<u>Original Paper</u>

A Novel Approach Using Serious Game Data to Predict the WISC-V Processing Speed Index in Children With Attention-Deficit/Hyperactivity Disorder: Machine Learning Study

Jun-Su Kim[1*], MS; Yoo Joo Jeong[1,2,3*], BS; Seung-Jae Kim[1], MS; Su Jin Jun[1], BS; Jin-Yeop Park[1], BS; Hyang-Sook Hoe[1,2,3], PhD; Jeong-Heon Song[1], PhD

[1]AI-based Neurodevelopmental Diseases Digital Therapeutics Group, Korea Brain Research Institute (KBRI), Daegu, Republic of Korea
[2]Neurodegenerative Disease Group, Korea Brain Research Institute (KBRI), Daegu, Republic of Korea
[3]Department of Brain and Cognitive Sciences, Gyeongbuk Institute of Science & Technology, Daegu, Republic of Korea
*these authors contributed equally

Corresponding Author:

Jeong-Heon Song, PhD
AI-based Neurodevelopmental Diseases Digital Therapeutics Group
Korea Brain Research Institute (KBRI)
61, Cheomdan-ro
Daegu 41062
Republic of Korea
Email: jhsong@kbri.re.kr

Related Article:

This is a corrected version. See correction statement in: https://games.jmir.org/2025/1/e88500

Abstract

Background: The processing speed index (PSI) of the Korean Wechsler Intelligence Scale for Children-Fifth Edition (K-WISC-V) is highly correlated with symptoms of attention-deficit/hyperactivity disorder (ADHD) and is an important indicator of cognitive function. However, restrictions on the frequency of testing prevent short-term PSI assessments. An accessible, objective technique for predicting PSI scores would enable better short-term monitoring and intervention for children with ADHD.

Objective: To enable objective and accessible monitoring of cognitive function beyond traditional clinical assessments, this study aimed to develop a machine learning model that predicts the PSI scores of children with ADHD using behavioral data from serious games.

Methods: Sixty-eight children (6-13 y of age) with ADHD were recruited, and after excluding incomplete data, 59 participants were included in the final analysis. The participants completed an initial PSI assessment using the K-WISC-V followed by 25 minutes of engagement with serious game content. Data from the game sessions were used to train machine learning models, and the models' performance in predicting PSI scores was evaluated using the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percent error (MAPE), with K-fold cross-validation (k=4) applied to ensure robustness.

Results: Among the individual machine learning models, support vector regression (SVR) had the best performance, with the lowest RMSE of 11.288, MAE of 7.874, and MAPE of 7.375%. The best overall performance was achieved by the ensemble integrating AdaBoost, Elastic Net, and SVR, which recorded the lowest RMSE of 10.072, MAE of 6.798, and MAPE of 6.611%. The predictive accuracy of this ensemble model was highest for PSI scores near the mean value of 100, demonstrating its reliability for clinical applications.

Conclusions: The developed PSI prediction model has the potential to serve as an objective and accessible tool for monitoring cognitive function in children with ADHD. As a complement to traditional assessments, this approach allows continuous tracking of symptom changes and can support more personalized treatment planning in both clinical and everyday settings, which may improve accessibility and adherence. However, the findings need to be validated in larger, more diverse populations, and the long-term feasibility of using serious games in clinical and educational settings must be further examined.

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that progresses throughout childhood and continues into adulthood [1]. Accurate assessment and management are essential for determining the best time to intervene and improve treatment outcomes, but the diagnosis of ADHD and other neurodevelopmental disorders often relies on subjective clinical evaluations, which can result in inconsistent or inaccurate diagnoses [2-4]. Additionally, symptom evaluation outside clinical environments, such as at home or school, typically depends on observation rather than objective measurements [5]. Processing speed (PS), that is, the ability to respond to a given stimulus within a limited time frame, may offer valuable insights for treatment and monitoring [6]. Individuals with ADHD exhibit slower PS, which has been shown to hinder an individual's ability to process and respond to information efficiently, negatively affecting academic performance, adaptability, anxiety levels, and social competence [7,8]. Lower scores on PS measures, such as the processing speed index (PSI), are correlated with greater severity of attention deficits [9].

The PSI concept was first introduced in the Third Edition of the Wechsler Intelligence Scale for Children (WISC-III), and a formal composite score was established in the Fourth Edition (WISC-IV) [10,11]. The WISC is one of the most widely used and highly validated tools for assessing cognitive function in children, making its index scores highly reliable [12-14]. The latest version, WISC-V, includes the PSI as one of its indices and is frequently used with children in both clinical and educational settings [12]. Due to its strong psychometric properties, the PSI is considered a reliable indicator of PS. Beyond its utility in diagnosis, the PSI is also associated with daily functioning and provides a useful way to quantify ADHD symptoms, highlighting its importance in managing ADHD. However, PSI assessment has two key limitations. First, for effective PSI measurement, trained professionals, such as clinical psychologists or physicians, must administer and interpret specialized tests. Second, repeated testing can lead to practice effects, which may reduce data validity [15-17]. Consequently, the recommended usage interval for WISC is 1-to 2.6 years, making it inappropriate for continuous monitoring [18-20].

To overcome the methodological limitations of measuring the PSI, this study aims to develop a machine learning model that predicts PSI scores in children with ADHD. The model uses behavioral characteristic data collected from Neuro-World, an artificial intelligence-based serious game classified as Software as a Medical Device (SaMD). Serious games and other digital therapeutics (DTx) have emerged as promising tools for treating neurodevelopmental disorders like ADHD [21-24]. DTx allows continuous, location-independent monitoring, improving adherence and enabling real-time tracking [25,26]. These advantages of DTx may also be useful for cognitive assessment. For example, a recent study demonstrated the feasibility of using mobile platforms for assessing PS and working memory via momentary assessment methods [27]. Building on this, large-scale population studies have used mobile cognitive games to evaluate PS across different age groups and cognitive domains [28]. Although digital tools have shown promise for assessing PS through mobile technologies, most have not been validated against established neuropsychological indices such as the PSI. Moreover, while some research has used machine learning models to predict mild cognitive impairment (MCI) based on Mini-Mental State Examination (MMSE) scores, no study has specifically focused on predicting the PSI from behavioral data [29,30]. In addition, preliminary studies have primarily focused on detecting age-related differences or demonstrating feasibility rather than predicting clinically meaningful cognitive metrics.

Based on the well-established association between ADHD severity and a lower PSI [8], as well as prior evidence of improvements in WISC-V scores following Neuro-World training [31], this study aimed to monitor and track cognitive function related to PS. A systematic evaluation of multiple machine learning algorithms identified the model that exhibited the highest level of performance while maintaining the accuracy and reliability required for PSI prediction from game-derived behavioral data.

Methods

Participants

This study was conducted with pediatric participants aged 6-13 years with a diagnosis of ADHD. A total of 34 children with ADHD were recruited between December 2021 and December 2022 (IRB no. 2021-10-080), and an additional 34 children diagnosed with ADHD were recruited between December 2022 and March 2023 (CRIS Registration no. KCT0009326). However, 4 participants were excluded as they refused to participate. The total sample size of the study was 64 participants: 51 (79.7%) male and 13 (20.3%) female participants. The average age of the participants was 8.5 (SD 1.7) years. The participants were all South Korean and were recruited from a single clinical institute. In this study, demographic factors such as socioeconomic status and the education level of the legal guardian were not collected.

Ethical Considerations

This study was approved by two different institutional review boards (IRBs) due to the different recruitment phases. The initial phase of the study was reviewed and approved by the IRB of Keimyung University Dongsan Hospital (IRB no. 2021-10-080), which is certified by the Korea Laboratory Accreditation Scheme (KOLAS), a nationally recognized accreditation body in South Korea. The subsequent phase of the study was registered with the Korea Clinical Research Information Service (CRIS Registration no. KCT0009326), which is accredited by the World Health Organization (WHO). Participants were recruited from the Department of Psychiatry at Keimyung University Dongsan Hospital. Before study enrollment, written informed consent was obtained from the legal guardians of all participants, and the study procedures were explained to the participants to secure their assent. The consent form specified that the collected data could be used not only for the primary study but also for future research and secondary analyses without requiring additional consent. All collected data were deidentified and anonymized to ensure confidentiality. As compensation, participants were provided with coverage for transportation costs, and their legal guardians received professional psychological assessment results and evaluation reports. No identifiable images of participants are included in this manuscript or supplementary materials, and no personal information was disclosed.

Experimental Methodology

To minimize disruptions to routine that might affect performance on attention-based activities, we requested that the participants follow their usual daily routines and take their prescribed medications as scheduled. After the participants completed consent forms, they underwent a psychological assessment using the K-WISC-V, which took 1 hour. The participants were provided with a 10-minute break after the psychological assessment and performed gameplay with the developed serious game content for 25 minutes. The overall clinical trial process is presented in Figure 1.

Figure 1. Clinical trial process. A total of 64 children diagnosed with ADHD participated in this study. After completing consent forms, the participants underwent a 1-hour psychological assessment using K-WISC-V. After the K-WISC-V assessment, the children performed a digital assessment using Neuro-World software for 25 minutes. Artificial intelligence–generated image (Generator: ChatGPT, OpenAI; July 14, 2025; Requestor: Jun-Su Kim). ADHD: attention-deficit/hyperactivity disorder; K-WISC-V: Korean Wechsler Intelligence Scale for Children-Fifth Edition.
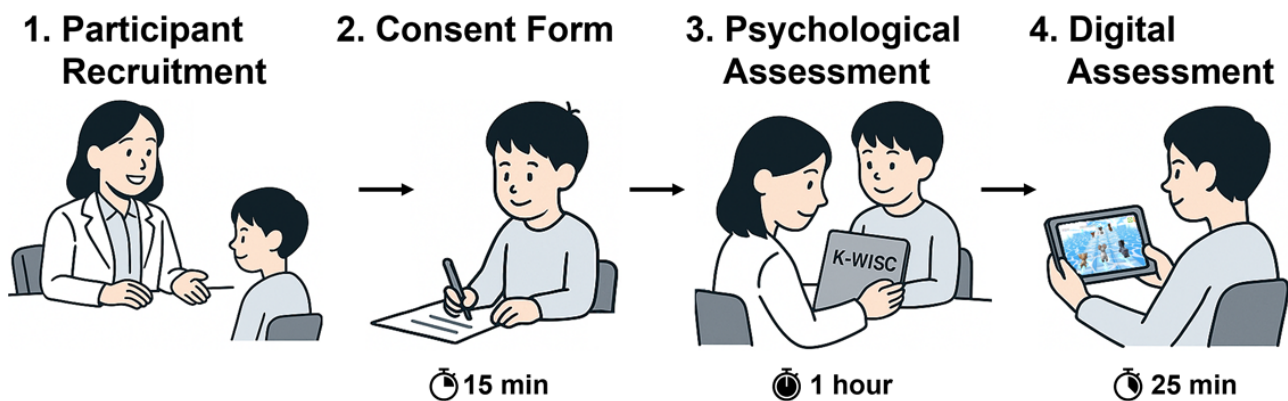
To ensure an objective setting for gameplay, all participants performed the serious game content independently in a quiet environment free from external distractions and noise. To ensure participant comfort, a pause button was provided, which allowed participants to stop if they experienced dizziness or required a break.

K-WISC-V: Processing Speed Index (PSI)

WISC is widely recognized as a comprehensive instrument for assessing cognitive functioning in children [31]. The WISC-V has been translated and culturally adapted for use in Korea, and this version is known as the Korean Wechsler Intelligence Scale for Children-Fifth Edition (K-WISC-V). The K-WISC-V is targeted to children and adolescents (aged 6 to 16 years and 11 mo) and includes core indices for the verbal comprehension index (VCI), perceptual reasoning index (PRI), working memory index (WMI), and PSI [32, 33]. Each assessment is conducted individually, with scores calculated based on established criteria and then converted to age- and gender-specific scaled scores. The index scores are standardized with a mean of 100 and an SD of 15, enabling comparisons of an individual's abilities with those in the same age group [34].

PSI, a critical component of the K-WISC-V, assesses visual PS and psychomotor efficiency by measuring reaction time, spatial navigation, discrimination, and accuracy. Attention and hand-eye coordination are the major contributing factors to the PSI [35]. The PSI captures learning efficiency, attentional control, and executive functioning, which are related to various cognitive dimensions in children. Higher PSI scores indicate faster and more accurate visual decision-making, reflecting greater efficiency in time management for learning tasks and assessments [36]. Symbol search and coding, the two core subtests of PSI, are described in Table 1.

Table 1. Detailed description of core PSI subtests.

Subtest	Description
Symbol Search	A target symbol is given to the participants, who are asked to identify and mark the matching symbols in a given set within a limited time. This task assesses visual search speed and coordination-discrimination ability.
Coding	A set of numbers is matched with symbols, and participants are required to write down the symbols corresponding to a given set of numbers within a limited time. This task evaluates visuomotor coordination, associative learning ability, and visual short-term memory.

Serious Game Contents for PS Measurement

PS refers to the ability to quickly complete simple tasks while maintaining focus. Cognitive function assessments typically use a brief neuropsychological task that requires approximately 3-5 minutes [37], but neuropsychological tasks used to evaluate sustained attention generally require continuous engagement for 15 to 20 minutes [37]. Based on these observations, we designed serious game sessions with a total duration of 25 minutes. Several recent studies have found that interventions using the software Neuro-World, developed by Woorisoft Co, Ltd (Republic of Korea), improve decision-making ability, overall cognitive PS, working memory, and ADHD-related symptoms, including attention deficit, hyperactivity, and internalization and externalization disorder, in children with ADHD [38-40]. Building on these findings, Neuro-World was used in this study to monitor and track cognitive function related to PS in pediatric participants diagnosed with ADHD.

As shown in Figure 2, Neuro-World comprises 5 independent serious game content modules. Each module takes approximately 5 minutes to complete; thus, the whole digital assessment procedure involves a total of 25 minutes of engagement. The structure of the assessment was designed to assess both cognitive performance and sustained attention within a single session. The sequence of content modules was randomized, and the next content module began automatically after completion of the previous content module. To address potential device or platform bias, all participants completed the game-based assessments using standardized Android tablets (≥9.7-inch screens) with identical touch screen interfaces. No external input devices were used. Table 2 summarizes the technical specifications of the devices and software.

Figure 2. Five Neuro-World serious game content modules used to evaluate processing speed. Content 1: identifying and moving animals; content 2: rearranging sequences; content 3: avoiding obstacles; content 4: dual obstacle avoidance; and content 5: memory matching.

Table 2. Device and software specifications used in the Neuro-World-based digital assessment.

Component	Specification
Item name	Cognitive Therapy Software
Model name	Neuro-World DTx-ADHD

Component	Specification
Operating system	Android
Device type	Tablet (9.7 inches or larger)
Input method	Touch screen

Behavioral Feature Extraction From the Serious Game

The following methodologies were used to extract behavioral characteristics from the 5 distinct content modules. First, the complexity of each content module was dynamically adjusted based on the participant's performance:

- Play speed: As the difficulty level increased, the speed of content progression accelerated, intensifying the sense of urgency during task performance.
- Number of items to remember: The cognitive load was gradually increased by requiring the participants to memorize a larger number of items.

Previous studies have shown that the variability of response time is strongly correlated with ADHD [41-43]. Based on these observations, the game content was designed to gradually reduce the effective time available for problem-solving as the difficulty increased, requiring participants to process information more quickly and efficiently.

Second, a series of data were collected during the cognitive assessment using the Neuro-World content modules: $x1$, total number of trials conducted in one content module; $x2$, accuracy rate; $x3$, total score; $x4$, total number of screen touches; $x5$, unnecessary actions during content module performance, calculated as the difference between the total number of screen touches and the number of touches required for correct answers; $x6$, highest level of performance; $x7$, average level of performance; $x8$, SD of performance level; $x9$ and $x10$, average time to touch the screen after a task was presented (ie, response time) and its SD; $x11$ and $x12$, average time to the final touch required to complete each game (ie, task completion consistency) and its SD; and $x13$ and $x14$, average decision-making time, calculated as the difference between the final and first touch times and its SD.

Exploratory studies have suggested that individuals with ADHD exhibit deficits in sustained attention [44-47]. Based on these findings, we divided the total content performance into 2 phases and aimed to quantitatively assess whether the participants' performance improved over time or deteriorated due to reduced concentration, thereby capturing changes in their cognitive task-solving processes. The 14 variables were recorded separately in the 2 phases: $x1{:}14$ (initial phase) and $x15{:}28$ (later phase). The 28 features were extracted from each of the 5 content modules, resulting in a total of 140 variables ($x1{:}140$) [48].

Data Preprocessing

Machine learning models, including support vector regression (SVR), random forest, and AdaBoost, generally require a sample size that is at least 10- 20 times larger than the number of input features to ensure stable learning performance [49]. In this study, the initial features consisted of 140 variables extracted from serious game performance data. To minimize the risk of overfitting and enhance the stability of the prediction model, dimensionality reduction was required. Accordingly, to extract the appropriate variables from the data to enhance the model's predictive performance, the correlation-based feature selection (CFS) method was applied. CFS is a filter-based method that identifies an optimal set of variables by maximizing the correlations of the variables with the PSI while minimizing correlations between variables [50]. The effectiveness of the selected variable set in CFS is measured using <Equation 1>. In this equation, *Merits* represent the cumulative correlation between the variables and the external variable; k denotes the number of variables; $\overline{r_{zi}}$ represents the average correlation between variables and the external variable; and $\overline{r_{ii}}$ denotes the average intercorrelation between variables [51-53]. The value of *Merits* corresponds to the usefulness of the variable set for prediction; a higher value indicates a stronger correlation with the target class and reduced redundancy among variables. A higher value of *Merits* therefore signifies not only that the variable set is more closely aligned with the target class but also that it contributes to improved model performance by reducing irrelevant or redundant information.

$$Merits = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{ii}}}} \tag{1}$$

Using the CFS method, 6 variables were selected from 140 content variables. In addition, the variables gender and age were included, resulting in a total of 8 variables used as inputs for the machine learning models in the analysis. The selected variables are listed in Table 3.

Table 3. List of input variables for the machine learning model.

Index	Variable (xn)	Descriptions
1	Con1_answer_rate_back ($x16$)	Accuracy rate in the later phase of content 1
2	Con2_touch_count_diff_back ($x47$)	Number of unnecessary touches in the later phase of content 2
3	Con3_user_touch_count_back ($x74$)	Total number of touches in the later phase of content 3
4	Con5_total_score_front ($x115$)	Total score in the initial phase of content 5

Index	Variable (x_n)	Descriptions
5	Con3_first_touch_time_mean_front (x_{65})	Mean response time to the first touch in the initial phase of content 3
6	Con1_decision_time_mean_front (x_{13})	Mean decision-making time in the initial phase of content 1
7	Gender	Participant's gender
8	Age	Participant's age

Machine Learning Models Used for Predicting the PSI

To predict the PSI scores of children diagnosed with ADHD, we collected performance data from serious games, trained various machine learning models using the selected variables, and compared the predictive performance of the models. We utilized 4 different machine learning models: random forest, AdaBoost, SVR, and Elastic Net. The characteristics of each model are described below.

Random forest is an algorithm that enhances prediction performance by ensemble learning with multiple decision trees [54]. The random forest algorithm constructs regression trees, processes an input vector composed of feature values corresponding to specific training instances, and averages the predicted outputs from each tree [55]. After creating several training data subsets, a bagging method is applied to reduce correlations among different decision trees to reduce variance [56]. This bagging characteristic enhances the predictive accuracy and stability of the random forest algorithm [54].

AdaBoost (Adaptive Boosting) is a boosting algorithm that iteratively trains a series of weak learners and combines them to form a robust ensemble model [57]. Unlike other boosting methods, AdaBoost applies the relative instead of absolute error, giving more weight to examples with lower values [58]. Additionally, AdaBoost continues to run even if the error rate exceeds 0.5 and supports user-defined iterations, allowing performance improvement over initial iterations [58].

SVR solves regression problems by extending binary classification through convex optimization formulations [59]. SVR generalizes support vector machines to adapt to regression problems and introduces an ε-insensitive region (tube) to reformulate the optimization problem [60]. Unlike other multivariate regression models, support vector machines can learn nonlinear functions with the kernel trick, which maps independent variables in a higher dimensional feature space [61]. The ε-insensitive region approximates continuous functions well and balances model complexity with prediction errors [60].

Elastic Net, proposed by Zou and Hastie in 2005, was developed to estimate the regression coefficient ß in situations where p>n [62]. Elastic Net combines the concepts of Lasso and Ridge Regression to estimate ß [63]. The l2-norm penalty tends to reduce coefficients toward zero while retaining all predictors in the Elastic Net model, potentially leading to multicollinearity, whereas the l1-norm penalty automatically selects key information and continuously reduces redundant information [64]. By combining both penalties, Elastic Net effectively mitigates multicollinearity and enables the group selection of correlated features [65,66].

Machine Learning Model Training and Hyperparameter Optimization for PSI Prediction

To develop an optimal PSI prediction model, we trained the machine learning model and specifically optimized the hyperparameters of each model. No power analysis was conducted because of the exploratory nature of the study, but the use of dimensionality reduction and cross-validation mitigated overfitting and supported generalizability. To prevent issues such as limited generalizability due to the small sample size, we implemented a robust machine learning framework incorporating dimensionality reduction, k-fold cross-validation, hyperparameter tuning via grid search, and random shuffling to mitigate overfitting. Of the initial dataset comprising 64 participants, 5 were excluded because of missing data. Thus, 59 datasets were available for analysis, divided into a training set of 47 (80%) and an independent test set of 12 (20%).

Given the relatively small sample size of this machine learning algorithm study, we also used k-fold cross-validation to ensure robust model validation. The number of folds (k) can be flexibly determined based on the dataset size and the model's structural complexity and inherent characteristics [67-69]. Following this rationale, we experimented with different k values (k=3, 4, 5, and 6) and chose k=4 to achieve a balanced distribution among the training, validation, and test sets. In each fold, approximately 35-36 samples were used for training and 11-12 samples were used for validation. This configuration ensured that the size of the validation set closely matched the final test set, allowing for more effective parameter tuning under conditions similar to final testing.

Generalizability was further enhanced by performing randomized shuffling in each fold (shuffle=True), ensuring that the final test set remained completely unseen throughout training and tuning. A grid search was performed within each fold to optimize hyperparameters, and the optimal settings for each model are presented in Table 4. For example, the random forest model achieved the best performance with "max_features=log2", "n_estimators=50", "max_depth=10", and "min_samples_split=2".

Table 4. Hyperparameter grids and selected optimal values for each machine learning model.

Model and hyperparameter	Grid search range	Optimal value
Random forest		
max_features	['sqrt', 'log2']	log2
n_estimators	[50, 100, 200, 300]	50
max_depth	[3,5,7,10]	10
min_samples_split	[2-5]	2
AdaBoost		
n_estimators	[50, 100, 200, 300]	200
max_depth	[3-10]	5
learning_rate	[0.01, 0.05, 0.1, 0.5, 1.0]	1.0
Elastic Net		
alpha	[0.1, 0.5, 1.0, 2.0, 5.0]	0.1
l1_ratio	[0.1, 0.3, 0.5, 0.7, 0.9]	0.9
Support vector regression		
C	[0.1, 1, 10, 100]	10.0
epsilon	[0.01, 0.1, 1]	1.00
kernel	['linear', 'rbf', 'poly']	linear

Results

The SVR Machine Learning Model Effectively Predicts PSI Scores From the Serious Game Performance Outcomes of Children Diagnosed With ADHD

The predictive performance of the 4 machine learning models was compared to determine which was most effective in predicting the PSI scores of children diagnosed with ADHD from their Neuro-World content performance outcomes (Table 5). SVR demonstrated the best overall performance on the training set, achieving a root mean squared error (RMSE) of 13.107, mean absolute error (MAE) of 9.933, and mean absolute percent error (MAPE) of 9.792% (Table 5). The Elastic Net model exhibited the second highest performance on the training set, with an RMSE of 13.458, MAE of 10.580, and MAPE of 10.352% (Table 5). Random forest and AdaBoost had higher errors on the training set, with RMSE values of 14.307 and 14.269, respectively.

Table 5. Comparison of machine learning model performance for predicting processing speed index (PSI) scores of children diagnosed with attention-deficit/hyperactivity disorder (ADHD).

Model	Training set			Test set		
	RMSE[a]	MAE[b]	MAPE[c] (%)	RMSE	MAE	MAPE (%)
Random forest	14.307	11.813	11.828	15.517	13.165	12.314
AdaBoost	14.269	11.676	11.738	14.511	12.092	11.324
Elastic Net	13.458	10.58	10.352	12.488	8.328	7.993
SVR[d]	13.107	9.933	9.792	11.288	7.874	7.375

[a]RMSE: root mean squared error.

[b]MAE: mean absolute error.

[c]MAPE: mean absolute percentage error.

[d]SVR : support vector regression.

We then investigated the performance of each model on the test set and found that SVR continued to outperform the other models, with an RMSE of 11.288, MAE of 7.874, and MAPE of 7.375% (Table 5). As with the training set, Elastic Net was second, with an RMSE of 12.488, MAE of 8.328, and MAPE of 7.993%, demonstrating the ability of this model to provide accurate predictions (Table 5). Random forest and AdaBoost again displayed weaker performance. Random forest had an RMSE of 15.517, MAE of 13.165, and MAPE of 12.314%, while AdaBoost had an RMSE of 14.511, MAE of 12.092, and MAPE of 11.324% (Table 4). Overall, SVR emerged as the most effective model for predicting PSI across both the training and test sets.

To improve the interpretability of the machine learning model, SHAP (SHapley Additive exPlanations) analysis was conducted to assess the contribution of each feature to PSI prediction. Figure 3 presents a summary plot of SHAP analysis results for all machine learning models based on the top 8 input variables for each model (detailed descriptions of these 8 input variables are provided in Table 3). Across all models, "Con5_total_score_front", "Con1_answer_rate_back", "age", and "Con1_decision_time_mean_front" consistently emerged

as the main prediction variables. While the relative importance of individual features varied across models, the behavioral performance indices derived from the serious games, including response rate, touch time, and decision time, consistently demonstrated a high degree of predictive efficacy. Gender had a negligible impact on all models. The results of the study support the feasibility of using behavioral performance metrics derived from digital games to predict the PSI scores of children diagnosed with ADHD.

Figure 3. SHAP plot for each machine learning model: (A) Random forest, (B) AdaBoost, (C) Elastic Net, and (D) support vector regression (SVR). Each model represents the top 8 input variables contributing to processing speed index (PSI) prediction. Each dot indicates the SHAP value for an individual participant, and the color indicates the initial value (red: high, blue: low). Features are aligned in order of absolute average SHAP value, and higher values indicate a larger effect of the factor on model output. SHAP: SHapley Additive exPlanations.
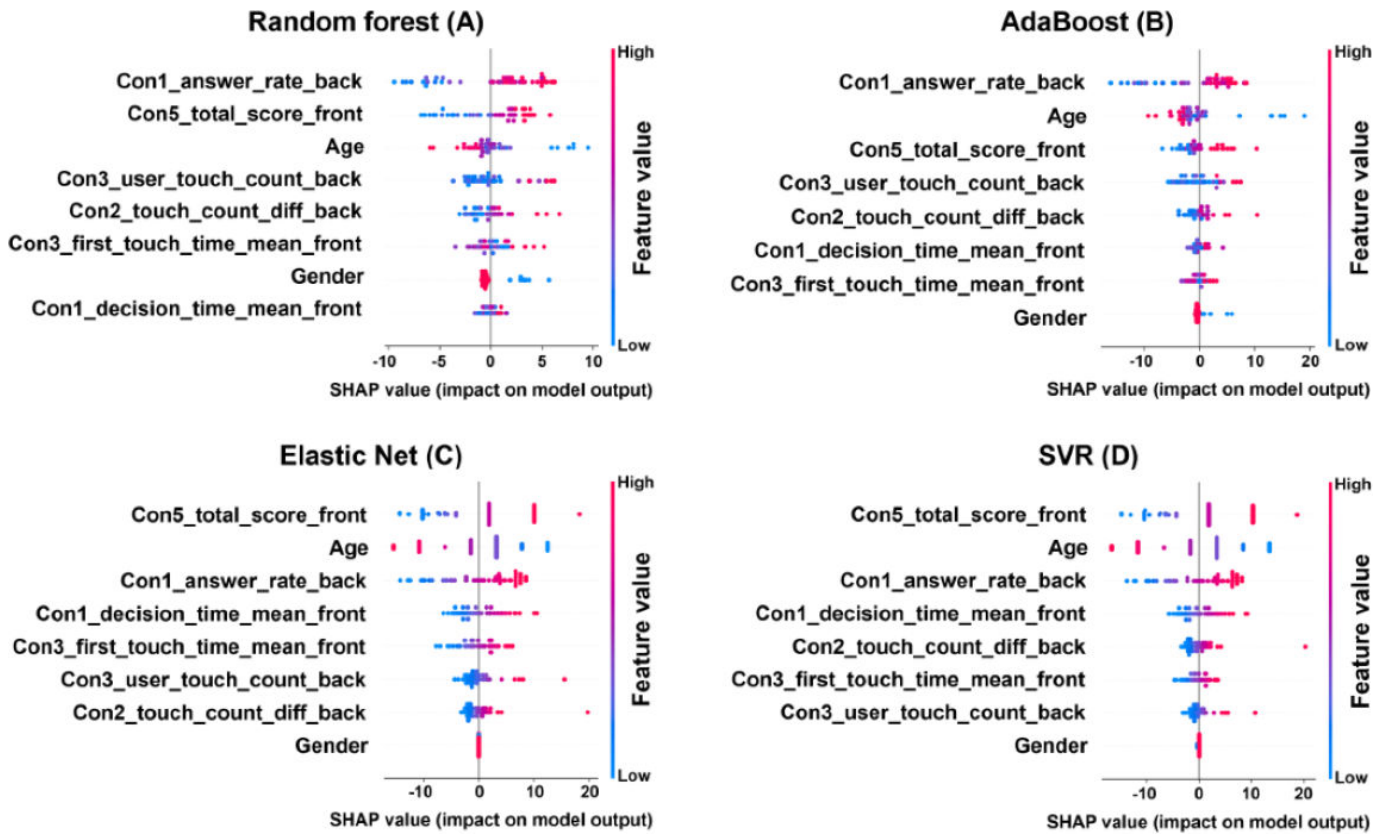
The Ensemble of AdaBoost and Elastic Net Model Exhibits High Training Performance for Predicting PSI Scores of Children With ADHD

The ensemble technique combines the prediction results of multiple individual machine learning models to achieve superior accuracy and stability performance compared with a single machine learning model [70]. Therefore, to enhance predictive performance by leveraging the strengths of various machine learning models, we applied an ensemble technique using the performance dataset of children with ADHD playing serious game content modules. Specifically, ensemble combinations of random forest, AdaBoost, Elastic Net, and SVR were applied to the training set. The top 5 model combinations with the lowest RMSE scores are presented in Table 6. The combination of AdaBoost and Elastic Net achieved the lowest RMSE of 12.738, indicating the strongest PSI score predictive performance. This combination also demonstrated competitive accuracy, with an MAE of 10.116 and MAPE of 10.086% (Table 6). By contrast, the combination of random forest and Elastic Net had the lowest MAE and MAPE scores of 9.878 and 9.816%, respectively, along with an RMSE of 12.781 (Table 6). These findings showed that combinations of tree-based models (AdaBoost, random forest) with a regression-based model (Elastic Net) outperformed the individual models by leveraging complementary strengths, improving overall predictive performance.

Table 6. Top 5 ensemble model combinations ranked by RMSE performance on the training set for predicting PSI[a] scores of children with ADHD[b].

Combination	Training set		
	RMSE[c]	MAE[d]	MAPE[e] (%)
[AdaBoost, Elastic Net]	12.738	10.116	10.086

Combination	Training set		
	RMSE[c]	MAE[d]	MAPE[e] (%)
[AdaBoost, Elastic Net, SVR[f]]	12.770	10.013	9.938
[Random forest, Elastic Net]	12.781	9.878	9.816
[Random forest, AdaBoost, Elastic Net, SVR]	12.802	10.135	10.071
[Random forest, Elastic Net, SVR]	12.821	9.930	9.829

[a]PSI:processing speed index.
[b]ADHD: attention-deficit/hyperactivity disorder.
[c]RMSE: root mean squared error.
[d]MAE: mean absolute error.
[e]MAPE: mean absolute percentage error.
[f]SVR: support vector regression.

The Ensemble of AdaBoost, Elastic Net, and SVR Exhibits High Testing Performance for Predicting PSI Scores of Children With ADHD

The performance of the ensemble machine learning combinations was then tested using the remaining 12 datasets of Neuro-World content performance of children with ADHD. The combination of AdaBoost, Elastic Net, and SVR resulted in the lowest RMSE of 10.072, with MAE and MAPE values of 6.798 and 6.611%, respectively, indicating high performance on the test set (Table 7). The combination of AdaBoost and Elastic Net, which showed the highest training performance for predicting PSI scores, exhibited the second highest RMSE of 10.089, with MAE and MAPE values of 7.126 and 6.946%, respectively (Table 7). By contrast, the combination of random forest and Elastic Net, which performed strongly on the training set and had the lowest MAE of 9.878, exhibited higher error rates on the test set, with an MAE of 7.788. Together, these results suggest that the PSI prediction accuracy of the ensemble model of AdaBoost and Elastic Net, which provided consistent performance on both the training and test sets, was enhanced by the inclusion of SVR when applied to the test set.

Table 7. Test set results of the top 5 ensemble model combinations from the training set for predicting PSI[a] scores of children with ADHD[b].

Combination	Test set		
	RMSE[c]	MAE[d]	MAPE[e] (%)
[AdaBoost, Elastic Net]	10.089	7.126	6.946
[AdaBoost, Elastic Net, SVR[f]]	10.072	6.798	6.611
[Random forest, Elastic Net]	10.432	7.788	7.578
[Random forest, AdaBoost, Elastic Net, SVR]	10.384	7.411	7.272
[Random forest, Elastic Net, SVR]	10.203	7.202	7.035

[a]PSI; processing speed index.
[b]ADHD: attention-deficit/hyperactivity disorder.
[c]RMSE: root mean squared error.
[d]MAE: mean absolute error.
[e]MAPE: mean absolute percentage error.
[f]SVR: support vector regression.

Comparison of the PSI Score Predictions of the Ensemble Model of AdaBoost, Elastic Net, and SVR With Actual PSI Scores

Since we found that the ensemble model of AdaBoost, Elastic Net, and SVR exhibited high test performance for predicting the PSI scores of children with ADHD, we visualized the prediction scores and actual PSI values on a graph (Figure 4). The x-axis presents the actual PSI scores, while the y-axis presents the scores predicted using the ensemble model of AdaBoost, Elastic Net, and SVR. The data points on the dashed line ($y = x$) indicate accurate prediction of the actual PSI values. The regression line, expressed as $y = 0.58x + 43.72$, reflects the relationship between the predicted and actual PSI scores.

Figure 4. Comparison of actual processing speed index (PSI) values with those predicted by ensemble application. The dashed line (y=x) represents perfect correspondence of the predicted and actual PSI values, and the red line (y=0.58x+43.72) is the regression line of the relationship between the predicted and actual PSI scores.
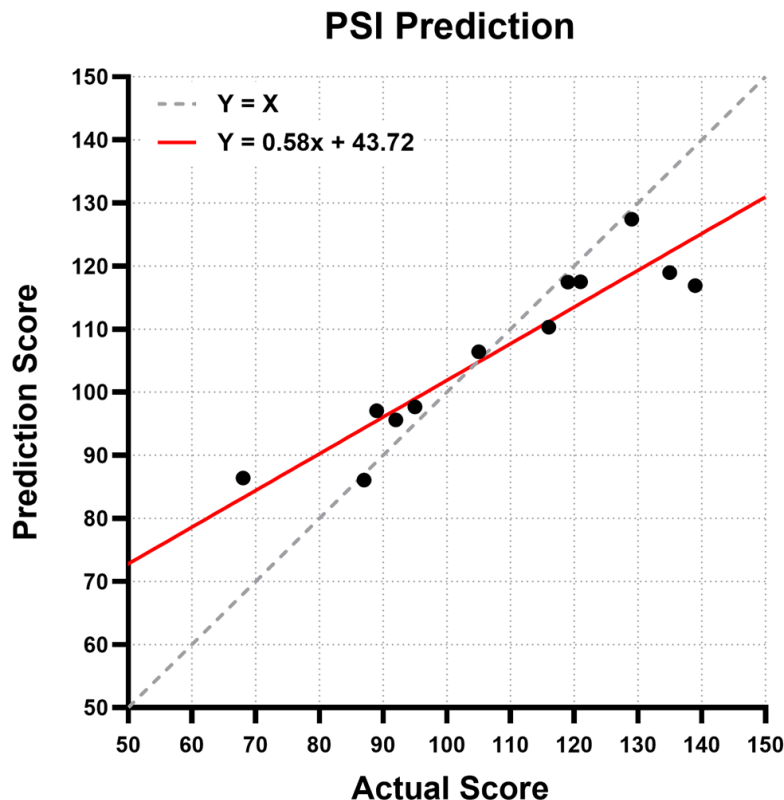
PSI scores typically follow a normal distribution with a mean value of 100, and the intersection of the regression line with the $y = x$ line near a PSI score of 100 suggests higher prediction accuracy for PSI values close to the mean. This alignment highlights the model's strong performance in accurately predicting PSI scores within the most densely distributed range. Furthermore, most data points closely follow the $y = x$ line, underscoring the model's reliability and robust predictive capability.

Discussion

Principal Findings

Objective methods for tracking ADHD symptoms remain limited in both clinical and nonclinical settings. Due to the lack of standardized measurement tools, symptom monitoring often relies on observation, especially in home and school environments. Although the results of WISC-V assessments are closely linked to daily functioning and thus can be used to quantify symptoms and inform management strategies, such clinical assessments can only be conducted by qualified professionals and cannot be repeated for approximately 2 years. Because traditional clinical assessment tools can only be readministered after a considerable period following the initial psychological evaluation, it is difficult to continuously and objectively monitor symptom improvement.

While serious games can be effective means of monitoring progress between hospital visits, most are primarily designed for training and rely on subjective evaluations or intrinsic performance metrics such as task duration and accuracy rates, which have a limited correlation with actual symptom improvement.

To overcome these limitations, this study aimed to integrate serious game performance data with clinically relevant assessment outcomes for ADHD. Specifically, we developed a machine learning-based algorithm to predict PSI scores from behavioral data collected from 59 children aged 6-13 years diagnosed with ADHD and evaluated the algorithm's performance. We applied the CFS method to extract 6 variables from the serious game performance data, which were input along with gender and age into machine learning models for analysis. Given the relatively small sample size, we applied cross-validation with varying k values (3, 4, 5, and 6) to enhance the robustness of model evaluation. Among these, k=4 provided the most stable performance without evidence of overfitting. Among the individual machine learning models, SVR was the most effective in predicting the PSI scores of children with ADHD. The combination of AdaBoost and Elastic Net model had the highest training performance for predicting PSI scores, while the ensemble of AdaBoost, Elastic Net, and SVR model had the highest test performance. Ultimately, we developed a model capable of predicting PSI scores using game

performance data, enabling estimation of the PSI comparable to that obtained from expert-administered WISC-V assessments (Figure 5).

Figure 5. Limitations of traditional digital therapeutics (DTx) monitoring and the need for effective tracking.
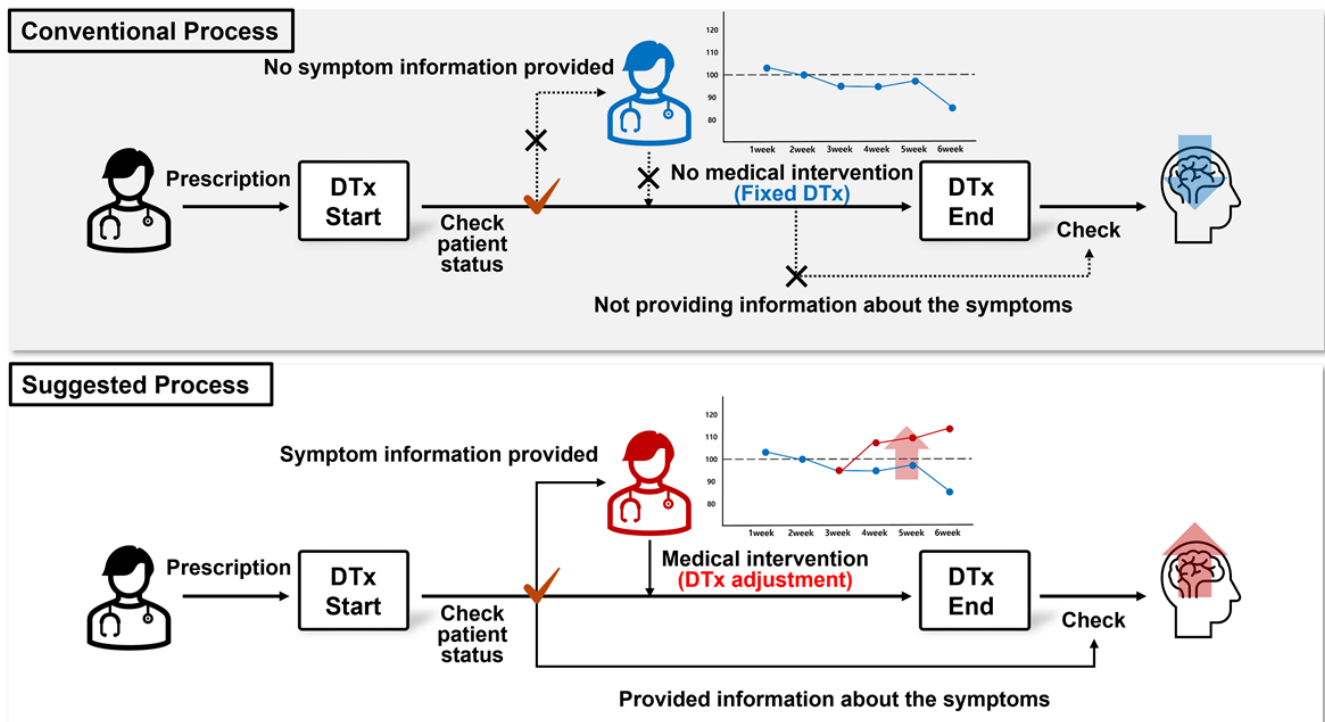
How can our PSI prediction model be applied in clinical practice or ADHD interventions? The predicted PSI scores can be used to assess symptom improvement using clinically meaningful metrics without time or location constraints. Thus, the model may be feasibly integrated into ADHD care to support ongoing monitoring of cognitive function by complementing or partially replacing standardized assessments such as the full WISC-V, allowing more frequent and accessible tracking of cognitive changes. For instance, regular evaluation of PS using the model may give clinicians continual insights into the treatment responses of pediatric patients with ADHD, enabling timely and precise treatment adjustments. Prior studies have demonstrated the clinical value of machine learning-based predictive models. For instance, a cardiovascular disease risk prediction model developed using machine learning outperformed traditional models in terms of accuracy [71]. Another risk prediction model for cardiovascular disease also demonstrated improved predictive accuracy, particularly in high-risk subgroups such as individuals with diabetes [72]. Machine learning has been shown to enhance both discrimination and calibration in cardiovascular risk prediction, underscoring its potential to improve personalized care in clinical settings [73]. Collectively, these findings support the applicability of machine learning-based PSI prediction for not only diagnosing but also managing and monitoring ADHD.

Our comparison of individual machine learning models revealed that SVR had the highest accuracy for predicting PSI scores, followed by Elastic Net, while random forest and AdaBoost had higher errors on both the training and test sets. Although SVR achieved the highest accuracy in both the training and test evaluations, the 3 other algorithms also exhibited strengths. It is difficult for a single model to achieve sufficient predictive stability and generalizability, and previous studies have shown that leveraging multiple models can substantially improve predictive performance [74]. For example, when multiple models were combined to analyze digital therapeutic log data, accuracy comparisons yielded determination coefficients (R^2) ranging from 0.82 to 0.86 [74]. These results support the idea that algorithmic diversity is a key factor in the enhanced efficacy of ensemble methods. When we adopted an ensemble approach in this study, we found that the combination of AdaBoost and Elastic Net outperformed SVR, and the best overall performance was achieved by the ensemble integrating AdaBoost, Elastic Net, and SVR, which was superior to any individual model. These improvements suggest that integrating diverse algorithmic techniques—such as adaptive boosting, regularized linear regression, and complementary model tuning—can significantly improve the reliability of PSI score predictions, a critical factor in ADHD symptom tracking.

The regression line of the PSI prediction accuracy of our model intersected y=x at a PSI score of approximately 100, indicating stable predictive performance in the clinically significant range (approximately 85-115). This is particularly important for effective monitoring and timely interventions, underscoring the practical applicability of our ensemble model. Similarly, HT Jung et al found that the regression

line for MMSE score prediction for MCI crossed y=x near the critical cutoff of 27, indicating high accuracy in distinguishing normal cognition from MCI as well as stability in clinically meaningful ranges [48].

Our findings suggest that the PSI prediction model using serious games can contribute to the quantitative and objective tracking of ADHD symptoms. The ensemble approach effectively compensates for the limitations of individual models, improving predictive stability and accuracy. This method holds promise for clinical use to support continuous patient monitoring and optimization of treatment plans. Furthermore, by integrating this tool into everyday environments such as homes or schools, clinicians may gain valuable insights into symptom tracking and treatment progress beyond those attainable by traditional approaches.

Limitations

Notwithstanding its contributions, this study has several limitations.

First, this study included 64 participants (59 after excluding incomplete data), a relatively small sample size for developing a generalizable PSI prediction model. Riley et al [75] suggest sample sizes ranging from 100 to several hundred participants to ensure adequate generalizability in clinical machine learning. Although cross-validation was applied to minimize bias, the limited sample size may still restrict reproducibility. In future work, we will validate the generalizability of our prediction model by using a more substantial dataset.

Second, this study focused on pediatric participants aged 6 to 13 years with ADHD, all recruited from South Korea. As the sample was limited geographically and demographically, it is possible that testing of our prediction model on datasets collected from participants from other geographical locations, socioeconomic levels, or backgrounds may yield different outcomes. Although standardized diagnostic procedures were implemented to ensure internal validity, future studies incorporating more diverse populations are needed to improve external validity.

Third, this study used serious game performance data as an alternative to traditional assessment tools. However, repeated use of the serious game may introduce practice effects similar to those observed in the WISC-V assessments. To mitigate practice effects, we used randomization techniques to modulate game complexity in each session. The game uses adaptive mechanisms to adjust difficulty based on individual performance, providing a balanced experience. Nevertheless, there is a paucity of research on the long-term use of game-based assessments and their feasibility. It is imperative to assess the extent of the learning effect and recalibrate the model in the future.

Future Work

To further enhance our PSI prediction model and facilitate its integration into clinical practice, future studies will use larger and more diverse datasets encompassing varied geographic and demographic populations. Expanding the dataset will also strengthen external validity and improve applicability across broader clinical contexts. In addition, longitudinal studies assessing the long-term reliability of the PSI prediction model by tracking PSI changes over time could further validate its clinical utility in monitoring ADHD-related cognitive function and informing personalized interventions. Finally, incorporating physiological signals, including multimodal data from electroencephalogram, electrocardiogram, and functional near-infrared spectroscopy, may enhance the robustness of the PSI prediction model and advance its potential for medical application in ADHD monitoring and intervention.

Conclusion

This study introduces a novel approach for predicting K-WISC-V PSI scores using data from serious games designed specifically for children with ADHD. The experimental results showed that among the individual models, SVR had the highest performance, followed by Elastic Net, AdaBoost, and random forest. Furthermore, the ensemble method combining AdaBoost and Elastic Net provided stable and reliable predictive accuracy, highlighting the potential of model integration to enhance overall performance. Our findings suggest that this approach can complement existing subjective methods of treatment monitoring to facilitate continuous symptom tracking and longitudinal monitoring and contribute to the optimization of treatment planning. Integrating PSI prediction with serious game content may enhance patient accessibility and provide a quantitative and reliable tool for clinicians to objectively track and assess a patient's condition in their everyday environment. Future research should aim to validate these findings using larger datasets including patients of diverse ages and ADHD symptom severities. Incorporating data from participants with varied demographic and clinical characteristics will enhance the model's generalizability and applicability in real-world clinical environments.

Data Availability

The datasets generated and analyzed in this study are not publicly available as they are restricted to the scope originally approved by the institutional review board (IRB). In addition, even for deidentified datasets with all personal information removed, additional IRB review and approval would be required to assess the feasibility of data sharing; therefore, the data cannot be shared at this stage.

Authors' Contributions

JSK: conceptualization, formal analysis, investigation, visualization, writing (original draft, review and editing); YJJ: formal analysis, writing (original draft, review and editing); SJK: formal analysis, visualization, writing (original draft); SJJ: formal analysis, investigation; JYP: formal analysis, investigation; HSH: supervision, writing (review and editing); JHS: conceptualization, methodology, supervision, visualization, writing (review and editing)

Conflicts of Interest

None declared.

References

1. Mahone EM, Denckla MB. Attention-deficit/hyperactivity disorder: a historical neuropsychological perspective. J Int Neuropsychol Soc. Oct 2017;23(9-10):916-929. [doi: 10.1017/S1355617717000807] [Medline: 29198277]
2. A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. The MTA Cooperative Group. Multimodal Treatment Study of Children with ADHD. Arch Gen Psychiatry. Dec 1999;56(12):1073-1086. [doi: 10.1001/archpsyc.56.12.1073] [Medline: 10591283]
3. Felt BT, Biermann B, Christner JG, Kochhar P, Harrison RV. Diagnosis and management of ADHD in children. Am Fam Physician. Oct 1, 2014;90(7):456-464. [Medline: 25369623]
4. Chang TM, Wu RC, Yang RC, et al. Objective diagnosis of ADHD through movement analysis by using a smart chair with piezoelectric material. Pediatr Neonatol. Jan 2023;64(1):46-52. [doi: 10.1016/j.pedneo.2022.06.007] [Medline: 36089537]
5. Hall CL, Bellato A, Kirk JD, Hollis C. The clinical utility of QbTest in supporting the assessment and monitoring of attention-deficit/hyperactivity disorder (ADHD): what do paediatricians need to know? Paediatr Child Health (Oxford). Sep 2023;33(9):259-264. [doi: 10.1016/j.paed.2023.06.006]
6. Cook NE, Braaten EB, Surman CBH. Clinical and functional correlates of processing speed in pediatric attention-deficit/ hyperactivity disorder: a systematic review and meta-analysis. Child Neuropsychol. Jul 2018;24(5):598-616. [doi: 10. 1080/09297049.2017.1307952] [Medline: 28345402]
7. Rommelse N, Luman M, Kievit R. Slow processing speed: a cross-disorder phenomenon with significant clinical value, and in need of further methodological scrutiny. Eur Child Adolesc Psychiatry. Oct 2020;29(10):1325-1327. [doi: 10. 1007/s00787-020-01639-9] [Medline: 32915272]
8. Shanahan MA, Pennington BF, Yerys BE, et al. Processing speed deficits in attention deficit/hyperactivity disorder and reading disability. J Abnorm Child Psychol. Oct 2006;34(5):585-602. [doi: 10.1007/s10802-006-9037-8] [Medline: 16850284]
9. Anker E, Ogrim G, Heir T. Verbal working memory and processing speed: correlations with the severity of attention deficit and emotional dysregulation in adult ADHD. J Neuropsychol. Mar 2022;16(1):211-235. [doi: 10.1111/jnp.12260] [Medline: 34218514]
10. Kaufman AS. Intelligent Testing with the WISC-III. John Wiley & Sons; 1994. ISBN: 0471578452
11. Calhoun SL, Mayes SD. Processing speed in children with clinical disorders. Psychol Schs. Apr 2005;42(4):333-343. [doi: 10.1002/pits.20067]
12. Dale BA, Finch WH, Shellabarger KAR, Davis A. Wechsler Intelligence Scale for Children, Fifth Edition profiles of children with autism spectrum disorder using a classification and regression trees analysis. J Psychoeduc Assess. Oct 2021;39(7):783-799. [doi: 10.1177/07342829211025924]
13. McGill RJ, Ward TJ, Canivez GL. Use of translated and adapted versions of the WISC-V: Caveat emptor. Sch Psychol Int. Jun 2020;41(3):276-294. [doi: 10.1177/0143034320903790]
14. Wilson CJ, Bowden SC, Byrne LK, Vannier LC, Hernandez A, Weiss LG. Cross-national generalizability of WISC-V and CHC broad ability constructs across France, Spain, and the US. J Intell. Aug 7, 2023;11(8):159. [doi: 10.3390/ jintelligence11080159] [Medline: 37623542]
15. Hagemeister C. How useful is the power law of practice for recognizing practice in concentration tests? Eur J Psychol Assess. Jan 2007;23(3):157-165. [doi: 10.1027/1015-5759.23.3.157]

16. Lievens F, Reeve CL, Heggestad ED. An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. J Appl Psychol. Nov 2007;92(6):1672-1682. [doi: 10.1037/0021-9010.92.6.1672] [Medline: 18020804]

17. Roediger HL III, Butler AC. The critical role of retrieval practice in long-term retention. Trends Cogn Sci (Regul Ed). Jan 2011;15(1):20-27. [doi: 10.1016/j.tics.2010.09.003]

18. Canivez GL, Watkins MW. Long-term stability of the Wechsler Intelligence Scale for Children--Third Edition. Psychol Assess. 1998;10(3):285-291. [doi: 10.1037//1040-3590.10.3.285]

19. Ryan JJ, Glass LA, Bartels JM. Stability of the WISC-IV in a sample of elementary and middle school children. Appl Neuropsychol. Jan 2010;17(1):68-72. [doi: 10.1080/09084280903297933] [Medline: 20146124]

20. Watkins MW, Canivez GL, Dombrowski SC, et al. Long-term stability of Wechsler Intelligence Scale for Children-fifth edition scores in a clinical sample. Appl Neuropsychol Child. 2022;11(3):422-428. [doi: 10.1080/21622965.2021.1875827] [Medline: 33556254]

21. Ahuja J, Jaka S, Saeed S, Gunturu S. Download to heal: navigating the pixelated path of digital therapeutics in psychiatric care. Braz J Psychiatry. Sep 27, 2024. [doi: 10.47626/1516-4446-2024-3802] [Medline: 39329484]

22. Suravarapu ST, Yetukuri K. Software programmes employed as medical devices. World J Curr Med Pharm Res. 2023:67-74. [doi: 10.37022/wjcmpr.v5i3.268]

23. Craven MP, Groom MJ. Computer games for user engagement in attention deficit hyperactivity disorder (ADHD) monitoring and therapy. Presented at: 2015 International Conference on Interactive Technologies and Games (iTAG); Oct 22-23, 2015; Nottingham, United Kingdom. [doi: 10.1109/iTAG.2015.9]

24. Shah A, Kraemer KR, Won CR, Black S, Hasenbein W. Developing digital intervention games for mental disorders: a review. Games Health J. Aug 2018;7(4):213-224. [doi: 10.1089/g4h.2017.0150] [Medline: 30106642]

25. Kollins SH, Childress A, Heusser AC, Lutz J. Effectiveness of a digital therapeutic as adjunct to treatment with medication in pediatric ADHD. NPJ Digit Med. Mar 26, 2021;4(1):58. [doi: 10.1038/s41746-021-00429-0] [Medline: 33772095]

26. Sverdlov O, van Dam J, Hannesdottir K, Thornton-Wells T. Digital therapeutics: an integral component of digital innovation in drug development. Clin Pharmacol Ther. Jul 2018;104(1):72-80. [doi: 10.1002/cpt.1036] [Medline: 29377057]

27. Daniëls NEM, Bartels SL, Verhagen SJW, Van Knippenberg RJM, De Vugt ME, Delespaul P. Digital assessment of working memory and processing speed in everyday life: feasibility, validation, and lessons-learned. Internet Interv. Mar 2020;19(100300):100300. [doi: 10.1016/j.invent.2019.100300] [Medline: 31970080]

28. Bonnechère B. Evaluation of processing speed of different cognitive functions across the life span using cognitive mobile games. Games Health J. Apr 2022;11(2):132-140. [doi: 10.1089/g4h.2021.0144] [Medline: 35180366]

29. Zhu F, Panwar B, Dodge HH, et al. COMPASS: A computational model to predict changes in MMSE scores 24-months after initial assessment of Alzheimer's disease. Sci Rep. Oct 5, 2016;6(1). [doi: 10.1038/srep34567]

30. Farzana S, Parde N. Exploring MMSE score prediction using verbal and non-verbal cues. Presented at: Interspeech 2020; Oct 25-29, 2020; Shanghai, China. [doi: 10.21437/Interspeech.2020-3085]

31. Twomey C, O'Connell H, Lillis M, Tarpey SL, O'Reilly G. Utility of an abbreviated version of the stanford-binet intelligence scales (5th ed.) in estimating "full scale" IQ for young children with autism spectrum disorder. Autism Res. Mar 2018;11(3):503-508. [doi: 10.1002/aur.1911] [Medline: 29282895]

32. Kadiyala PK. Mnemonics for diagnostic criteria of DSM V mental disorders: a scoping review. Gen Psychiatr. 2020;33(3):e100109. [doi: 10.1136/gpsych-2019-100109] [Medline: 32478284]

33. Audras-Torrent L, Miniarikova E, Couty F, et al. WISC-V profiles and their correlates in children with autism spectrum disorder without intellectual developmental disorder: report from the ELENA cohort. Autism Res. May 2021;14(5):997-1006. [doi: 10.1002/aur.2444] [Medline: 33247546]

34. Salonen J, Slama S, Haavisto A, Rosenqvist J. Comparison of WPPSI-IV and WISC-V cognitive profiles in 6-7-year-old Finland-Swedish children - findings from the FinSwed study. Child Neuropsychol. Jul 2023;29(5):687-709. [doi: 10.1080/09297049.2022.2112163] [Medline: 35996965]

35. Sattler JM, Dumont R. Assessment of Children: WISC-IV and WPPSI-III Supplement. Jerome M. Sattler; 2004. ISBN: 0-9702671-1-8

36. Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychol Rev. Jul 1996;103(3):403-428. [doi: 10.1037/0033-295x.103.3.403] [Medline: 8759042]

37. Tucha L, Fuermaier ABM, Koerts J, et al. Sustained attention in adult ADHD: time-on-task effects of various measures of attention. J Neural Transm (Vienna). Feb 2017;124(Suppl 1):39-53. [doi: 10.1007/s00702-015-1426-0] [Medline: 26206605]

38.   Kim SC. K-WISC-V processing speed index analysis to verify the effectiveness of ADHD symptom improvement using pediatric digital content. Appl Sci (Basel). May 2024;14(9):3792. [doi: 10.3390/app14093792]

39.   Kim SC, Lee H, Lee HS, Kim G, Song JH. Adjuvant therapy for attention in children with ADHD using game-type digital therapy. Int J Environ Res Public Health. Nov 14, 2022;19(22):14982. [doi: 10.3390/ijerph192214982] [Medline: 36429699]

40.   Kim SC, Lee H suk. Effect of game-based cognitive training programs on cognitive learning of children with intellectual disabilities. Appl Sci (Basel). Sep 2021;11(18):8582. [doi: 10.3390/app11188582] [Medline: 36003951]

41.   Adams ZW, Roberts WM, Milich R, Fillmore MT. Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? Psychol Assess. Jun 2011;23(2):427-436. [doi: 10.1037/a0022112] [Medline: 21443365]

42.   Cross-Villasana F, Finke K, Hennig-Fast K, et al. The speed of visual attention and motor-response decisions in adult attention-deficit/hyperactivity disorder. Biol Psychiatry. Jul 15, 2015;78(2):107-115. [doi: 10.1016/j.biopsych.2015.01.016] [Medline: 25773661]

43.   Kofler MJ, Rapport MD, Sarver DE, et al. Reaction time variability in ADHD: a meta-analytic review of 319 studies. Clin Psychol Rev. Aug 2013;33(6):795-811. [doi: 10.1016/j.cpr.2013.06.001] [Medline: 23872284]

44.   Epstein JN, Conners CK, Sitarenios G, Erhardt D. Continuous performance test results of adults with attention deficit hyperactivity disorder. Clin Neuropsychol. May 1998;12(2):155-168. [doi: 10.1076/clin.12.2.155.2000]

45.   Epstein JN, Erkanli A, Conners CK, Klaric J, Costello JE, Angold A. Relations between continuous performance test performance measures and ADHD behaviors. J Abnorm Child Psychol. Oct 2003;31(5):543-554. [doi: 10.1023/a:1025405216339] [Medline: 14561061]

46.   López J, López V, Rojas D, et al. Effect of psychostimulants on distinct attentional parameters in attentional deficit/hyperactivity disorder. Biol Res. 2004;37(3):461-468. [doi: 10.4067/s0716-97602004000300010] [Medline: 15515970]

47.   Stins JF, Tollenaar MS, Slaats-Willemse DIE, et al. Sustained attention and executive functioning performance in attention-deficit/hyperactivity disorder. Child Neuropsychol. Jun 2005;11(3):285-294. [doi: 10.1080/09297040490916938] [Medline: 16036452]

48.   Jung HT, Lee H, Kim K, et al. Predicting cognitive impairment level after a serious game-based therapy in chronic stroke survivors. In: Jung HT, Lee H, Kim K, editors. Presented at: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); May 19-22, 2019; Chicago, IL, USA. [doi: 10.1109/BHI.2019.8834484]

49.   Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas M. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol Rev. Dec 2015;71:804-818. [doi: 10.1016/j.oregeorev.2015.01.001]

50.   Hall MA. Correlation-Based Feature Selection for Machine Learning. The University of Waikato; 1999.

51.   Ghiselli EE. Theory of Psychological Measurement. McGraw-Hill; 1964.

52.   Hogarth RM, Jungermann H, editors. Methods for aggregating opinions. In: Decision Making and Change in Human Affairs Theory and Decision Library. Vol 16. Springer, Dordrecht; [doi: 10.1007/978-94-010-1276-8_16]

53.   Zajonc RB. A note on group judgements and group size. Human Relations. May 1962;15(2):177-180. [doi: 10.1177/001872676201500206]

54.   Breiman L. Random forests. Mach Learn. Oct 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]

55.   Hastie T. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer; 2009. [doi: 10.1007/978-0-387-84858-7]

56.   Zeini HA, Al-Jeznawi D, Imran H, Bernardo LFA, Al-Khafaji Z, Ostrowski KA. Random forest algorithm for the strength prediction of geopolymer stabilized clayey soil. Sustainability. Jan 2023;15(2):1408. [doi: 10.3390/su15021408]

57.   Freund Y, Schapire RE, editors. Experiments with a new boosting algorithm. In: ICML'96: Proceedings of the Thirteenth International Conference on International Conference on Machine Learning. Morgan Kaufmann Publishers Inc; 1996:148-156.

58.   Solomatine DP. AdaBoost.RT: a boosting algorithm for regression problems. Presented at: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat No 04CH37541); Jul 25-29, 2004; Budapest, Hungary. [doi: 10.1109/IJCNN.2004.1380102]

59.   Vapnik V. Statistical Learning Theory. Vol 2. John Wiley Sons; 1998:831-842. ISBN: 978-0-471-03003-4

60.   Awad M, Khanna R. Support vector regression. In: Efficient Learning Machines: Theories, Concepts, Applications for Engineers System Designers. Springer; 2015:67-80. [doi: 10.1007/978-1-4302-5990-9_4]

61.   Poirier C, Lavenu A, Bertaud V, et al. Real time influenza monitoring using hospital big data in combination with machine learning methods: comparison study. JMIR Public Health Surveill. Dec 21, 2018;4(4):e11361. [doi: 10.2196/11361] [Medline: 30578212]

62.   Zou H, Hastie T. Regularization and variable selection via the Elastic Net. J R Stat Soc Series B Stat Methodol. Apr 1, 2005;67(2):301-320. [doi: 10.1111/j.1467-9868.2005.00503.x]

63.   Gana R. Ridge regression and the Elastic Net: how do they do as finders of true regressors and their coefficients? Mathematics. 2022;10(17):3057. [doi: 10.3390/math10173057]

64.   Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol. Jan 1, 1996;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]

65.   Zhang Z, Lai Z, Xu Y, Shao L, Wu J, Xie GS. discriminative elastic-net regularized linear regression. IEEE Trans Image Process. Mar 2017;26(3):1466-1481. [doi: 10.1109/TIP.2017.2651396] [Medline: 28092552]

66.   Gopukumar D, Ghoshal A, Zhao H. Predicting readmission charges billed by hospitals: machine learning approach. JMIR Med Inform. Aug 30, 2022;10(8):e37578. [doi: 10.2196/37578] [Medline: 35896038]

67.   Mladenova T, Valova I. Comparative analysis between the traditional k-nearest neighbor and modifications with weight-calculation. Presented at: 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT); Oct 20-22, 2022; Ankara, Turkey. [doi: 10.1109/ISMSIT56059.2022.9932693]

68.   Li JS. Asymptotics of k-fold cross validation. J Artif Intell Res. 2023;78:491-526. [doi: 10.1613/jair.1.13974]

69.   Moon J, Ke FF, Sokolikj Z. Automatic assessment of cognitive and emotional states in virtual reality-based flexibility training for four adolescents with autism. Brit J Educational Tech. Sep 2020;51(5):1766-1784. [doi: 10.1111/bjet.13005]

70.   Rokach L. Ensemble-based classifiers. Artif Intell Rev. Feb 2010;33(1-2):1-39. [doi: 10.1007/s10462-009-9124-7]

71.   Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE. 2017;12(4):e0174944. [doi: 10.1371/journal.pone.0174944] [Medline: 28376093]

72.   Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. PLoS One. 2019;14(5):e0213653. [doi: 10.1371/journal.pone.0213653] [Medline: 31091238]

73.   Cho SY, Kim SH, Kang SH, et al. Pre-existing and machine learning-based models for cardiovascular risk prediction. Sci Rep. Apr 26, 2021;11(1):8886. [doi: 10.1038/s41598-021-88257-w] [Medline: 33903629]

74.   Kim Y, Yang J, Rim D, Oh U. Prediction of cognitive ability utilizing a machine learning approach based on digital therapeutics log data. Int J Adv Smart Converg. 2023;12(2):17-24. [doi: 10.7236/IJASC.2023.12.2.17]

75.   Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med. Aug 30, 2021;40(19):4230-4251. [doi: 10.1002/sim.9025] [Medline: 34031906]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
CFS: correlation feature selection
IRB: institutional review board
K-WISC-V: Korean Wechsler Intelligence Scale for Children-Fifth Edition
MAE: mean absolute error
MAPE: mean absolute percentage error
PSI: processing speed index
RMSE: root mean squared error
SVR: support vector regression
WISC: Wechsler Intelligence Scale for Children