DOI 10.15641/ghi.v2i2.809 *Global Health Innovation, 2019, 2(2), article #2* Research Article

GLOBAL HEALTH INNOVATION

OPEN UCT PUBLICATIONS

ADHD Screening Tool: Investigating the effectiveness of a tablet-based game with machine learning

Romano Swarts*, Pieter Fourie, Dawie van den Heever

Biomedical Engineering Research Group, Department of Mechanical and Mechatronic Engineering, Stellenbosch University, Cape Town, South Africa

Received 8 August 2019; Accepted 15 November 2019; Published online 27 November 2019

Abstract

This study investigated the ability of a low-cost, mobile device-based game incorporating machine learning to screen participants between ages six and twelve years for attention-deficit/hyperactivity disorder inattentive subtype ($ADHD_I$). Relevant information from literature was incorporated into the game in light of the DSM-V diagnostic criteria. The game has seven back-to-back segments with unique layouts and a visual theme. The ADHD Screening Tool presents a novel patient-testing interface with a cloud-based machine learning classifier (MLC) integrated with a consensus algorithm. The game was tested with 39 clinically diagnosed participants ($ADHD_I$ and non-ADHD). Out of nine classifiers tested, the locally-deep support vector machine gave the best results: using leave-one-out cross-validation, this MLC classified data from five game segments for 38 participants with sensitivity of 92.9% and specificity of 82.9%. By making use of the consensus algorithm, the 39th participant was correctly classified according to the clinical diagnosis. The MLC and consensus algorithm were able to classify 39 participants with a sensitivity of 100% and specificity of 87.5%. To overcome participant class imbalance, the synthetic minority oversampling technique (SMOTE) was implemented on game segment data. The SMOTE two-class LDSVM yielded sensitivity of 90.7% and specificity of 94.4%. The study used an internet-connected, commercially available tablet.

Keywords: ADHD, screening, machine learning, game

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorders, distinctly characterised by a persistent pattern of inattentive, hyperactive or impulsive behaviour. Predominantly identified in early childhood, the persistent behavioural patterns associated with ADHD often continue into adolescence and adulthood and are associated with varying degrees of functional impairment across multiple settings (Polanczyk et al., 2014; Sharp et al., 2009; Preuss et al., 2006; Díaz-Orueta et al., 2017).

Multiple key developments take place in the brain during the growth stages from infancy to adolescence. These changes are primarily determined by genetics but are also influenced by environmental and social interactions. Key dependent relationships, such as parents and grandparents, play a vital role through these developmental stages. Although studies have revealed genetic overlaps with ADHD, the aetiology of ADHD remains unknown (Polanczyk et al., 2014; Kessler et al., 2006; Sharp et al., 2009).

The diagnosis of ADHD has thus far been based on clinical evaluations, coupled with parent and teacher questionnaires. Consequently, much criticism has arisen regarding the subjective nature of ADHD diagnosis. Over and under-diagnosis of ADHD has been widely debated, driven by variations in world-wide prevalence and broadening diagnostic criteria (Thomas et al., 2015). Although there has been a rise in the reported number of ADHD cases, it is still unclear whether this rise can be attributed to changes in diagnostic methods or whether there are other environmental factors increasingly playing a significant role (Rettner, 2014; Preuss et al., 2006; Díaz-Orueta et al., 2017). Another consideration is the cost of the diagnostic process, the cumulative fee for which can cover (Gualtieri & Johnson, 2005): consultations with clinical psychologists, paediatricians and other medical practitioners to evaluate carer/parent and teacher questionnaires and academic performance; provision and administration of neuropsychological test batteries and screening tools; and contact sessions with the child. Teachers are most often the first to make recommendations to carers/parents regarding ADHD, based upon observed classroom behaviour of children who make it difficult for other students to perform or teachers to cope.

*Corresponding author: swarts.rom@gmail.com

However, the lack of knowledge and understanding of ADHD often leads to teachers developing negative views of the learners they refer for assessment (Perold et al., 2010).

Determining an accurate, homogenous and repeatable method for the identification of ADHD symptoms is a vital step to better healthcare and ADHD assessment, diagnosis and treatment. The greatest challenge is the subjective nature of ADHD referrals and diagnosis, which has the potential to result in the over- or under-diagnosis of ADHD in children and adolescents. This challenge is accompanied by the costs associated with the diagnostic process. We present the design and preliminary testing of a low-cost ADHD screening tool that is capable of objective, quantitative screening for ADHD inattentive subtype ($ADHD_I$) in children between the ages of six and 12 years.

Methods

Game Design

A new tablet-based game was designed and developed using the Unreal Engine (UE) platform (v.4.18.0). The device chosen to run the game was an NVIDIA Shield K1 Tablet. As shown in Figure 1, the game was designed to have seven back-to-back segments with unique layouts and a visual theme. Each segment was designed with a specific purpose and served either as a reference, to deliver auditory and/or visual distractors, or to simply extend time. The visual distractors introduced in segments three and four are presented in the form of rocks falling and explosion dust clouds, with the auditory distractors in segments two and four are presented as explosive noises. The game incorporates a panda character travelling on a cart through an underground mine. The mine theme was specifically introduced to force sustained attention in the dark setting.

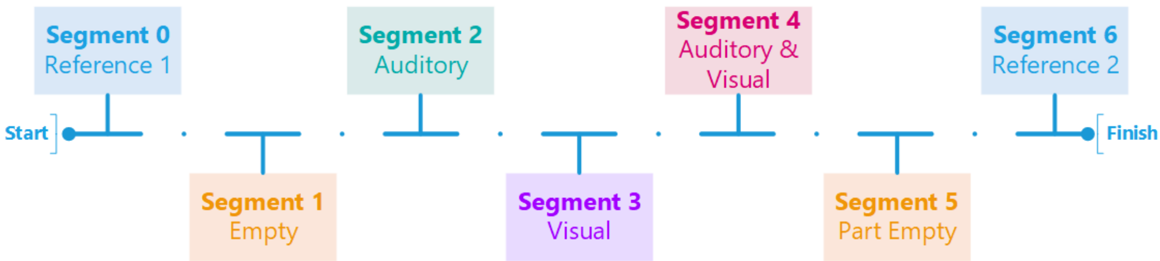

Figure 1. Game segment layout.

Table 1 indicates the various assets included in each of the seven game segments. Segment assets have been placed at random throughout each segment with the purpose of encouraging joystick engagement for effective navigation through the mine. The random placement of assets also served to strengthen the ability of the machine learning classifier (MLC) to generalise between segments (McDermott et al., 2016; Hussain, 2018).

Table 1. Layout of each game segment.

Segment	Pink gems	Obstacles	Auditory distractors	Visual distractors	Kamikaze gem	Tiles
0	87	53	0	0	0	89
1	0	0	0	0	0	89
2	73	57	6	0	0	89
3	80	52	0	8	0	89
4	61	56	3	4	0	89
5	0	12	0	0	1	89
6	87	53	0	0	0	89

The objective of the task for each segment was to reach the end of the mine as fast as possible. The dark setting was implemented to limit the visual stimuli presented to the participant at any given point within a segment. The controlled line of sight and the irregular presentation of response-stimuli, in the form of obstacles to be avoided and gems to be collected, were designed to limit anticipatory responses. The participant was able to control the game character with a joystick and had the choice of three positional lanes (left, middle or right). Figure 2 shows a screenshot of the game, with arrows indicating the location of the joystick and various buttons used during gameplay. The buttons are used collectively to effectively navigate through the mine, with the torch enabling a greater line of sight when used. In order to use the torch, a participant is required to collect gems, which result in fuel for the torch. The torch fuel gradually depletes with use.

Of the nine DSM-V diagnostic criteria for ADHD inattention subtype, seven were incorporated into the game. These seven criteria were selected as they were the easiest to incorporate, and are summarised into the following six data parameters:

1. *Following instructions*: Participants were required to complete a tutorial level of the game prior to screening gameplay. The tutorial included visual annotations to indicate the desired interactions.

2. *Task completion*: Measures such as pausing the game, pause duration or even exiting formed part of the recorded feature set.
3. *Sustained attention*: Sustained attention was incorporated by automated game avatar movement, requiring a participant to stay engaged with the task at hand and controlling when distracting stimuli are introduced.
4. *Forgetfulness*: This goes together with attention and following instructions and relates to the collection of in-game gems which provides fuel to switch the avatar's torch on or off, increasing and decreasing the range of sight respectively.
5. *Mistakes made*: Making mistakes results in a visual and auditory response from the avatar. Additionally, the speed of the avatar is also negatively impacted when an obstacle collision occurs. Obstacles present in various forms, requiring avoidance by either moving around or jumping over them.
6. *Distractibility*: Distractions have been introduced in the form of visual and auditory stimuli, in various combinations, with the intention of distracting the participant's attention from the main gameplay activity.

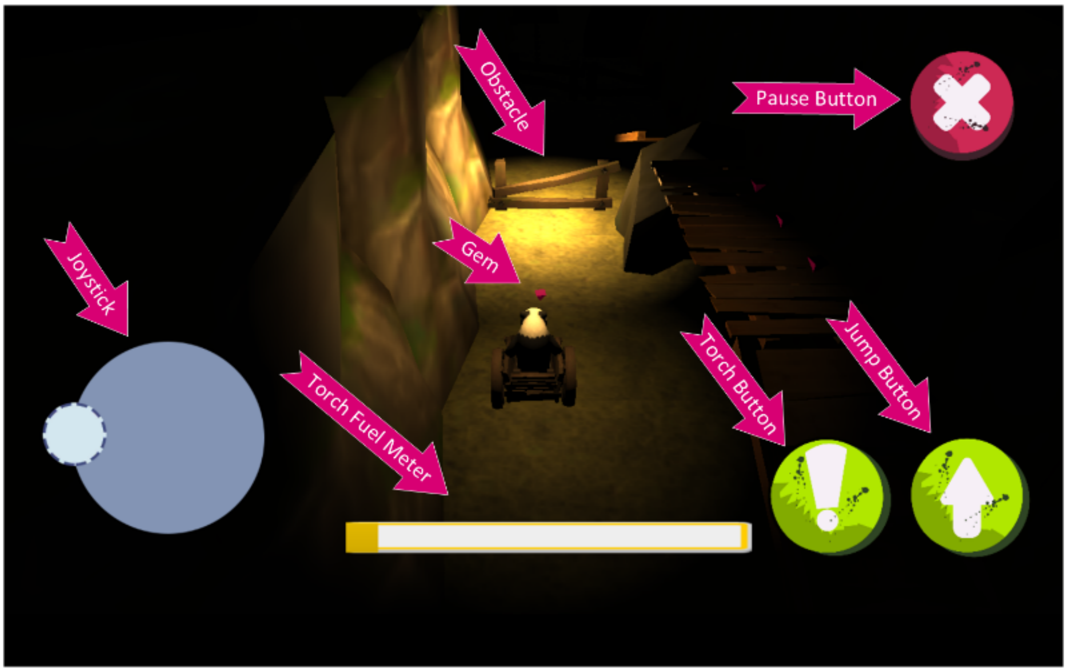

Figure 2. Screenshot of game with button layout.

Participant movement data were collected during gameplay by capturing the 3-axis accelerometer output at set intervals throughout the game. These data points were then incorporated into the complete feature set. Figure 2 illustrates the game performance features used in classification. The first-order features are collected during gameplay and the second-order features are calculated post gameplay for input to the MLC.

The game has been developed with a panda character moving through a mine. All the scenarios, including the aesthetic presentation of the game, have been constructed to control the level of stimulus a participant receives. The goal of the game is for the Panda to complete it in the shortest time possible, while avoiding obstacles and simultaneously collecting as many gems as possible. The game presents unfamiliar and unexpected challenges, forcing the player's sustained attention. The game is divided into seven segments (±one minute each). Each segment presents a unique set of challenges, aimed at forcing errors when levels of attention are poor. Furthermore, the game is intended to be culturally unbiased, without language and speech; the only fields to be captured are participant data.

Statistical features were incorporated according to findings in the literature (D'Souza & Kavitha, 2017; Altun & Barshan, 2010; Li et al., 2018). Accelerometer features were included based on findings investigating the significance of movement in ADHD participants during a Go/No-Go task (Li et al., 2016).

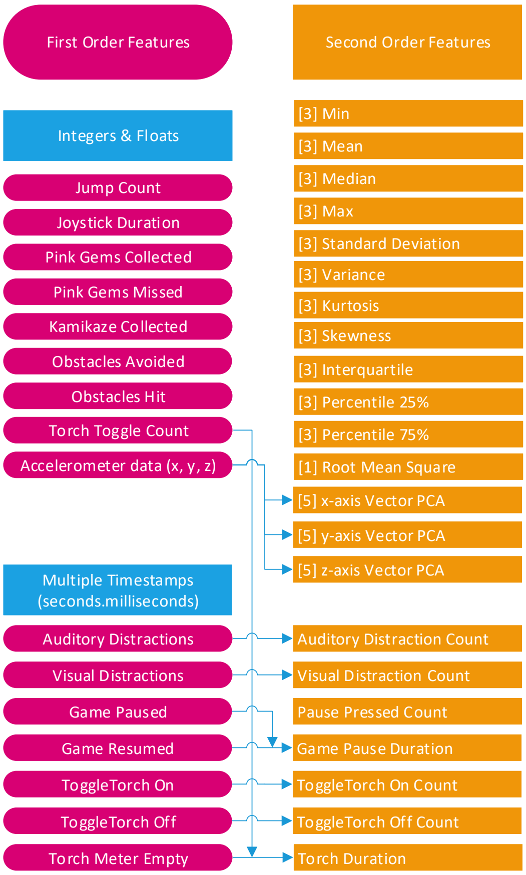

Figure 3. Game performance feature set.

Participants

Participation was voluntary and parental consent was sought with the aid of information leaflets. Prior to taking part in the study, participants were briefed and required to sign an assent form. The ADHD status of each participant was validated by a clinical psychologist and paediatrician prior to proceeding with the clinical study. The study protocol was approved by Stellenbosch University's Health Research Ethics Committee (HREC) and was conducted in accordance with the ethical guidelines and principles of the international Declaration of Helsinki, South African Guidelines for Good Clinical Practice and the Medical Research Council (MRC) Ethical Guidelines for Research.

Inclusion criteria

The following criteria were used to determine whether a participant was eligible for the study:

• Between six and 12 years of age;

• Meet the DSM-V criteria for $ADHD_I$, based on the gold standard for clinical ADHD assessment; alternatively, meet the criteria for non-ADHD based on the ADHD DSM-V criteria. The SWAN rating scale was used as an additional evaluation measure;

• Performance medication naïve;

• Written, informed consent from parents; and

• Assent from the participant.

Exclusion criteria

The following criteria were used to determine whether a participant was not eligible for the study:

• The presence of other subtypes of ADHD ($ADHD_H$ or $ADHD_C$);

• Co-existing severe psychiatric conditions (e.g. autism spectrum disorder) or known sensorineural deficits (e.g. blindness or deafness);

• History of epileptic seizures; or

• Known intellectual disability.

Testing of participants in this study took place at a single clinical site in Cape Gate, Cape Town. The testing room was selected for its lack of distractions, containing only two chairs and two cabinets.

Classifiers

Game data were captured during gameplay testing sessions. In order to determine the best MLC, a selection of nine classifiers was evaluated. Given the two-class classification problem, these classifiers were all of the two-class variety and were selected based on their availability on the cloud interface, Azure Machine Learning Studio, and findings from the literature. The following make up the list of nine: averaged perceptron (AP), support vector machine (SVM), decision jungle (DJ), boosted decision tree (BDT), decision forest (DF), neural network (NN), logistic regression (LR), Bayes point machine (BPM) and the locally deep support vector machine (LDSVM). The BDT, DJ, DF, LR, SVM and LDSVM classifiers commonly appeared throughout the literature when evaluating sensor and movement data (Moraru et al., 2010; Silverstein et al., 2016; D'Souza & Kavitha, 2017; Altun & Barshan, 2010).

Two approaches were evaluated to determine the best MLC for participant classification. The first approach created seven individual classifiers for each gameplay segment. The second approach discussed in more detail below, was to determine the best classifier by training it on all the gameplay segments combined and making use of a consensus algorithm to establish a single participant classification. This latter approach was implemented due to the limited number of data samples and the poor performance of the first approach.

The implemented approach iteratively adjusted the hyperparameters of each classifier using a 10-iteration randomised parameter sweep and was compared with the standard classifier to achieve the largest area under the curve (AUC). Due to the imbalance of the sample groups, the AUC was selected as the target performance metric. The AUC summarises the diagnostic efficiency of a classifier (Li et al., 2016; Berger & Cassuto, 2014). A leave-one-out cross-validation (LOOCV) was implemented during training. The LOOCV was implemented as it is widely used in practice to obtain a reliable measure of the test error for the overall performance estimation. Additionally, it provides a good indication of how a model will generalise to an independent dataset and was implemented during training to avoid problems with overfitting (Mao et al., 2014; Zhang & Wang, 2016).

The consensus algorithm employed has the following form:

$$C_f = \sum_{i=0}^{n} c_i$$

where i represents the position of the segment in the game, c represents the classification of the segment ($Neurotypical = -1,\ ADHD_I = 1$), n represents the total number of segments included in the analysis, and C_f represents the final consensus classification. The consensus confidence score, C_c, is a percentage value indicating the degree of consensus and takes the following form:

$$C_c = \frac{|C_f|}{n} \times 100$$

For participant classification, the following is applicable:

$$if\ C_f > 0; Classification = ADHD_I$$
$$if\ C_f < 0; Classification = Neurotypical$$

Data Processing

Data were packaged and directly sent to a secure database upon completion of gameplay. The data were organised in a table structure. A unique identifier was assigned to each participant and their respective data for the seven segments. These data files underwent further processing to generate the second order features. Once feature generation was complete, the data were used to train the MLCs.

Results

A sample size calculation determines the necessary population size for statistical significance. Following consultation at the Stellenbosch University Centre for Statistical Consultations and using a two-proportion, paired sample McNemar's test, a participant population size for classifying participants as either $ADHD_I$ or non-ADHD was determined for a statistical power goal of 0.9. McNemar's test was performed using *Statistica 13* software. The required sample size was 76.

Due to recruiting challenges, only 45 participants were enlisted for this study. Six sets of participant data were excluded after testing due to medication use, signs of co-existing ADHD subtypes and incomplete data. Despite knowledge of certain exclusionary factors disclosed before participation, no child was denied playing the game if they arrived for testing. The final test group comprised 39 participants: 31 $ADHD_I$ (25 male, 6 female) and eight non-ADHD control group (6 male, 2 female) between the ages of six and 12 years, with a mean age of 9.02 (SD = 1.88).

A 190-fold cross-validation model was used for each classifier. The 190-fold samples are made up of data from segments 0, 2, 3, 4, 6 for 38 of the 39 participants. Data from segments 1 and 5 were removed during the training of the classifiers as these segments served to strain the sustained attention of the participants. Additionally, given the layout of these two segments, very little quantitative discriminatory value was found. This was confirmed by the increase in the classification performance of the classifiers after the removal of these two segments. Data from the five gameplay segments of the 39th participant were kept as hold-out samples for manual verification.

To classify a participant as either neurotypical or ADHD$_I$, of the nine evaluated MLCs, the three MLCs with the highest AUC and sensitivity-to-specificity ratio were selected for further evaluation. The sensitivity indicates the number of correctly identified ADHD$_I$ participants with respect to the total number of ADHD$_I$ participants. The specificity indicates the number of correctly identified neurotypical participants with respect to the total number of neurotypical participants. These two parameters can be used to determine a crude estimate of the AUC by taking the average of their combined values, as calculated from the confusion matrix, at a given threshold (Powers, 2011; Sokolova et al., 2006). However, a more accurate result is obtained by integrating over the ROC curve (DeLong et al., 1988). The results in Table 2 indicate the performance of the consensus algorithm implemented on the top three performing MLCs. The corresponding 95% confidence intervals (CI) for these MLCs are presented in Table 3.

Table 2. Performance metrics for the top three adjusted MLMs (190 participant samples).

Two-Class Classifier	Accuracy %	Sensitivity %	Specificity %	AUC
Boosted Decision Tree	80.5	96.7	20	0.806
Neural Network	78.4	82.7	62.5	0.836
LDSVM	91.1	92.9	82.9	0.942

Table 3. 95 % Confidence intervals (CI) for the top three MLMs.

Two-class classifier	95 % CI (Sensitivity)	95 % CI (Specificity)
Boosted Decision Tree	[93.8 %, 99.5 %]	[7.6 %, 32.4 %]
Neural Network	[76.4 %, 89 %]	[49.2 %, 75.8 %]
LDSVM	[88.9 %, 96.9 %]	[70.4 %, 95.3 %]

As seen in Tables 2 and 3, the top performing MLC is the Two-Class LDSVM. Additionally, Table 4 and Figure 3 show the confusion matrix and receiver operating characteristic (ROC) curve for the two-class LDSVM classifier, respectively; these represent the performance of the classifier before the implementation of the consensus algorithm. Table 5 illustrates the tuned hyperparameters of the two-class LDSVM classifier, normalising the dataset with a MinMax normalizer before training.

Table 4. Confusion matrix for the 190 samples of the 38 participants.

		True Condition	
		ADHD	Non-ADHD
Test prediction	ADHD	144 (true positive)	11 (false positive)
	Non-ADHD	6 (false negative)	29 (true negative)

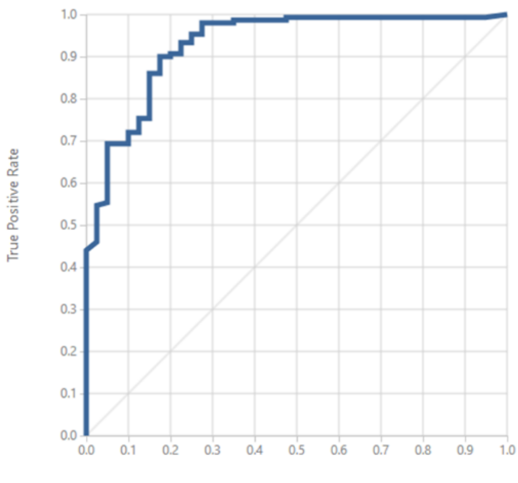

Figure 3. MLC ROC curve using the 190 samples of the 38 participants.

Table 5. Tuned Two-Class LDSVM Hyperparameters.

Field	Value
Tree Depth	2
Lambda W	0.08536786
Lambda Theta	0.0647911057
Lambda Theta Prime	0.061967507
Sigma	0.447105616
Num Iterations	12483
Normalizer Type	MinMax
Allow Unknown Levels	True
Random Number Seed	12345

A cloud-based model of the top performing classifier (two-class LDSVM) was then created with an application programming interface (API) for manual model verification with the 39th participant's segment data. The responses from the API for the five individual gameplay segments can be seen in Table 6. The final consensus classification, C_f, indicates the stepwise implementation of the consensus algorithm.

Table 6. Classifier response for each segment of the 39th participant.

Segment	Classification	Probability score	C_f
0	1	0.9570	+1
1	-	-	-
2	1	0.8538	+1
3	0	0.2024	-1
4	1	0.8743	+1
5	-	-	-
6	1	0.9967	+1
Actual Diagnosis	1		3

As illustrated in Table 6, substituting the individual gameplay segment classifications from the API response into the consensus algorithm provides a final consensus classification (C_f) and consensus confidence (C_c) for the 39th participant as follows:

$$C_f = 3 \geq 0,\ ADHD_I$$
$$C_c = 80\ \%$$

By making use of the consensus algorithm, the 39th participant was correctly classified according to the actual participant diagnosis. Additionally, the consensus confidence indicates a strong degree of certainty for the classification. The consensus algorithm was then implemented to evaluate all the incorrectly classified gameplay segments highlighted in Table 4. Findings indicated that five out of the 17 segments belonged to a single participant. Each of the 12 remaining incorrect classifications of the MLC belonged to different participants and were insufficient to "swing the vote" once the consensus algorithm was implemented. The final classification output after implementing the consensus algorithm for all 39 participants can be seen in the confusion matrix, Table 7.

Table 7. Confusion matrix for all 39 participants.

		True Condition	
		ADHD	Non-ADHD
Test prediction	ADHD	31 (true positive)	1 (false positive)
	Non-ADHD	0 (false negative)	7 (true negative)

The performance metrics for the ADHD Screening Tool classifying the 39 participants, can be seen in Table 8. The 95 % CI (Sensitivity) is [100 %, 100 %] and 95 % CI (Specificity) is [64.6 %, 100 %].

Among the LDSVM performance metrics, Delta is the proportion difference in the event of interest between the first and the second measurement standards, the first being the gold standard (clinical assessment) and the second the ADHD Screening Tool. Eta represents the total population proportion of disagreement between the two measurement standards. As determined from the confusion matrix in Table 7, the ADHD Screening Tool achieved Delta and Eta values of 0.025 (Table 8). Compared to the Delta and Eta values of the proposed sample size determined using McNemar's test, the ADHD Screening Tool achieved Delta and Eta values which were six and eight times smaller, respectively. The findings are statistically significant ($p < 0.05$), with a p-value of 1.07E-16.

Table 8. ADHD Screening Tool performance metrices for LDSVM (39 participants). The Delta and Eta values calculated from McNemar's test (proposed sample size 76) are included.

Accuracy (%)	Sensitivity (%)	Specificity (%)	Delta (δ)	Eta (η)	Delta (δ) for proposed sample size	Eta (η) for proposed sample size
97.4	100	87.5	0.025	0.025	0.15	0.2

As additional analysis, and to overcome the participant class imbalance, the synthetic minority oversampling technique (SMOTE) was implemented. This statistical technique increases the number of instances in a dataset to achieve balanced classes. The algorithm generates new instances of the minority class only by taking samples from the minority class feature space and the nearest neighbours. This over-sampling technique is frequently used to improve the performance of traditional classifiers when using imbalanced datasets, with overfitting being combated by increasing the number of nearest neighbours from which to sample (Guo et al., 2018).

Of the 195 samples in this study (data from segments 0, 2, 3, 4, 6 for 39 participants), 155 constituted the $ADHD_I$ class. The remaining 40 samples constituted the neurotypical class and were used as input to the SMOTE analysis until a ratio of 155:160 ($ADHD_I$: neurotypical) samples was achieved. Forty nearest neighbours were used to achieve the least possible overfitting influence, with a stratified 10-fold cross validation during training. The 160 neurotypical samples could be interpreted as equating to 32 neurotypical participants. The resultant classifier confusion matrix minus data from participant 39, is shown in Table 9. The corresponding classifier hyperparameters are shown in Table 10.

Table 9. SMOTE classifier confusion matrix for 310 samples.

		True Condition	
		ADHD	Non-ADHD
Test prediction	ADHD	136 (true positive)	9 (false positive)
	Non-ADHD	14 (false negative)	151 (true negative)

Table 10. Tuned SMOTE two-class LDSVM hyperparameters.

Field	**Value**
Tree Depth	2
Lambda W	0.0290210117
Lambda Theta	0.0638889
Lambda Theta Prime	0.0301830862
Sigma	0.37583825
Num Iterations	14263
Normalizer Type	MinMax
Allow Unknown Levels	True
Random Number Seed	12345

Across the 10 folds, the classifier achieved an accuracy standard deviation of 4.8% and an AUC standard deviation of 2%. The AUC for the 310 samples was 0.98. The resulting performance parameters are shown in Table 11.

Table 11. SMOTE two-class LDSVM performance metrices.

Accuracy (%)	Sensitivity (%)	Specificity (%)	Delta (δ)	Eta (η)
92.6	90.7	94.4	0,016	0,074

Discussion

Analysis of the confusion matrix shown in Table 4 indicates that the top classifier incorrectly classified 17 of the 190 samples. Five of the 17 incorrectly classified samples belong to a single participant, following the intervention of the consensus algorithm. The remaining incorrectly classified samples of Table 4 were incapable of influencing the consensus algorithm to incorrectly classify a larger number of participants. For the purposes of this study, false positive classifications are preferred over false negative classification, as the cost of a false positive outweighs the challenges as a result of a false negative classification, especially within the classroom setting (Topkin et al., 2015).

In light of findings from the literature highlighting the significance of omission errors in the presence of visual and a combination of visual and auditory distractors, the ADHD Screening Tool found segment three of the game (visual distractors only) to be a poor discriminator (Berger & Cassuto, 2014). However, both the auditory and a combination of auditory and visual distractors were found to be strong discriminators for Participant 39 of this study (Figure 1 and Table 6). A comparison of the performance of existing tools as presented in the literature with the results from the ADHD Screening Tool is presented in Table 12.

The ADHD Screening Tool that has been presented, outperforms the tools described in the literature. The estimated time taken to receive classification feedback for a participant using the ADHD Screening Tool testing ranges between 8-15 minutes given the variance in the participant's performance and requires no interpretation. The MOXO-CPT takes 15.2 minutes to complete irrespective of the participant's performance, with the Integrated Visual and Auditory Continuous Performance Test (IVA + CPT) and the Test of Variables of Attention (T.O.V.A) taking approximately 20 minutes and 21.6 minutes to complete, respectively. Groundskeeper takes approximately 31 minutes to complete. Additional time is required for a trained professional to administer the test, as well as interpret the results of the MOXO-CPT, T.O.V.A, Groundskeeper and IVA + CPT. The MOXO-CPT, T.O.V.A, Groundskeeper and IVA+CPT are all different forms of continuous performance tests incorporating the similar go/no-go mechanisms. The ADHD Screening Tool is a continuous interactive interface which incorporates a range of interdependent time-related performance measures which is unique to each participant. This configuration continuously monitors and records a participant's interactions, including interactions between game-driven stimuli. The ADHD Screening Tool makes use of a single internet connected mobile device, enabling portability for quick and easy testing. The MOXO-CPT, T.O.V.A, Groundskeeper and IVA+CPT platforms all make use of bulky hardware made up of various combinations of components such as internet connected computers, external speakers, keyboards and an additional external button is used for the T.O.V.A. The Groundskeeper game requires a computer in addition to the game cubes to run the game (Berger et al., 2013; Kim et al., 2015; Greenberg et al., 2018; Heller et al., 2013)

The ADHD Screening Tool performs most similarly to the clinical assessment (gold standard) for the given study population. It is also able to provide participant classification feedback in a shorter time than both the gold standard and the tools identified in the study literature.

Table 12. Comparison of classification performance with existing technology.

References	Tool	Sample	Performance Measures (%)
(Berger et al., 2017)	MOXO	798 children (Mean Age = 9.27 yrs.): 339 ADHD; 459 non-ADHD	Sensitivity = 86.5* Specificity = 86.2*
(Schatz et al., 2001)	T.O.V.A	48 children: (Mean Age = 11.1 yrs.): 28 ADHD; 20 non-ADHD	Sensitivity = 85.7 Specificity = 70
(Heller et al., 2013)	Groundskeeper	52 children: (Mean Age = 13.65 yrs.): 26 ADHD; 26 non-ADHD	Sensitivity = 76.9 Specificity = 80.8
(Kim et al., 2015)	IVA + CPT	157 children (Mean Age = 9.25 yrs.): 85 ADHD; 72 non-ADHD	Sensitivity = 72.9 Specificity = 70.9
Current Device	ADHD Screening Tool	39 children (Mean Age = 9.02 yrs.): 31 ADHD; 8 non-ADHD	Sensitivity = 100 Specificity = 87.5

* Averaged for age range between seven and 12. Values were sampled at optimal sensitivity and specificity.

The study limitations include the small control group and the small total number of participants and this should be addressed in future studies for greater confidence in the results. The implementation of the SMOTE analysis attempts to jointly mitigate the participant imbalance and highlight the impact that a balanced dataset could have on the performance of the classifier. Even though computational methods were introduced to combat classifier training bias, the classification of a single participant carries a large weight in the performance measures of the ADHD Screening Tool. The small sample size resulted in only the data of participant 39 being used as hold-out samples for manual testing and validation of the classifier and consensus algorithm.

This study served as a pilot and the findings support the feasibility of the tool. Results of this study show that a mobile device-based tool can be used to discriminate between $ADHD_I$ and non-ADHD participants in reduced time compared to conventional methods. Future work should seek to include a larger study population, investigating the ability of the tool to discriminate between all three subtypes of ADHD and the non-ADHD control group. In comparison with the aforementioned tools, the ADHD Screening Tool is considered a low-cost setup as it simply requires an internet connected mobile or tablet device to capture participant data.

Acknowledgements

The authors would like to thank Rose-Hannah Brown for her guidance, Professor Daan Nel and Professor Martin Kidd for statistical assistance, as well as the school, parents and children who were involved in the clinical study.

References

Altun, K. & Barshan, B. 2010. Human Activity Recognition Using Inertial/Magnetic Sensor Units. *Human Behavior Understanding*, 6219(1): 38–51. https://doi.org/10.1007/978-3-642-14715-9_5.

Berger, I. & Cassuto, H. 2014. The effect of environmental distractors incorporation into a CPT on sustained attention and ADHD diagnosis among adolescents. *Journal of Neuroscience Methods*, 222(1): 62–68. https://doi.org/10.1016/j.jneumeth.2013.10.012.

Berger, I., Slobodin, O., Aboud, M., Melamed, J. & Cassuto, H. 2013. Maturational delay in ADHD: evidence from CPT. *Frontiers in Human Neuroscience*, 7(1): 691–702. https://doi.org/10.3389/fnhum.2013.00691.

Berger, I., Slobodin, O. & Cassuto, H. 2017. Usefulness and Validity of Continuous Performance Tests in the Diagnosis of Attention-Deficit Hyperactivity Disorder Children. *Archives of Clinical Neuropsychology*, 32(1): 81–93. https://doi.org/10.1093/arclin/acw101.

D'Souza, W.T. & Kavitha, R. 2017. Human Activity Recognition Using Accelerometer and Gyroscope Sensors. *International Journal of Engineering and Technology*, 9(2):1171–1179. DOI: 10.21817/ijet/2017/v9i2/170902134.

DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L. 1988. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 44(3):837–845.

Díaz-Orueta, U., Garcia-López, C., Crespo-Eguílaz, N., Sánchez-Carpintero, R., Climent, G. & Narbona, J. 2017. AULA virtual reality test as an attention measure: Convergent validity with Conners' Continuous Performance Test. *Child Neuropsychology*, 20(3):328–342. https://doi.org/10.1080/09297049.2013.792332.

Greenberg, L.M., Holder, C., Kindschi, C.L. & Dupuy, T.R. 2018. *Test of Variables of Attention Clinical Manual*. Available: http://www.tovatest.com/.

Gualtieri, C.T. & Johnson, L.G. 2005. ADHD: Is Objective Diagnosis Possible? *Psychiatry (Edgmont)*, 2(11):44–53. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2993524/.

Guo, H., Zhou, J. & Wu, C.-A. 2018. Imbalanced Learning Based on Data-Partition and SMOTE. *MDPI*, 9(9):238. http://www.mdpi.com/2078-2489/9/9/238.

Heller, M.D., Roots, K., Srivastava, S., Schumann, J., Srivastava, J. & Hale, T.S. 2013. A Machine Learning-Based Analysis of Game Data for Attention Deficit Hyperactivity Disorder Assessment. *Games for Health Journal*, 2(5):291–298. http://online.liebertpub.com/doi/abs/10.1089/g4h.2013.0058.

Hussain, L. 2018. Detecting epileptic seizure with different feature extracting strategies using robust machine learning classification techniques by applying advance parameter optimization approach. *Cognitive Neurodynamics*, 12(3):271–294. https://doi.org/10.1007/s11571-018-9477-1.

Kessler, R.C., Adler, L., Barkley, R., Biederman, J., Conners, C.K., Demler, O., Faraone, S. V., Greenhill, L.L., Howes, M.J., Secnik, K., Spencer, T., Ustun, T.B., Walters, E.E. & Zaslavsky, A.M. 2006. The Prevalence and Correlates of Adult ADHD in the United States: Results From the National Comorbidity Survey Replication. *The American Journal of Psychiatry*, 163(4):716–723. https://ajp-psychiatryonline-org.ez.sun.ac.za/doi/full/10.1176/ajp.2006.163.4.716.

Kim, J., Lee, Y., Han, D., Min, K., Kim, D. & Lee, C. 2015. The utility of quantitative electroencephalography and Integrated Visual and Auditory Continuous Performance Test as auxiliary tools for the Attention Deficit Hyperactivity Disorder diagnosis. *Clinical Neurophysiology*, 126(3):532–540. http://dx.doi.org/10.1016/j.clinph.2014.06.034.

Li, F., Shirahama, K., Nisar, M.A.N., Köping, L. & Grzegorzek, M. 2018. Comparison of feature learning methods for human activity recognition using wearable sensors. *MDPI*, 18(2):1–22. https://www.mdpi.com/1424-8220/18/2/679.

Li, F., Zheng, Y., Smith, S.D., Shic, F., Moore, C.C., Zheng, X., Qi, Y., Liu, Z. & Leckman, J.F. 2016. A preliminary study of movement intensity during a Go/No-Go task and its association with ADHD outcomes and symptom severity. *Child and Adolescent Psychiatry and Mental Health*, 10(1):1–10. http://dx.doi.org/10.1186/s13034-016-0135-2.

Mao, W., Mu, X., Zheng, Y. & Yan, G. 2014. Leave-one-out cross-validation-based model selection for multi-input multi-output support vector machine. *Neural Computing and Applications*, 24(2):441–451. https://doi.org/10.1007/s00521-012-1234-5.

McDermott, A.F., Rose, M., Norris, T. & Gordon, E. 2016. A Novel Feed-Forward Modeling System Leads to Sustained Improvements in Attention and Academic Performance. *Journal of Attention Disorders*, 1(1):1–14. https://doi.org/10.1177/1087054715623044.

Moraru, A., Pesko, M., Porcius, M., Fortuna, C. & Mladenic, D. 2010. Using machine learning on sensor data. *Journal of Computing and Information Technology*, 18(4):341–347. https://doi.org/10.2498/cit.1001913.

Perold, M., Louw, C. & Kleynhans, S. 2010. Primary school teachers' knowledge and misperceptions of attention deficit hyperactivity disorder (ADHD). *South African Journal of Education*, 30(3):457–473. https://hdl.handle.net/10520/EJC32240.

Polanczyk, G. V., Willcutt, E.G., Salum, G.A., Kieling, C. & Rohde, L.A. 2014. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *International Journal of Epidemiology*, 43(2):434–442. https://doi.org/10.1093/ije/dyt261.

Powers, D.M.W. 2011. Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63.

Preuss, U., Ralston, S.J., Baldursson, G., Falissard, B., Lorenzo, M.J., Rodrigues Pereira, R., Vlasveld, L., Coghill, D., Curatolo, P., Dalsgaard, S., Döpfner, M., Hervas, A., Le Heuzey, M.F., Nøvik, T.S., Rasmussen, P., Riley, A.W., Rothenberger, A., Spiel, G. & Steinhausen, H.C. 2006. Study design, baseline patient characteristics and intervention in a cross-cultural framework: Results from the ADORE study. *European Child and Adolescent Psychiatry*, 15(Suppl 1):I4–I14. https://doi.org/10.1007/s00787-006-1002-0.

Rettner, R. 2014. Autism's Rise: Researchers Look at Why Cases Are Increasing. Available: http://www.livescience.com/48586-autism-prevalence-increase-reporting.html.

Schatz, A.M., Ballantyne, A.O. & Trauner, D.A. 2001. Sensitivity and Specificity of a Computerized Test of Attention in the Diagnosis of Attention-Deficit/Hyperactivity Disorder. *Assessment*, 8(4):357–365. https://doi.org/10.1177/107319110100800401.

Sharp, S.I., Mcquillin, A. & Gurling, H.M.D. 2009. Genetics of attention-deficit hyperactivity disorder (ADHD). *Neuropharmacology*, 57(7–8):590–600. https://doi.org/10.1016/j.neuropharm.2009.08.011.

Silverstein, M., Hironaka, L.K., Feinberg, E., Sandler, J., Pellicer, M., Chen, N. & Cabral, H. 2016. Using Clinical Data to Predict Accurate ADHD Diagnoses Among Urban Children. *Clinical Pediatrics*, 55(4):326–332. https://doi.org/10.1177/0009922815591882.

Sokolova, M., Japkowicz, N. & Szpakowicz, S. 2006. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. *AI 2006: Advances in Artificial Intelligence*, 4304(1):1015–1021. https://doi.org/10.1007/11941439_114.

Thomas, R., Sanders, S., Doust, J., Beller, E. & Glasziou, P. 2015. Prevalence of Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-analysis. *Pediatrics*, 135(4):e994–e1001. https://doi.org/10.1542/peds.2014-3482.

Topkin, B., Roman, N. V. & Mwaba, K. 2015. Attention Deficit Disorder (ADHD): Primary school teachers' knowledge of symptoms, treatment and managing classroom behaviour. *South African Journal of Education*, 35(2):1–8. https://hdl.handle.net/10520/EJC171011.

Zhang, J. & Wang, S. 2016. A fast leave-one-out cross-validation for SVM-like family. *Neural Computing and Applications*, 27(6):1717–1730. https://doi.org/10.1007/s00521-015-1970-4.