy (NIRS; Yasumura et al., 2017), and a combination of subjective and objective measures of ADHD (Emser et al., 2018). ML was also used to predict methylphenidate response in youth with ADHD using environmental, genetic, neuroimaging, and neuropsychological data (Kim et al., 2015). Although these models showed promising results in discriminating children and adults with ADHD from controls or other clinical conditions (e.g., autism spectrum disorders), their limited availability, high costs, and invasiveness hindered their widespread use.

The Current Study

Behavioral diagnosis of ADHD is a time-consuming, multi-informant procedure that can be complicated by the overlaps in symptomatology. This complexity may lead to delayed diagnosis and treatment (Duda et al., 2016). Given the variation in causes and behavioral consequences of ADHD, there is no single test used to diagnose the disorder. Therefore, a diagnostic model of ADHD based on CPT performance holds the promise of enhancing, perhaps complementing, behavioral assessment. This study aimed to apply a ML-based model using a CPT for the prediction of ADHD.

MATERIALS AND METHODS

Participants and Procedure

Participants were 458 children aged 6–12 years (mean = 8.68, SD = 1.77), 267 were boys (59%) and 191 girls (41%). Of them, 213 children were diagnosed with ADHD, and 245 were typically developed, children. No age differences were found between the ADHD and the non-ADHD groups (M = 8.62, SD = 1.83, and M = 8.72, SD = 1.71, respectively, p = 0.94). However, the rate of boys was significantly higher in the ADHD group than in the non-ADHD group (67% vs. 51%, respectively; $\chi^2_{(1,458)}$ = 10.84, p < 0.001).

Participants in the ADHD group were clinic-referred children recruited from out-patient pediatric clinics of a Neuro-Cognitive Centre, based in a tertiary care university hospital. Children were referred for ADHD evaluation by their pediatrician, general practitioner, teacher, mental health professional, or by their parents. All participants in the ADHD group met the criteria for ADHD, according to DSM-5 (American Psychiatric Association, 2013), as assessed by a certified pediatric neurologist. The diagnostic procedure included an interview with the patient and parents, medical/neurological examination as described by the American Academy of Pediatrics (AAP) clinical practice guidelines (Wolraich et al., 2011), and completing ADHD symptoms scales (DuPaul et al., 2016). All children were drug naïve.

Participants in the control group were randomly recruited from regular primary schools. Inclusion criteria for participants in the control group were: (1) the child scored below the clinical cut off point for ADHD symptoms on ADHD DSM Scales (American Psychiatric Association, 2013; DuPaul et al., 2016); and (2) an absence of academic or behavioral problems based on parents' and teachers' reports. Exclusion criteria for all participants were: intellectual disability, chronic use of medications, and primary psychiatric diagnosis (e.g., depression, anxiety, and psychosis).

All children were administered with the MOXO-CPT. The test was administered to children with ADHD during the process of clinical evaluation. In the non-ADHD group, the test was delivered by a member of the research team at the child's school or home.

All participants agreed to participate in the study, and their parents provided written informed consent to the study, approved by the Helsinki Committee (IRB) of Hadassah-Hebrew University Medical Center Jerusalem, Israel. Participants were not compensated for their participation in the study.

Measures

CPT performance—the current study used the MOXO-CPT[1] version (Berger and Goldzweig, 2010). The MOXO-CPT (Neuro-Tech Solutions Limited) is a standardized computerized test designed to diagnose ADHD-related symptoms. The MOXO-CPT task requires the child to sustain attention over a

---

[1]The term "MOXO" derives from the world of Japanese martial arts and means a "moment of lucidity." It refers to the moments preceding the fight, when the warrior clears his mind from distracting, unwanted thoughts, and feelings.

continuous stream of stimuli and to respond to a prespecified target. However, in contrast to other existing CPTs, the test includes visual and auditory stimuli serving as measurable distractors. The test's validity and utility in distinguishing children and adolescents with ADHD from their typically developing peers were demonstrated in previous studies (Berger et al., 2017; Slobodin et al., 2018).

The test consisted of eight stages (levels). Each level consisted of 53 trials (33 target and 20 non-target stimuli) and lasted 114.15 s. The total duration of the test was 15.2 min. In each trial, a stimulus (target or non-target) was presented in the middle of the computer screen for durations of 0.5, 1, or 3 s and was followed by a "void" of the same duration (**Figure 1**). This method enabled us to distinguish accurate responses performed in "good timing" (quick and correct responses to the target performed during stimulus presentation) from accurate but slow responses (correct responses to the target performed after the stimulus presentation; during the void period). These two aspects of timing correspond to the two different deficiencies typical to ADHD; responding quickly and responding accurately (National Institute of Mental Health, 2012). The child was instructed to respond to the target stimulus as quickly as possible by pressing the space bar once and only once. The child was also instructed not to respond to any other stimuli but the target, and not to press any other key but the space bar.

Both target and non-target stimuli were cartoon pictures free of letters or numbers. Also, the test included six different environmental distractors, each of them could appear as pure visual (e.g., three birds moving their wings), pure auditory (e.g., birds singing), or as a combination of visual and auditory stimuli (birds moving their wings and singing simultaneously). Each distractor was presented on the screen for a different duration ranging from 3.5 to 14.8 s, with a constant interval of 0.5 s between two distractors.

For each child, four CPT indices were recorded: attention (number of correct responses to target stimuli, including the rate of omission errors), Timeliness (correct responses to target stimuli conducted on accurate timing), Hyperactivity (a measure of motor activity) and Impulsiveness (responses to non-target stimuli, including the rate of commission errors). Test administration time, namely, the day of the week (DoW) and time of day (ToD), was also recorded. A full description of the MOXO-CPT test is provided in **Supplementary Material (Appendix 1)**.
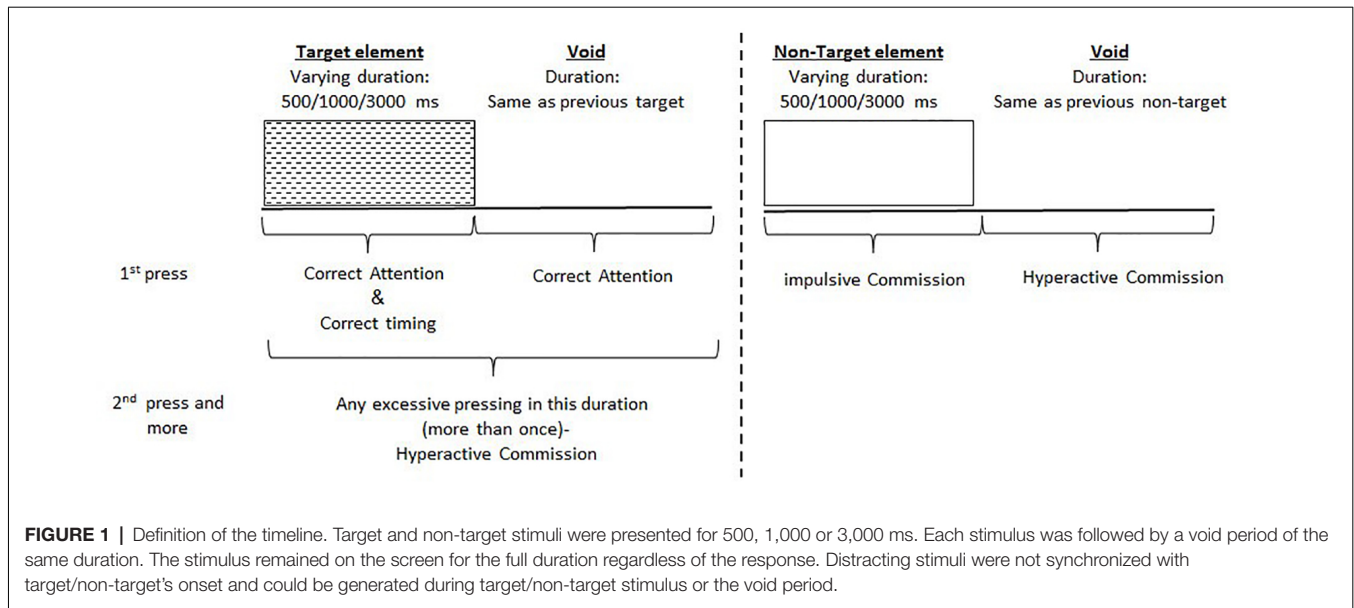
Data Analysis

Data analysis included four stages: (a) data exploration and processing; (b) data partition and manipulation of training; (c) model fitting; and (d) ADHD prediction and evaluation.

(1) Data exploration and processing—following the descriptive data analysis outlined in the method section, two sources of inherent biases were evident, which we denoted: across the level and within level imbalances.

Across level imbalances—demographic and test time variables were imbalanced. For example, the total number of children that performed the MOXO-CPT during Mondays

FIGURE 1 | Definition of the timeline. Target and non-target stimuli were presented for 500, 1,000 or 3,000 ms. Each stimulus was followed by a void period of the same duration. The stimulus remained on the screen for the full duration regardless of the response. Distracting stimuli were not synchronized with target/non-target's onset and could be generated during target/non-target stimulus or the void period.

was significantly higher than the number of children that performed the test during any other DoW. The lowest number of tests was conducted on Saturdays. Imbalanced designs concerning the number of observations per level were shown to be highly susceptible to statistical biases, such as heteroscedasticity (Milliken and Johnson, 2009). Moreover, our data showed significant differences in MOXO-CPT performance across various DoW, ToD, and child's ages, suggesting that across level imbalances may be associated with observed between or within-group differences in MOXO-CPT performance. **Figure 2** illustrates different hyperactivity levels in children without ADHD as a function of the time of the day they performed MOXO-CPT. As seen in **Figure 2**, the level of hyperactivity significantly decreased with the time of the day ($F = 8.26$, $p < 0.01$).

Within level imbalance—diagnostic class within demographic and test time levels were also shown to be imbalanced. For example, the ADHD group significantly differed from the non-ADHD group in their gender distribution, with more boys in the ADHD group than in the non-ADHD group ($\chi^2_{(1,458)} = 10.84$, $p < 0.001$). Moreover, for some factor levels, such as DoW and ToD, the data included only children from one diagnostic group, but not the other; administration of the MOXO-CPT during the weekend was evident only among the non-ADHD group. Also, all children in the ADHD group (with one exception) performed the MOXO-CPT during the morning hours.
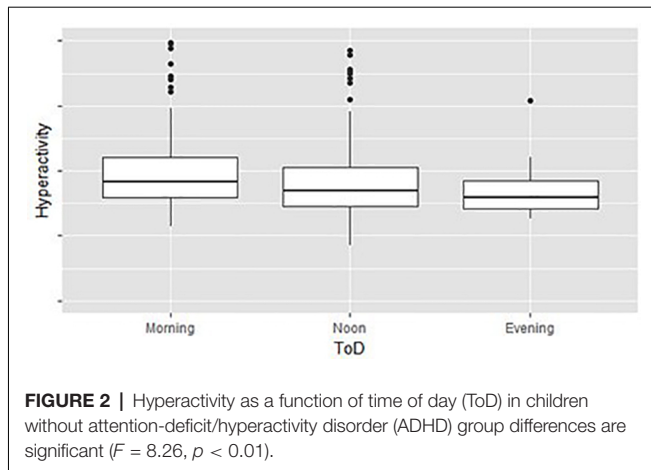
Within-level imbalance results in a statistical bias that is closely related to the renowned self-selection bias (Heckman, 1990), a biased caused by participants choosing themselves into treatment groups rather than assigned randomly. In the context of the current study, the time of MOXO-CPT administration was not randomized, and group affiliation was not matched, leading to differences in gender, age, DoW, and ToD distribution between groups. This imbalance may

impose a significant prediction bias if demographic and test time controls were not matched before deploying an ML model on the CPT data.

To quantify the statistical biases in our data, we conducted a retrospective factorial fitting (Loy et al., 2002). This technique sorts observations into groups according to their factor levels—a total of 7 [age] × 2 [gender] × 7 [DoW] × 3 [ToD] = 194 groups. Following the procedure proposed by Yahav et al. (2016), we then merged groups in which the impact of ADHD on test performance was statistically equal, thus reducing the number of groups to 28 groups. We then computed the number of records in each group: n_i (across-level imbalance), and the fraction of children with ADHD $p_i(ADHD)$ (within-level imbalance). Groups in which the fraction of children with ADHD equaled to either 0 or 1, were removed from the dataset, as the impact ADHD on test performance within these groups was unquantifiable. The processed data contained 445 observations (97% of the unprocessed data).

(2) Data partition and manipulation of training—as customary in predictive analysis (Shmueli, 2010), we partitioned the data randomly to training and holdout. The training set consisted of 60% of the records and was used for model training. The holdout set held the rest of the records (40%) and was used for model evaluation.

We manipulated the training set to correct the inherent biases. Specifically, we bootstrapped the inflated number of training records ($\tilde{N} = 5{,}000$) using bootstrap sampling with repetition, following the principles of the Synthetic Minority Oversampling Technique (SMOTE). SMOTE algorithm is a kind of random oversampling algorithm which generates new synthetic samples by analyzing neighbors of minority samples (Chawla et al., 2002; He and Garcia, 2009). We set the oversampling probability per record as inverse to its appearance in the data concerning

FIGURE 2 | Hyperactivity as a function of time of day (ToD) in children without attention-deficit/hyperactivity disorder (ADHD) group differences are significant ($F = 8.26$, $p < 0.01$).

levels ($1 - n_i/N$) and diagnosis [$1 - p_i(ADHD)$] for children with ADHD, and $p_i(ADHD)$ for children without ADHD. This sampling procedure generated a training data set that was: (1) synthetically large enough to allow the use of robust ML techniques that operate on large amounts of data; and (2) balanced within- and across-levels, thus was free of statistical biases. The holdout set remained untouched, to allow a fair evaluation of the prediction model.

(3) Model fitting—in this stage, we trained an ML prediction model f to the training set that mapped all or a subset of the test performance measures (Attention, Timeliness, Hyperactivity, and Impulsiveness) and additional control variables (age, gender, DoW, and ToD) to the diagnosis class:

$$p(\text{diagnosis} = ADHD) = f(\text{test performance, controls})$$
(1)

Specifically, f in our analysis is either a random forest (Hothorn et al., 2006) or Neural Network with cross-validation on 100 folds[2].

(4) ADHD prediction and evaluation—in this stage, we used the training model to predict the detection of ADHD in holdout records. That is, we computed the accuracy, sensitivity, and specificity of our ML model comparing to clinical diagnosis. We repeated steps 3 and 4 (data partition and training manipulation, model fitting, and ADHD detection and evaluation) 100 times, generating a different training-validation random split as each repetition, to compute conference intervals for the accuracy measures. The detection procedure is summarized in **Figure 3**.

RESULTS

We examined different subsets of MOXO-CPT performance indices (Attention, Timeliness, Hyperactivity, and

---

[2]We have also conducted analyses with additional ML models, including kNN, Logistic regression and Adaptive boosting. The best performance was achieved by Neural Network (overall accuracy of 0.89, compared to 0.68 of the benchmark model), and the lowest by logistic regression (0.83, compared to 0.61).

Impulsiveness) and control variables (age, gender, Dow and ToD) as ADHD predictors in children who were diagnosed with ADHD, using the "gold standard" clinical criteria. The gold standard of ADHD diagnosis was based on the DSM-V criteria for ADHD (American Psychiatric Association, 2013) and the AAP clinical practice guideline (Wolraich et al., 2011).

As a benchmark, we trained function f in Equation (1) on a training set derived from the unprocessed (original) data. Since the unprocessed data contained inherent statistical biases caused by level-imbalance (within and across), it was impossible to use the control variables as reliable predictors in the benchmark model.

The results of the random forest and the Neural Network (NN) techniques are presented in **Table 1**; t-tests analyses for paired samples revealed that the differences between the benchmark model and the proposed models (in both random forest and NN techniques) were significant ($p < 0.05$). However, no significant differences were found between the results of the random forest technique and the NN technique ($p > 0.05$). As seen in the table, NN provides, on average, higher rates of accuracy and sensitivity to ADHD, compared to the random forest technique. However, the confidence intervals in the NN model are higher, indicating non-robust cases and higher risk prediction errors. Given the increased robustness of the random forest technique and the fact that it was not significantly different from NN in all observed measures (overall accuracy, sensitivity, and specificity), it was chosen as the preferred method.

As shown in the table, the best performance of the random forest technique (over 87% accuracy, 89% sensitivity and 84% specificity) was achieved by applying the proposed method to predict ADHD, using the four CPT performance indices and four controls as predictors. This performance was 34% higher than the best-achieved accuracy of the benchmark[3]. Prediction solely based on MOXO-CPT performance, without the use of additional controls, had an accuracy of 81% (84% sensitivity and 79% specificity), an improvement of 24.6% compared to the benchmark. Among the four CPT performance indices, Impulsiveness under the benchmark model had the highest ability to rule out ADHD in children (Specificity = 89%, significantly higher than all other models).

As an alternative to the SMOTE rebalancing technique, we also examined the impact of the under-sampling technique to balance between the control and the ADHD groups. This rebalancing technique revealed an overall accuracy of 73% (%95 CI = 63 to 83%), a sensitivity rate of 79% (%95 CI = 57 to 100%), and a specificity rate of 66% (%95 CI = 42 to 90%). As seen, there are considerable differences between the results of the two rebalancing techniques. While both techniques addressed the four potentially confounding variables (DoW, time of the day, gender, and age) in the same manner (same size, unprocessed), they differ in their ability to capture information about the majority class. Specifically, the disadvantage of the under-sampling technique is that removing modules may cause the training data to lose important information related to

---

[3]The 34% improvement was calculated with respect to 65% performance level of the benchmark.

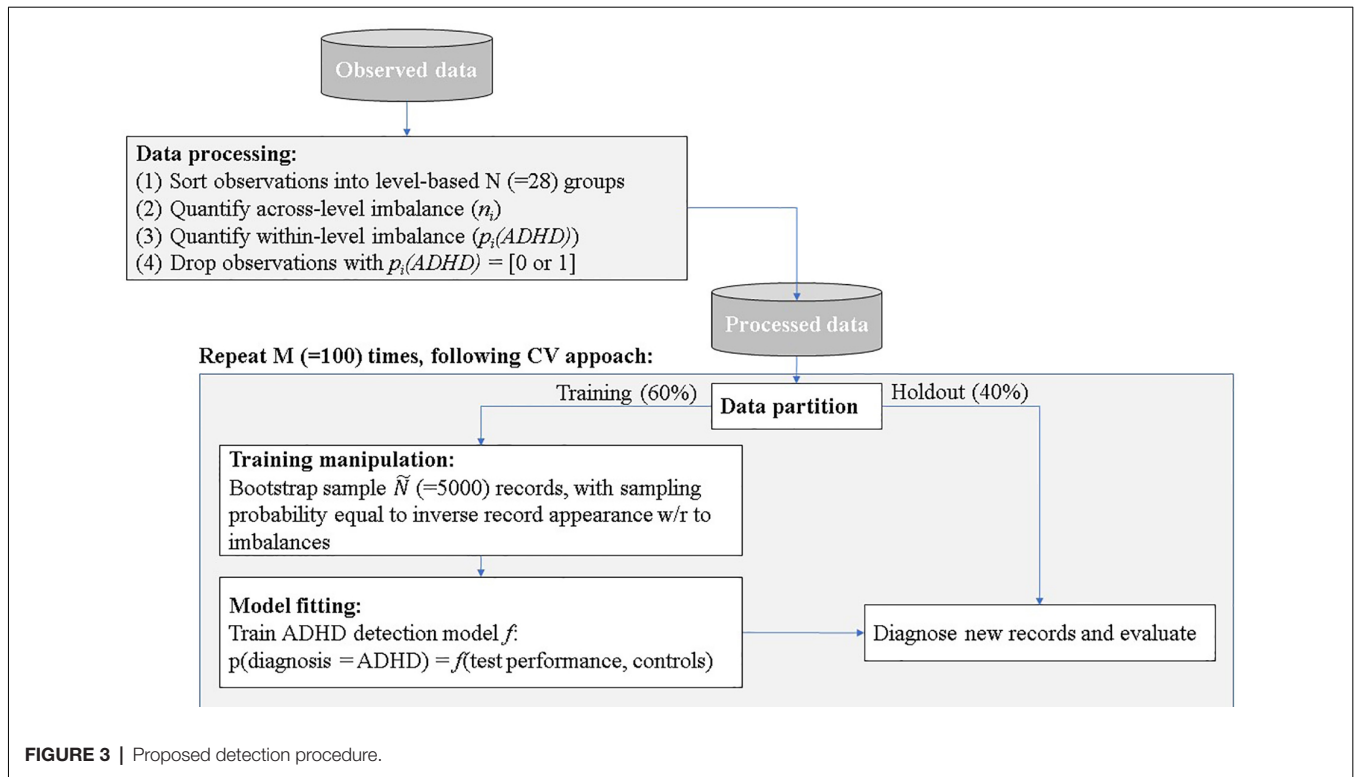**FIGURE 3 |** Proposed detection procedure.

TABLE 1 | Accuracy, sensitivity and specificity of machine learning (ML) in attention-deficit/hyperactivity disorder (ADHD) classification.

Model's name	Diagnosis predictors	Mean performance (SD)[c]					
		Overall accuracy (%)	95% CI	Sensitivity%	95% CI	Specificity (%)	95% CI
Benchmark	MOXO	MOXO-CPT scores:	65	59%, 71%	65	39%, 91%	65	43%, 87%
Proposed model f = Random forest	MOXO	MOXO-CPT scores[a]	81	77%, 85%	84	78%, 90%	79	71%, 87%
	MOXO w/Controls	MOXO-CPT scores considering control variables[b]	87	81%, 93%	89	83%, 95%	84	76%, 92%
Proposed model f = Neural network	MOXO w/Controls	MOXO-CPT scores considering control variables[b]	89	79%, 99%	95	73%, 100%	76	26%, 100%

[a]In this prediction model, all four MOXO-CPT indices (z-scores) were used as independent variables. [b]In this prediction model, all four MOXO-CPT indices (z-scores) and four control variables (age, gender, time of day, day of week) were used as independent variables. [c]The differences between the benchmark model and the proposed models (random forest method and NN) were all significant ($p < 0.05$). No significant differences were found between the results of random forest technique and the NN technique ($p > 0.05$).

the majority class. The SMOTE technique was proposed to combat this disadvantage by creating artificial data based on the feature space (rather than the data space) similarities from the minority modules (Tantithamthavorn et al., 2018). The advantage of the SMOTE is that it leads to no information loss and ensures that even small-size confounding effects are not overlooked.

A comparison of the model's performance as a function of data size is presented in **Supplementary Material (Appendix 2)**. The models are all trained on a balanced training set and predicted into a non-balanced validation set. Notably, the increase in the training size was synthetic and resulted from repeating the same records several times, following

the bootstrap technique. The validation score remained constant (40%).

Table 2 presents the random forest feature importance. In the current study, we used the accuracy-based importance as a measure of variable importance in the random forest. In this measure, each tree has its out-of-bag sample that is used to calculate the importance of a specific variable. In the first step, the prediction accuracy of the out-of-bag sample is measured. Then, the values of the variable in the out-of-bag-sample are randomly shuffled, as all other variables are kept unchanged. Finally, the decrease in prediction accuracy on the shuffled data is measured. The mean decrease in accuracy across all trees is reported. Variables with high importance

TABLE 2 | Random forest feature importance.

Predicting variables	Random forest feature importance
Z-score timeliness	0.170
Z-score hyperactivity	0.049
Z-score attention	0.041
Z-score impulsiveness	0.041
Gender	0.037
Day of the week	0.039
Child's age	0.028
Time of the day	0.004

These results are not identical to variable coefficient, and the interaction between the variables could not be observed in the RF model.

have a significant impact on the outcome values (Breiman, 2001, 2002). Notably, the displayed values indicate the relative importance of each feature when making a prediction and not absolute importance.

DISCUSSION

The current study applied a machine learning-based predictive model using CPT indices for the prediction of ADHD in children aged 6–12 years.

Our findings demonstrated that the ML proposed model performed better and had a higher accuracy than the benchmark approach that used clinical data only. We also found that the CPT total score (including all four indices: Attention, Timeliness, Hyperactivity, and Impulsiveness), and all four control variables (age, gender, DoW, and ToD), provided the most salient information for discriminating between children with ADHD and their typically developed peers. This model had an accuracy rate of 87%, a sensitivity rate of 89%, and a specificity rate of 84%. Using this model increased the performance by 34% compared to the benchmark approach.

The improvement in the model's prediction accuracy after quantifying for cross-level and within-level imbalances suggests that such statistical biases may affect the discriminative validity of the CPT. These findings are in line with previous research that pointed to the influence of age (Berger et al., 2013; Slobodin et al., 2018) and gender (for review, Hasson and Fine, 2012) on CPT performance. Although studies focusing on the effect of time administration on CPT performance are currently scarce, there is evidence to support the importance of accounting for inter-day and intraday variations when comparing CPT performance of children with ADHD to that of their typically developed peers (van der Heijden et al., 2010). For example, Imeraj et al. (2012) showed that children with ADHD (with or without a co-morbid oppositional defiant disorder) significantly differed from healthy controls in their cortisol profiles across the day. Such variations may underline group differences in arousal mechanisms and may also affect CPT performance (Wang et al., 2017).

Comparing our results to other ML methods previously used to discriminate between ADHD and controls suggests that an ML model based on CPT data holds the promise of discriminating children with ADHD from controls, even when compared to more invasive or expensive approaches. For example, Yasumura et al. (2017), who used near-infrared spectroscopy to quantify

the change in prefrontal cortex oxygenated hemoglobin during reversed Stroop task, found an overall discrimination rate of 86.25%, with a sensitivity of 88.71% and a specificity of 83.78%. Likewise, using MRI data, Peng et al. (2013) found an ADHD prediction accuracy of 90.18%. Recently, Emser et al. (2018) developed an ML prediction model of ADHD based on subjective and objective measures of ADHD, including a CPT (Quantified Behavior Test for adolescents and adults). Their results showed that the objective measures had an overall 78% accuracy and that the combined model accuracy of the objective and subjective measures was 86.7%.

Using ML CPT-based model to predict ADHD offers several clinical and practical advantages. First, this model provides an easy-to-administer, affordable, non-invasiveness measure of ADHD-related symptoms. Second, it has an extremely fast discrimination speed and satisfactory high classification accuracy. In particular, the observed high sensitivity rates (89%) may improve clinicians' ability and confidence in ruling out ADHD. Excluding ADHD when it is not present is very important given the complicated, time-consuming, and expensive process of ADHD diagnosis (Hall et al., 2017). Third, ML models can handle various demographic and procedural variables, that may hinder CPT's discriminative utility, to make an objective prediction (Emser et al., 2018). Finally, the current ML model is based on CPT performance that is obtained under the presence of environmental distractors, thus enables the assessment of the child's cognitive performance in an ecologically-valid environment (Barkley, 2015).

While this study does not support the viability of solely CPT-based algorithms for establishing a diagnosis of ADHD, it presents a step towards the goal of precision medicine in psychiatry. The ML model proposed here may inform the development of a decision-support module that utilizes the best-performing model, thus improving the quality of care for ADHD (Carroll et al., 2013).

Several limitations of this study should also be considered. The current study only examined standard CPT variables, including inattention, RT, hyperactivity, and impulsivity. However, ADHD is associated with additional cognitive deficits that may affect CPT performance, such as distractibility and fatigue over time (Pelham et al., 2011; Bioulac et al., 2012). Future investigation should use these additional variables with the current classification method to further improve ADHD classification performance. Also, our sample was limited to clinically-referred children with a definite ADHD diagnosis while excluding children with suspected ADHD (whose symptoms may not reach the threshold for diagnosis). Given that the time gap between initial suspicion and diagnosis could reach a year or even more, the ability of the ML model to provide preliminary risk evaluation and/or pre-clinical screening is of considerable significance in terms of timely intervention (Duda et al., 2016). Another limitation of the study is related to its reliance on a single recruitment center for children with ADHD. Although the current sample was derived from a tertiary care university hospital that provides services to the general population, this fact limits our ability to generalize our results to different populations. Also, there is limited information about

participants' demographic, cognitive, and clinical characteristics, such as IQ level and ADHD subtypes, which can be associated with CPT performance (Mahone et al., 2002; Collings, 2003). Given our limited information about participants' cognitive and personal characteristics, there was no way to rule out the possibility that some participants were not motivated to optimize their CPT performance.

Our model is also limited by the relatively small number of candidate classification features that were tested in the adaptive models (four CPT indices, two demographic variables, and two condition variables). Future ML research in ADHD prediction should expand this examination to include a broader range of clinical, behavioral, and demographic variables, including education, ethnicity, socio-economic status, psychiatric-co-morbidities, and medication use. Finally, while the predictive accuracy of ML models can be satisfactory, a naive implementation of ML without careful validation may have adverse consequences. An ML algorithm may replicate past decisions, including biases around ethnicity and gender, that may have affected the clinical judgment. Therefore, model extrapolation should be avoided until such biases are corrected (Saria et al., 2018).

CONCLUSIONS

Previous studies using standard, traditional analyses of CPT data provided evidence for the ability of the test to differentiate between children with ADHD and their typically developing peers. Nevertheless, these approaches were limited in their ability to optimally predict ADHD, to draw a causal inference, and to include multiple variables in a given analysis (Saxe et al., 2017). Thus, developing reliable validation techniques for both clinical and research implications of CPT is of high importance.

Our result showed that ML diagnostic model could predict ADHD in children with high accuracy based on CPT performance indices. This model performed better than any achieved benchmark model to CPT. Using an ML model based on CPT may provide good classification accuracy for supporting ADHD diagnoses in children and encourages the use of the CPT as a quick, cost-effective, and accurate decision-making tool in the ADHD diagnosis process (Hollis et al., 2018).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Helsinki Committee (IRB) of Hadassah Hebrew University Medical Center, Jerusalem, Israel. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

IB recruited the patients and performed clinical evaluations. IY performed data analyses. OS, IY, and IB contributed equally to the development of study design, integration of findings, and writing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.560021/full#supplementary-material.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders Text Revision (DSM-V).* 5th Edn. Washington, DC: American Psychiatric Association.

Barkley, R. A. (2015). *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment.* 4th Edn. New York, NY: The Guilford Press.

Berger, I. (2011). Diagnosis of attention deficit hyperactivity disorder: much ado about something. *Isr. Med Assoc. J.* 13, 571–574.

Berger, I., and Goldzweig, G. (2010). Objective measures of attention-deficit/hyperactivity disorder—a pilot study. *Isr. Med Assoc. J.* 12, 531–535.

Berger, I., Slobodin, O., Aboud, M., Melamed, J., and Cassuto, H. (2013). Maturational delay in ADHD: evidence from CPT. *Front. Hum. Neurosci.* 7:691. doi: 10.3389/fnhum.2013.00691

Berger, I., Slobodin, O., and Cassuto, H. (2017). Usefulness and validity of continuous performance tests in the diagnosis of attention-deficit hyperactivity disorder children. *Arch. Clin. Neuropsychol.* 32, 81–93. doi: 10.1093/arclin/acw101

Bioulac, S., Lallemand, S., Rizzo, A., Philip, P., Fabrigoule, C., and Bouvard, M. P. (2012). Impact of time on task on ADHD patient's performances in a virtual classroom. *Eur. J. Paediatr. Neurol.* 16, 514–521. doi: 10.1016/j.ejpn.2012.01.006

Breiman, L. (2002). *Manual on Setting Up, Using and Understanding Random Forests v3.1.* Available online at: https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf. Accessed August 20, 2020.

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Carroll, A. E., Bauer, N. S., Dugan, T. M., Anand, V., Saha, C., and Downs, S. M. (2013). Use of a computerized decision aid for ADHD diagnosis: a randomized controlled trial. *Pediatrics* 132, e623–e629. doi: 10.1542/peds.2013-0933

Centers for Disease Control and Prevention (2018). *Attention-Deficit/Hyperactivity Disorder (ADHD).* Available online at: https://www.cdc.gov/ncbddd/adhd/data.html. Accessed August 20, 2020.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *J. Artif. Intell. Res* 16, 321–357. doi: 10.1613/jair.953

Collings, R. D. (2003). Differences between ADHD inattentive and combined types on the CPT. *J. Psychopathol. Behav. Assess.* 25, 177–189. doi: 10.1023/A:1023525007441

Danielson, M. L., Bitsko, R. H., Ghandour, R. M., Holbrook, J. R., Kogan, M. D., and Blumberg, S. J. (2018). Prevalence of parent-reported ADHD diagnosis and associated treatment among U.S. children and adolescents, 2016. *J. Clin. Child Adolesc. Psychol.* 47, 199–212. doi: 10.1080/15374416.2017.1417860

Deshpande, G., Libero, L. E., Sreenivasan, K. R., Desphande, H. D., and Kana, R. K. (2013). Identification of neural connectivity signatures of autism using machine learning. *Front. Hum. Neurosci.* 7:670. doi: 10.3389/fnhum.2013.00670

Duda, M., Ma, R., Haber, N., and Wall, D. P. (2016). Use of machine learning for behavioral distinction of autism and ADHD. *Transl. Psychiatry* 6:e732. doi: 10.1038/tp.2015.221

DuPaul, G. J., Reid, R., Anastopoulos, A. D., Lambert, M. C., Watkins, M. W., and Power, T. J. (2016). Parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms: factor structure and normative data. *Psychol. Assess.* 28, 214–225. doi: 10.1037/pas0000166

Edwards, M. C., Gardner, E. S., Chelonis, J. J., Schulz, E. G., Flake, R. A., and Diaz, P. F. (2007). Estimates of the validity and utility of the conner's CPT in the assessment of inattentive and/or hyperactive impulsive behaviors in children. *J. Abnorm. Child Psychol.* 35, 393–404. doi: 10.1007/s10802-007-9098-3

Emser, T. S., Johnston, B. A., Steele, J. D., Kooij, S., Thorell, L., and Christiansen, H. (2018). Assessing ADHD symptoms in children and adults: evaluating the role of objective measures. *Behav. Brain Func.* 14:11. doi: 10.1186/s12993-018-0143-x

Faraone, S. V., Sergeant, J., Gillberg, C., and Biederman, J. (2003). The worldwide prevalence of ADHD: is it an american condition? *World Psychiatry* 2, 104–113.

Hall, C. L., Valentine, A. Z., Groom, M. J., Walker, G. M., Sayal, K., Daley, D., et al. (2016). The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring ADHD in children: a systematic review. *Eur. Child Adolesc. Psychiatry* 25, 677–699. doi: 10.1007/s00787-015-0798-x

Hall, C. L., Valentine, A. Z., Walker, G. M., Ball, H. M., Cogger, H., Daley, D., et al. (2017). Study of user experience of an objective test (QbTest) to aid ADHD assessment and medication management: a multi-methods approach. *BMC Psychiatry* 17:66. doi: 10.1186/s12888-017-1222-5

Halperin, J. M., Bédard, A. C., and Curchack-Lichtin, J. T. (2012). Preventive interventions for ADHD: a neurodevelopmental perspective. *Neurotherapeutics* 9, 531–541. doi: 10.1007/s13311-012-0123-z

Hasson, R., and Fine, J. G. (2012). Gender differences among children with ADHD on continuous performance tests: a meta-analytic review. *J. Atten. Disord.* 16, 190–198. doi: 10.1177/1087054711427398

Hatton, C. M., Paton, L. W., McMillan, D., Cussens, J., Gilbody, S., and Tiffin, P. A. (2019). Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare. *J. Affect. Disord.* 246, 857–860. doi: 10.1016/j.jad.2018.12.095

He, H., and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE. Trans. Knowl. Data Eng.* 21, 1263–1284. doi: 10.1109/TKDE.2008.239

Heckman, J. (1990). Varieties of selection bias. *Am. Econ. Rev.* 80, 313–318.

Hollis, C., Hall, C. L., Guo, B., James, M., Boadu, J., Groom, M. J., et al. (2018). The impact of a computerised test of attention and activity (QbTest) on diagnostic decision-making in children and young people with suspected attention deficit hyperactivity disorder: single-blind randomised controlled trial. *J. Child Psychol. Psychiatry* 59, 1298–1308. doi: 10.1111/jcpp.12921

Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat.* 15, 651–674. doi: 10.1198/106186006X133933

Huang-Pollock, C. L., Karalunas, S. L., Tam, H., and Moore, A. N. (2012). Evaluating vigilance deficits in ADHD: a meta-analysis of CPT performance. *J. Abnorm. Psychol.* 121, 360–371. doi: 10.1037/a0027205

Imeraj, L., Antrop, I., Roeyers, H., Swanson, J., Deschepper, E., Bal, S., et al. (2012). Time-of-day effects in arousal: disrupted diurnal cortisol profiles in children with ADHD. *J. Child. Psychol. Psychiatry* 53, 782–789. doi: 10.1111/j.1469-7610.2012.02526.x

Kim, J. W., Sharma, V., and Ryan, N. D. (2015). Predicting methylphenidate response in ADHD using machine learning approaches. *Int. J. Neuropsychopharmacol.* 18:pyv052. doi: 10.1093/ijnp/pyv052

Loy, C., Goh, T. N., and Xie, M. (2002). Retrospective factorial fitting and reverse design of experiments. *Total. Qual. Manag.* 13, 589–602. doi: 10.1080/0954412022000002009

Mahone, E. M., Hagelthorn, K. M., Cutting, L. E., Schuerholz, L. J., Pelletier, S. F., Rawlins, C., et al. (2002). Effects of IQ on executive function measures in children with ADHD. *Child Neuropsychol.* 8, 52–65. doi: 10.1076/chin.8.1.52.8719

Milliken, G. A., and Johnson, D. E. (2009). *Analysis of Messy Data Volume 1: Designed Experiments.* 2nd Edn. New York, NY: CRC Press.

Na, K. S. (2019). Prediction of future cognitive impairment among the community elderly: a machine-learning based approach. *Sci. Rep.* 9:3335. doi: 10.1038/s41598-019-39478-7

National Institute of Mental Health. (2012). *Attention Deficit Hyperactivity Disorder.* Available online at: http://www.nimh.nih.gov/health/publications/attention-deficit-hyperactivity-disorder/complete-index.shtml. Accessed August 20, 2020.

Nigg, J. T., Willcutt, E. G., Doyle, A., and Sonuga-Barke, E. J. S. (2005). Causal heterogeneity in attention-deficit/hyperactivity disorder: do we need neuropsychologically impaed subtypes? *Biol. Psychiatry* 57, 1224–1230. doi: 10.1016/j.biopsych.2004.08.025

Nikolas, M. A., Marshall, P., and Hoelzle, J. B. (2019). The role of neurocognitive tests in the assessment of adult attention-deficit/hyperactivity disorder. *Psychol. Assess.* 31, 685–698. doi: 10.1037/pas0000688

Pan, X.-X., Ma, H.-W., and Dai, X.-M. (2007). Value of integrated visual and auditory continuous performance test in the diagnosis of childhood attention deficit hyperactivity disorder. *Zhongguo Dang Dai Er Ke Za Zhi* 9, 210–212.

Pelham, W. E., Waschbusch, D. A., Hoza, B., Gnagy, E. M., Greiner, A. R., Sams, S. E., et al. (2011). Music and video as distractors for boys with ADHD in the classroom: comparison with controls, individual differences and medication effects. *J. Abnorm. Child Psychol.* 39, 1085–1098. doi: 10.1007/s10802-011-9529-z

Peng, X., Lin, P., Zhang, T., and Wang, J. (2013). Extreme learning machine-based classification of ADHD using brain structural MRI data. *PLoS One* 8:e79476. doi: 10.1371/journal.pone.0079476

Rousseau, C., Measham, T., and Bathiche-Suidan, M. (2008). DSM IV, culture and child psychiatry. *J.Can. Acad. Child Adolesc. Psychiatry* 17, 69–75.

Sakai, K., and Yamada, K. (2019). Machine learning studies on major brain diseases: 5-year trends of 2014–2018. *Jpn. J. Radiol.* 37, 34–72. doi: 10.1007/s11604-018-0794-4

Saria, S., Butte, A., and Sheikh, A. (2018). Better medicine through machine learning: what's real and what's artificial? *PLoS Med.* 15:e1002721. doi: 10.1371/journal.pmed.1002721

Saxe, G. N., Ma, S., Ren, J., and Aliferism, C. (2017). Machine learning methods to predict child posttraumatic stress: a proof of concept study. *BMC Psychiatry* 17:223. doi: 10.1186/s12888-017-1384-1

Sen, B., Borle, N. C., Greiner, R., and Brown, M. (2018). A general prediction model for the detection of ADHD and Autism using structural and functional MRI. *PLoS One* 13:e0194856. doi: 10.1371/journal.pone.0194856

Shmueli, G. (2010). To explain or to predict? *Statist. Sci.* 25, 289–310. doi: 10.1214/10-STS330

Slobodin, O., Cassuto, H., and Berger, I. (2018). Age-related changes in distractibility: developmental trajectory of sustained attention in ADHD. *J. Attent. Disord.* 22, 1333–1343. doi: 10.1177/1087054715575066

Slobodin, O., and Masalha, R. (2020). Challenges in ADHD care for ethnic minority children: a review of the current literature. *Transcult. Psychiatry* 57, 468–483. doi: 10.1177/1363461520902885

Sonuga-Barke, E. J., and Halperin, J. M. (2010). Developmental phenotypes and causal pathways in attention deficit/hyperactivity disorder: potential targets for early intervention? *J. Child Psychol. Psychiatry* 51, 368–389. doi: 10.1111/j.1469-7610.2009.02195.x

Tantithamthavorn, C., Hassan, A. E., and Matsumoto, K. (2018). The impact of class rebalancing techniques on the performance and interpretation of defect prediction models. *IEEE Trans. Softw. Eng.* 43, 1–18. doi: 10.1109/tse.2018.2876537

Tenev, A., Markovska-Simoska, S., Kocarev, L., Pop-Jordanov, J., Müller, A., and Candrian, G. (2014). Machine learning approach for classification of ADHD adults. *Int. J. Psychophysiol.* 93, 162–166. doi: 10.1016/j.ijpsycho.2013.01.008

Toplak, M. E., West, R. F., and Stanovich, K. E. (2013). Practitioner review: do performance-based measures and ratings of executive function assess the same construct? *J. Child Psychol. Psychiatry* 54, 131–143. doi: 10.1111/jcpp.12001

van der Heijden, K. B., de Sonneville, L. M., and Althaus, M. (2010). Time-of-day effects on cognition in preadolescents: a trails study. *Chronobiol. Int.* 27, 1870–1894. doi: 10.3109/07420528.2010.516047

Vieira, S., Gong, Q.-Y., Pinaya, W. H. L., Scarpazza, C., Tognin, S., Crespo-Facorro, B., et al. (2019). Using machine learning and structural neuroimaging

to detect first episode psychosis: reconsidering the evidence. *Schizophr. Bull.* 46, 17–26. doi: 10.1093/schbul/sby189

Wang, L.-J., Huang, Y.-S., Hsiao, C.-C., and Chen, C.-K. (2017). The trend in morning levels of salivary cortisol in children with ADHD during 6 months of methylphenidate treatment. *J. Attent. Disord.* 21, 254–261. doi: 10.1177/1087054712466139

Wolraich, M., Brown, L., Brown, R. T., DuPaul, G., Earls, M., Feldman, H. M., et al. (2011). ADHD: clinical practice guideline for the diagnosis, evaluation and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics* 128, 1007–1022. doi: 10.1542/peds. 2011-2654

Yahav, I., Shmueli, G., and Mani, D. (2016). A tree-based approach for addressing self-selection in impact studies with big data. *MIS. Quart.* 40, 819–848. doi: 10.25300/misq/2016/40.4.02

Yasumura, A., Omori, M., Fukuda, A., Takahashi, J., Yasumura, Y., Nakagawa, E., et al. (2017). Applied machine learning method to predict children with ADHD using prefrontal cortex activity: a multicenter study in Japan. *J. Atten. Disord.* 1087054717740632. doi: 10.1177/10870547177 40632