A Novel Machine Learning-based Approach for Using Serious Games Data to Predict the WISC-V Processing Speed Index of Children with Attention Deficit/Hyperactivity Disorder

Jun-Su Kim, Seung-Jae Kim, Yoo Joo Jeong, Su Jin Jun, Jin-Yeop Park, Hyang-Sook Hoe, Jeong-Heon Song

Submitted to: Journal of Medical Internet Research
on: March 04, 2025

Table of Contents

A Novel Machine Learning-based Approach for Using Serious Games Data to Predict the WISC-V Processing Speed Index of Children with Attention Deficit/Hyperactivity Disorder

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

[1] Korea Brain Research Institute Daegu KR
[2] Daegu Gyeongbuk Institute of Science & Technology Daegu KR

Corresponding Author:
Jeong-Heon Song PhD

Korea Brain Research Institute
61, Cheomdan-ro
Daegu
KR

Abstract

Background: The Processing Speed Index (PSI) of the Korean Wechsler Intelligence Scale for Children - Fifth Edition (K-WISC-V) is highly correlated with ADHD symptoms and serves as an important indicator of cognitive function. However, restrictions on how frequently PSI testing can be performed prevent PSI determination on a daily or short-term basis. Therefore, an accessible and objective technique for predicting PSI scores is needed to help improve monitoring and intervention for children with ADHD.

Objective: To overcome the limitations of K-WISC-V and collect PSI scores in the short term, this study aimed to develop a machine learning-based technique to use data collected from serious games to predict the PSI scores of children with ADHD.

Methods: Sixty children (6-12 years old) with ADHD were recruited. The participants completed an initial PSI assessment using K-WISC-V followed by 25 min 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: Sixty children (6-12 years old) with ADHD were recruited. The participants completed an initial PSI assessment using K-WISC-V followed by 25 min 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.

Conclusions: The PSI prediction model using serious games data shows promise for providing an objective assessment tool for ADHD. By accurately predicting PSI scores, this approach could complement traditional subjective evaluations, allowing for continuous tracking of patient status and optimized treatment planning. The use of serious games offers a novel, engaging way to gather meaningful cognitive data outside traditional clinical settings, potentially improving accessibility and adherence. However, the results remain to be validated in larger, more diverse populations, and the long-term feasibility of using serious games in clinical and educational settings needs to be explored. Clinical Trial: The initial phase of this study (IRB No. 2021-10-080) was approved by the Institutional Review Board of Kye Myung University Hospital, certified by KOLAS in South Korea. The subsequent phase was registered with the Korea Clinical Research Information Service (CRIS Registration No. KCT0009862), accredited by WHO. Further details are provided in the Methods section.

(JMIR Preprints 04/03/2025:73408)
DOI: https://doi.org/10.2196/preprints.73408

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
- ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
- Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
- Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
- No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

A Novel Machine Learning-based Approach for Using Serious Games Data to Predict the WISC-V Processing Speed Index of Children with Attention Deficit/ Hyperactivity Disorder

Jun-Su Kim[1], MS | Seung-Jae Kim[1], MS | Yoo Joo Jeong [1,2,3], BS | 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), 61, Cheomdan-ro, Daegu, Republic of Korea, 41062; [2] Department of Neural Development and Disease, Korea Brain Research Institute (KBRI), 61, Cheomdan-ro, Daegu, Republic of Korea, 41062; [3]Department of Brain and Cognitive Sciences, Daegu Gyeongbuk Institute of Science & Technology, Daegu, Republic of Korea, 42988.

* Corresponding author:
Hyang-Sook Hoe: Department of Neural Development and Disease, AI-based Neurodevelopmental Diseases Digital Therapeutics Group, Korea Brain Research Institute (KBRI) 61, Cheomdan-ro, Daegu, Republic of Korea, Tel: 82-53-980-8310; sookhoe72@kbri.re.kr

Jeong-Heon Song: AI-based Neurodevelopmental Diseases Digital Therapeutics Group, Korea Brain Research Institute (KBRI), 61, Cheomdan-ro, Daegu, 41062, Republic of Korea, Tel: 82-53-980-8310, Fax: 85-53-980-8309; jhsong@kbri.re.kr

E-mail addresses of authors
Jun-Su Kim: ingkkings@kbri.re.kr
Seung-Jae Kim: tmd0212@kbri.re.kr
Yoo Joo Jeong: yoojoo930@kbri.re.kr
Su Jin Jun: wjstnwls0122@kbri.re.kr
Jin-yeop Park: jo3505@kbri.re.kr
Hyang-Sook Hoe: sookhoe72@kbri.re.kr
Jeong-Heon Song: jhsong@kbri.re.kr

Abstract

Background: The Processing Speed Index (PSI) of the Korean Wechsler Intelligence Scale for Children - Fifth Edition (K-WISC-V) is highly correlated with ADHD symptoms and serves as an important indicator of cognitive function. However, restrictions on how frequently PSI testing can be performed prevent PSI determination on a daily or short-term basis. Therefore, an accessible and objective technique for predicting PSI scores is needed to help improve monitoring and intervention for children with ADHD.

Objective: To overcome the limitations of K-WISC-V and collect PSI scores in the short term, this study aimed to develop a machine learning-based technique to use data collected from serious games to predict the PSI scores of children with ADHD.

Methods: Sixty children (6-12 years old) with ADHD were recruited. The participants completed an initial PSI assessment using K-WISC-V followed by 25 min 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: SVR had the best performance among the individual models, with the lowest RMSE of 11.288 and an MAE and MAPE of 7.874 and 7.375%, respectively. The ensemble combining AdaBoost and Elastic Net achieved consistent results on the test set, with an RMSE of 10.089, MAE of 7.126, and MAPE of 6.946%. The best overall performance was achieved by the ensemble integrating AdaBoost, Elastic Net, and SVR, which recorded the lowest RMSE of 10.072, an MAE of 6.798, and a MAPE of 6.611%. The predictive accuracy of this model was highest for PSI scores near the mean value of 100, demonstrating its reliability for clinical applications.

Conclusions: The PSI prediction model using serious games data shows promise for providing an objective assessment tool for ADHD. By accurately predicting PSI scores, this approach could complement traditional subjective evaluations, allowing for continuous tracking of patient status and optimized treatment planning. The use of serious games offers a novel, engaging way to gather meaningful cognitive data outside traditional clinical settings, potentially improving accessibility and adherence. However, the results remain to be validated in larger, more diverse populations, and the long-term feasibility of using serious games in clinical and educational settings needs to be explored.

Trial Registration: The initial phase of this study (IRB No. 2021-10-080) was approved by the Institutional Review Board of Kye Myung University Hospital, certified by KOLAS in South Korea. The subsequent phase was registered with the Korea Clinical Research Information Service (CRIS

Registration No. KCT0009862), accredited by WHO. Further details are provided in the Methods section.

Keywords: ADHD, Prediction, Machine Learning, Ensemble, Symptom Tracking, Processing Speed, Serious Games

Introduction

Attention deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that progresses throughout childhood and continues into adulthood [1]. Assessment and management play a critical role in determining the optimal timing for interventions and enhancing treatment effectiveness [3, 4], but the diagnosis of mental health conditions such as ADHD is predominantly based on subjective judgments from patients, families, and healthcare providers [2]. Moreover, evaluating patient symptoms outside clinical settings, such as at home or school, often relies on observation rather than objective measures [3].

Individuals with ADHD exhibit slower processing speeds (PS), that is, the ability to respond to a given stimulus within a limited time frame. Slower PS impairs the ability to process and respond to information efficiently [5, 6], and reduced PS has been shown to negatively affect academic performance, adaptability, anxiety, and social competence. Thus, PS is a critical indicator of ADHD and can provide valuable insights into symptom severity[7]. The Processing Speed Index (PSI), a measure of PS, was initially introduced in the Wechsler Intelligence Scale for Children, Third Edition (WISC-III) [8], and has been a primary assessment tool since the Fourth Edition (WISC-IV) [9]. PSI is closely linked to daily functioning and offers a valuable method for quantifying symptoms, highlighting its importance in managing ADHD. However, PSI has several limitations. Effective PSI measurement requires trained professionals, such as clinical psychologists or physicians, to administer and interpret specialized tests. In addition, repeated assessments can lead to practice effects, reducing data validity [10-12]. The recommended usage interval for WISC is 1 to 2.6 years, making it difficult to collect PSI measurements continually [13-15].

In the present study, machine learning was used to develop a new approach designed to overcome the limitations of PSI. The machine learning-based PSI prediction model uses behavioral data from children diagnosed with ADHD collected from Neuro World, a serious game type of software as a medical device (SaMD) powered by artificial intelligence (AI). The predictive performance of various machine learning models was compared, and an optimized ensemble model for PSI prediction was created.

Methods

Recruitment

The initial phase of this study was reviewed and approved by the Institutional Review Board (IRB) of Kye Myung University Hospital (IRB No. 2021-10-080), which is certified by KOLAS (Korea Laboratory Accreditation Scheme), 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. KCT0009862), accredited by WHO. Participants were recruited from the Department of Psychiatry at Kye Myung University Hospital. A total of 30 children with ADHD were recruited between December 2021 and December 2022 (IRB No. 2021-10-080), and an additional 30 children diagnosed with ADHD were recruited between December 2022 and March 2023 (CRIS Registration No. KCT0009862). The entire clinical procedure is illustrated in Figure 1.

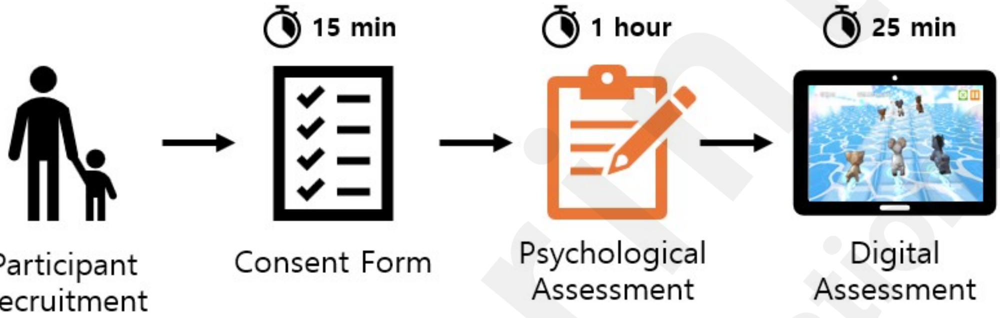


Figure 1. Clinical trial process. A total of 60 children diagnosed with ADHD were recruited. After completion of consent forms by the participants, a psychological assessment was conducted for 1 hour using K-WISC-V. After the K-WISC-V assessment, the children performed a digital assessment using Neuro World software for 25 minutes.

Experimental Methodology

After the participants completed consent forms, they underwent a psychological assessment of PSI using 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 functional game content for 25 minutes. To ensure participant comfort, a pause button was provided, allowing 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 [16]. 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). K-WISC-V is targeted to children and adolescents (aged 6 to 16 years and 11 months) [17] and includes core indices for the Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and PSI [18]. Each assessment is conducted individually, with scores calculated based on established criteria and converted to age- and gender-specific scaled scores. The index scores are standardized with a mean of 100 and a standard deviation of 15 [19], enabling comparisons of an individual's abilities with those in the same age group.

PSI, a critical component of 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 PSI [20]. Symbol search and coding, two core subsets of PSI, are described in Table 1. 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 [21].

Table 1. Detailed description of PSI core subset

Task	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 are 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 Content for Processing Speed Measurement

PS indicates the ability to quickly complete simple tasks while maintaining focus. To evaluate PS of children diagnosed with ADHD, the software Neuro World developed by Woorisoft Co. Ltd (South Korea) was utilized (Figure 2). Neuro World consists of five different serious game content modules designed for rapid completion. Each content module requires 5 minutes of game engagement. During the digital assessment, participants were provided with Neuro World for 25 minutes. The sequence of content modules was randomized, and the next content module began automatically after completion of the previous content module. The difficulty of each content module was dynamically adjusted based on the participant's performance as follows:

- 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 variability of response time is strongly correlated with ADHD [22-24]. Therefore, 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.

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, Content 5 – Memory matching.

Various data were collected during the performance of the Neuro World content modules to evaluate cognitive performance: $x1$, 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; $x8$, standard deviation of performance level; $x9$ and $x10$, average time to touch the screen after a task was presented (i.e., response time) and its standard deviation; $x11$ and $x12$, average time to the final touch required to complete each game (i.e., task completion consistency) and its standard deviation; and $x13$ and $x14$, average decision-making time, calculated as the difference between the final and first touch times, and its standard deviation.

Previous studies have demonstrated that ADHD exhibit deficits in sustained attention [25-28]. Based on these findings, we divided the total content performance into two phases and aimed to quantitatively assess whether 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 two 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 ($x1{:}140$) [29].

Data Preprocessing

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 PSI while minimizing correlations between variables [30]. The effectiveness of the selected variable set in CFS is measured using <Equation 1>. In this equation, *Merits* represents the cumulative correlation between the variables and the external variable; k denotes the number of variables; r'_{zi} represents the average correlation between variables and the external variable; and r'_{ii} denotes the average intercorrelation between variables [31-33]. 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. *Merits* not only signifies that the variable set is more closely aligned with the target class but also contributes to improved model performance by reducing irrelevant or redundant information.

$$Merits = \frac{k r'_{zi}}{\sqrt{k + k(k-1) r'_{ii}}} \tag{1}$$

Using the CFS method, six variables were selected from 140 content variables. In addition, the variables gender and age were included, resulting in a total of eight variables used as inputs for the machine learning models in the analysis. The selected variables are listed in Table 2.

Table 2. List of input variables for the machine learning model

Index	Variable (xn)
1	Content1_answer_rate_back ($x16$)
2	Content2_touch_count_diff_back ($x47$)
3	Content3_user_touch_count_back ($x74$)
4	Content5_total_score_front ($x115$)
5	Content3_first_touch_time_mean_front ($x65$)

6	Content1_decision_time_mean_front (x13)
7	Gender
8	Age

Machine Learning Models Used for Predicting 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. Specifically, we utilized 4 different machine learning models: Random Forest, AdaBoost, Support Vector Regression (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 [34]. 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 [35]. After creating several training data subsets, a bagging method is applied to reduce correlations among different decision trees to reduce variance [36]. This bagging characteristic enhances the predictive accuracy and stability of the Random Forest algorithm [34].

AdaBoost (Adaptive Boosting) is a boosting algorithm that iteratively trains a series of weak learners and combines them to form a robust ensemble model [37]. Unlike other boosting methods, AdaBoost applies the relative instead of absolute error, giving more weight to examples with lower values [38]. Additionally, AdaBoost continues to run even if the error rate exceeds 0.5 and supports user-defined iterations, allowing performance improvement over initial iterations [38].

Support Vector Regression (SVR) is a model that solves regression problems by extending binary classification through convex optimization formulations [39]. SVR generalizes Support Vector Machines (SVM) to adapt to regression problems and introduces an ε-insensitive region (tube) to reformulate the optimization problem[40]. Compared to other multivariate regression models, SVM can learn nonlinear functions with the kernel trick, which maps independent variables in a higher dimensional feature space [41]. The ε-insensitive region approximates continuous functions well and balances model complexity with prediction errors [40].

Elastic Net, proposed by Zou and Hastie in 2005, was developed to estimate the regression coefficient β in situations where p > n [42]. Elastic Net combines the concepts of Lasso and Ridge Regression to estimate β [43]. The l2-norm penalty tends to reduce coefficients towards 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 [44]. By combining both penalties, Elastic Net effectively mitigates multicollinearity and enables the group selection of correlated features [45, 46].

Results

SVR machine learning model effectively predicts PSI scores from the serious game performance outcomes of children diagnosed with ADHD

The predictive performance of the four 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 3). Among the 60 initial datasets of the participants, one was excluded due to missing values, resulting in a total of 59 datasets used for analysis. Among the 59 datasets, 47 (80%) were used as training data, while the remaining 12 (20%) were used as test data for each machine learning model. Given the limited sample size, we utilized K-fold cross-validation with the number of folds (k) set to 4 to enhance the prediction performance of each model.

SVR demonstrated the best overall performance on the training dataset, 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 3). The Elastic Net model exhibited the second highest performance on the training dataset, with an RMSE of 13.458, MAE of 10.580, and MAPE of 10.352% (Table 3). Random Forest and AdaBoost had higher errors on the training dataset, with RMSE values of 14.307 and 14.269, respectively.

We then investigated the performance of each machine learning model on the test dataset 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 3). Parallel with the performance on the training dataset, Elastic Net was second, with an RMSE of 12.488, MAE of 8.328 and MAPE of 7.993%, demonstrating its ability to provide accurate predictions (Table 3). 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 3). Overall, SVR emerged as the most effective model for predicting PSI across both the training and test datasets.

Table 3. Comparison of machine learning model performance for predicting PSI scores of children diagnosed with ADHD

	Train			Test		
Model	RMSE	MAE	MAPE(%)	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
SVR	13.107	9.933	9.792	11.288	7.874	7.375
Elastic Net	13.458	10.58	10.352	12.488	8.328	7.993

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 [47]. 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, we applied ensembles of the four machine learning models used in this study (Random Forest, AdaBoost, Elastic Net, and SVR) to the training dataset, and the results of the top five combinations of machine learning models with the highest RMSE performance are presented in Table 4. 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 4). 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 4). Collectively, these findings demonstrate that the combination of AdaBoost and Elastic Net outperformed the individual models by leveraging complementary strengths, improving overall predictive performance.

Table 4. Top 5 ensemble model combinations ranked by RMSE performance on the training set for predicting PSI scores of children with ADHD

Combination	Training		
	RMSE	MAE	MAPE(%)
[AdaBoost, Elastic Net]	12.738	10.116	10.086
[AdaBoost, Elastic Net, SVR]	12.770	10.013	9.938
[Random Forest, Elastic Net]	12.781	9.878	9.816
[Random Forest, AdaBoost,	12.802	10.135	10.071

Elastic Net, SVR]			
[Random Forest, Elastic Net, SVR]	12.821	9.930	9.829

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 combination 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 dataset (Table 5). 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 5). By contrast, the combination of Random Forest and Elastic Net, which performed strongly on the training dataset and had the lowest MAE of 9.878, exhibited higher error rates on the test dataset, with an MAE of 7.788. Together, these results suggest that the ensemble model of AdaBoost and Elastic Net, which provided consistent performance on both the training and test datasets, was enhanced by the inclusion of SVR to accurately predict the PSI scores of children with ADHD from the test dataset.

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

Combination	Test		
	RMSE	MAE	MAPE(%)
[AdaBoost, Elastic Net]	10.089	7.126	6.946
[AdaBoost, Elastic Net, SVR]	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

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 of AdaBoost, Elastic Net, and SVR Model 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 3). 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 $y=0.58x+43$, reflects 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, the majority of data points closely follow the $y=x$ line, underscoring the model's reliability and robust predictive capability.

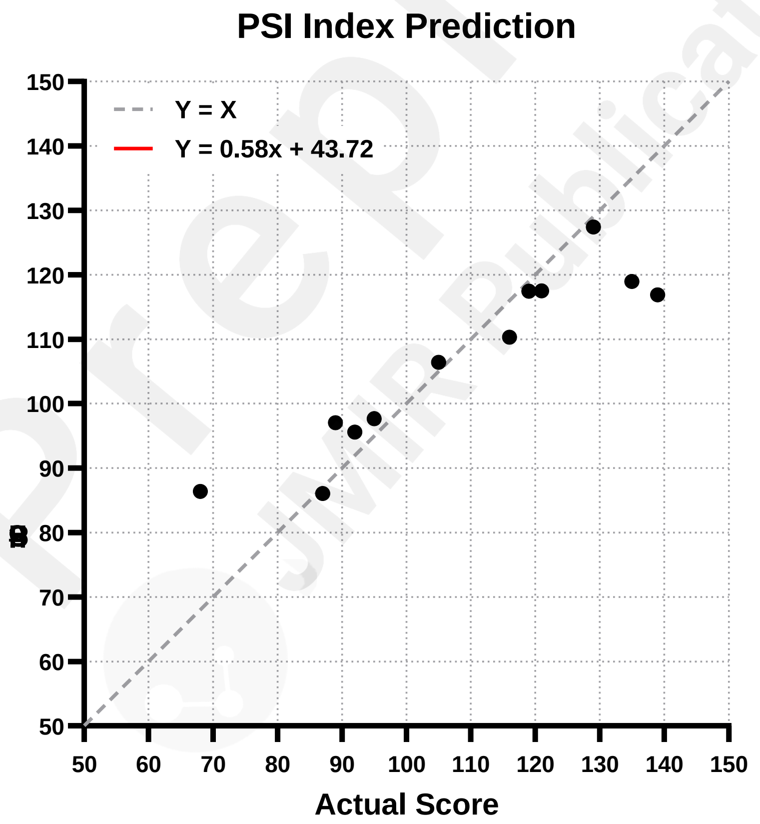


Figure 3. Comparison of actual PSI values with those predicted by ensemble application. The dashed line $(y=x)$ represents matching of the predicted values with the actual PSI values, and the red line ($y=0.58x+43$¿ is the regression line of the relationship between the predicted and actual PSI scores.

Discussion

In this study, we collected 60 datasets of serious game performance of children aged 6 to 12 diagnosed with ADHD. We applied the CFS method to extract six variables from the serious game performance data, which were input along with gender and age for analysis by machine learning models. Among the individual machine learning models, SVR was most effective in predicting the PSI scores of children with ADHD based on performance outcomes. 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.

Objective methods for monitoring ADHD symptoms remain limited in both clinical and non-clinical settings. In home and school environments, symptom evaluation is often based on observation due to the lack of standardized measurement tools. Even in clinical settings, reevaluations typically occur only after a significant period following the initial psychological assessment, making it difficult to objectively track patient improvement. PSI scores, which are closely linked to daily functioning, offer a valuable method for quantifying symptoms and informing ADHD management. PSI is part of WISC-V, a primary assessment tool for diagnosing ADHD. It is difficult to use PSI to repeatedly assess children diagnosed with ADHD because of its recommended usage interval and the lack of professionals qualified to conduct the assessment. To overcome the limitations of PSI score collection, in the present study, we aimed to develop a model for predicting PSI scores based on performance on serious games.

Serious games have primarily been used for long-term training and are an effective means of monitoring progress between hospital visits. However, existing monitoring methods mainly rely on subjective evaluations or content-related factors, such as task duration and accuracy rates, which are weakly correlated with actual symptom improvement and underlying issues. To overcome these shortcomings, this study utilized serious game performance data to predict psychological assessment indices that are practically used in clinical settings. By leveraging these predictions, symptom improvement can be assessed using clinically relevant measures. Additionally, the predicted indices can serve as valuable reference data for clinicians, allowing timely intervention during psychological treatment. This approach offers a more objective, quantitative method for ADHD symptom tracking.

In this study, we compared and combined various machine learning models to enhance the accuracy of PSI predictions for children with ADHD. We found that SVR achieved the highest accuracy, followed by Elastic Net, while Random Forest and AdaBoost had higher errors on both the training and test datasets. These results reinforce the importance of selecting models based on their predictive stability and generalizability. While SVR achieved the highest accuracy in both the training and test evaluations, the other algorithms—Elastic Net, Random Forest, and AdaBoost—exhibited strengths. Previous studies have shown that leveraging multiple models can substantially improve predictive performance [48]. 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 [48]. Our analysis of the predictive stability and accuracy of ensemble approaches showed that the ensemble combination of AdaBoost and Elastic Net outperformed SVR. 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 effectiveness of ensemble algorithms is widely recognized; for example, a study identified a strong correlation between ensemble accuracy and composition methods [49], while another study demonstrated that combining deep neural networks with traditional machine learning improves neuroimaging-based disorder prediction [50]. These results support the idea that algorithmic diversity plays a key role in enhancing ensemble efficacy.

The regression line 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, (REF) et al. found that the regression line for Mini-Mental State Examination (MMSE) score prediction for mild cognitive impairment (MCI) crossed y=x near the critical cutoff of 27, indicating high accuracy in distinguishing normal cognition from MCI and stability in clinically meaningful ranges [29]. Future research will explore the long-term reliability of the model for PSI score prediction. Longitudinal studies tracking PSI changes over time could further validate the model's clinical utility, ensuring its effectiveness in monitoring ADHD-related cognitive function and informing personalized interventions.

Our findings suggest that the PSI prediction model utilizing 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, supporting continuous patient monitoring and optimization of treatment plans. By integrating these tools into everyday environments, such as homes or schools, clinicians may gain valuable insights into cognitive function and treatment efficacy beyond traditional assessments.

Limitations

The present study is subject to several limitations. First, the small sample size may limit the generalizability of the PSI score prediction model, and studies with larger datasets are required to validate our findings. Second, the current study utilized the serious game performance of children diagnosed with ADHD on their first visit (25 min/1 session) to predict the PSI score. It is possible that long-term intervention with serious games (e.g. 25 min/day, 5 times/week, for 4 weeks) may result in different PSI score prediction models, which we will address in a future study.

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 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 integrating models to enhance overall performance. These findings suggest that this approach can complement existing subjective methods of treatment evaluation, facilitate continuous symptom tracking, and optimize treatment planning. The integration of PSI prediction with serious game content improves accessibility for participants and offers a quantitative, reliable assessment tool, enabling clinicians to objectively evaluate a patient’s condition in everyday settings (Figure 4). Future research should aim to validate these findings using larger datasets including patients of diverse ages and ADHD symptom severities. Expanding the study to incorporate varied demographic and clinical characteristics will enhance the model’s generalizability and applicability in real-world clinical environments.

Authors’ Contributions

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

JHS: conceptualization, supervision, visualization, writing (review and editing)

Acknowledgements

This work was supported by the KBRI basic research program through KBRI funded by the Ministry of Science, ICT & Future Planning (grants 24-BR-02-03, 23-BR-02-12, 24-BR-03-07, 24-BR-05-02, 24-BR-03-01, and 24-BR-03-05). This work was also supported by the Korea Ministry of Science and ICT's Special Account for Regional Balanced Development for Commercialization supervised by the NIPA (National IT Industry Promotion Agency) to support AI-based digital medical devices for neurodevelopmental disorders (H0301-24-1001).

Conflicts of Interest

The authors declare that they have no competing interests.

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

References

1. Mahone EM, Denckla MB. Attention-deficit/hyperactivity disorder: a historical neuropsychological perspective. Journal of the International Neuropsychological Society. 2017;23(9-10):916-29.
2. Taliaz D, Souery D. A New Characterization of Mental Health Disorders Using Digital Behavioral Data: Evidence from Major Depressive Disorder. J Clin Med. 2021 Jul 14;10(14):3109. PMID: 34300275. doi: 10.3390/jcm10143109.
3. Duffy FF, Chung H, Trivedi M, Rae DS, Regier DA, Katzelnick DJ. Systematic use of patient-rated depression severity monitoring: is it helpful and feasible in clinical psychiatry? Psychiatric Services. 2008;59(10):1148-54.
4. Jacobi F, Hofler M, Meister W, Wittchen HU. [Prevalence, detection and prescribing behavior in depressive syndromes. A German federal family physician study]. Nervenarzt. 2002 Jul;73(7):651-8. PMID: 12212528. doi: 10.1007/s00115-002-1299-y.
5. Rommelse N, Luman M, Kievit R. Slow processing speed: a cross-disorder phenomenon with significant clinical value, and in need of further methodological scrutiny. European Child Adolescent Psychiatry: Springer; 2020. p. 1325-7.
6. Shanahan MA, Pennington BF, Yerys BE, Scott A, Boada R, Willcutt EG, et al. Processing speed deficits in attention deficit/hyperactivity disorder and reading disability. J Abnorm Child Psychol. 2006 Oct;34(5):585-602. PMID: 16850284. doi: 10.1007/s10802-006-9037-8.
7. 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. 2018 Jul;24(5):598-616. PMID: 28345402. doi: 10.1080/09297049.2017.1307952.
8. Kaufman AS. Intelligent testing with the WISC-III: John Wiley & Sons; 1994. ISBN: 0471578452.
9. Calhoun SL, Mayes SD. Processing speed in children with clinical disorders. Psychology in the Schools. 2005 Apr;42(4):333-43. PMID: WOS:000228052800001. doi: 10.1002/pits.20067.
10. Hagemeister C. How useful is the power law of practice for recognizing practice in concentration tests? European Journal of Psychological Assessment. 2007;23(3):157-65.
11. Lievens F, Reeve CL, Heggestad ED. An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. J Appl Psychol. 2007 Nov;92(6):1672-82. PMID: 18020804. doi: 10.1037/0021-9010.92.6.1672.
12. Roediger HL, 3rd, Butler AC. The critical role of retrieval practice in long-term retention. Trends Cogn Sci. 2011 Jan;15(1):20-7. PMID: 20951630. doi: 10.1016/j.tics.2010.09.003.
13. Canivez GL, Watkins MW. Long-term stability of the Wechsler Intelligence Scale for Children—Third Edition. Psychological Assessment. 1998;10(3):285.
14. Ryan JJ, Glass LA, Bartels JM. Stability of the WISC-IV in a sample of elementary and middle school children. Appl Neuropsychol. 2010 Jan;17(1):68-72. PMID: 20146124. doi: 10.1080/09084280903297933.
15. Watkins MW, Canivez GL, Dombrowski SC, McGill RJ, Pritchard AE, Holingue CB, et al. Long-term stability of Wechsler Intelligence Scale for Children-fifth edition scores in a clinical sample. Appl Neuropsychol Child. 2022 Jul-Sep;11(3):422-8. PMID: 33556254. doi: 10.1080/21622965.2021.1875827.
16. Twomey C, O'Connell H, Lillis M, Tarpey SL, O'Reilly G. Utility of an abbreviated version of the stanford-binet intelligence scales in estimating 'full scale'IQ for young children with autism spectrum disorder. J Autism Research. 2018;11(3):503-8.
17. Kadiyala PK. Mnemonics for diagnostic criteria of DSM V mental disorders: a scoping

review. Gen Psychiatr. 2020;33(3):e100109. PMID: 32478284. doi: 10.1136/gpsych-2019-100109.
18. Audras-Torrent L, Miniarikova E, Couty F, Dellapiazza F, Berard M, Michelon C, et al. WISC-V profiles and their correlates in children with autism spectrum disorder without intellectual developmental disorder: Report from the ELENA cohort. 2021;14(5):997-1006.
19. 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. 2023 Jul;29(5):687-709. PMID: 35996965. doi: 10.1080/09297049.2022.2112163.
20. Sattler JM, Dumont R. Assessment of children: WISC-IV and WPPSI-III supplement. 2004.
21. Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychol Rev. 1996 Jul;103(3):403-28. PMID: 8759042. doi: 10.1037/0033-295x.103.3.403.
22. Adams ZW, Roberts WM, Milich R, Fillmore MT. Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? Psychological assessment. 2011;23(2):427.
23. Cross-Villasana F, Finke K, Hennig-Fast K, Kilian B, Wiegand I, Muller HJ, et al. The Speed of Visual Attention and Motor-Response Decisions in Adult Attention-Deficit/Hyperactivity Disorder. Biol Psychiatry. 2015 Jul 15;78(2):107-15. PMID: 25773661. doi: 10.1016/j.biopsych.2015.01.016.
24. Kofler MJ, Rapport MD, Sarver DE, Raiker JS, Orban SA, Friedman LM, et al. Reaction time variability in ADHD: a meta-analytic review of 319 studies. Clin Psychol Rev. 2013 Aug;33(6):795-811. PMID: 23872284. doi: 10.1016/j.cpr.2013.06.001.
25. Epstein JN, Conners CK, Sitarenios G, Erhardt D. Continuous performance test results of adults with attention deficit hyperactivity disorder. Clinical Neuropsychologist. 1998 May;12(2):155-68. PMID: WOS:000075574200004. doi: DOI 10.1076/clin.12.2.155.2000.
26. 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. 2003 Oct;31(5):543-54. PMID: 14561061. doi: 10.1023/a:1025405216339.
27. López J, López V, Rojas D, Carrasco X, Rothhammer P, García R, et al. Effect of psychostimulants on distinct attentional parameters in attentional deficit/hyperactivity disorder. Biological Research. 2004;37(3):461-8.
28. Stins JF, Tollenaar MS, Slaats-Willemse DI, Buitelaar JK, Swaab-Barneveld H, Verhulst FC, et al. Sustained attention and executive functioning performance in attention-deficit/hyperactivity disorder. Child Neuropsychol. 2005 Jun;11(3):285-94. PMID: 16036452. doi: 10.1080/09297040490916938.
29. Jung H-T, Lee H, Kim K, Kim B, Park S, Ryu T, et al., editors. Predicting cognitive impairment level after a serious game-based therapy in chronic stroke survivors. 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2019: IEEE.
30. Hall MA. Correlation-based feature selection for machine learning: The University of Waikato; 1999.
31. Ghiselli EE. Theory of psychological measurement. 1964 ed.
32. Hogarth RM, editor. Methods for aggregating opinions. Decision Making and Change in Human Affairs: Proceedings of the Fifth Research Conference on Subjective Probability, Utility, and Decision Making, Darmstadt, 1–4 September, 1975; 1977: Springer.
33. Zajonc RB. A note on group judgements and group size. Human Relations. 1962;15(2):177-80.
34. Breiman L. Random forests. Mach Learn. 2001 Oct;45(1):5-32. PMID: WOS:000170489900001. doi: Doi 10.1023/A:1010933404324.
35. Hastie T. The elements of statistical learning: data mining, inference, and prediction.

Springer; 2009.
36. 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. 2023 Jan;15(2):1408. PMID: WOS:000916357100001. doi: ARTN 1408
10.3390/su15021408.
37. Freund Y, Schapire RE, editors. Experiments with a new boosting algorithm. icml; 1996: Citeseer.
38. Solomatine DP, Shrestha DL, editors. AdaBoost. RT: a boosting algorithm for regression problems. 2004 IEEE international joint conference on neural networks (IEEE Cat No 04CH37541); 2004: IEEE.
39. Vapnik V. Statistical learning theory. John Wiley Sons google schola. 1998;2:831-42.
40. Awad M, Khanna R, Awad M, Khanna R. Support vector regression. Efficient learning machines: Theories, concepts, applications for engineers system designers. 2015:67-80.
41. Poirier C, Lavenu A, Bertaud V, Campillo-Gimenez B, Chazard E, Cuggia M, et al. Real Time Influenza Monitoring Using Hospital Big Data in Combination with Machine Learning Methods: Comparison Study. JMIR Public Health Surveill. 2018 Dec 21;4(4):e11361. PMID: 30578212. doi: 10.2196/11361.
42. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2005;67(2):301-20.
43. Gana R. Ridge Regression and the Elastic Net: How do they do as Finders of True Regressors and their Coefficients? J Mathematics. 2022;10(17):3057.
44. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1996;58(1):267-88.
45. Zhang Z, Lai Z, Xu Y, Shao L, Wu J, Xie GS. Discriminative Elastic-Net Regularized Linear Regression. IEEE Trans Image Process. 2017 Mar;26(3):1466-81. PMID: 28092552. doi: 10.1109/TIP.2017.2651396.
46. Gopukumar D, Ghoshal A, Zhao H. Predicting Readmission Charges Billed by Hospitals: Machine Learning Approach. JMIR Med Inform. 2022 Aug 30;10(8):e37578. PMID: 35896038. doi: 10.2196/37578.
47. Rokach L. Ensemble-based classifiers. Artif Intell Rev. 2010 Feb;33(1-2):1-39. PMID: WOS:000273807600001. doi: 10.1007/s10462-009-9124-7.
48. Kim Y, Yang J, Rim D, Oh UJIjoasc. Prediction of Cognitive Ability Utilizing a Machine Learning approach based on Digital Therapeutics Log Data. 2023;12(2):17-24.
49. Sesmero MP, Iglesias JA, Magán E, Ledezma A, Sanchis A. Impact of the learners diversity and combination method on the generation of heterogeneous classifier ensembles. Applied Soft Computing. 2021 Nov;111:107689. PMID: WOS:000711075800010. doi: ARTN 107689
10.1016/j.asoc.2021.107689.
50. Nguyen D, Nguyen H, Ong H, Le H, Ha H, Duc NT, et al. Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer's disease. 2022;13:255-63.

Figure Legend

Figure 1. Clinical trial process. A total of 60 children diagnosed with ADHD were recruited. After completion of consent forms by the participants, a psychological assessment was conducted for 1 hour using K-WISC-V. After the K-WISC-V assessment, the children performed a digital assessment using Neuro World software for 25 minutes.

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, Content 5 – Memory matching.

Figure 3. Comparison of actual PSI values with those predicted by ensemble application. The dashed line ($y=x$) represents matching of the predicted values with the actual PSI values, and the red line ($y=0.58x+43$¿ is the regression line of the relationship between the predicted and actual PSI scores.

Figure 4. Limitations of Traditional DTx Monitoring and Need for Effective Tracking

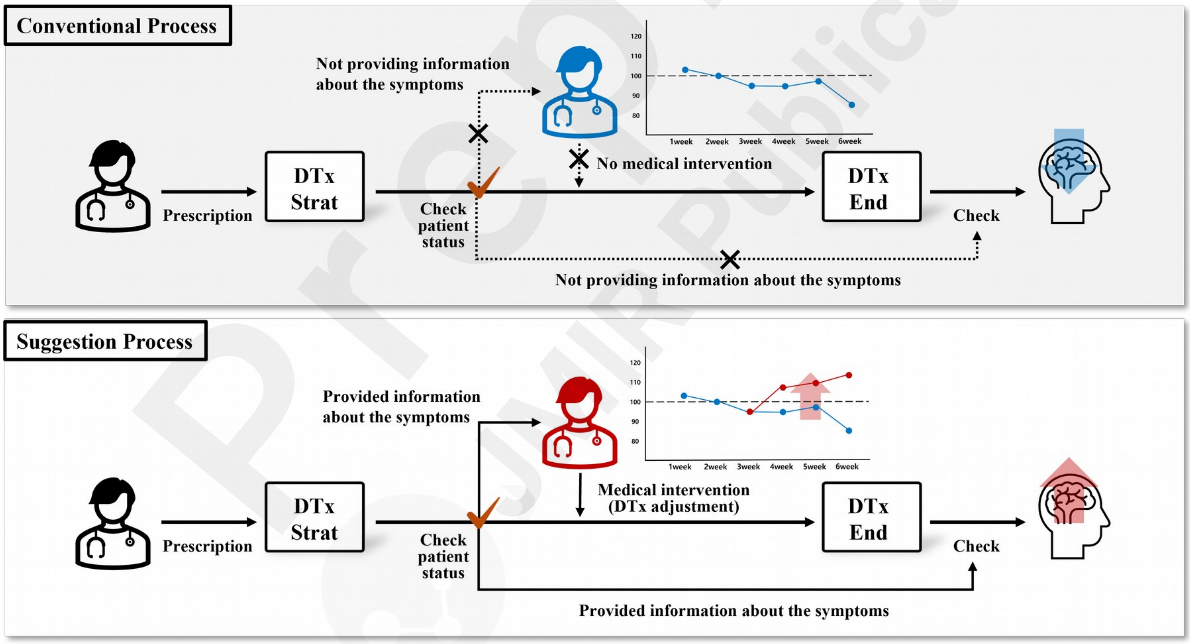


Table Legend

Table 1. Detailed description of PSI core subset

Table 2. List of input variables for the machine learning model

Table 3. Comparison of machine learning model performance for predicting PSI scores of children diagnosed with ADHD

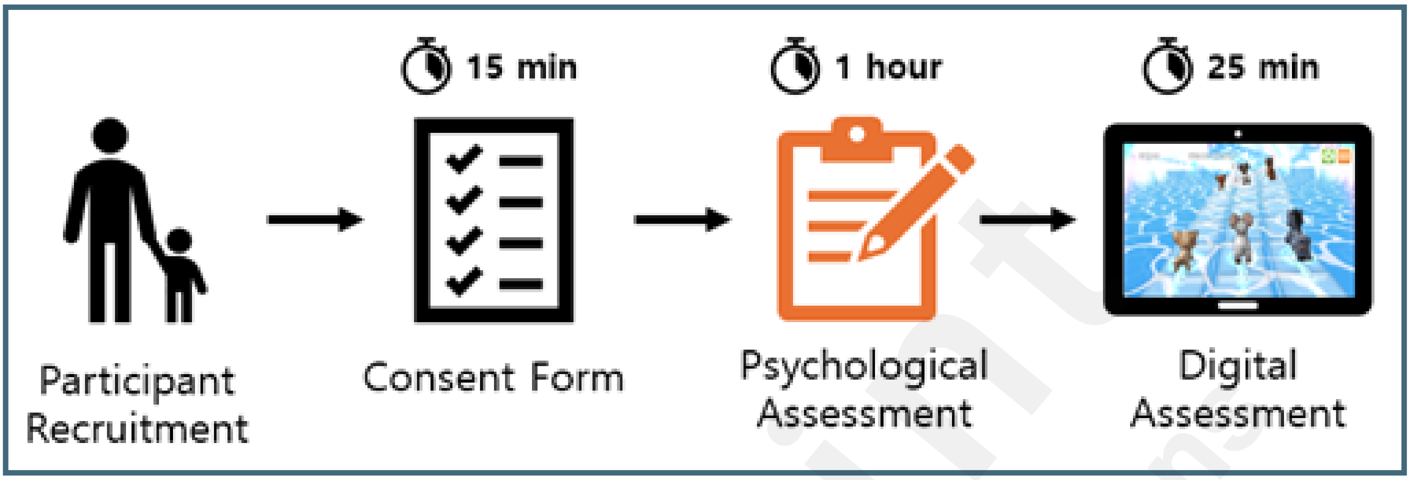
Table 4. Top 5 ensemble model combinations ranked by RMSE performance on the training set for predicting PSI scores of children with ADHD

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

Supplementary Files

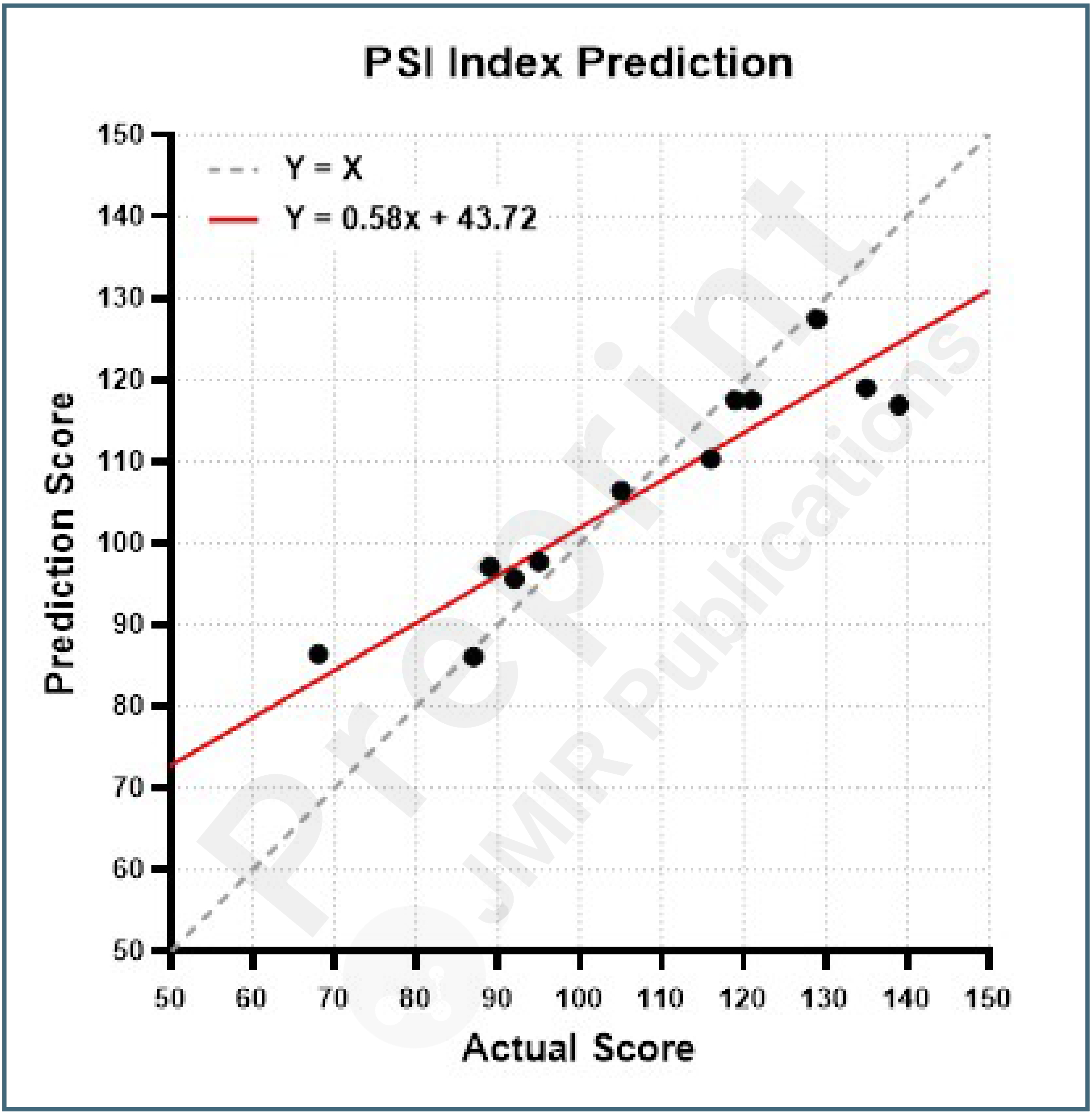
Figures

Clinical trial process. A total of 60 children diagnosed with ADHD were recruited. After completion of consent forms by the participants, a psychological assessment was conducted for 1 hour using K-WISC-V. After the K-WISC-V assessment, the children performed a digital assessment using Neuro World software for 25 minutes.



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, Content 5 – Memory matching.

Comparison of actual PSI values with those predicted by ensemble application. The dashed line (y=x) represents matching of the predicted values with the actual PSI values, and the red line (y=0.58x+43) is the regression line of the relationship between the predicted and actual PSI scores.



Limitations of Traditional DTx Monitoring and Need for Effective Tracking.