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

by
Romano Swarts

Thesis presented in partial fulfilment of the requirements for the degree of Master of Engineering (Mechatronic) in the Faculty of Engineering at Stellenbosch University

Supervisor: Prof, Pieter Rousseau Fourie
Co-supervisor: Prof, Dawie van den Heever

April 2019

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

April 2019

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

Plagiaatverklaring / *Plagiarism Declaration*

1 Plagiaat is die oorneem en gebruik van die idees, materiaal en ander intellektuele eiendom van ander persone asof dit jou eie werk is.
Plagiarism is the use of ideas, material and other intellectual property of another's work and to present is as my own.

2 Ek erken dat die pleeg van plagiaat 'n strafbare oortreding is aangesien dit 'n vorm van diefstal is.
I agree that plagiarism is a punishable offence because it constitutes theft.

3 Ek verstaan ook dat direkte vertalings plagiaat is.
I also understand that direct translations are plagiarism.

4 Dienooreenkomstig is alle aanhalings en bydraes vanuit enige bron (ingesluit die internet) volledig verwys (erken). Ek erken dat die woordelikse aanhaal van teks sonder aanhalingstekens (selfs al word die bron volledig erken) plagiaat is.
Accordingly all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism.

5 Ek verklaar dat die werk in hierdie skryfstuk vervat, behalwe waar anders aangedui, my eie oorspronklike werk is en dat ek dit nie vantevore in die geheel of gedeeltelik ingehandig het vir bepunting in hierdie module/werkstuk of 'n ander module/werkstuk nie.
I declare that the work contained in this assignment, except otherwise stated, is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this module/assignment or another module/assignment.

Studentenommer / *Student number*	**Handtekening / *Signature***
Voorletters en van / *Initials and surname*	**Datum / *Date***

ABSTRACT

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

R. Swarts

Department of Mechanical and Mechatronic Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Thesis: MEng (Mechatronic)
April 2019

This study investigated the effectiveness of a tablet-based game that incorporated machine learning to screen participants between the ages of six and twelve years for ADHD inattentive subtype. Prior to the design and development of the ADHD screening tool, a thorough investigation of the literature was conducted. Additionally, existing ADHD screening tools and cognitive training tools were identified. This research project implemented lessons learned from the literature, as well as input from medical professionals and the DSM-V diagnostic criteria. The ADHD screening tool presents a patient-testing interface in the form of a tablet-based game with a cloud-based machine learning classifier. The cloud-based classifier is integrated with an algorithm, and together they can discriminate between ADHD and non-ADHD patients with a sensitivity of 100 % and specificity of 87.5 %. The device used for testing was a single, internet connected, commercially available tablet. No additional hardware is required.

UITTREKSEL

ADHD Keuring Instrument: Ondersoek die effektiwiteit van 'n tablet-gebaseerde speletjie met masjienleer

R. Swarts

Departement Meganiese en Megatroniese Ingenieurswese,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid-Afrika.

Tesis: MIng (Megatronies)
April 2019

Hierdie studie het ondersoek ingestel om die effektiwiteit van 'n tablet-gebaseerde speletjie om deelnemers tussen die ouderdomme van ses en twaalf jaar vir ADHD-onoplettende subtipe te evalueer. Voor die ontwerp en ontwikkeling van die ADHD keuring instrument was 'n deeglike ondersoek ingestel om die literatuur te ondersoek. Daarbenewens was die bestaande ADHD keuring instrumente en kognitiewe opleidingsinstrumente geïdentifiseer. Hierdie navorsingsprojek het lesse van uit die literatuur geïmplementeer, sowel as insette van mediese professionele en die DSM-V diagnostiese kriteria. Die ADHD evalueringsinstrument bied 'n pasiënt-toets in die vorm van 'n tablet-gebaseerde speletjie met 'n wolk-gebaseerde masjienleer klassifiseerder. Die wolk-gebaseerde klassifiseerder is geïntegreer met 'n algoritme, en saam kan hulle onderskei tussen ADHD en nie-ADHD pasiënte met 'n sensitiwiteit van 100 % en spesifisiteit van 87.5 %. Die toestel wat gebruik was vir toetsing is 'n enkele, internet-gekoppelde, kommersieel beskikbare tablet. Geen bykomende hardeware word benodig nie.

ACKNOWLEDGEMENTS

Above all, I thank God for the opportunity, provision and strength. For the path You have set out before me, and the people I meet on the way, may it all be to Your glory, from now until that day. I would also like to express my sincere gratitude to the following people:

- To my Mom, for the sacrifice, encouragement and support. Without you, none of this would have been possible.
- To Professor Pieter Fourie for the means, perspective and unwavering vision to make a difference.
- To Dr Dawie van den Heever, thank you for the light-hearted but valuable time and effort you've invested in me.
- To Mr Atkinson, your input and resourcefulness was one of the pillars in seeing this project become a reality. Thank you.
- To Mrs. Rose-Hannah Brown, thank you for your generosity and sacrifice. Your kindness will not be forgotten.
- To the parents and teachers who made sacrifices and went out of their way, thank you.
- To my editor in chief, may the results be worth the sleepless nights. Your time and efforts have been invaluable.

DEDICATION

To the billion in Africa.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Abbreviations

ADHD	Attention-Deficit/Hyperactive Disorder
ADHD-C	Combined subtype
ADHD-CH	Combined subtype plus Hyperactive-impulsive subtype
ADHD-H	Hyperactive-impulsive subtype
ADHD-I	Inattentive subtype
ADORE	Attention-Deficit Hyperactivity Disorder Observational Research in Europe
AMLS	Azure Machine Learning Studio
ANX	Anxiety Disorder
APA	American Psychiatric Association
API	Application Programming Interface
AUC	Area Under the Curve
CD	Conduct Disorder
CHRNB	Children's Halstead-Reitan Neuropsychological Test Battery
CPT	Continuous performance test
DEP	Depression Disorder
DSM	Diagnostic and Statistical Manual of Mental Disorders
EEG	Electroencephalograph
FDA	Food and Drug Administration
FFM	Feed-Forward Modelling
GB	Gigabyte
GPU	Graphics Processing Unit
HREC	Health Research Ethics Committee
ICD	International Classification of Mental and Behavioural Disorders
LD	Learning Disabilities
LDSVM	Locally-Deep Support Vector Machine
LNNB	Laria-Nebraska Neuropsychological Test Battery
LOOCV	Leave-One-Out Cross Validation
LSB	Level Streaming Blueprints
ML	Machine Learning
MLC	Machine Learning Classifier
ODD	Oppositional Defiant Disorder
PCA	Principal Component Analysis
RAM	Random Access Memory
RFC	Random Forest Classifier
RINB	Reitan-Indiana Neuropsychological Test Battery
ROC	Receiver Operating Characteristic

RT	Response Time
RTV	Response Time Variability
SDA	Standard Deviation of accuracy
SNAP	Swanson, Nolan and Pelham
SPD	Sensory Processing Dysfunction
SPD_{+IA}	SPD subgroups
SVM	Support Vector Machine
SWAN	Strengths and Weaknesses of Attention-Deficit/Hyperactivity-symptoms and Normal-behaviours
TOVA	Test of Variables of Attention
TUI	Tangible User Interface
UE	Unreal Engine
USA-HIPPA	Health Insurance Portability and Accountability Act
VADPRS	Vanderbilt ADHD Parent Rating Scale
VADTRS	Vanderbilt ADHD Diagnostic Rating Scale
VR	Virtual Reality
WCED	Western Cape Education Department
WISC	Wechsler Intelligence Scale for Children

Symbols

C_c	Percentage of consensus……..	[%]
C_f	Final consensus classification……...	[-]
N	Number of samples	[-]
c	Classification of segment……...	[-]
c_i	Classification of segment i…….	[-]
i	Segment in-game position…….	[-]
n	Total number of segments included in the analysis……...	[-]
p_i	Confidence score for segment i……	[-]
r	Pearson's correlation……..	[-]
y	Actual diagnosis……..	[-]
$\bar{y}$	Correlated feature……	[-]

1 INTRODUCTION

1.1 Background

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. [1–4]

Multiple key developments take place in the brain during the growth stages of infancy (zero to two years), toddlers (three to five years), school age (six to 12 years) and adolescence (13 to 18 years). 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. [1, 2, 5, 6]

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. As a result, over and under-diagnosis of ADHD has been widely debated, driven by variations in world-wide prevalence and broadening diagnostic criteria [7]. Rosenberg et al. suggest in this regard that the development of ADHD biomarkers, which reflect pathological understanding of the disorder, and which can be used as an identification tool, could combat diagnostic subjectivity [8].[1]

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 [3–5]. Another consideration is the cost of the diagnostic process, the cumulative fee of which can include: clinical psychologists, paediatricians and other medical practitioners to evaluate carer/parent and teacher questionnaires and academic performance, providing and administering neuropsychological test batteries and screening tools, and contact sessions with the child [12]. Teachers are most often the first to make recommendations to carers/parents for ADHD, based upon observed classroom behaviour of children who make it difficult for other students to perform or teachers to cope. However, the lack of knowledge and understanding of ADHD often leads to teachers developing negative views of the learners they refer for assessment. [13]

[1] The term “biomarker” refers to a broad subcategory of medical signs. It is an objective indication of a medical state which is observed from outside the patient and can be measured accurately and reproducibly. [120]

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 in this 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.

The aim of this project, therefore, is to develop an 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. Such a device, with related software, must be capable of capturing a quantitative feature set during participant testing.

1.2 Motivation

It is common knowledge that private medical services are costly. Given the current diagnostic process of ADHD mentioned above, it follows that early, accurate screening for ADHD could help to prevent these high diagnostic costs. It would also help to ensure that children identified by the tool could be referred for diagnosis and receive treatment as early as possible. An ADHD screening tool could be used to ascertain the effectiveness of existing or new stimulant or non-stimulant type medication or treatment, to monitor the degree of severity of ADHD, as well as to help carers and parents to monitor dosage effects and allow strict control over ADHD medication. These benefits, in turn, decrease the need for the patient to frequently visit a mental healthcare professional. A portable diagnostic tool could be utilised in rural and remote areas within South Africa, as well as abroad. The tool will serve as a method to aid proper diagnosis by providing quantitative output. The tool could also be used to conduct population studies in order to ascertain the incidence of ADHD for clinical or statistical research purposes.

The focus of this study will specifically be to determine the ability and effectiveness of the ADHD screening tool to distinguish between ADHD and non-ADHD participants. The study entails the design and development of a portable ADHD screening tool that is easy to administer by a layman without in-depth knowledge of ADHD. The screening tool has been designed to enable cares, parents and teachers to identify children with potential ADHD-I during the early developmental stages of children's lives. The device is intended to provide feedback to the administrator so that children identified by the tool can be referred to a clinical psychologist or paediatrician for an official ADHD evaluation according to the gold standard (discussed in section 2.2 below). The findings of this study will either reject or not reject the null hypothesis found in section 1.4 below.

1.3 Statement of the Problem

ADHD screening tool: Investigating the effectiveness of a tablet-based game with machine learning.

1.4 Statement of Hypotheses

The null hypothesis for this research project is as follows:

There is no difference in the discriminating ability of the gold standard and that of a screening tool with machine learning when used to distinguish ADHD-I participants from a normal population group.

The alternate hypothesis states that:

There is a difference in the discriminating ability of the gold standard and that of a screening tool with machine learning when used to distinguish ADHD-I participants from a normal population group.

1.5 Aims and Objectives

The main aims of this research project are as follows:

1. To research and develop a portable, screening tool that incorporates machine learning to screen participants with potential ADHD inattentive subtype; and
2. To test the feasibility of the screening tool by taking recordings of ADHD and non-ADHD participants identified by clinical psychologists, teachers and parents, and comparing the results with existing technology.

The development of a portable ADHD screening tool will need to meet the following project objectives:

1. Development of a game to capture ADHD-I features;
2. The screening tool should be portable, accessible and easy to administer;
3. The cost of the screening tool should be relatively affordable;
4. The screening tool should contain a wireless data sharing capability to safely store participant data online;
5. The screening tool should have the capability of reporting screening feedback; and
6. The development of machine learning algorithms to classify a participant as either neurotypical or having ADHD-I.[2]

1.6 Thesis Outline

Chapter Two:

This chapter comprises the literature review for this study, which provides background information of ADHD, its subtypes, prevalence and genetics. The chapter also discusses the ADHD diagnostic gold standard, existing diagnostic technology and machine learning.

---

[2] Neurotypical - Exhibiting or characteristic of typical neurological development [121].

Chapter Three:

Chapter Three addresses the hardware and software components used to implement the research aims, and also discusses the design specifications of the ADHD screening tool.

Chapter Four:

This chapter describes the research methodology of the project, comprising the ethics statement, the study design and the clinical study.

Chapter Five:

In this chapter, results from the clinical study are analysed and presented by using statistical methods found in the literature.

Chapter Six:

This chapter provides an overview of the project cost and safety considerations. Results from the clinical study are also discussed and compared with existing technology.

Chapter Seven:

The concluding chapter provides a summary of the work completed in relation to the project objectives. Project limitations, lessons learned and recommendations for future work are also discussed.

2 LITERATURE REVIEW

The aim of this chapter is to identify the need for an assistive diagnostic method for ADHD. This will be achieved by presenting an overview of ADHD, followed by a discussion of the important ADHD rating scales and neuropsychological assessments used to complement the diagnosis of ADHD. This chapter will then discuss existing technology, followed by a comparison with this study.

2.1 ADHD

ADHD is one of the most common, highly heritable, neurobiological, developmental disorders, prevalent predominantly in children. The disorder is characterised primarily by symptoms of developmentally inappropriate levels of inattentiveness or hyperactivity and impulsivity, and is one of the most thoroughly researched medical conditions. [2, 3, 14, 15]

The field of neurobiology provides insight to the relationship between ADHD and certain regions of the brain. As shown in Figure 1, ADHD impacts the frontal and parietal cortexes, basal ganglia, cerebellum and the corpus collosum [16–18]. Purper-Ouakil et al. highlight that these regions are involved in the functional network relating to ADHD. Furthermore, findings indicate that alterations in brain structures exist with neural networks possibly being combined in ADHD, leading to organised brain phenotypes. [18]

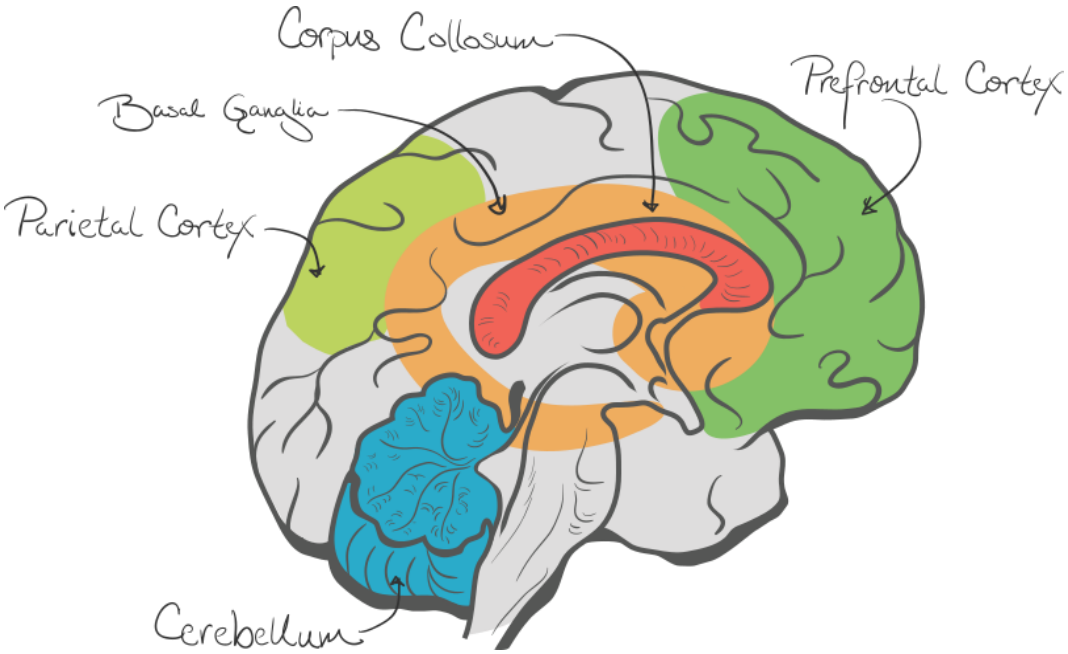


Figure 1: Schematic representation of the brain regions involved in attention [19].

2.1.1 Subtypes

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-V), ADHD has three subtypes, including ADHD-I, ADHD-H and the ADHD-C of both inattentive and hyperactive-impulsive symptoms (Figure 2).

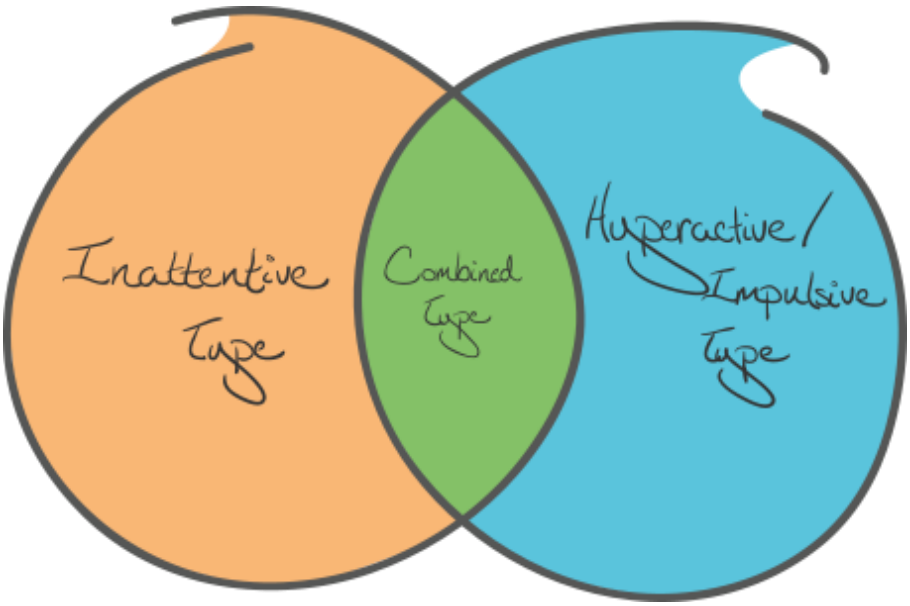


Figure 2: ADHD subtypes [19].

With the exception of ADHD-C, which has a combined 18-point classification criteria, the classification criteria for ADHD-I and ADHD-H are comprised of nine points each. As shown in Table 1, each criterion is a description of the specific behavioural symptom linked to the specific ADHD subtype. [20]

Table 1: Symptoms of ADHD according to the DSM-V criteria [20].

	Inattentive	***Hyperactive/Impulsive***
1	Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).	Often fidgets with or taps hands or feet or squirms in seat.
2	Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).	Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place).
3	Often does not seem to listen when spoken to directly (e.g., mind seems elsewhere,	Often runs about or climbs in situations where it is inappropriate. (Note: In

	even in the absence of any obvious distraction).	adolescents or adults, may be limited to feeling restless.)
4	Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., starts tasks but quickly loses focus and is easily sidetracked).	Often unable to play or engage in leisure activities quietly.
5	Often has difficulty organizing tasks and activities (e.g., difficulty managing sequential tasks; difficulty keeping materials and belongings in order; messy, disorganized work; has poor time management; fails to meet deadlines).	Is often “on the go,” acting as if “driven by a motor” (e.g., is unable to be or uncomfortable being still for extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).
6	Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework; for older adolescents and adults, preparing reports, completing forms, reviewing lengthy papers).	Often talks excessively.
7	Often loses things necessary for tasks or activities (e.g., school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).	Often blurts out an answer before a question has been completed (e.g., completes people’s sentences; cannot wait for turn in conversation).
8	Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).	Often has difficulty waiting his or her turn (e.g., while waiting in line).
9	Is often forgetful in daily activities (e.g., doing chores, running errands; for older adolescents and adults, returning calls, paying bills, keeping appointments).	Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people’s things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).

Results published by Grizenko et al. [21] in 2010, establish significant differences by comparing ADHD-I with ADHD-C and ADHD-H (collectively, ADHD-CH). These two participant groups were combined for the purpose of statistical analysis due to the many significant differences found between ADHD-I and ADHD-H. The study included 371 participants between six and 12 years of age and evaluated the level of co-existing disorders, treatment response, and possible etiological factors.

Pertinent to this study, Grizenko et al. highlight important categorical differences between participants with ADHD-I compared to the ADHD-CH group. The study compared participants in terms of their levels of co-existence, treatment responses, and possible etiological factors. The findings indicate significant differences between the three subtype groups with regard to age, gender distribution, severity of symptoms and co-existing disorders. ADHD-I had the highest mean age at 9.6, with the largest female-to-male ratio at 29.2 %. The lowest externalizing symptomatology score was also found when compared to the ADHD-CH group, with an internalizing symptomatology score between ADHD-C and ADHD-H groups. Additionally, a higher frequency for the co-existence of CD was found in the ADHD-CH group. [22] Grizenko, et al. conclude that differences

between ADHD-I and ADHD-CH groups raise the possibility that the two may be two separate disorders.

A study conducted by Park, et al. in 2014 further supports the findings above and shows a significant difference between ADHD-I and ADHD-C groups when evaluating the severity of symptoms, comorbidity, environmental risk factors and neuropsychological characteristics. Groups involved were compared in terms of genetic, perinatal, and developmental risk factors, as well as clinical and neuropsychological characteristics. The study recruited 147 diagnosed participants, between six and 15 years of age, with a control group of 502 participants without ADHD. Findings indicated that the ADHD-C group showed more severe externalizing symptoms, as well as more deficits when completing a continuous performance test (CPT). The study also highlighted a greater likelihood of comorbid disorders for this group. [23]

A different viewpoint is offered in a study conducted by Lemiere, et al. in 2010. The study made use of the TEA-Ch test battery to determine the difference in everyday attention between ADHD-I and ADHD-C groups. This test battery includes aspects of everyday attention relating to selective attention, sustained attention and attention control. The study recruited 140 participants but concluded that the results showed few differences across tasks and did not provide much support for the value of distinction between the two groups in predicting difficulties in everyday attention. However, the study also confirms the age and gender distribution findings of the studies discussed above. [24]

2.1.2 Prevalence

Research conducted by Polanczyk et al. in 2007, reveals a worldwide, pooled prevalence for ADHD in youth (18 years and younger) to be an estimated 5.29 %. Employing a meta-regression analysis of 102 studies between 1978 and 2005, the study comprised a total of 171 756 subjects from regions worldwide (North America, Europe, Asia, South America, Oceania, Middle East and Africa). The prevalence estimate includes significant variability, however, this variability is acceptable considering the geographic origin of the studies, diagnostic criteria used, information sourcing methods, as well as the requirement of impairment for diagnosis [25].

According to the American Psychiatric Association (APA) in 2013, the worldwide cross-cultural prevalence of ADHD was about 5 % in children and about 2.5 % in adults (age 18 and older) [20, 26]. All respondents were evaluated per the DSM-V criteria. Considering the proactive mindset toward diagnostic readiness, the prevalence statistics for ADHD could only be seen to increase in accuracy, given that ADHD is currently underdiagnosed and undertreated [27, 28].

Findings from a further study conducted by Polanczyk et al. in 2014 can be seen in Figure 3. The study systematically reviewed 135 studies published between 1985 and 2012 and addressed the worldwide pooled prevalence of ADHD in youth (18 years and younger). Results for the period can be interpreted from the graph to be about 6.8 % and the average prevalence percentage for 2012 at about 5.3 % [1].

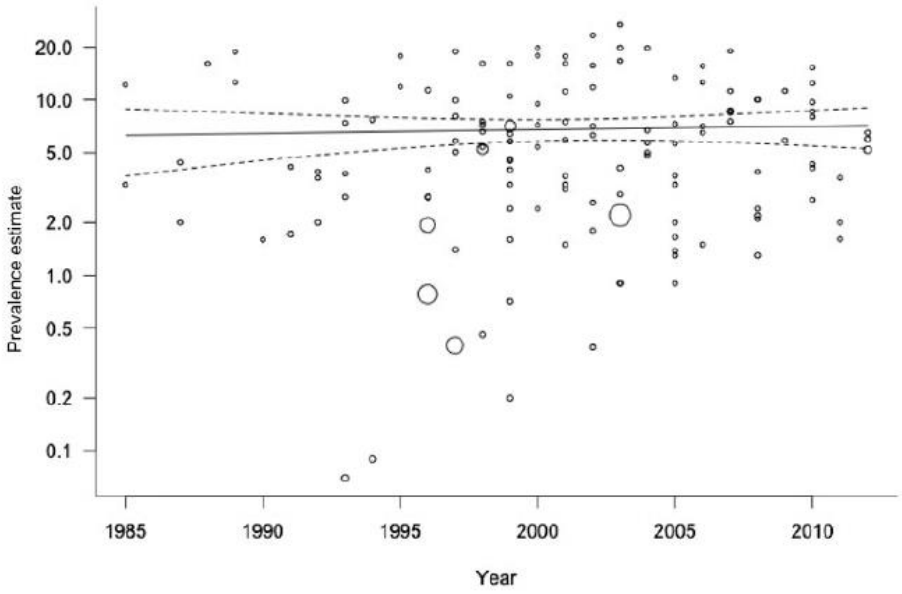


Figure 3: Worldwide average pooled prevalence estimates for ADHD in youth between 1985 and 2012 [1].

Results published by Thomas et al. in 2015 can be seen in Figure 4 and provide greater insight into the worldwide pooled prevalence for ADHD. The study reviewed a total of 175 studies published between 1977 and 2013, comprising 1 023 071 subjects under the age of 18. Findings indicate an estimated worldwide pooled prevalence of 7.1 % for the specified period. Although findings differ considerably from the study by Polanczyk et al. study in 2007, these differences can be attributed largely to the specified language restrictions of Polanczyk et al., as well as the 83 extra studies included by Thomas et al. [7]

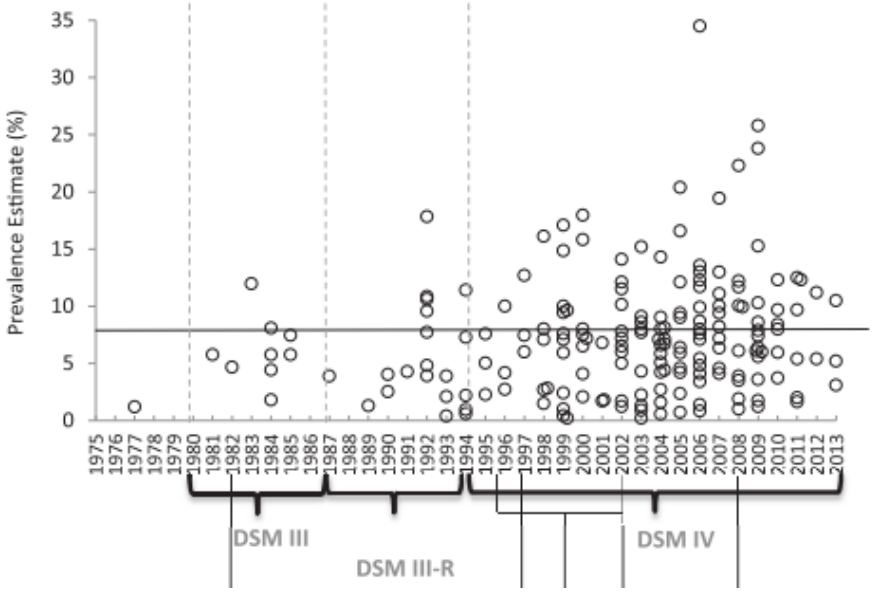


Figure 4: Worldwide average pooled prevalence estimates for ADHD in youth between 1977 and 2013 using the different DSM versions available [7].

Although ADHD is more commonly diagnosed during childhood, it has become recognised increasingly in adults too [6, 29]. Harrison et al. estimated in 2007 that the disorder affects between 2 and 4 % of the college student population in Canada [15], with Fayyad et al. publishing a worldwide, average prevalence in adults of 3.4 % (1.2 - 7.3 % for respondents aged 18 to 44) in 2007 [29]. Even though the diagnosis of ADHD has become more frequent in adults, it is a challenging diagnosis to make. This is arguably due to the insufficient nature of experimental and empirical evidence to provide the necessary diagnostic insight. Diagnosis for ADHD in adults can only be confirmed after multiple clinical sessions and depends largely on the individual's recollection of whether symptoms were met during their childhood. However, recollection quite often tends to be unreliable [20]. Diagnosis also depends on the individual's ability to self-report any symptoms currently present, which typically cannot be done with a great degree of accuracy. Therefore, accurate diagnosis additionally relies on consulting informants who have observed the individual in various settings [15, 20].

2.1.3 Genetics

Numerous studies have highlighted the presence of a genetic link in patients with ADHD [2, 21, 30]. However, for the purposes of this study, it is not the genetic link which is important but the severity of the symptomatology of ADHD-I. This study is therefore focused on identifying and quantifying the expression of ADHD-I symptoms.

2.2 The Gold Standard

There is currently no single tool used for the diagnosis of ADHD. In all diagnostic evaluations, there are rather four bases to consider [31, 32]. For diagnosis in youth, the first of these is a complete clinical and psychosocial assessment by specialist psychiatrists, paediatricians or trained health care professionals. The assessment evaluates symptoms and behaviour of the patient in the different settings and domains of everyday life. Secondly, it considers a subject's full developmental and psychiatric history. Third, observer reports, such as rating scales completed by parent and teacher, can be used for additional insight to symptom prevalence and severity [33]. Finally, the patient's mental state can be evaluated by making use of neuropsychological tests [34]. Cumulatively, these are tools that aid in the clinical diagnosis of ADHD [32, 35]. The literature refers to the use of them collectively as "the gold standard".

2.2.1 Co-existence

In light of the variability of the worldwide prevalence of ADHD, as well as the factors influencing its diagnosis, the proper diagnostic methodology is crucial in order to follow the correct course of treatment [25]. As discussed in part 2.2 above, a plethora of measures is utilised to diagnose ADHD. This is currently necessary as there are many other disorders that display similar symptoms to ADHD. Comorbid disruptive behaviour disorders in children with ADHD, such as oppositional defiant disorder (ODD) and conduct disorder (CD), have been well established for several

decades.[3] However, more recent research has identified the emergence of further ADHD comorbid disorders alongside these, namely, anxiety disorder (ANX) and depression disorder (DEP) [2, 31, 36, 37]. Therefore, the first step in providing the best treatment would be to accurately identify the most appropriate disorder classification, as it is often the case that the symptoms displayed by a patient are more accurately explained by a criteria of another disorder [32]. This misclassification of expressed symptoms within diagnosis presents a crucial challenge for repeatability and accuracy, often leading to misdiagnosis of patients.

A cross-cultural study conducted by the ADHD Institute in 2006, concluded that the diagnosis of suspected ADHD patients is generally a complex and involved process. One of the main factors influencing the correct diagnosis of a patient is the presence of comorbidities. Comorbidities have commonly been associated with ADHD for all age groups [3], with a high degree of comorbidity, specifically between ADHD and other disorders [2, 37–39].

“Comorbidities” is a common medical term. However, the word “comorbid” proves to be problematic when applied to psychopathology because its value is derived contextually.[4] In other words, the word “comorbid” is used to explain a state of being when dealing with well-validated disease entities, whose pathology and aetiology are understood.[5] It is therefore more appropriate to use the terms “co-existence” or “co-variance” in psychopathology, specifically when dealing with ADHD, since its aetiology is not clearly understood [21, 40]. Generally, it is also appropriate to use in the evaluation of clinical ratings in contrast to in-depth evaluations of psychiatric disorders [41].

Findings from an ADORE cohort study (N = 1 478), conducted by Steinhausen et al. in 2006, highlight the research implications of co-existing disorders. The study included children aged between six to 18 years, with a mean age of nine years (SD 2.5), sourced from 10 different European countries. Data samples were collected during six periods over a span of two years [3]. The findings presented by Steinhausen et al. suggest that co-existence of psychiatric problems with ADHD has serious clinical practice implications with regards to proper treatment. Another study that observed the impact on quality of life of the ADORE ADHD patients, highlighted the negative effects of ADHD on psychosocial development and quality of life in children with ADHD [42]. For the purpose of this research project, it is important to note the implications presented by co-existing disorders on clinical practice, and the need for adequate treatment guidelines [43] and intervention schemes [37, 44]. The ADORE study found that the co-existence of psychiatric

---

[3] Comorbidities – “The extent to which two pathological conditions occur together in a given population.” [122]

[4] Psychopathology – “1. The science concerned with the pathology of the mind and behaviour. 2. The science of mental and behavioural disorders, including psychiatry and abnormal psychology.” [123]

[5] Aetiology – “1. The science and study of the causes of disease and their mode of operation. 2. The science of causes, causality; in common usage, the cause itself” [124].

disorders associated with ADHD was significant. Findings from the study varied considerably between countries, however a grouped distribution highlights important overlapping symptoms, namely ODD (67 %), CD (46 %), anxiety (44 %), co-ordination problems (33 %), depression (32 %), tics (8 %) and Tourette Syndrome (1 %) [27].

Although the degree of overlapping symptoms vary, there is a strong case to be made for differentiating between normal groups and groups diagnosed with ADHD subtypes, as well as normal groups and groups diagnosed with ADHD subtypes with co-existing disorders [27]. A study by Grizenko et al., titled: “Is the Inattentive Subtype of ADHD Different from the Combined/Hyperactive Subtype?”, highlights that a better understanding of the differences between subtypes may help physicians in making a clearer diagnoses, as well as develop a clearer, more adequate treatment plan [21].

Further research and integration of rating scales and neuropsychological assessments could be the key difference-maker when evaluating patients. This suggestion is validated by several studies discussed below.

2.2.2 Classification Systems

There are two main systems of classification for diagnosing neurodevelopmental disorders and, specifically, ADHD. Firstly, there is the latest version of the American Psychiatric Association’s DSM criteria (DSM-V) [20]. Secondly, there is the International Classification of Mental and Behavioural Disorders 10th revision (ICD-10) which also forms part of the diagnostic criteria [45]. This study follows the diagnostic criteria for ADHD per the DSM-V criteria.

The DSM criteria system was selected due to the harmonisation of its classification of disorders with that of the ICD. Furthermore, the DSM criteria was able to accurately identify a broader group of children with the disorder when comparing ADHD (per the DSM-IV) with its ICD-10 equivalent (hyperkinetic disorder) for the same neurodevelopmental disorder group [46]. The DSM criteria system is globally accepted and has been used widely in research studies.

The DSM-V diagnostic system states that diagnosis should be based on a patients’ exhibition of a persistent pattern of negative symptoms relating to inattention and/or hyperactivity-impulsivity, which interferes with daily functioning and development. These symptoms are required to be present in at least two settings – for example, at home, school or work – before the age of 12 and for an uninterrupted period of at least six months [20, 34].

According to the DSM-V diagnostic criteria, it is possible to diagnose a child when at least six of the possible 18 criteria are met from either the inattention or the hyperactivity-impulsivity group. Older adolescents and adults (age 17 and older)

must meet at least five group criteria to be diagnosed with ADHD.[6] However, it is also possible for an individual to be classified as being in "partial remission" if there is a decrease in the number of diagnostic criteria met over an uninterrupted period of six months. Finally, the DSM-V system requires that an individual's current state of symptom severity be specified as either mild, moderate or severe [20, 26, 31].[7]

2.2.3 Rating Scales

The rating scales discussed in this section were selected based on their purpose, age group application, content, standardisation strength and psychometric properties, as well as their evidence for reliability, clinical utility and validity. Frequency and range of use in clinical practice were also considered. [47–49]

Vanderbilt ADHD Diagnostic Rating Scale (VADRS)

Based on the DSM-V criteria, the VADRS includes specific parent and teacher rating scales. The teacher rating scales were first introduced in 1998 and were followed by the introduction of the parent rating scales in 2003.

The effectiveness of the VADRS depends largely on the feedback accuracy and interpretation of parents in completing the Vanderbilt ADHD Diagnostic Parent Rating Scale (VADPRS), as well as that given by teachers in the Vanderbilt ADHD Diagnostic Teacher Rating Scale (VADTRS). These scales contain 55 and 43 questions for parents and teachers, respectively. The VADPRS was designed to evaluate symptoms expressed in the home setting, whereas the VADTRS evaluates symptoms expressed at school. The VADPRS includes all 18 symptoms for ADHD as specified by the DSM criteria. Furthermore, the scale expands on the word "often", as used by the DSM criteria, and employs a 4-point rating scale to capture the frequency of each symptom (0 = never, 1 = occasionally, 2 = often, 3 = very often). [50, 51]

The VADRS has specifically been designed to discriminate between children with and without ADHD, as well as ADHD's respective subtypes and possible co-existing disorders for youth aged six to 12 years. The clinical utility of these scales has repeatedly been validated in literature, and has been found to be reliable and well validated with normative data across sex and age. [49, 50]

A study published in 2016 by Silverstein et al. [52] sought to determine whether clinical data, used as a supplement to parent rating scale reports of ADHD symptoms, could be useful in predicting ADHD diagnosis according to the DSM-IV criteria. The study included 156 children between six and 12 years of age from urban regions. As stated above, it is important to note that the DSM criteria for ADHD diagnosis requires that symptoms present in at least two settings, for example, in the home and at school [20]. This means that both the parent and teacher rating scale reports form a crucial part in ADHD diagnosis. However, it is

[6] Adolescents (age 12 to 17); Adults (age 18 and older).

[7] Explanation of severity states found in the DSM-V criteria.

often the case that clinicians proceed to diagnose patients without the inclusion of the teacher rating scale report [53]. The absence of these teacher rating scales can lead to a significantly higher rate of misdiagnosis, which was identified by Silverstein et al. as problematic. Results from predictive models created by Silverstein et al. indicated that the models could correctly predict a positive ADHD diagnosis 56 % of the time based solely on a positive Vanderbilt parent rating scale report. The maximum predictive capability of these models was 84 %, which was achieved by incrementally adding fields of clinical data. As fields of clinical data were added, the predictive ability continued to increase: child's age (68 %), grade retention (78 %), anxiety and depression symptoms (81 %), ODD symptoms (83 %), a parent with a history of substance abuse (84 %). Here it is important to note the impact of predictive diagnostic accuracy without the addition of teacher rating scale reports but more so, the impact of including specific fields of clinical data.

SNAP Questionnaire

The Swanson, Nolan and Pelham (SNAP-IV) questionnaire is a rating scale derived verbatim from the DSM symptom list for ADHD. As the first of several questionnaires to incorporate the DSM symptoms for ADHD in a rating scale format, for use by both parents and teachers, the SNAP questionnaire has been demonstrated to discriminate effectively between children with and without ADHD. The original SNAP-III questionnaire was developed for use with the DSM-III criteria for ADHD. Since the initial conceptualisation, the rating scale has seen updated revisions with each DSM revision release. [54]

The SNAP-IV rating scale quantifies the presentation of 90 items on a four-point scale by making use of a frequency scoring system, ranging between 0 and 3 (0 = not at all, 1 = just a little, 2 = quite a bit, 3 = very much). After the calculation of each item's frequency, the symptom-severity dimension (ADHD subscale scores) are calculated for each ADHD subtype by adding the specific set of frequency score items related to that subtype. Finally, the ADHD subscale scores are compared to population norms in order to assist in diagnostic classification. [51]

All 90 impairment items are manifestations of either inattention, hyperactivity and impulsivity, or ODD. Additionally, 10 of the 90 items evaluated by the scale evaluate the presence and severity of impairment in the classroom setting. Although the SNAP questionnaire is not designed to formally diagnose co-existing ADHD disorders, the scale makes basic provisions for multiple disorders in a few of the questionnaire items. Specific co-existing disorders are investigated when an item frequency score of either two or three is recorded. [49]

SWAN Rating Scale

A study conducted by Brites et al., which analysed 61 articles concerning the development and application of the SWAN rating scale, stated that many rating scales are "too categorical". The study further stated that rating scales often only report on the presence or absence of a specific problem. This approach neglects the variance in cultural tolerance and evaluation of disruptive and socially unacceptable behaviour. ADHD patients with mild symptoms could therefore pass unnoticed and be undetected in an initial clinical assessment [55–58].

Consequently, data from these rating scales may not be ecologically valid. Furthermore, Brites et al. highlighted the importance of utilising dimensional profile scales for ADHD. Dimensional discrimination is a form of behavioural analysis that evaluates behavioural disorders while minimising social, cultural and statistical biases [58].

Overcoming bias present in previous rating scales, Swanson et al. [58] demonstrate the shortcomings of previous methods, which resulted from the application of reduced summary scores that assume the behaviour patterns of normal population groups. The implementation of assumption-based statistical cut-offs result in highly skewed outcomes [51].

A new rating scale - the SWAN rating scale - has subsequently been conceptualised and developed. This scale was modelled on the SNAP-IV rating scale, developed in collaboration with Swanson to overcome the shortfalls of the categorical/physiological SNAP model [51]. The SWAN scale reflects the distribution of attention skills as well as the severity of existing behavioural symptoms within a population. In comparison with the SNAP-IV [59] and other rating scales [60, 61], the SWAN can be used for ADHD evaluations with a reduced risk of bias. Furthermore, the SWAN scale gave a more accurate distribution profile of behaviour scores by making use of a grading system.

An additional feature of the SWAN is that the participants' behaviour scores are required to be compared with the average cultural age and behaviour expected thereof. The SWAN has reworked items on the SNAP scale to overcome outcome skewness and correct the tendency to over-identify extreme cases [51]. The scale consequently moves away from the pathological signs and symptoms of ADHD and addresses 30 measurable, behavioural items. These indices include focused attention, impulsive behaviour inhibition during prolonged mental effort tasks and daily activities, as well as anxiety control, to name a few. The grading system for each item is scored from minus three (below average) to plus three (above average), with zero being normal and based upon the population average [58, 59]. In conclusion, when considering the use of rating scales in diagnosis, it is important to evaluate the ecological value they add [58].

2.2.4 Neuropsychological Assessments

This section presents the case for the validity of neuropsychological testing techniques. It presents an overview of the establishment of the method, followed by a discussion and presentation of case studies regarding the validity of existing tests. The overview at the end of this section highlights the significance of neuropsychological assessments for the purposes of this research.

Background

According to Hartlage and Long [62], the field of neuropsychology in the 1940s was predominantly concerned with brain dysfunction. It was Ward Halstead who made one of the most influential contributions to the field. His contribution was published in the book, *Brain and Intelligence*, in 1947*,* presenting an approach for measuring biological bases for intellective functions.

Halstead's student, Ralph Reitan, then went on to develop Halstead's evaluations, later adding contributions of his own. Reitan's work went on to firmly establish and validate a comprehensive, yet sensitive standardised scientific and experimental test battery for use in neuropsychological assessments of brain dysfunctions [62]. It is this assessment that is known today as the Halstead-Reitan Neuropsychological Test Battery (HRB). More research has been conducted using this battery than any other single neuropsychological battery. [63]

Following the establishment of the HRB, Reitan conducted numerous studies relating to adults with verifiable brain injuries. He concluded from the findings of these studies that the HRB is not only able to identify but also able to differentiate between brain dysfunctions resulting from a range of aetiologies. [62]

Validity of Tests

During the 1970s and 1980s, the field of neuropsychology slowly became an established area of speciality in America, both scientifically and professionally. An increasing focus on the realm of adult brain injury within this field paved the way for similar research to be conducted on children. [62]

It was clear from the growing body of research in the field that specific behavioural and learning problems were related to known brain damage in adults. The expansion of the neuropsychological community to children's problems stemmed from a growing interest in the aetiology of central processing dysfunctions in children. Questions which consequently arose concerned the application of adult findings to children, the most appropriate diagnostic approach for children, as well as the suitability of selected tests when evaluating juvenile patients. The most obvious issue presented was the misclassification of children as "brain injured". Due to the lack of an existing external criterion by which to validate neuropsychological examinations, it was easy to incorrectly classify a child presenting a central nervous system dysfunction. [62]

Modified versions of the original HRB were developed in response to these questions. For children aged five to eight and nine to 14 years, the Reitan-Indiana Neuropsychological Test Battery (RINB) and Children's Halstead-Reitan Neuropsychological Test Battery (CHRNB) were established, respectively [64–66]. Additionally, the mid-1970s welcomed the Laria-Nebraska Neuropsychological Test Battery (LNNB), specifically developed for children eight to 12 years of age [64, 67].

Although many of the earlier studies discussed above used neuropsychological tests to discriminate between patients with problems related to brain damage, the usefulness is not limited to this application alone. Philip [68] states that many neuropsychiatric conditions are complex in nature with the potential to bring about changes in mood or motivational states. He further states that these changes result in secondary impacts on cognitive functioning that are just as real as those caused by brain injury.

Lovejoy et al. [69] state that the role of neuropsychologists within the field of independent neuropsychological evaluation is becoming increasingly valuable. Much of this value is attributed to the utility of neuropsychological techniques,

which have the ability to highlight functional impairments in psychological, psychiatric and cognitive domains. Furthermore, these techniques also offer a degree of discrimination between disabilities [70].

Neuropsychological assessments are said to be useful for various assessments. These include, but are not limited to, assessments purposed for the prediction of potential function, diagnosis, differential diagnosis or measurement of treatment response. The usefulness of neuropsychological assessments is highlighted in its ability to evaluate specific cognitive domain sets, which makes it possible to correlate a suspected condition in a patient with cognitive domain deficits known to exist for that specific condition [68].

Neuropsychological tests are frequently used after evaluating a subject per the DSM-V diagnostic criteria in order to quantify the impact of ADHD on the individual's cognitive functioning. These tests measure specific psychological functions, including intelligence, memory, language and executive functioning that are known to be linked with a particular brain structure. Currently, these tests are only supplementary and remain independent due to the fact that they are not yet able to diagnose ADHD reliably. They do, however, provide valuable insight for cognitive functioning and have been found to be useful in the diagnosis of learning disorder (LD), as well as defining strengths and weaknesses within the LD population. [66], [71]

Currently three subtypes of ADHD exist (see section 2.1.1 above). However, according to Jensen et al., sufficient data has been gathered to warrant delineation of ADHD into two further sub-classifications: (a) ADHD aggressive subtype and (b) ADHD anxious subtype [72]. Subtype (a) includes aggression and CD and has been validated by findings from neuropsychological studies [27]. For the purposes of this study, the official ADHD subtype definitions will be used, as stated by DSM-V classification criteria [20].

A study conducted by Sharp et al. found that neuropsychological testing of children revealed a style of impulsive and inaccurate responses in subjects with the 7R allele.[8] These genetic factors are not explained by the DSM-V ADHD criteria. Furthermore, task results revealed that children with the 7R allele had significantly more incorrect responses, coupled with shorter average reaction times, than those children without the allele but with an ADHD diagnosis. Children with the 7R allele also displayed higher levels of activity when compared to ADHD children without the allele. It is important to note that the number of ADHD symptoms presented by both these groups of children did not differ significantly. Moreover, and importantly for purposes of this thesis, results revealed that both groups of children were more neuropsychologically impaired than the comparison normal control group. [30]

More specific to this study is one of the most commonly used neuropsychological assessments in the diagnosis of ADHD: the continuous performance test (CPT). Designed to measure impulsivity, sustained attention and selective attention, CPTs

[8] Allele – "One member of a pair or series of genes that occupies a specific position on a specific chromosome" [125].

are considered aiding tools in the ADHD diagnostic procedure. However, many researchers still raise questions relating to its limited sensitivity, specificity and ecological validity.

CPTs usually involve the rapid presentation of visual or auditory stimuli (typically numbers, letters or figures) in the centre of a screen (usually a computer screen) for a predetermined period to induce sustained attention. The purpose of this rapid presentation of stimuli is to measure impulsive behaviour, as well as any lack of focus. Selective attention is measured by the participants' ability to focus on the relevant task or activity at hand whilst ignoring extraneous stimuli, often included in the form of visual or auditory distractors. A single press of a response button is required as soon as a target stimulus is presented. No press should occur when a non-target stimulus is presented. Each target/non-target stimulus is presented for a predetermined period, followed by a "void" period before the next stimulus is presented. [73–75] The absence of a response to a presented "target" stimulus is considered an "omission error", typically considered a measure of attention, and a response to a "non-target" stimulus is considered a "commission error", typically considered a measure of impulsivity. Additional measures often included in CPT's are the number of correct responses, response time (RT) and response time variability (RTV). [35] For the purposes of this study, the discriminating value of CPTs as discussed in section 2.4 will be incorporated.

2.3 Machine Learning

The utility of incorporating machine learning to discriminate autonomously between population groups has commonly become recognised in the last century. Studies conducted by D'souza et al. and Li et al. have also implemented machine learning techniques to identify differences in human movement. These studies have with great success been able to recognise specific human activity using a general population. [76–78] A study conducted by Silverstein et al. investigated the predictive capabilities of machine learning techniques to accurately predict ADHD diagnosis among urban children. It is interesting, but not essential to this study that results from the study by Silverstein et al. found a correct positive prediction ability of 84 %. [52]

Given any specific problem, multiple machine learning techniques exist to provide a solution. Certain techniques will perform better than others on certain data sets, but it is often the case that the converse is also true when presented with an alternative data set. The onus therefore rests on the accurate identification of the problem, followed by an investigation of the available techniques suited for problems of that nature. Machine learning techniques are commonly categorised into one of several task categories, namely supervised learning, semi-supervised learning, active learning, unsupervised learning or reinforcement learning. These categories can also be described from an application perspective of the machine learning technique. Application categories are commonly known as classification, regression, clustering, density estimation and dimensionality reduction. The inter-link between task and application is found in the problem the machine learning technique is addressing. [79]

2.4 Existing Technology Review

This section discusses both diagnostic assistive tools, as well as cognitive training tools. The aim in discussing these methods is to highlight their significance in identifying ADHD participants, as well as to draw from their strengths and weaknesses later in the chapter.

2.4.1 MOXO

MOXO is a standardised, computerised, internet-based CPT, designed to aid in the diagnosis of ADHD symptoms [80]. Created by Neurotech Solutions Ltd., MOXO was developed to quantify four performance indices, namely attention, hyperactivity, impulsivity and timing. The innovation is focused on providing accurate measurement of responses as this forms a crucial part of the CPT system. [73, 80, 81]

Two versions of the MOXO system have been developed. One version targets youth (aged six to 12 years), and another is aimed at adolescents and adults (aged 13 to 70 years), referred to as Groups A and B, respectively. Both versions present participants with continuous stimuli in the form of target/non-target stimuli, as seen in Figure 5, with the addition of visual and auditory environmental distractors for certain levels, as seen in Figure 6. Testing takes an average of 15.2 minutes for Group A (53 trials per level) and 18.2 minutes for Group B (59 trials per level). [73] This research project will focus on Group A and the related design specification for that MOXO version.

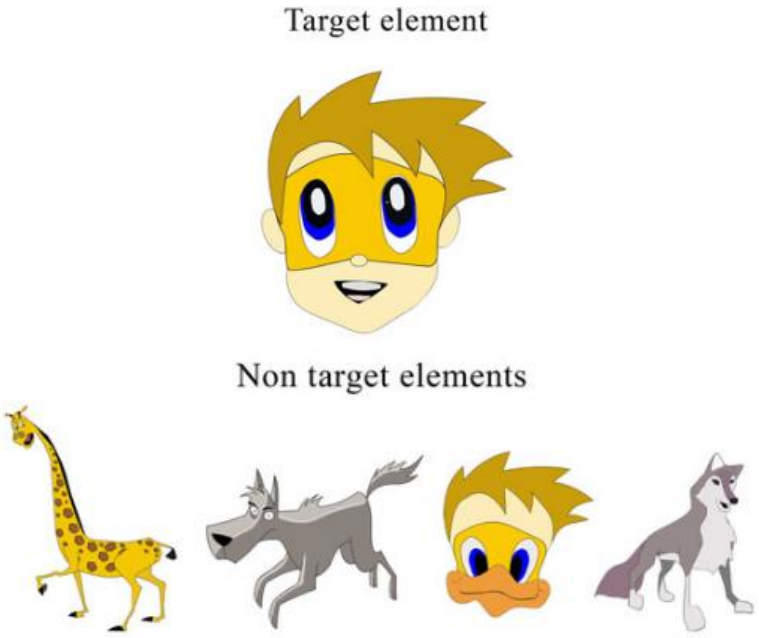


Figure 5: MOXO target and non-target stimuli [35].

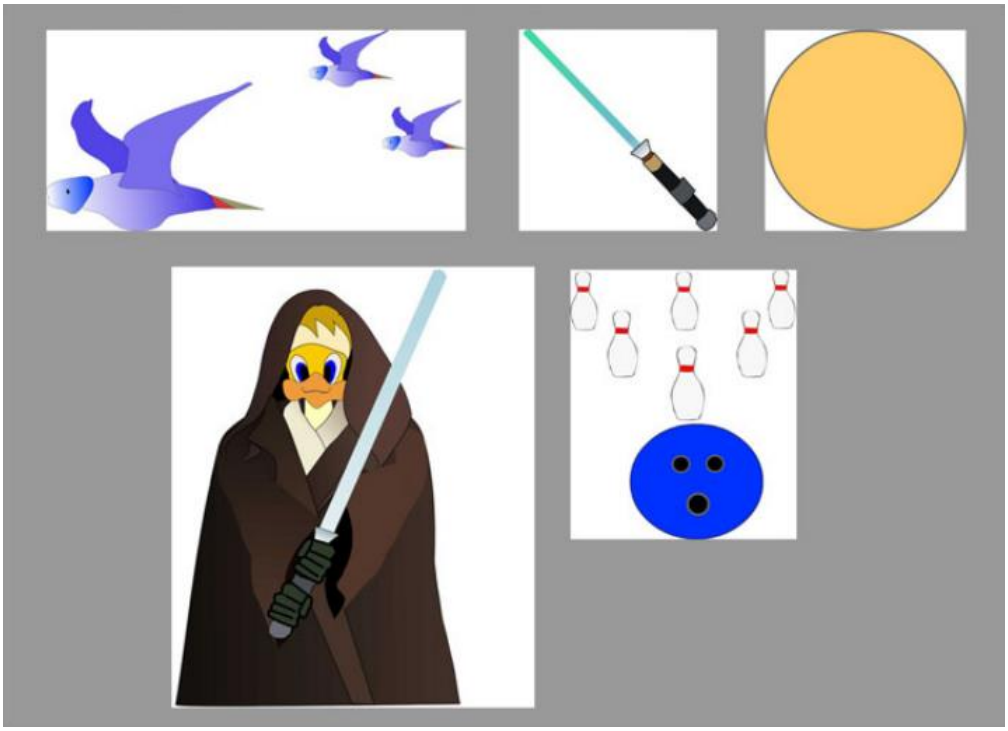


Figure 6: MOXO visual distractors set [35].

Berger et al., serving on the scientific advisory board of Neurotech Solutions Ltd., conducted a study in 2014 to investigate the ability of CPTs to distinguish between ADHD and non-ADHD control participants using MOXO. The study included 176 adolescents (aged 13 to 18 years) and indicates statistical significance in omission errors to distinguish between the two groups. Each participant met the DSM-IV-TR criteria.[9] According to Berger et al., the findings emphasised the importance of incorporating distractors and integrating a set of attention parameters when measuring attention indices with CPTs. Additionally, visual distractors and a combination of visual and auditory distractors were found to more accurately distinguish between groups than auditory distractors alone. Although the author states that the addition of distractors improved the sensitivity and specificity, no percentage values were given for these parameters. However, a test efficiency score was given as the AUC = 0.890 for the addition of distractors. Data analysis was conducted using SAS software. [73]

Another study published by Berger et al. in 2017 further strengthens these findings. This later study investigated the usefulness and validity of CPTs, specifically MOXO, in the diagnosis of ADHD in children [35]. These findings indicate MOXO's ability to distinguish between children with ADHD and children without, based on the four performance metrices (attention, timing, impulsivity and hyperactivity), and revealed that ADHD participants consistently performed worse than their control peers. As was the case in the results of the study conducted in 2014, visual

[9] DSM-IV-TR is a text revision of the DSM-IV criteria and was published in July 2000, preceding the DSM-V which was published in May 2013 [126].

distractors used with a combination of visual and auditory distractors were found to be more significant at distinguishing between groups than auditory distractors alone [73].

Significantly, the 2017 study made use of cut-off values to attain the optimal sensitivity and specificity values. The averaged sensitivity and specificity for the age group ranging between seven and 12 was calculated to be 86.5 % and 86.2 %, respectively. Cut-off values were based on the risk-benefit ratio to achieve the lowest false-positive and false-negative classifications. Berger et al. state that important information may be lost when sensitivity and specificity are defined by selecting a single cut-off value from a continuous variable. The study also did not indicate the repeatability of MOXO, and parameters for the deviation of accuracy between population test sets were also not investigated in this study. Subtypes of ADHD were not specified for this study (only a “general” ADHD population was used), which must be kept in mind when addressing the research question of this thesis. The study included 798 (493 boys: 305 girls) participants, aged between seven and 12 years. Of the 798 participants, 339 were diagnosed with ADHD and 459 participants formed part of the non-ADHD control group.

The MOXO interface is accessible through an internet-connected computer and provides a performance graph to visualise a subject’s performance throughout the duration of the tests. As seen in Figure 7, participants interact with the MOXO interface by making use of the spacebar on a standard computer keyboard, coupled with an internet connected computer, and speakers to relay auditory distractors. All information is secured according to USA-HIPPA confidentiality laws and regulations.[10] [82]

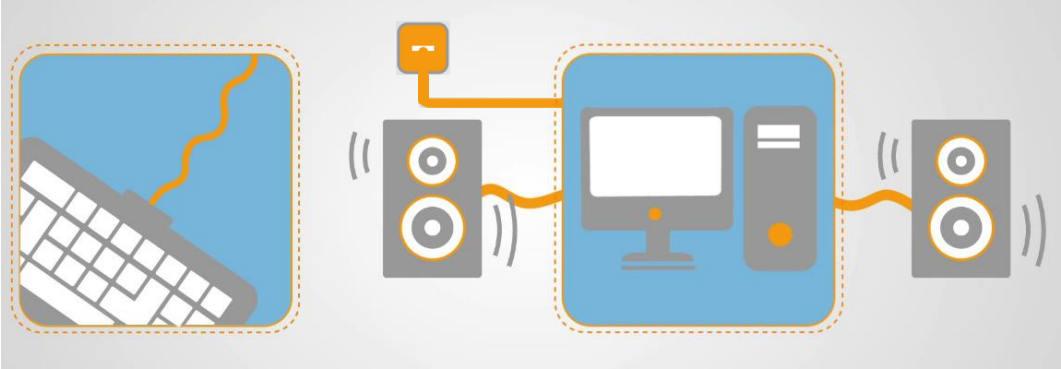


Figure 7: Required equipment to run the MOXO system [83].

2.4.2 T.O.V.A

The Test of Variables of Attention (T.O.V.A) is a computerised measure of attention and inhibitory control. Developed by The TOVA Company to aid in the diagnosis of ADHD in children and adults, the T.O.V.A is said to be an objective, accurate and FDA cleared CPT. Furthermore, The TOVA Company states that the T.O.V.A

[10] USA-HIPPA confidentiality laws and regulations [127]

provides a culture and language free interface with the task developing a sufficient level of fatigue and boredom. This forces participants to pay close attention in order to make the correct responses. [84]

Like MOXO, the T.O.V.A makes use of a computer system, as seen in Figure 8, to present both target and non-target stimuli, but participants are required to deliver responses to stimuli by pressing an external button. Task duration varies for different age groups: 10.9 minutes (four to five years of age) and 21.9 minutes (six years and older), with stimuli presented for a predetermined period, followed by a predetermined interval period. Figure 9 shows the target stimuli are presented in the form of a small monochromatic square with a hole near the top of the square and non-target stimuli have the hole presented near the bottom of the square. The T.O.V.A does not incorporate any distractors. [84, 85]

Figure 8: The T.O.V.A system with external hand-held button [84].

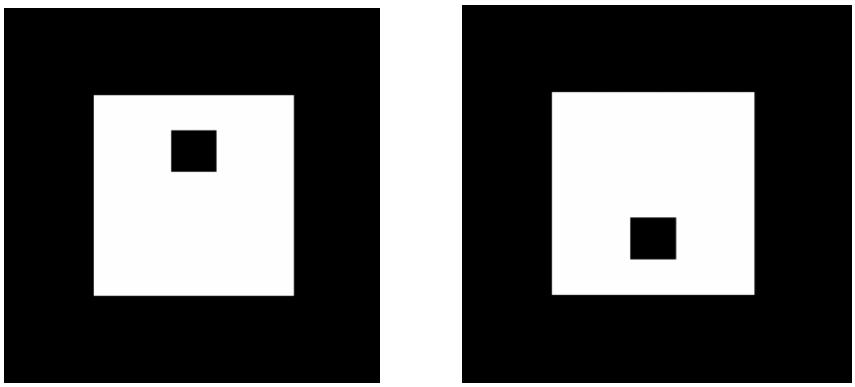


Figure 9: T.O.V.A target (left) and non-target (right) stimuli [86].

The TOVA Company specifically opted for the use of a hand-held button (micro-switch) for user response. The company points out that this solution significantly affects test reliability with an insignificant response error of +/-1 millisecond compared to the measurement error of +/- 28 milliseconds of CPTs that make use of a computer keyboard for their response mechanism. [84]

A study conducted by Anguera et al. in 2017 administered the T.O.V.A to assess the sustained attention and impulsivity of their participants. Findings from this

assessment indicated an estimated sensitivity of 85 % for the use of the T.O.V.A as a predictor of ADHD. No specificity percentage was given. [87] A study published by Reddy et al. in 2010 provides a test-retest reliability percentage for omission, commission, RT and RTV ranging between 51 % and 82 %. Furthermore, the study indicates the validity of the T.O.V.A by highlighting the sensitivity and specificity determined by two different test samples: 1 - sensitivity (0.61 to 0.76), specificity (0.80 to 0.90); 2 - sensitivity (0.61 to 0.73), specificity (0.73 to 0.94). [88]

Variables measured by the T.O.V.A include mean RT, omission errors, commission errors, RTV, number of multiple responses to a single stimulus and anticipatory responses. An ADHD score is graphically reported at the end of the test which is compared to the participants age and gender specific group. [84, 85]

2.4.3 AULA Nesplora and Connors' CPT

The AULA Nesplora (AULA) is a virtual reality (VR) based continuous performance test used to evaluate attention processes in children and aid in the diagnostic assessment of ADHD. The AULA, meaning "Classroom" in Spanish, is based on the CPT methodology which attempts to provide an ecologically valid testing environment by making use of a VR test to immerse participants in a classroom setting. The AULA system is operated using a computer, a VR headset with motion sensors, headphones and an external response button (see Figure 10). [89]

Testing scenarios present "target" and "non-target" stimuli in the setting of a three-dimensional classroom, as seen in Figure 11, and offer the participant the perspective of being seated at a classroom desk, facing a blackboard.

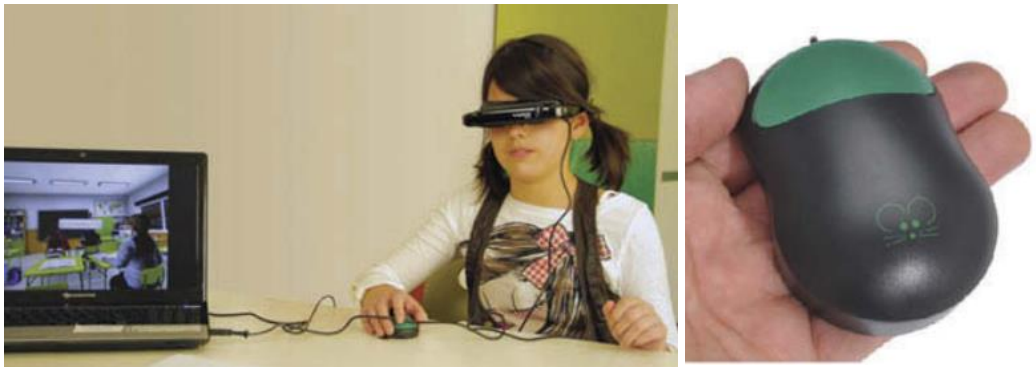


Figure 10: An illustration of the AULA system setup without the headphones [4].

Figure 11: A screenshot of the AULA test VR projection [4].

The Connor's CPT is a commonly used computer-based test designed to asses attention problems. [90] Research conducted by Díaz-Orueta et al. in 2014 finds that the AULA and the Connors' both present visual stimuli for the same duration (250 milliseconds). However, the AULA's addition of auditory stimuli in the form of words varies auditory stimuli presentation time. Auditory stimuli presentation time varies further depending on the language of test administration. Therefore, this language limitation also results in a change of the inter-stimulus interval. Both of which requires redevelopment of the platform to accommodate the language of application. The AULA test is comprised of two 10-minute testing sessions. One testing session contains distractions (auditory, visual, and a combination of both) and the other contains none. According to Díaz-Orueta et al., distractions are randomised and ecological in nature. The study found significant differences between the abilities of the AULA test and the Connor's CPT to differentiate between ADHD children with and without pharmacological treatment. [4] This can be likened to the ability to differentiate between neurotypical children and children with ADHD.

Due to the contrasting natures of the two tests, comparative measures are limited. In contrast to the AULA, Connors' CPT lacks distractions and contains only "target" and "non-target" visual stimuli. Furthermore, the AULA varies the presentation time of each stimuli, which is conversely fixed for the Connors' CPT. For the AULA, the variation of the length of auditory stimuli is influenced further given the language of application. However, the study found significant differences in measures relating to inattention, impulsivity, motor speed, processing speed, and quality of attention focus. For the shared non-target stimuli paradigm of the AULA and the Connors', a significant correlation ($p < 0.01$) was found between auditory and visual stimuli. Additionally, convergent validity between the two tests were still significant ($p < 0.5$) when changing the AULA test paradigm to the target stimuli task. [4]

A study published by Rodríguez et al. 2018 compared the ability of both the T.O.V.A and the AULA to identify ADHD in children aged between six and 16 years

of age. The study included 338 participants, split between CPTs: T.O.V.A (N = 172) and AULA (N = 166). Participants included all three subtypes of ADHD, with a non-ADHD control group of 101 participants, split between CPTs: T.O.V.A (N = 59) and AULA (N = 42). Results from the study indicated greater ADHD prediction accuracy by AULA than the T.O.V.A. The AULA was also able to better distinguish between participants with ADHD and participants without. These results are achieved by comparing the attention variables (omissions, commissions, RT and RTV) provided by both CPTs. The model based on the variables of the AULA was able to correctly classify 56.6 % of the four groups within the sample, whereas the model based on the T.O.V.A variables was only able to correctly classify 33.7 %. Both models highlighted the significance of omissions, commissions and RT. The strength of the AULA model may in part be attributed to the virtual classroom but that too has its limitations. [89] It should also be noted that the AULA also makes use of motion sensors to correlate with its attention parameters.

2.4.4 CogCubed

CogCubed Inc. is a Minneapolis-based start-up that develops games aimed at assisting in the remediation of cognitive disorders, including ADHD. CogCubed is a derivative of the Sifteo Cubes system developed at MIT in 2011. Sifteo Cubes is a gaming platform made up of small cubes with screens. The cube system wirelessly communicates between its individual cubes as the participant interacts with them. Moving, stacking or tapping any cube in relation to another is the core mechanism of gameplay. In response to user input, individual cube screens may in response change the images presented for the game to progress. Due to the nature of the interaction, the system is known as a “tangible user interface” (TUI). [91, 92]

Using the Sifteo platform, Cogcubed developed a game called Groundskeeper, seen in Figure 12 below. The game system employs elements that exercise the skills affected by ADHD. Groundskeeper gameplay closely resembles that of the old arcade game, Whack-a-Mole that has been translated into a TUI [91]. Gameplay consists of four-cubes and 17 game sessions, each with different types and levels of distractions. Each session lasts 90 seconds and requires the participant to use one cube, with a mallet on the screen, to hit only a gopher when it is presented on any of the remaining three cube screens. Each session is followed by a 20-second interval. Successfully “hitting” the gopher means tapping the mallet cube against the gopher cube and receiving an auditory “bonk” noise as feedback.

Figure 12: An illustration of CogCubed gameplay [93].

As the game progresses, visual, auditory and spatial distractors are added. The visual distractors take the place of the gopher and are not to be hit. Distractor levels are either no-distractor, low-distractor or high-distractor, with a low visual distractor taking the form of a bird on one of the three remaining cube screens. A high visual distractor presents a large rabbit. Visual distractors appear on each screen at random intervals of either 1 000, 1 500 or 3 000 milliseconds. A low auditory distraction presents the occasional tweeting noise and a high auditory distraction presents more frequent tweeting. For no spatial distractors, the three interacting cubes are placed in a vertical line. For a low spatial distraction, the three interacting cubes are set diagonally, two inches apart, and three inches apart for a high spatial distractor.[11] Session zero, one and 16 have no distractors, with session zero serving as a practice session. Session one is used to measure a participant's initial ability and session 16 serves as a control comparison to measure learning and endurance. The cubes communicate wirelessly with one another but in order for data to be collected, they need to be synced with a computer that runs the game. [94]

According to Ampel, the cubes are capable of producing data on variables like fidgeting [92]. Although the game has been designed as a cognitive training tool to help players improve their cognitive skills, another version is being developed as a diagnostic tool. The cubes produce objective data for symptoms presented by means of built-in sensors. Upon completion of the game, captured data is transformed into ADHD features and processed using machine learning

[11] 1 inch equates to 2.54 centimeters.

algorithms. These algorithms are then used to develop diagnostic models for the assessment of ADHD.

A study was conducted in 2013 by Heller et al. including a total of 52 American participants aged six to 17 years, with ADHD (N = 26) and without (N = 26). Of the data samples collected, 33 variables were identified and sampled at a frequency of 10 Hz. Evaluation of these sample results after Groundskeeper gameplay indicated a high level of accuracy according to the F-measure, representing the percentage of correct diagnostic prediction for ADHD inattention subtype (78 %), ADHD-C (75 %), ANX (71 %) and DEP (76 %). More specific to this study was the sensitivity (77.8 %) and specificity (80 %) of the ADHD-I model. [94][12]

2.4.5 Akili

Founded by and subsidiary to PureTech Health, Akili follows a proprietary neuroscience approach to assess and adaptively target improvements in cognitive control for groups with cognitive disorders and executive function deficits. Akili's flagship product, Project EVO™ ADHD as seen in Figure 13, was developed from the principles used in NeuroRacer, a previous cognitive intervention. [95] EVO is specifically developed as an iOS compatible application which aims to improve cognitive and disease symptoms through an at-home tablet-based digital interface. Furthermore, the project aims to use its approach of proprietary interference processing therapy to improve the symptoms of inattention, working memory and executive function. Results of this approach are pending the STARS-ADHD study evaluation. [87, 96, 97] Additionally, the company is also in the process of developing several other products for screening and treatment of cognitive deficiencies such as Autism, Depression, Alzheimer's disease, and traumatic brain injury [96].

Project EVO assessment is a tablet-based cognitive training tool that makes use of a wireless internet connection to transfer gameplay data. The assessment is divided into three tasks: a perceptual discriminatory task (essentially a CPT), a visuomotor tracking task, and finally a multitasking task, which tests the two tasks simultaneously. [87]

[12] F-measure is the measure diagnosis accuracy in this case [94]. However, concern has been raised for the measure exhibiting bias [112, 128].

Figure 13: EVO game launch-screen [98].

By making use of adaptive psychometric staircase algorithms to adapt the level of difficulty to each participant, EVO is able to avoid testing-based disparities and compare performance differences. This allows EVO to determine an individualised level of performance for each participant by dynamically altering the level of difficulty, trial-by-trial, to maintain a participants' performance accuracy at 80 %, combatting age or instrumentation biases. [87]

EVO was designed in line with the findings of Anguera et al. [95] These findings demonstrated that custom-designed games, such as NeuroRacer, which incorporates attention and goal management with the addition of interference, serve as a powerful tool for cognitive remediation [87]. A further study using the EVO intervention to test the use and effectiveness of mobile apps for depression found that EVO had beneficial effects on the cognitive control abilities of adults 18 years and older [99].

The multitasking required by the EVO intervention limits a participant from succeeding if attention is focused on a single task only. To perform adequately, participants must carefully distribute their attention across the multiple tasks as the game continues, in order to push the overlap of task boundaries. This method is employed to teach the brain to properly prioritise sensory input and strengthen neural pathways. EVO is a runner-based game with a character speeding down a river. Different stimuli are presented as the character progresses and responses to these stimuli are required.

According to Anguera et al. participants of the EVO training intervention are exposed to different visual "worlds" as their performance improves during training

runs. These different worlds are intended to immerse a participant in a different EVO universe that requires a greater level of engagement and compliance, demanding +/- 80 % accuracy of participants. During these runs, participants are given visual and auditory feedback indicative of their performance. Results of the first Project EVO pilot study experiment conducted with 62 children suffering from SPD was released in April 2017. The aim of the experiment was to characterise attentional abilities. It is important to note that the SPD patients were divided into two subgroups exceeding the cut score for inattention or hyperactivity (SPD_{+IA}). SPD_{+IA} participants exhibited the same performance impairment and response variability in selective and sustained attention, as well as in goal management when compared to ADHD patients. However, when compared to the typically developing control, results of the study showed significantly more differences between SPD_{+IA} and the typically developing control than SPD and typically developing control in features such as RT and RTV. [87]

Following the first Project EVO pilot study experiment, a second experiment was conducted to determine whether selective attention, sustained attention, and goal management could be trained within the respective groups. Only 57 participants were selected from the first experiment and subjected to attention-based training. Participants were required to complete seven sessions of 30-minutes each per day, five days a week for a four-week period. According to Anguera et al., results showed statistically significant improvements in real world function using the Vanderbilt Assessment post training. Improvements in inattention behaviours also remained stable for the SPD_{+IA} nine months post-training. Furthermore, it was reported after training that 33 % of the SPD_{+IA} participants who initially met the Vanderbilt criteria for inattention, no longer did. Anguera et al. additionally highlights a significant difference between the RT and RTV when comparing the T.O.V.A to the EVO. [87]

2.4.6 CogoLand and ATENTIVmynd™

CogoLand is a computerised 3D graphic game product, developed by Atentiv LLC. The overall intervention employs FFM by interfacing a Bluetooth enabled EEG headband with the computer-based game. The system serves as a nonpharmacological intervention that focuses on the treatment of children with ADHD (eight to 12 years of age), as well as cognitive skills training for the improvement of academic performance [100]. As seen in Figure 14, the CogoLand test setup includes a computer, an EEG headband including three frontal sensors, as well as speakers.

Figure 14: Participant engaged in the CogoLand training game [101].

The approach used by the CogoLand team was to develop and use an individualised EEG profile for attention for each individual participant. This calibration profile forms the baseline by which the game is played and according to which the participant learns. However, calibration is a lengthy process as the calibrating game mechanism employs filter banks in the form of spatial pattern filtering to capture a range of EEG rhythms for each participant prior to testing. These rhythms are used to determine individual-specific patterns to discriminate between attentive and inattentive states. [101]

The CogoLand gameplay mechanism utilises the EEG headband and the participants' attentive and inattentive states to control the movement speed of the game avatar. These states form part of a range for strength of concentration. The EEG readings are transformed into a concentration score ranging from zero (minimum attention) to 100 (maximum attention) according to the participants' calibration profile. A score of zero results in the slowest movement of the avatar, and a score of 100 the fastest. [100, 101]

Findings from a study conducted by Lim, et. al. included statistically significant improvements of inattentive symptoms in the ADHD-I population, based on the behavioural rating scales of parents. The study included a combination of 20 ADHD-C and ADHD-I participants between six and 12 years of age, of which results for 14 participants were complete. Each participant underwent a training schedule including three sessions per week over a period of eight weeks. Each training session consisted of 30 minutes, including breaks. Following alternate training sessions, participants were required to complete worksheets serving as performance measures for comparison of sessions. These worksheets took

approximately 10 minutes to complete. Patients were tested again at week 20. The primary outcome measure was the difference in ADHD Rating Scale score between the baseline score (week zero) and weeks eight and 20, respectively. Additionally, the EEG data from week eight and 20 were also compared but showed no statistical significance. [101]

As seen in Figure 15, ATENTIVmynd employs the same EEG interface but replaces the computer-based game component with a tablet-based game component instead. Published details of the tablet-based EEG intervention are highly restricted but according to Qian et al. results indicate a significant reduction in ADHD symptoms as reported by the ADHD Rating Scale scores when applying the intervention to ADHD participants for a course of eight weeks. [102]

Figure 15: ATENTIVmynd tablet-based EEG game interface [103].

2.4.7 Comparison of Existing Technology with the Present Study

The author evaluated additional ADHD tools in the literature review. These tools comprised the Integrated Visual and Auditory Continuous Performance Test and the Quotient ADHD Test (a computer-based CPT with an integrated motion tracking system) [100]. However, due to the similarity of operating mechanisms to tools already discussed or otherwise falling outside of the scope of this study, these additional tools were not discussed in depth.

The current study takes note of significant findings of previous research, as well as the existing technology discussed in this chapter. Implementing omission and commission measures within the feature set is of value for further development, as is clear from the existing technology used to discriminate between ADHD and non-ADHD individuals. The value of distractors, specifically visual distractors and a combination of both visual and auditory distractors has repetitively been found to accurately discriminate between these two groups. The existing technology shows consistent use of Continuous Performance Tests. As with T.O.V.A., the continuous

nature of the tool is designed to induce and test participants' levels of fatigue and boredom. This is crucial for identifying key symptoms of ADHD relating to attention. As seen with T.O.V.A and AULA, limiting the inclusion of language and cultural biases strengthens the possibility for accessibility and adoption globally. Analysis of AULA also highlights the value of including motion sensors in the feature set.

Finally, the core and common mechanism of the technology discussed in this chapter relates to "target" and "non-target" stimuli, which demand specific responses or inhibitions. The studies discussed in this chapter reveal that this is a crucial component of any tool designed to discriminate between ADHD-I and non-ADHD individuals.

Conversely, all the existing technology identified either require multiple components (except Akili: Project EVO) or are too bulky and impractical for portable applications. Many of the tools are expensive and therefore generally inaccessible globally. Many of them also require results to be interpreted professionally. It is thus possible to conclude that existing tools arguably restrict rather than enable parents, teachers and carers from attaining crucial insight into the neuropsychological state of the children in their care.

Based on the conclusions drawn from existing technology and research, the current study will explore the possibility of the design and development of a portable, compact, tablet-based intervention with the aim of deploying the intervention in schools, clinical practices and households as an early screening tool for ADHD-I. It is important that the tool be designed to be operated by teachers, clinicians, parents or any other layman administrator, and without the need for specialist knowledge of ADHD. The intervention is packaged as single hardware component with the aim of expanding compatibility to multiple tablet devices. This approach prioritises accessibility and portability with software, cloud-computing and machine learning doing most of the heavy lifting. The new tool is intended to improve on the pitfalls of the existing technology and must be capable of accurately screening and discriminating between ADHD-I and non-ADHD youth, between six and 12 years of age.

Previous tools rely heavily on limited feature sets and place great emphasis on features such as RT and RTV to discriminate between population groups. The current study aims to incorporate a more natural approach to stress simple and sustained attention whilst remaining inclusive of these significant features. The inclusion of distractors, the value of which was also highlighted by previous studies, will be incorporated in the current study. Design elements and game mechanics of the cognitive training tools were also considered in this study.

The current study also intends to investigate the possibility of training a MLC to learn the complex functional relationship between the gameplay feature set and ADHD participants, as well as non-ADHD participants. Features extracted from a tablet-based game will serve as the input to the MLC. The MLC output will be indicative of a participants' ADHD inattentive state.

3 SYSTEM DESIGN

The primary aim of this thesis is to investigate the possibility of developing a screening tool in the form of a tablet-based game, which uses machine learning to assist with the diagnosis of ADHD-I. Chapter Three provides an overview of the tool that was designed and developed in line with the findings of Chapter Two. The primary components of the tool include hardware and software. These two elements are discussed in greater detail below, as well as the screening specifications.

3.1 Hardware

This study utilised two core hardware components, namely a tablet and a laptop. The laptop provided the platform on which the ADHD screening tool software (discussed later in this chapter) was developed, and additionally functioned as a tool for processing and analysis of data. The ADHD screening tool itself takes the form of a game installed on a tablet, which provided the screening interface to be used by study participants. These two hardware components are depicted in Figure 16 below.

Figure 16: ADHD screening tool hardware components.

3.1.1 NVIDIA Shield K1 Tablet

The present study made use of two NVIDIA Shield K1 tablets. This tablet runs the Android operating system and was selected based on its internal components,

performance specifications and affordability at the time of purchase. Developed specifically for gaming, the NVIDIA Shield K1 tablet provided the freedom for fast prototyping, as well as efficient game optimisation. The device specifications can be seen in Table 2 below.

Table 2: NVIDIA Shield K1 tablet Specifications [104].

PROCESSOR	NVIDIA® Tegra® K1 192 core Kepler GPU 2.2 GHz ARM Cortex A15 CPU with 2 GB RAM
DISPLAY	8-inch 1920x1200 multi-touch Full-HD display
AUDIO	Front facing stereo speakers with built-in microphone
WIRELESS	802.11n 2x2 MIMO 2.4 GHz and 5 GHz Wi-Fi, Bluetooth 4.0 LE, GPS / GLONASS
MOTION SENSORS	3-axis gyro, 3-axis accelerometer, 3-axis compass
WEIGHT AND SIZE	Weight: 356 g, Height: 221 mm Width: 126 mm, Depth: 9.2 mm
BATTERY	19.75-Watt Hours

In order to ensure maximum accessibility, it was important that the tablet utilised in this study included components available across multiple devices. The four components in question comprised a multi-touch display, speakers, wireless connectivity and motion sensors (specifically a three-axis accelerometer). Additional valuable features include the low weight of the device and a relatively good battery life. This study aimed to make adequate use of these components as they were present in most mobile devices at the time of this study, and considered to be present in future Android supported devices. [105]

3.1.2 Laptop

The laptop used for purposes of this study was a Gigabyte I7 Sabre, with internal components upgraded for performance. The laptop components included 16 GB DDR4 RAM, Intel i7-7700HQ CPU and a GTX 1050 GPU. The laptop is also equipped with Wi-Fi functionality for remote access to participant data.

3.2 Software

The tablet-based game designed and developed to assist with screening of ADHD involves various software programmes. Software was required to design a game, store data online, process gameplay data, and implement machine learning. This section presents an overview of all the software elements used during the game development process and how these software platforms are interfaced.

3.2.1 Unreal Engine

Game development was conducted on the Unreal Engine (UE) platform, which is one of the most powerful game creation engines currently available. At the time this research was conducted, the platform could be used free of charge for non-

commercial purposes. Unreal Engine has a rapid prototyping interface, a well-established support community and a vast marketplace for cost-free development assets. Figure 17 illustrates the working interface for map level development with droppable assets (top left), folder hierarchy (bottom left), active project assets (top right) and properties of the selected active project asset (bottom right).

Figure 17: Unreal Engine design interface [91].

3.2.2 Google Cloud Storage

This research project made use of a Google database instance for anonymous data storage. A Google cloud storage was selected due to its free tier offering and robust uptime of instances.

3.2.3 Python and Jupyter Notebook

The gameplay data collected during this research was processed with Python 3.6 and Jupyter Notebook (see Figure 18 below). Jupyter Notebook is an open-source web application developed to support interactive data science and scientific computing. It is typically used for cleaning and transforming data, statistical modeling, data visualisation, machine learning, as well as many other use cases [106]. This setup was chosen for its ease of use and debugging, extensive libraries, as well as its support documentation and online community.

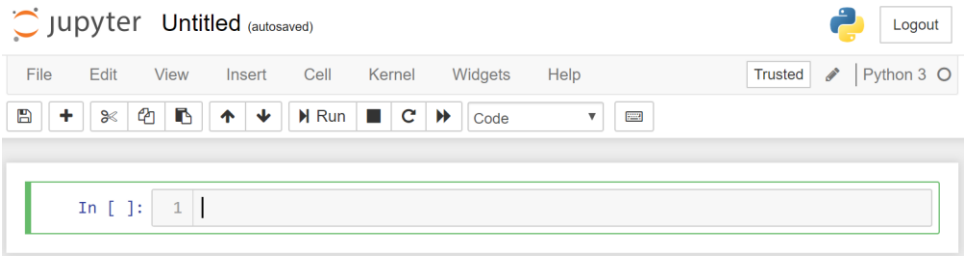


Figure 18: An illustration of the Jupyter Notebook interface using Python 3.6.

3.2.4 Azure Machine Learning Studio

The Azure Machine Learning Studio (AMLS) was used for simultaneous training of multiple machine learning models. AMLS is an online machine learning platform with integrated Python and Jupyter Notebook functionality. Accessible through the web browser, AMLS makes use of cloud-based computing. This platform makes it easy to import and process data, as well as effortlessly train and score machine learning models. A vast array of machine learning models is available, all of which are capable of specifying hyperparameters for fine-tuning. The studio offers web-API functionality, which enables the remote query of any trained machine learning model for classification feedback, thereby making it ideal for commercial implementation.

3.2.5 Interfacing the System

Unreal Engine was used to design and develop a tablet-based game for participant interfacing. The tablet utilised the gaming interface to capture quantitative game features during gameplay. The captured features were then sent to the Google database via wireless internet connection. Gameplay data files were then extracted to the laptop from the database and processed using Python and Jupyter Notebook to generate a new feature set for each participant. Processed data was then uploaded to AMLS to train machine learning models.

3.3 ADHD Screening Tool Specifications

This section discusses the design specifications of the screening tool system with a detailed focus of the tablet-based game design, as well as development in UE. Details of the database, Python processing scripts and AMLS are also provided.

3.3.1 Game Design

Layout

As shown in Figure 19, the game designed in this research study is made up of seven segments (zero to six), each of which present a combination of challenges. Segments are modelled on the DSM-V classification criteria for ADHD-I, and also draw from the existing technology discussed in section 2.4. Individual segments are considered inter-linked, mini-games, which are designed to have a duration of approximately one minute each. Each segment is followed by a subsequent segment after a brief three-second black loading screen. Segments zero and six are identical and serve as references for comparison. Segments two and four include auditory distractors, whereas segments three and four include visual distractors based on findings from the literature [73]. Segment one was designed as an empty mine tunnel, and segment five contains a few game assets toward the end of the segment to induce a level of boredom and fatigue [84, 86]. Game assets are the components that fill the segment (e.g. rocks, gems, obstacles, lights sources, etc.)

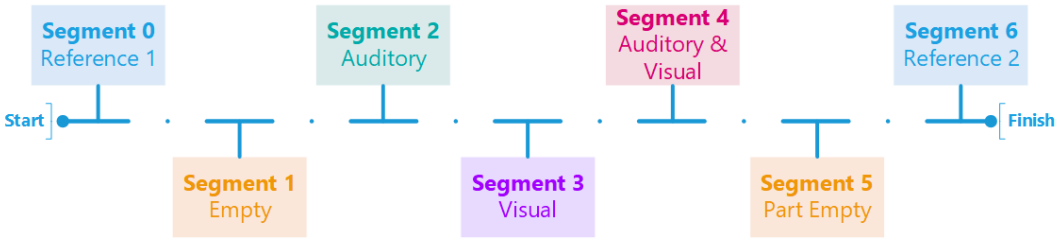


Figure 19: An illustration of the game segment layout.

Table 3 below 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 was also to strengthen the ability of the MLC to generalise well between segments [100, 107]. It is also important to note that the smallest difference in asset placement influences all the other game features.

Table 3: The layout of each game segment.

Game Layout						
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

Setting

As shown in Figure 20, each segment is played in the same setting, and involves a panda bear avatar travelling through a dark mine tunnel on a cart. The goal in each segment is to reach the end of the tunnel as fast as possible. The dark setting was chosen for the purpose of control by being able to limit the visual stimuli presented to the participant. The controlled line of sight and the irregular presentation of response-stimuli were designed to limit anticipatory responses. Additionally, the goal was to force a participant's sustained attention for good performance. Anticipatory responses are a common feature and challenge found in the literature [85]. Compared to the mechanism of the CPTs discussed in Chapter Two, the CPT mechanism employed in this study is unconventional. This

study offers a participant response-based mechanism that determines the rate of presented stimuli.

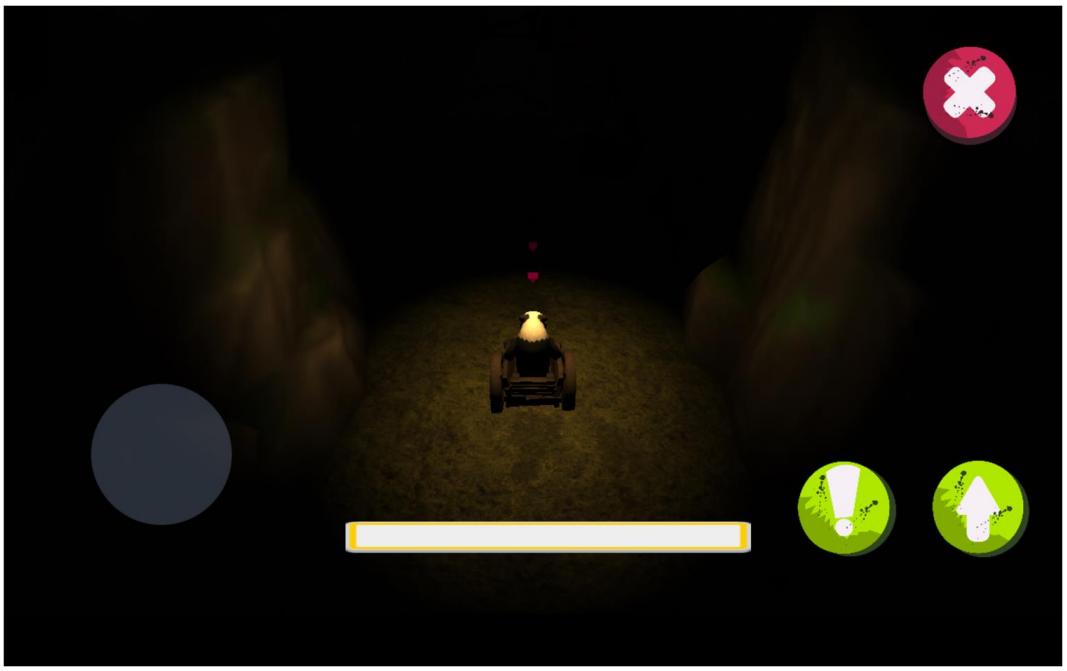


Figure 20: An illustration of the dark mine setting presented in the tablet-based game.

From the start of each segment, the stimuli presentation rate increases incrementally, from base speed to maximum speed, as the avatar progresses through the segment. Should the avatar collide with an obstacle, a time penalty is incurred: the speed of the avatar is reset to the base speed and the speed incrementor is reset. Stimuli presented aim to recreate a mine tunnel setting, and consists of boundary walls, ramps, obstacles, collectables, as well as auditory and visual distractors.

Controls and Logic

As seen in Figure 21, participants are required to make use of an on-screen joystick to control the side-to-side movement of the avatar. A joystick was selected as the moving mechanism to isolate and capture participant movement with the tablet's three-axis accelerometer during gameplay. This method was implemented based on results from the Quotient ADHD System, the AULA, as well as multiple studies investigating the movement intensity of participants during a Go/No-Go task and its association with ADHD [85, 100, 108]. Contrary to traditional methods of capturing accelerometer data, and to complement the CPT mechanism implemented in this study, each accelerometer sample is game avatar positional dependent and not time dependent. This was implemented for direct event-based comparison.

Accelerometer data is captured 2262 times for the total 89 game tiles traversed in each segment. This translates to 25.42 vector data samples per game tile. The fastest segment completion time, which includes starting from base speed and

without obstacle collisions, is 61 seconds. According to the frequency formula, represented by Equation 1, the sampling frequency is therefore: 37.08 Hz.

$$f = \frac{n}{T} \quad (1)$$

This results in an average traveling speed of 1.46 tiles per second. Due to undesired data artefacts at the beginning of the accelerometer vector data set, the set was reduced to a fixed 2230 vector data features for each of the three axes (x, y, z). Sampling frequency reduces to 24.32 Hz if the avatar travels at base speed for the entire duration of a segment. This results in an average traveling speed of 0.96 tiles per second.

To achieve the game goal, it is important to avoid stimuli presented in the form of obstacles and to respond to gem stimuli by collecting them. Collection of pink gems increases the fuel in the torch fuel meter on a unit basis. The collection of a kamikaze gem fills the torch fuel meter completely but requires an intentional sacrifice of at least two obstacle collisions in return. The kamikaze gem was only included once throughout the entire game and serves to challenge the participant's cognitive reasoning and decision making.

If the torch is toggled on with the torch button, the participant's line of sight in the tunnel is increased. The torch fuel meter then decreases at a constant rate as long as the torch is on, which simulates real-world consequences. The torch can be toggled off with the torch button in order to conserve fuel. The on-toggle of the torch increases the range of visibility in the tunnel and simultaneously decreases the pressure on RT by making it easier to avoid obstacles and collect gems. It follows therefore that the converse is also true, as seen by the range of visibility in Figure 20. Certain obstacles that the participant encounters can be avoided by making use of the jump button. Both the jump and torch buttons must be utilised by the participant to improve obstacle avoidance and gem collection during gameplay. This is an example of simple attention. The overall avoidance of obstacles and collection of gems requires the application of sustained attention. Combined, these result in greater overall performance to best achieve the game goal.

Compared to the CPTs discussed in Chapter Two, this game is designed to force responses from participants according to their performance by automatically and continuously moving the avatar through the mine at speeds that are influenced by game elements. Go/No-Go task stimuli are presented in the form of gems (to be collected) and obstacles (to be avoided). The RT feature and impulsivity are measured by the number of gems collected and missed, as well as the number of obstacle collisions and misses. The RTV feature is measured by the segment duration as any obstacle collisions result in a time penalty. Measurement of RT and RTV therefore employs a reinforced learning mechanism by rewarding the participant with torch fuel when gems are collected. It also penalises the participant for obstacle collisions by an auditory injury sound from the panda avatar, resetting the avatars' speed to zero, increasing the overall segment duration and further decreasing the torch fuel meter due to the avatar speed reset. The pause button presents the option to exit the game or return to the task.

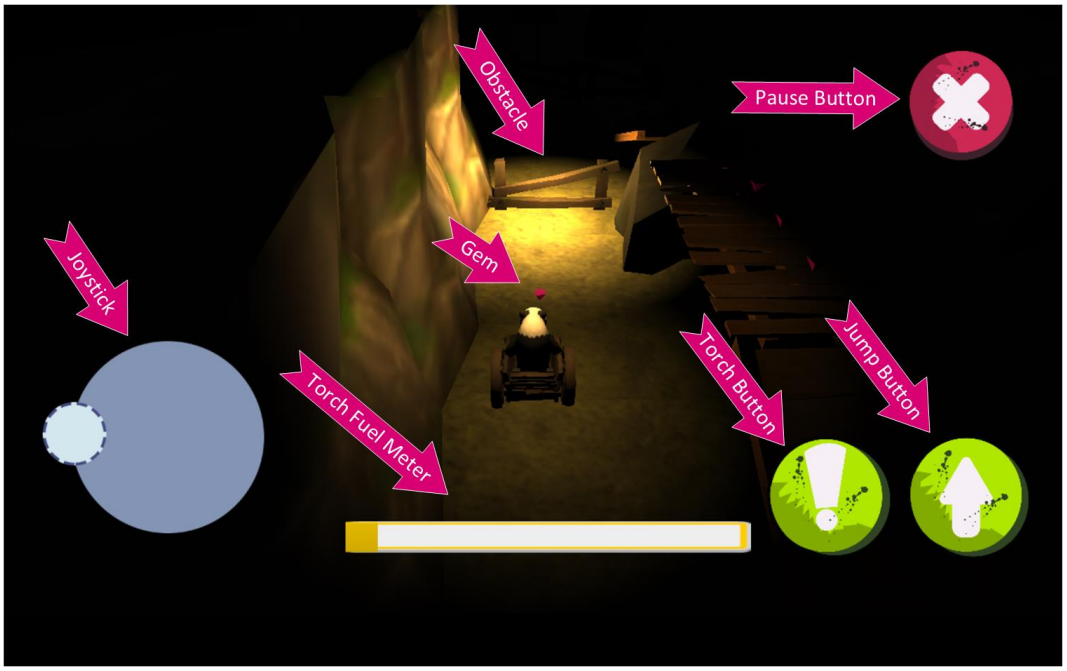


Figure 21: An illustration of the various game elements with the torch on.

The first-order features captured by the tablet during participant gameplay can be seen in Figures A.1 and A.2 in Appendix A (see pink pills). Second-order features are also shown (see orange rectangles). Data processing and second-order feature creation is performed by making use of Python scripts which will be discussed later in this chapter.

3.3.2 Game Development

This project made use of the UE blueprint visual scripting tool. This is a node-based scripting interface provided by the UE Editor [109]. Figure 22 provides a view of the Editor, as well as the nodes, events, functions and variables controlling amongst other things, the gameplay logic and interactions between in-game assets. The logic displayed details loading of the next game segment when the avatar collides with the “open level” in-game asset. The open level in-game asset was implemented using LSB. This method was most effective at successfully executing the game logic with minimal optimisation and installation file size.

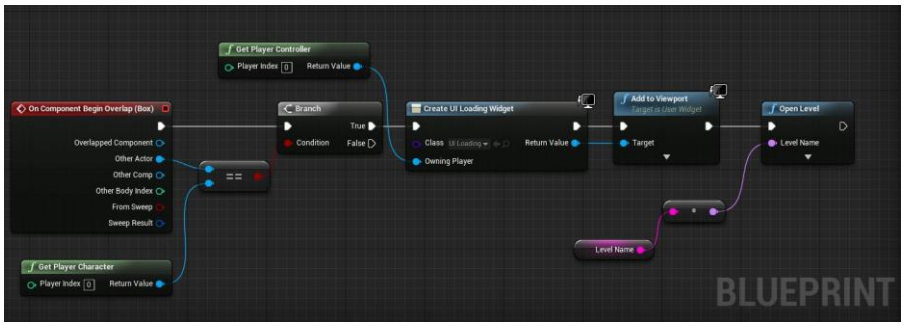


Figure 22: UE game loading logic using blueprint visual scripting.

Each game segment has a corresponding map-level file. This was used as a canvas to implement segment-related assets and logic, such as subsequent segment loading, obstacle collision, respawn points, gem collection, movement restriction, distractor triggers, button presses, auditory feedback and first-order feature capturing (see Figures A.1 and A.2 in Appendix A).

3.3.3 Database Structure

The Google database instance was initialised to receive and store participant gameplay data upon completion of the game or early game exit by participant. The database front-end is secured with password protected login. Upon successful login, participant data is displayed with unique session IDs, serving to identify each individual participant with a randomised number (see Figure 23). For the purposes of this research, the *SessionIDs* were manually linked to participants, by writing the corresponding *SessionID* next to the participant name on a printed name list after testing. Name lists were intentionally kept offline for the purpose of confidentiality. As shown in Figure 23, data from each segment is stored individually according to its specified *SegmentID* of each participant. Should a participant choose to exit the game during gameplay, all data up to that point is captured and uploaded to the database.

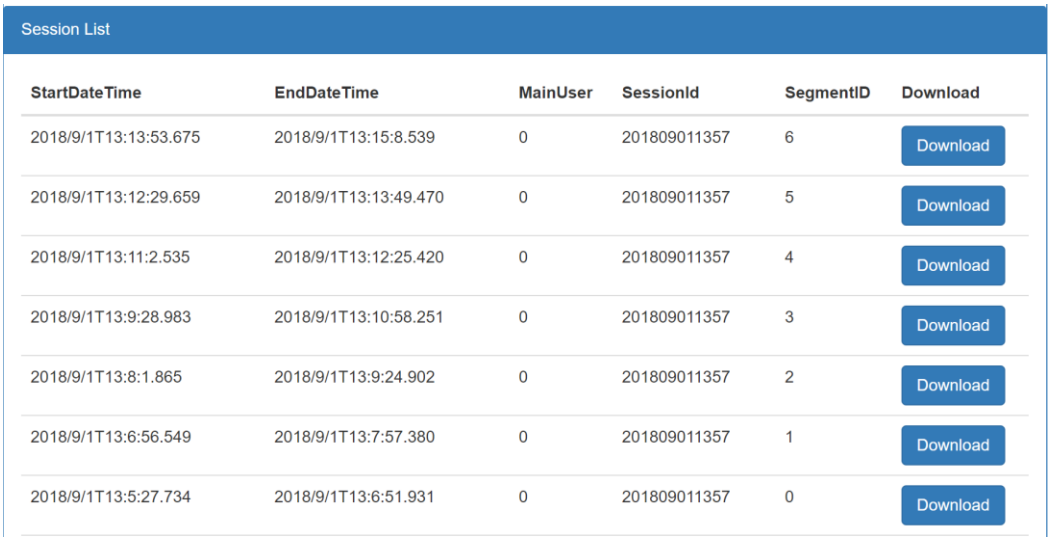


Figure 23: Database interface for extracting anonymous participant data.

3.3.4 Python Scripts and AMLS

Python was used to create data processing scripts to generate the second-order features (as seen in Figure A.1 and A.2 in Appendix A) for the training of MLCs. This scripting language was chosen due to the integration possibilities with AMLS. Although this study employed manual execution of data processing scripts, a commercial version of the ADHD screening tool can easily integrate the same data processing scripts with AMLS for automated execution and data processing. Python also serves as an integral tool to validate data integrity and structure. The

first-order and second-order features are discussed in greater detail in section 4.4 below.

AMLS was used to test a hold-out sample on the trained classifier through the available web-API service. This web-API functionality was successfully tested with the use of Python to demonstrate that it works and that it can easily be integrated into the table-based game.

3.3.5 Machine Learning

For the purposes of this study a two-class MLC model was implemented using a supervised learning approach. As discussed in Chapter Two, a MLC is one of many machine learning techniques used to categorise or class samples of data. Due to the two classes of participants in this study, namely ADHD and non-ADHD participants, a two-class MLC was used. Supervised learning is the training technique used for the two-class MLC to categorise the participant data samples. This learning technique gives the MLC access to the true diagnostic condition of the participants while the classifier is training how to categorise the participants.

Based on the research discussed in section 2.3 and given the binary or two-class participant group in this study, namely: ADHD and non-ADHD, this study will make use of a classification technique. Given that the diagnostic state of the participants is known, a supervised machine learning technique will be selected. The machine learning technique is therefore known as a two-class classifier trained using supervised learning. Multiple algorithms exist within this classifier group to perform the training and participant classification.

Individual Classifiers

Due to the uncertainty of the performance of a classifier when applied to a new dataset, a multi-classifier approach was implemented. Each game segment would have a corresponding MLC. The average from these individual classifiers would then constitute the final participant classification.

Consensus Classifier

As an alternative and to ensure accurate participant classification, a skeletal classifier model approach was taken. The skeletal classifier was designed to achieve a model with a strong ability to generalise on segment data, including inter-segment variation on all its features. The design resulted in the creation of a consensus algorithm that cumulatively evaluates game segments and provide a single consensus classification output, C_f, as well as a consensus confidence score, C_c. The consensus algorithm has of the following form:

$$C_f = \sum_{i=0}^{n} c_i \,, \tag{2}$$

$$C_c = \frac{|C_f|}{n} \times 100, \tag{3}$$

where i represents the position of the segment in the game, c represents the classification of the segment $(Normal = 1, Abnormal = -1)$, 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. For participant classification, the following is applicable:

$$if\ C_f > 0; Classification = Abnormal$$

$$if\ C_f < 0; Classification = Normal$$

This study will be evaluating both the individual and the consensus algorithm approach to determine the effectiveness of the ADHD screening tool.

4 METHODOLOGY

This chapter provides an overview of the study methodology employed to investigate the effectiveness of the ADHD diagnostic screening tool (discussed in Chapter Three).

4.1 Ethics Statement

The research conducted during this study was approved by the HREC of Stellenbosch University. Written, informed consent from parents, as well as assent from participants, was obtained prior to testing. The ethical reference number for this study is M17/05/019, with the proof of ethical clearance presented in Figure B.4 in Appendix B.

4.2 Study Design

This was a single-phase study, which aimed to enrol 76 participants between the ages of six and 12 years. Participants were required to play a tablet-based game once during a 20-minute testing session. The game was specifically designed and developed for the screening of ADHD-I. The consort flow diagram is shown in Figure 24.

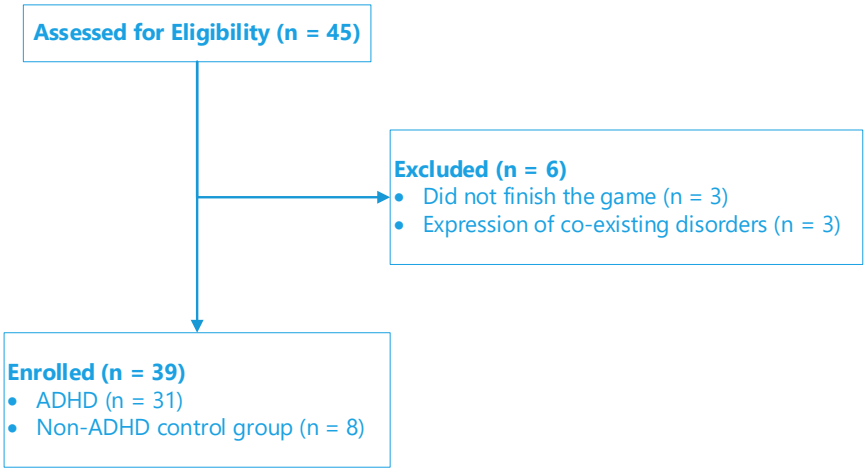


Figure 24: Consort flow diagram.

4.3 Participants

The following section discusses the sample size calculation, inclusion and exclusion criteria, as well as the methods employed to recruit study participants.

4.3.1 Sample Size Calculation

A sample size calculation determines the necessary population size for statistical significance. Following consultation at the Stellenbosch University Centre for

Statistical Consultations,[13] and using a two proportion, paired sample McNemar's test, a participant population size for classifying participants as either ADHD or non-ADHD was determined for a statistical power goal of 0.9. McNemar's test was performed using *Statistica 13* software. The test parameters can be seen in Table 4 below.

Table 4: McNemar's test parameters and values.

McNemar's Test	
Parameters	**Values**
Delta (δ)	0.15
Eta (η)	0.2
Type 1 error rate alpha (α)	0.05
Power Goal	0.9
Actual Power for Required N	0.9032
Require Sample Size (N)	76

The required sample size in comparison with the desired power curve is illustrated in Figure B.1, Appendix B. As seen in Table 4 above, the required sample size for statistical significance was determined to be 76, with a corresponding power goal of 0.9. The Delta parameter represents the population proportion difference in the event of interest between the first and second measurement standards i.e. true cases identified by the gold standard[14] compared to the ADHD screening tool designed and developed in this study. The Eta parameter represents the total population proportion of disagreement between the two measurement standards.

The null hypothesis and alternate hypothesis were identified in the following form:

$$H_0: \delta = 0,$$

$$H_1: \delta \neq 0, \text{ where}$$

$$\delta_s = \pi_1 - \pi_2. \qquad (4)$$

In respect of Equation 4, π_1 represents the number of true positives identified according to the gold standard, and π_2 represents the number of true positives identified by the ADHD screening tool.

[13] Statistical consultation for purposes of this study was provided by Professor Daan Nel and Professor Martin Kidd.

[14] This term is discussed in part 2.2 above.

4.3.2 Inclusion Criteria

All 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 inattention subtype, based on the gold standard for clinical ADHD assessment; alternatively, meet the criteria for neurotypical symptoms based on the ADHD DSM-V criteria, coupled with the SWAN rating scale;
- Performance medication naïve;
- Written, informed consent from parents; and
- Assent from the participant.

4.3.3 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 mental retardation.

4.3.4 Recruitment

ADHD-I participants

ADHD-I participants were recruited by a clinical psychologist.[15] These participants were selected following initial consultation and validation of eligibility according to the ADHD inattention subtype inclusion criteria.

Non-ADHD participants

Non-ADHD participants (the study control group) were recruited from schools within reasonable proximity of the study site. Five of the schools attended by participants were public schools as classified by Section 21 of the South African Schools Act.[16] One school was a private institution. Calls for participation were also distributed via friends and colleagues.

As required by law, permission was requested from and granted by the Western Cape Education Department prior to contacting the schools for participant

[15] Clinical psychologist Rose-Hannah Brown facilitated participant recruitment at the Cape Gate Therapy Centre.

[16] Act 84 of 1996.

recruitment. Teachers at the schools concerned identified neurotypical learners according to the SNAP rating scale. Invitations were then sent to parents of identified learners requesting their children's participation in this study. Of the six schools, two were willing to distribute invitations, two declined to partake and two neglected to respond to multiple enquiries. Participating schools collectively received 400 invitations.

The SNAP rating scale was used for its simple, categorical summation-score of symptoms. As seen in Appendix C, the invitation required parents to indicate their willingness to participate in the study, as well as provide their contact details on the return slip. Attached to the invitation was a SWAN rating scale questionnaire to be completed by parents who indicated their willingness to participate in the study. Return slips were collected from the schools. Responding parents were then contacted to schedule testing appointments and requested to bring along the completed SWAN rating scale.

4.4 Feature Extraction

Figure 25 shows the data flow and processing procedure from participant gameplay to the evaluation of participant performance metrics. After a participant completed the game phase, game data was automatically uploaded to the database in the *.json* file format. This data was then downloaded from the database and stored on the laptop. As previously discussed, Python was used to process the downloaded data. First-order features were extracted from the data file according to the *Feature Matrix* in Figures A.1 and A.2 in Appendix A. Second-order features were created from first-order features through simple mathematical computations and keeping track of more in-depth game logic. These mathematical computations were performed using Python to free the tablet device from the unnecessary computational workload during testing.

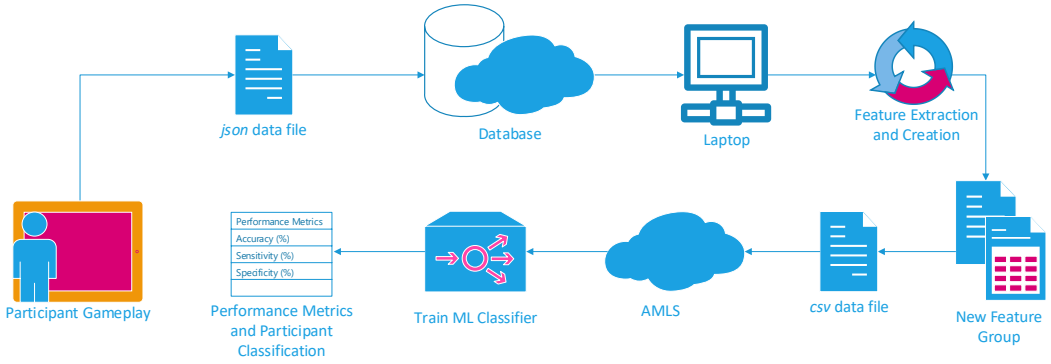


Figure 25: Flowchart indicating the data flow and preparation procedure for machine learning.

The principle component analysis (PCA) was performed on each axis individually as this resulted in a stronger classifier. Missing vector data values were replaced with a zero value as the accelerometer vector values ranged between negative and positive real numbers. This was identified as the best alternative substitute by evaluating the classifier performance metrics. Individual classes were generated for *Gender*, *Race* and individual game segments. This was implemented to remove weighting bias prior to the training of MLCs.

4.4.1 First-order Features:

The first-order features were captured by the tablet in their simplest form during gameplay to avoid unnecessary computational workload for the tablet during testing. Due to the number of samples collected, the most performance-intensive first-order feature is the accelerometer vector data captured by the tablet's built-in 3-axis sensor.

4.4.2 Second-order Features:

Figures A.1 and A.2 in Appendix A show the second-order features that were created by either transforming first-order features from integer values to classes, to count event occurrences captured by first-order features or to perform computations on first-order feature values.

Feature Conversion

Participant profile features such as *Gender*, *Race* and *Segment IDs* were transformed into classes to prevent the addition of a weighting for any specific value and to instead indicate a class difference. The *Diagnosis* feature was transformed into a single binary class feature, as only two mutually exclusive options exist. Multiple first-order features were captured by means of timestamps as events occurred during gameplay. These timestamps were either converted into a single binary feature (e.g. *Game Exit* to *Exit Pressed*) or used to calculate durations (e.g. *Start Time* and *End Time* to determine *Segment Duration*). Certain timestamps were simply used to determine the number of times a feature occurred (e.g. *Auditory Distractions* to attain *Auditory Distraction Count*).

Torch Duration

Multiple first-order features, of both integer and timestamp format, were required to calculate the *Torch Duration* second-order feature. These first-order timestamp features are, namely, *Torch Toggle On*, *Torch Toggle Off* and *Torch Meter Empty*. The requisite integer feature was *Torch Toggle Count*. The torch is a feature of the game that is unaffected by transitions between segments. Due to this continuous mechanism, multiple torch state conditions had to be checked for the array of timestamps.

Principle Component Analysis

Principle Component Analysis (PCA) was performed on the three-axis accelerometer data to reduce the dimensionality of the vector data features. This technique is commonly employed in practice, especially when working with small sample sizes or a small sample-to-feature data set. [110]

The PCA was performed on each axis individually (axes *x*, *y* and *z*) as this resulted in a stronger classifier. Missing vector data values were replaced with a zero value as the accelerometer vector values ranged between negative and positive real numbers. This was identified as the best alternative substitute by evaluating the classifier performance metrics. For all cases, a standard scaler was employed

before PCA, which was used to normalise the vector data (standard deviation = 1, mean = 0).

Statistical Features

Based on the approaches of both D'souza et al. and Li et al. [76, 78], a statistical feature set was created from the accelerometer data captured during gameplay. Figure A.1 and A.2 in Appendix A show the 34 second-order statistical features created from the three-axis accelerometer data. The *Root Mean Square* feature is calculated using all three axes.

4.5 Site description

Testing of participants in this study took place at a single clinical site in Cape Gate, Cape Town. The room used was selected for its lack of distractions and contained only two chairs and two cabinets. Figure 26 below shows the room layout, which was specifically set up so that the researcher would not to be a distraction for the participant during testing. As seen in Figure 27, the researcher sat in the chair behind the participant to oversee the testing session.

Figure 26: Test site layout.

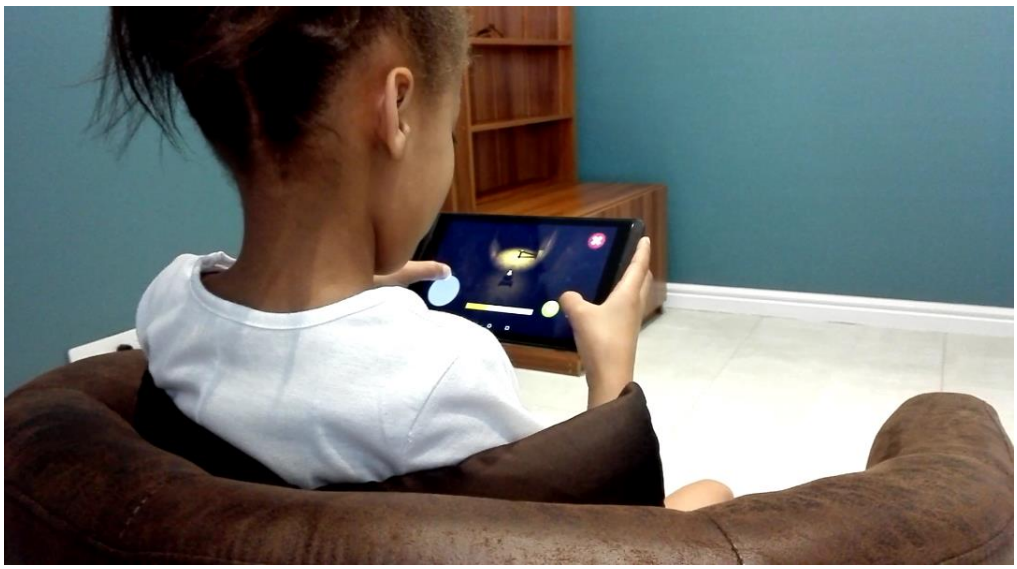


Figure 27: Perspective of participant from the researcher's chair.

4.6 Testing Procedure

Two NVIDIA Shield K1 tablets with a wireless internet connection were used during testing. Only one was utilised at any given time. The second tablet served as a backup in the event that the battery of the first reached a charge of 15 percent. To validate successful installations, the game and tutorial were installed and executed on the tablets prior to arrival at the test site. Additionally, the researcher used a laptop with a wireless internet connection to validate the authenticity of data uploaded to the database by both tablets.

4.6.1 Tutorial Phase

Prior to testing, participants were instructed to keep their hands and forearms from resting on or against any surface, and to keep the tablet suspended during gameplay. Tutorial gameplay then commenced, and participants were subtly reminded of the instruction when they erred. The tutorial systematically explains all the gameplay controls with the aid of in-tutorial visual cues. The researcher answered any additional questions as they arose. The tutorial level is made up of two segments, both of which are the same duration as the game segments.

4.6.2 Game Phase

Following the tutorial phase, participants played the entire game from start to finish. Participants were left to complete all seven game segments without external input. If participants asked the researcher questions, the researcher responded as quickly as possible and encouraged the participant to keep going. The fastest possible game completion time is just over seven minutes (61 seconds per segment), but poorer performing participants can take considerably longer. Upon completion of the game participants receive a score for the number of gems collected during gameplay of all seven segments.

4.7 Machine Learning

AMLS was used to develop and train the MLCs. Classifiers were trained, and parameters adjusted until the best performance was achieved. The two-class MLCs investigated included the Averaged Perception, Bayes Point Machine, Boosted Decision Tree, Decision Forest, Decision Jungle, Locally Deep Support Vector Machine, Logistic Regression, Neural Network, and a Support Vector Machine. Segments one and five were excluded from analysis as these were directed at influencing the participant (see Table 3 for game design).

4.7.1 Individual Classifiers

The individual classifier approach was to use a separate ML classifier for each game segment. This resulted in five classifiers across segments zero, two, three, four and six. All classifier techniques were trained and adjusted on segment zero to determine the optimal performing classifier. The optimal performing classifier was then selected to train on each of the five game segments individually. The performance of these classifiers is discussed in section 5.2.

4.7.2 Consensus Classifier

Nine ML classifiers were trained and adjusted on the five segments. The optimal performing classifier was selected for integration with the consensus algorithm which resulted in a final classification of the 39 participants. Filter-based feature selection was used to determine the most significant features according to Pearson's Correlation. Feature significance is discussed in section 5.3.1. Furthermore, according to the finding of a study conducted by Silverstein et al., a stepwise feature removal was implemented according to Pearson's correlation to improve model performance. [52] Figure 28 shows the participant classification procedure from participant gameplay to classification. This illustrates the use of the AMLS web-API which will be demonstrated in section 5.3.4.

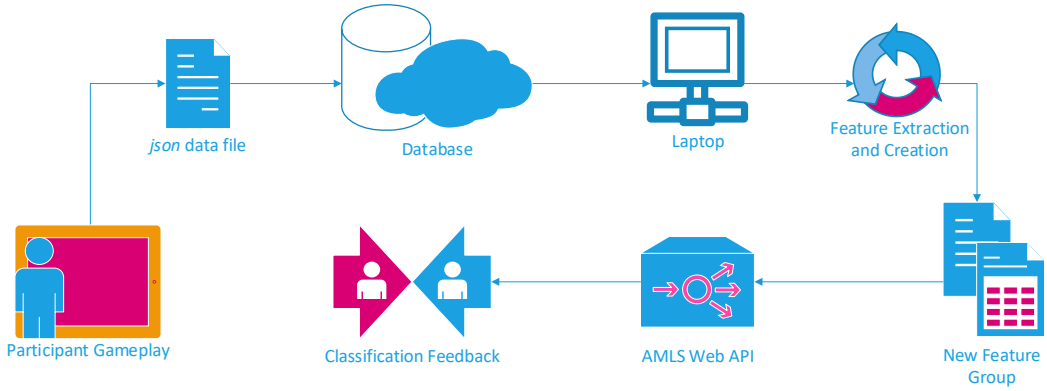


Figure 28: Flowchart indicating participant classification procedure from gameplay to classification.

4.8 Statistical Analysis

4.8.1 Feature Set

The Pearson's correlation coefficient, r, was utilised in this study to determine the strength of the linear relationship between each game feature and the ADHD diagnostic state of the participants. This statistical method is widely used in industry for its efficiency in declaring a linear relationship between two variables. [111] The value for the Pearson r coefficient can be calculated as seen below:

$$r = \frac{N\sum y\bar{y} - \sum y \sum \bar{y}}{\sqrt{N\sum y^2 - (\sum y)^2}\sqrt{N\sum \bar{y}^2 - (\sum \bar{y})^2}} \quad (5)$$

The variable N in Equation 5 represents the number of samples, y the actual diagnosis and $\bar{y}$ the feature being correlated. For each correlating relationship, r, ranges from [-1, 1], where minus one represents a strong negative linear relationship, one a strong positive linear relationship, and zero no relationship.

4.8.2 Patient Classification

Performance metrics used to illustrate the performance of the ADHD screening tool are discussed in this section. The ROC curve is a commonly used tool to visualise and comprehensively evaluate the performance of a classifier [112]. As seen in Figure 29, the graph is created by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) of a given test. Due to its common use in practice, the ROC curve and the corresponding AUC were used to gauge the performance of the classifier. The ROC curve provides a true and false positive rate for any given threshold value. Thresholds are specified according to the application of the classifier and are coupled with a trade-off of either sensitivity or specificity as the thresholds moves away from the optimal threshold.

The AUC is a measure of the accuracy of a diagnostic test measured on a scale from [0, 1]. A perfect classifier produces a ROC curve that crosses the coordinate points (1, 0). A classifier with a ROC curve that follows the diagonal line from (0, 0) to (1, 1) produces an AUC of 0.5 and is considered to produce a completely random output. The AUC grades a test's accuracy as: fail (0.5 - 0.6), poor (0.6 – 0.7), fair (0.7 – 0.8), good (0.8 – 0.9), excellent (0.9 - 1).

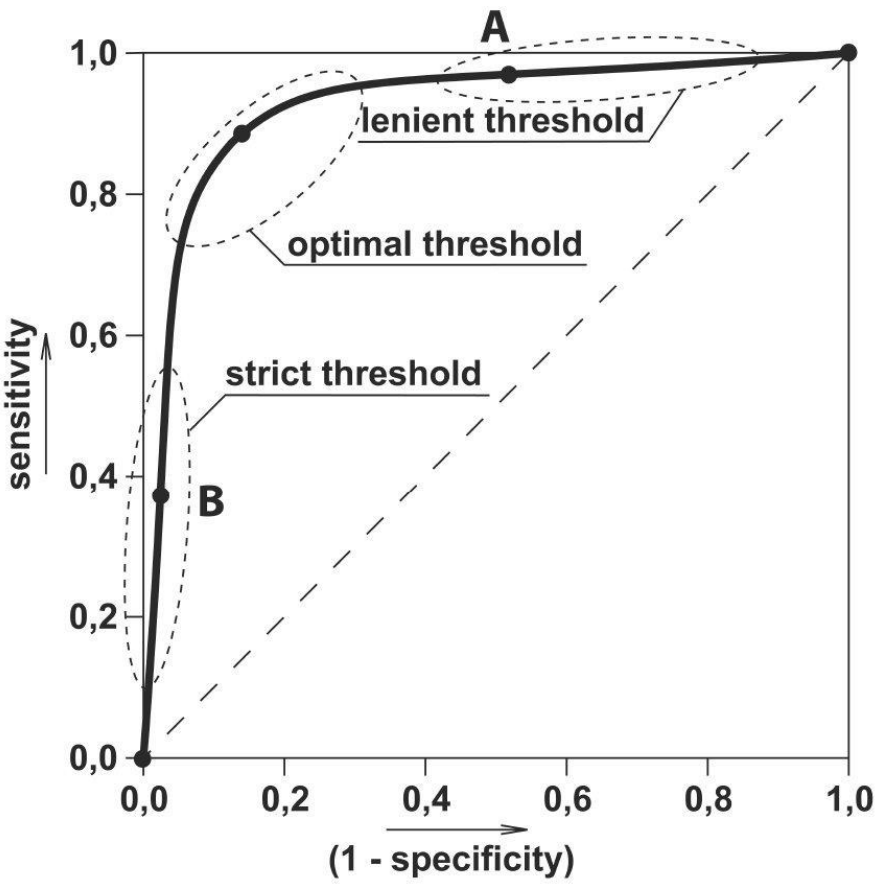


Figure 29: ROC curve with different thresholds [113].

As illustrated in Table 5, another metric to measure the performance of a classifier is a confusion matrix. A confusion matrix is a tool used to quantify the performance of a ML model for a specified threshold value. The default threshold value is 0.5 which corresponds to the optimal threshold. According to the ROC curve in Figure 29, a lower threshold will result in a lower sensitivity and higher specificity value.

Table 5: Confusion matrix structure.

		True Condition	
		ADHD	Non-ADHD
Test Prediction	ADHD	**A** (True Positive)	**B** (False Positive)
	Non-ADHD	**C** (False Negative)	**D** (True Negative)

The confusion matrix entries represent the classifiers classifications versus the gold standard. Matrix entries are true positive (A), true negative (B), false positive (C) and false negative (D). Entries are used to calculate the sensitivity, specificity and accuracy of the classifier:

$$Sensitivity = \frac{A}{(A + C)}\ x\ 100 \tag{6}$$

$$Specificity = \frac{D}{(D + B)}\ x\ 100 \tag{7}$$

$$Accuracy = \frac{A + D}{(A + B + C + D)}\ x\ 100 \tag{8}$$

$$AUC = \frac{Sensitivity + Specificity}{2} \tag{9}$$

The sensitivity indicates the number of correctly identified abnormal participants with respect to the total number of abnormal participants. The specificity indicates the number of correctly identified normal participants with regards to the total number of normal participants. The accuracy indicates the number of correctly identified normal and abnormal participants with regards to the total population.

5 RESULTS

5.1 Clinical Study

A total of 45 participants were recruited for this study. However, six sets of participant data were excluded after testing due to medication use, signs of co-existing ADHD subtypes and incomplete data.[17] Final test data therefore included 39 participants: 31 ADHD-I (25 M: 6 F) and eight non-ADHD control group (6 M: 2 F) between the ages of six and 12 years, with a mean age of 9.02 (SD = 1.88). The corresponding power curve can be seen in Figure B.2, Appendix B for the given sample size. A larger *Delta* value was selected to achieve a power goal of 0.9 for the 39 recruited participants. The McNemar's Test parameters are illustrated in Table 6 below.

Table 6: McNemar's Test parameters and values for study sample size.

McNemar's Test	
Parameters	**Values**
Delta (δ)	0.19
Eta (η)	0.2
Type 1 error rate alpha (α)	0.05
Power Goal	0.9
Actual Power for Required N	0.9057
Require Sample Size (N)	39

5.2 Individual Classifiers

The performance metrics of individual classifiers can be seen in Table 7 below. All classifiers were trained on data from segment zero, which included 39 samples. Classifier training included a 39-fold cross validation with a five-iteration randomised parameter sweep to attain the optimal classifier.

Table 7: Performance metrics for the adjusted classifiers trained on segment zero.

	Performance Metrics			
	%			
Two-Class Classifier	*Accuracy*	*Sensitivity*	*Specificity*	*AUC*
Averaged Perceptron	76.9	96.8	0	0.415
Support Vector Machine	76.9	96.8	0	0.440
Decision Jungle	79.5	96.8	12.5	0.488
Boosted Decision Tree	76.9	93.5	12.5	0.492
Decision Forest	82.1	100	12.5	0.532
Neural Network	76.9	93.5	12.5	0.544
Logistic Regression	79.5	96.8	12.5	0.613
Bayes Point Machine	76.9	96.8	0	0.605
LDSVM	76.9	90.3	25	0.694

[17] Researcher's note: Despite knowledge of certain exclusionary factors disclosed before participation, no child was turned away once they arrived for testing.

As seen from Table 7, the performance of the classifiers would be considered poor according to the scoring scale used for the AUC. The same classifiers were then trained on data from segment six to compare participant performance. According to Figure 19, segment six serves as a second reference for segment zero. Results from this comparison can be seen as a measure of the learning rate of participants (see Table 8). The top performing classifier for each segment can be seen in Table 9 below. Results for segments two, three and four can be found in Appendix D.

Table 8: Performance metrics for the adjusted classifiers trained on segment six.

	Performance Metrics			
	%			
Two-Class Classifier	*Accuracy*	*Sensitivity*	*Specificity*	*AUC*
Support Vector Machine	76.9	96.8	0	0.097
Averaged Perceptron	76.9	96.8	0	0.141
Bayes Point Machine	74.4	96.7	0	0.222
Logistic Regression	69.2	87.1	0	0.242
Decision Jungle	76.9	96.8	0	0.242
LDSVM	64.1	77.4	12.5	0.242
Neural Network	76.9	96.8	0	0.244
Decision Forest	76.9	96.8	0	0.310
Boosted Decision Tree	74.4	90.3	12.5	0.411

Table 9: Performance metrics of the top performing classifier for each of the five game segments.

Two-Class Classifier	*Seg.*	**Performance Metrics**			
		%			
		Accuracy	*Sensitivity*	*Specificity*	*AUC*
LDSVM	*0*	76.9	90.3	25	0.694
-	1	-	-	-	-
SVM	2	82.1	96.8	25	0.609
Boosted Decision Tree	3	84.6	96.8	37.5	0.714
Neural Network	4	74.4	93.5	0	0.512
-	5	-	-	-	-
Boosted Decision Tree	6	74.4	90.3	12.5	0.411

5.3 Consensus Classifier

Classification of participants in the clinical study was performed using features extracted during gameplay of the designed tablet-based game discussed in Chapter Three. All participants were either previously diagnosed as ADHD-I or classified as non-ADHD.

5.3.1 Input Data Structure

Of the 39 participants, the data set of one randomly selected participant was excluded from the training set of the classifiers as a hold-out sample and retained for post-testing classifier validation. The performance of the nine classifiers trained on the *Iteration 1* feature set (see Appendix E) can be seen in Table 10 below. All nine classifiers employed seven-fold cross-validation with a randomised evenly-partitioned, stratified sample split. The classifiers were trained on the five segments of data of the remaining 38 participants, a total of 190 samples. The final performance of the nine classifiers can be seen in Table 11. This follows a stepwise removal of the weakest correlated features identified by Pearson's r correlation. Classifiers are tabulated in ascending order of the AUC.

Table 10: Performance metrics for the nine unadjusted classifiers on the Iteration 1 feature set.

	Performance Metrics			
	%			
Two-Class Classifier	*Accuracy*	*Sensitivity*	*Specificity*	*AUC*
Support Vector Machine	78.4	99.3	0	0.515
Averaged Perceptron	74.7	92	10	0.526
Logistic Regression	78.9	100	0	0.536
Bayes Point Machine	76.8	92.4	0	0.553
Decision Jungle	78.9	95.3	17.5	0.591
Decision Forest	73.7	87.3	22.5	0.709
Boosted Decision Tree	81.1	96.7	22.5	0.702
Neural Network	77.9	85.3	50	0.817
LDSVM	82.6	92.7	45	0.819

Table 11: Performance metrics for the nine unadjusted classifiers on the Iteration 2 feature set.

	Performance Metrics			
	%			
Two-Class Classifier	*Accuracy*	*Sensitivity*	*Specificity*	*AUC*
Averaged Perceptron	71.6	88.7	7.5	0.518
Support Vector Machine	78.9	100	0	0.526
Logistic Regression	78.9	100	0	0.549
Bayes Point Machine	77.4	97.3	2.5	0.572
Decision Forest	78.9	92	30	0.654
Decision Jungle	78.9	96	15	0.692
Boosted Decision Tree	80.5	97.3	17.5	0.702
Neural Network	73.2	76.7	60	0.802
LDSVM	87.4	94.7	60	0.881

5.3.2 Refinement of Parameters

To classify a participant as either ADHD-I or non-ADHD, the three MLCs with the highest AUC and sensitivity-to-specificity ratio were selected for further evaluation. The hyperparameters of each classifier was iteratively adjusted using a 10-iteration

randomised parameter sweep and compared with the standard model parameter to achieve the largest AUC. Due to the imbalance of samples, the AUC was selected as the target performance metric.[18] A leave-one-out cross-validation was implemented, resulting in 190-fold cross-validation. The performance of the top three classifiers can be seen in Table 12. The corresponding 95 % confidence intervals (CI) for the classifiers are presented in Table 13.

Table 12: Performance metrics for the top three adjusted classifiers (190 participant samples).

	Performance Metrics			
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 13: 95 % Confidence intervals (CI) for the top three classifiers.

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 %]

The classification accuracy and standard deviation of accuracy (SDA) for the top performing classifiers are shown Figure 30 below.

[18] The AUC summarises the diagnostic efficiency of the classifier [73], [108].

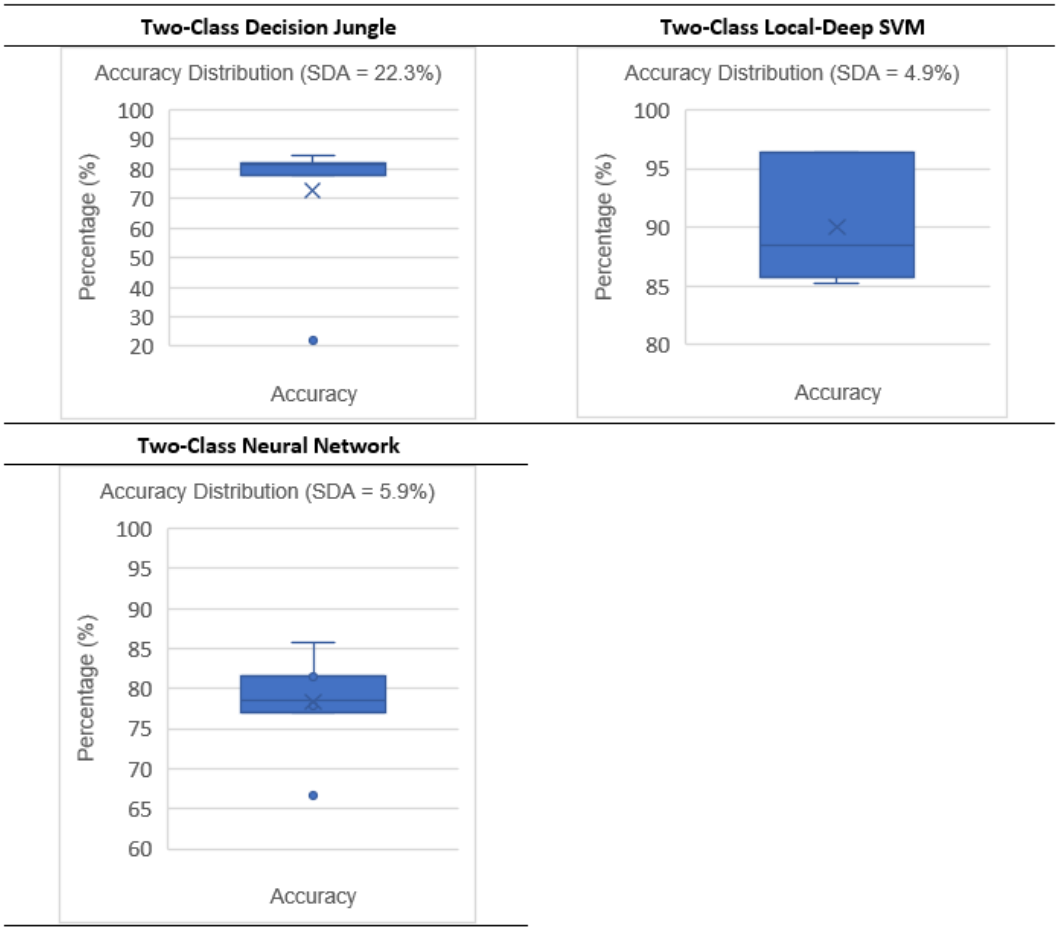


Figure 30: Accuracy and accuracy standard deviation of top three classifiers.

5.3.3 Top-performing Classifier

As seen in Table 10 to 14 and Figure 30, the top performing classifier is the Two-Class LDSVM. Table 14 shows the confusion matrix for the LDSVM classifier with Figure 31 illustrating the corresponding ROC curve.

Table 14: Clinical study 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)

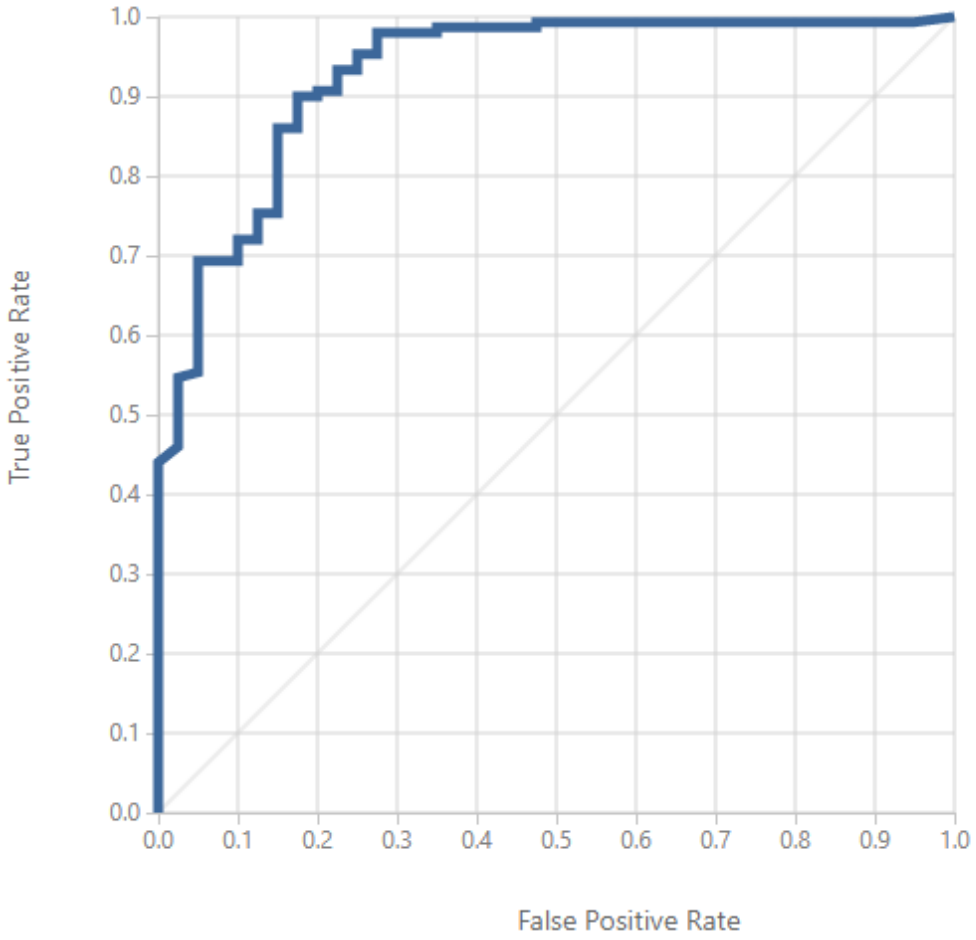


Figure 31: Clinical study ML model ROC curve using the 190 samples of the 38 participants.

5.3.4 Application Programming Interface (API)

The top performing classifier's AMLS web-API (discussed in 3.3.4 above) was then created and used for the classification of the 39th participant. The responses from the classifier for participant's five game segments can be seen under the heading "Classification" in Table 15. Each of these responses has a corresponding probability score, which is a measure of confidence of the MLC for the classification output. The average web-based response compute time of the AMLS web-API was 156 milliseconds, which amounts to 0.78 seconds per participant classification. This participant classification compute time excludes the data pre-processing time, which took at most an additional second with unoptimised Python code. The total time for classification output was therefore less than two seconds.[19]

[19] This time may, however, be affected by internet connectivity speed.

Table 15: Classifier response for each segment of the 39th participant.

	39th Participant Classification		
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

5.3.5 The Consensus Algorithm

Substituting the classifier responses from Table 16 into the consensus algorithm, the final consensus classification, C_f, and consensus confidence, C_c, for the 39th participant is as follow:

$$C_f = 3 \geq 0, Abnormal$$

$$C_c = 80\ \%.$$

As seen from the consensus classification, the 39th participant was correctly classified according to the actual diagnosis. Additionally, the confidence rating indicates a strong degree of certainty for the consensus classification. The final and cumulative consensus algorithm output for the 39 study participants can be seen in the confusion matrix, shown by Table 16.

Table 16: Clinical study ADHD screening tool 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)

5.4 Screening Tool Performance

The performance metrics for the ADHD screening tool classifier can be seen in Table 17 below. The 95 % CI (Sensitivity) is [100 %, 100 %] and 95 % CI (Specificity) is [64.6 %, 100 %].

Table 17: ADHD screening tool performance metrices.

LDSVM Performance Metrics				
Accuracy (%)	*Sensitivity* (%)	*Specificity* (%)	*Delta* (δ)	*Eta* (η)
97.4	100	87.5	0,025	0,025

6 DISCUSSION

6.1 Cost Analysis

Existing tools, as well as relevant and available hardware and software, were researched thoroughly before embarking on this study. Suitable components were selected with a minimalist perspective. The ADHD screening tool and its setup time, as well as the potential cost of small and commercial roll-out of the tool, were also considered. Additionally, consideration was given to maintenance of the tool infrastructure. Table 18 below provides a comparison of the cost associated with the existing tools, as well as the ADHD screening tool developed in this study.

Table 18: Cost comparison of various ADHD screening tools (rates as at 19-10-2018).

Tool	Description	Cost
MOXO - Requires an internet connected computer with external speakers and a keyboard.	1. Single Administered Test 2. Test and Retest Administered with a comparative analysis	1. R1 000 – 15min 2. R2 500
T.O.V.A - Requires an internet connected computer with external speakers and a keyboard.	T.O.V.A 8 Kit: Tests, USB device, microswitch, user's manual, installation CD and guide, USB flash drive, 5 free test credits and accessory cables.	R 12 574.75 ǂ
AULA - Requires an internet connected computer with a virtual reality headset, an external button and external speakers or headphones.	Virtual Reality System: AULA Nesplora + 2 Other Products (12 Months of access)	R 2 400 ǂ* (10 Uses) R 5 760 ǂ* (30 Uses) R 9 600 ǂ* (60 Uses) R 1 600 ǂ* pm (Unlimited Uses)
Current Tool - Requires a wireless internet connection.	ADHD Screening Tool: NVIDIA Shield K1 Tablet AMLS Subscription Google database Subscription	R5000 Free Free

* Requires an internet connected computer with external speakers and a keyboard.

ǂ $1 = R14.05 (19-11-2018)

ǂ* €1 = R16.00 (19-11-2018)

It is important to note that the R1 000 cost associated with the MOXO tool is for a 15-minute testing session, at an official MOXO test centre in Somerset West, Cape

Town.[20] The cost reflects a once-off visit. Should the effectiveness of medication need monitoring, parents can expect to pay R2 500 for two 15-minute MOXO test sessions, six months apart, including a comparative analysis. The use of the MOXO tool also requires travelling to the test centre once an appointment has been scheduled, as the MOXO system includes multiple bulky components. MOXO test results then need to be interpreted by a MOXO professional.

Similarly, the T.O.V.A requires computer hardware and speakers over and above the cost tabulated for the testing kit. The tabulated cost is for the international test kit, shipped to locations outside the United States of America (USA). The T.O.V.A kit nine which is only available within the USA costs the same as the tabulated T.O.V.A kit eight, plus an additional R477.70 ($34) for shipping. The T.O.V.A additionally charges R140.50 ($10) per test session conducted and test results need to be interpretation by T.O.V.A professionals. [114]

The only cost available for the AULA is for a combined package, as seen in Table 18. However, the tabulated cost only reflects the price for access to the service and excludes the hardware components required. The cost of a computer, speakers or headphones, as well as the virtual reality headset still need to be determined and added.

Comparing the required components as well as the cost associated with the service and use of each of the tabulated tools, the *Current Tool* incurs the least costs (give or take the fluctuating cost of components for MOXO, T.O.V.A and AULA). It is also important to note that given the design of the *Current Tool*, further game optimisation will enable the game to be played on multiple tablet-based devices, enabling access to ADHD screening for a larger population group. Should the *Current Tool* see commercial development, it would also enable layman to interpret gameplay results, as the feedback is simple to interpret.

6.2 Safety Analysis

The tool designed and developed in this study does not alter the design of the tablet used to run the software. Manufacturer safety guideline for usage of the device needs to be consulted before using a tablet device to run the ADHD screening tool game and tutorial. The NVIDIA Shield K1 safety documentation can be found on their website. [115]

6.3 Clinical Study

It is important to first address the sample size calculation shown in Table 6. The sample size indicated in Table 6 was not the initial sample size calculated for this study (see Table 4 for the desired sample size). However, the sample size calculated and tabulated in Table 6 was an accurate representation of the sample size recruited for this study. The *Delta* and *Eta* values of both Tables 4 and 6 for

[20] The contact details for the test centre can be found at this website [129].

the desired *Power Goal* of 0.9 will be revisited later in this section as they are important for statistical significance.

6.3.1 Individual Classifier

As seen from the performance metrics in Table 7, 8 and 9, the performance of all the classifiers is poor. This is most likely due to the limited, unbalanced number of ADHD and non-ADHD data samples. It is evident from the *Accuracy* metric in these tables that the classifiers are overfit to the ADHD population. Furthermore, the tables provide insight into the learning rate of participants as they compare the same segment layouts found in segments zero at the start of the game with that of segment six at the end of game. However, this deduction should be investigated with further testing as the training sample size was limited. Further valuable insight from these tables is that the accuracy measure is unreliable when a sample set is unbalanced. Due to the equation for Accuracy (Equation 8) relying on the number of correct classifications, the imbalance in samples easily distorts the result. A more reliable metric for an unbalanced dataset would be the AUC as it reflects the discriminatory ability of the classifier, including both the sensitivity and specificity performance measures (see Equation 9).

6.3.2 Consensus Classifier

The performance of classifiers as tabulated in Tables 10 and 11 are distinguished by two different sets of training data. Table 10 indicates the performance of nine unadjusted classifiers trained on the complete data set identified as *Iteration 1* in Appendix E. Table 11 indicates the performance of the nine unadjusted classifiers trained on the reduced data set identified as *Iteration 2* in Appendix E. The creation of the *Iteration 2* data set followed a similar approach as that employed by Silverstein et al. in their study using clinical data to predict a positive ADHD Diagnosis [52]. Pearson's correlation coefficients were determined for the *Iteration 1* data set and used to identify the strength of the linear relationship between ADHD diagnosis and each data set feature. Features with zero correlation were then removed apart from features which contribute to the overall game and participant profile. Included uncorrelated features comprise *TorchDuration*, *ExitPressed* and multiple *Race* features. Although certain features show stronger correlations than others, it is important to note that correlation does not imply causation. This simply means that although two features are positively correlated, an increase in one feature does not necessarily cause the increase in the other feature but other factors might be causing the increase. The important difference between these two tables is the improved specificity of the top two classifiers, with the top classifier also improving in sensitivity.

Analysis of Table 11 and 12 indicates the improved performance of the top three classifiers with regard to the diagnostic efficiency of the classifiers (AUC) trained on the *Iteration 2* data set. The hyperparameters of the classifiers were tuned to achieve the best AUC and sensitivity-to-specificity ratios with a 10-iteration parameter sweep. Additionally, classifiers employed a leave-one-out cross-validation (LOOCV) which resulted in a 190-fold cross-validation model. A LOOCV approach was taken to avoid overfitting and improve the classifiers ability to

generalise to an independent data set. Furthermore, it serves as a useful performance measure and classifier selection criteria. [116, 117]

The top performing classifiers can be seen from Table 12 with corresponding 95 % confidence intervals shown in Table 13. The accuracy distribution for the top three classifiers can be seen in Figure 30 with the SDA being the most import performance metric. The Two-Class Locally Deep Support Vector Machine has the strongest accuracy, measured by the standard deviation of the classifier's accuracy. The SDA is a measure of the robustness and repeatability of classification for the classifier following the 190-fold LOOCV.

The analysis of the confusion matrix shown in Table 14 indicates that the top classifier incorrectly classified 17 of the 190 samples. Taking a closer look and connecting each of the samples to the corresponding participant through the consensus algorithm, leads to the creation of the final ADHD screening tool classifier as seen in the confusion matrix presented by Table 16. The sequential use of the consensus algorithm, proposed in section 3.3.5, can be seen in Table 15 where the classification of each segment for participant 39 was determined. It is clear from Table 16 that five of the 17 incorrectly classified samples belong to a single participant. The other samples did not influence the consensus algorithm to incorrectly classify additional participants. For the purposes of this study, false positive classifications are preferred over false negative classification, as the cost of a false positive is better than the lifetime of unnecessary challenges faced as a result of a false negative classification.

The *Delta* and *Eta* values, which were highlighted earlier in this section, will now be evaluated. According to the definition, *Delta* is the proportion difference in the event of interest between the first and the second measurement standards. *Eta* represents the total population proportion of disagreement between the two measurement standards. As determined from the confusion matrix in Table 16, the ADHD screening tool achieved a *Delta* and *Eta* value of 0.025. Table 19 below compares the *Delta* and *Eta* values of the ADHD screening tool with those achieved using McNemar's Test for sample size calculation with a power goal of 0.9. Compared to the proposed sample size, the ADHD screening tool achieved *Delta* and *Eta* values which were six and eight times smaller, respectively. Compared to the actual sample size, the ADHD screening tool achieved *Delta* and *Eta* values which were 7.6 and eight times smaller, respectively. These findings are statistically significant and represent a sample size of 272 participants for the same power goal (see Figure B.3 in Appendix B).

Table 19: Comparison of Delta and Eta values for a power goal of 0.9.

Source	Reference	Sample Size	*Delta* (δ)	*Eta* (η)
McNemar's Test: Proposed Sample Size	Table 5	76	0.15	0.2
McNemar's Test: Actual Sample Size	Table 7	39	0.19	0.2
ADHD Screening Tool	Table 16	39	0.025	0.025

In contrast to findings from the literature, the ADHD screening tool found the game segment with visual distractors only to be a poor discriminator. However, both the auditory, as well as a combination of auditory and visual distractors were strong discriminators (see segment 3 in Table 15 and Figure 19). Additionally, as seen in Appendix E and according to findings from D'souza et al., Li et al. and Altun et al., the statistical features constituted some of the strongest correlated game features identified by Pearson's r coefficient. [76–78]. However, the strongest correlated feature was related to participant movement during gameplay. Movement features were included based on findings from Zheng et al. investigating the significance of movement in ADHD participants during a Go/No-Go task [108].

The analysis of the existing tools as well as the results from the ADHD screening tool are presented in Table 20 below. Performance measures for the AULA could not be validated from the literature.

Table 20: Comparison of classification performance with existing technology.

References	Tool	Sample	Performance Measures (%)
Berger et al. [35]	MOXO	798 children (Mean Age = 9.27 yrs.): 339 ADHD 459 non-ADHD	Sensitivity = 86.5* Specificity = 86.2*
Schatz et al. [118]	T.O.VA	48 children: (Mean Age = 11.1 yrs.): 28 ADHD 20 non-ADHD	Sensitivity = 85.7 Specificity = 70
Kim et al. [119]	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.): 31ADHD 8 non-ADHD	Sensitivity = 100 Specificity = 87.5

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

6.4 Game Design

The ADHD screening tool has been designed and developed to include a feature set and a machine learning algorithm that serves as a skeleton for any game layout

or visual overlay within certain limits. This was implemented to enable the possibility for dynamically changing game segments whilst still providing classification accuracy. Therefore, each game segment can be an interchangeable mini-game used in the overall classification of a participant.

Random placement of game assets was implemented to establish a framework according to which future games can be developed. In principle, the seven segments constitute seven mini-games with the same feature set but different values for each of the features (e.g. the number of obstacles, gems and distractions). By retaining elements such as the number of segment tiles and game logic, any segment can be replaced by a different visual overlay (e.g. a car on a racetrack at night). If the segment measures the same features, that segment could potentially be used to classify participants according to the same underlying skeletal MLC that was trained on the five segments designed and developed in this study. Due to the classifier's ability to generalise well across the five gameplay segments included during the training of the classifier, there is scope to affirm the skeletal feature and game structure suggested. The probability score presented in Table 15, shows that the classifier generalises well across all gameplay segments, save one.

7 CONCLUSION

7.1 Overview

This project aimed to research and develop a low-cost, portable, gaming device capable of screening participants with potential ADHD-I. The prototype consisted of a Nvidia Shield K1 tablet and a web-based MLC. The tablet was used to run a game interface which was designed for ADHD-I and to record game data features during participant gameplay. Ethical clearance was acquired from the HREC of Stellenbosch University to conduct the clinical study at Cape Gate Therapy Centre. The test population included 39 participants, 31 of which were diagnosed as ADHD and eight as neurotypical. The ADHD state of each participant was validated by a clinical psychologist and paediatrician prior to proceeding with the clinical study. Clinical data from participants were used to train multiple MLCs to discriminate between ADHD and non-ADHD participants.

Following the identification of Pearson r correlated features, feature reduction, hyperparameter tuning and LOOCV, the best performing classifier was used in conjunction with the consensus algorithm to determine the classification of participants. The combination of the tablet-based game, top performing classifier and the consensus algorithm collectively represents the ADHD screening tool. The clinical study produced an ADHD screening tool classification accuracy, sensitivity and specificity of 97.4 %, 100 % and 87.5 %, respectively. The AUC can be considered above the 0.942 which was achieved by the ADHD screening tool without the addition of the consensus algorithm.

7.2 Objectives

The project objectives are summarised in Table 21 below. All the stated objectives were completed.

Table 21: Project objectives summary.

Objective	Description	Status
1. Develop a game to capture ADHD inattentive subtype features	The developed game evaluated and included relevant findings from the DSM-V criteria as well as relevant findings from the literature.	Complete
2. The screening tool should be portable, accessible and easy to administer	The selected device is operated like any android based tablet on the market. The developed game is easy to administer as the game tutorial visually explains all the controls.	Complete

3. The cost of the screening tool should be relatively cheap	Compared to the cost of existing tools, the tool developed in this study bears a greater and continual return for investment.	Complete
4. The screening tool should contain a wireless data sharing capability to safely store participant data online	The tablet is equipped with Wi-Fi technology and automatically uploads gameplay data to a secure database after gameplay has completed.	Complete
5. The screening tool should have the capability of reporting screening feedback.	The screening tool can easily provide classification feedback as was demonstrated by the web-API usage. This can easily be built into the game.	Complete
6. The development of machine learning algorithms to classify a participant as either neurotypical or having ADHD inattentive subtype.	Multiple machine learning classifiers were trained, the top performing of which had excellent performance. Coupled with the consensus algorithm the performance pf the classifier was increased.	Complete

7.3 Limitations

Due to the fact that this study depended on children between the ages of six and 12, one of the greatest limitations was sourcing enough participants. Although over 400 invitations were distributed to six schools, the poor response was unexpected. The fact that the study required human participants, the ethical approval process was time consuming. Over and above acquiring ethical approval from the HREC, additional permission had to be sought from the Western Cape Education Department to send invitations for participation to parents via school children of Public Schools. Much time was spent waiting for the return of the invitation return slips. The clinical study was also highly dependent on the availability of the test site, which could only be used for testing on Saturdays, and was not always available. The limited number of participants prevented this study from investigating the effectiveness of individual classifiers for each segment and integrating them with the consensus algorithm. The structure of the MEng Research degree also resulted in less time being available to work on this research study.

7.4 Lessons Learned

1. One of the most important lessons learned during this study is that the recruitment of human participants, especially children, can take a lot of time. It is best to start early and target as many sources as possible;
2. Things don't always go according to plan, plan accordingly;

3. A proactive attitude achieves more than choosing to be politely and remaining passive;
4. Children are impressionable, treat them kindly and be the example they might not have at home.

7.5 Future Recommendations

7.5.1 Game Design

Future studies can test the effect of different visual overlays with the same feature set incorporated in this study. A larger segment set with greater variation in the number of assets could also be investigated to increase the generalisation ability of the classifier. A game could specifically be designed and developed to target the ADHD-H and a comparison could be drawn with the classification strength of the tool developed in this study. A future study could investigate the implementation of a locally hosted classifier on the tablet device for classifications in rural or remote areas with no internet connection.

7.5.2 Machine learning

Due to the limited number of samples in this study, a good performing classifier could not be trained for each segment individually. Future studies can investigate the effectiveness of using individual classifiers for each segment and compare that with the single classifier and consensus algorithm.

7.5.3 The Consensus Algorithm

Due to the imbalance of participants groups, the consensus algorithm used in this study cannot be coupled with a confidence rating as desired. Although the consensus algorithm in Equation 2 reflects the true output of the ADHD screening tool classifier, the consensus confidence C_c (Equation 3) can still be refined to smaller confidence intervals. The following algorithm consensus classification C_f is a refined version of Equation 2. However, this algorithm is only valid for a balanced training sample of normal and abnormal participants.

$$C_f = \sum_{i=0}^{n} c_i \, p_i \,, \tag{10}$$

$$C_c = \frac{|C_f|}{n} \times 100, \tag{11}$$

where i represents the number of the segment, c represents the classification of the segment $(Normal = 1, Abnormal = -1)$, p represents the confidence score of the segment classification, n represents the total number of segments included in

the analysis, and C_f represents the final consensus classification. For participant classification, the following is applicable:

$$if\ C_f > 0; Classification = Abnormal$$

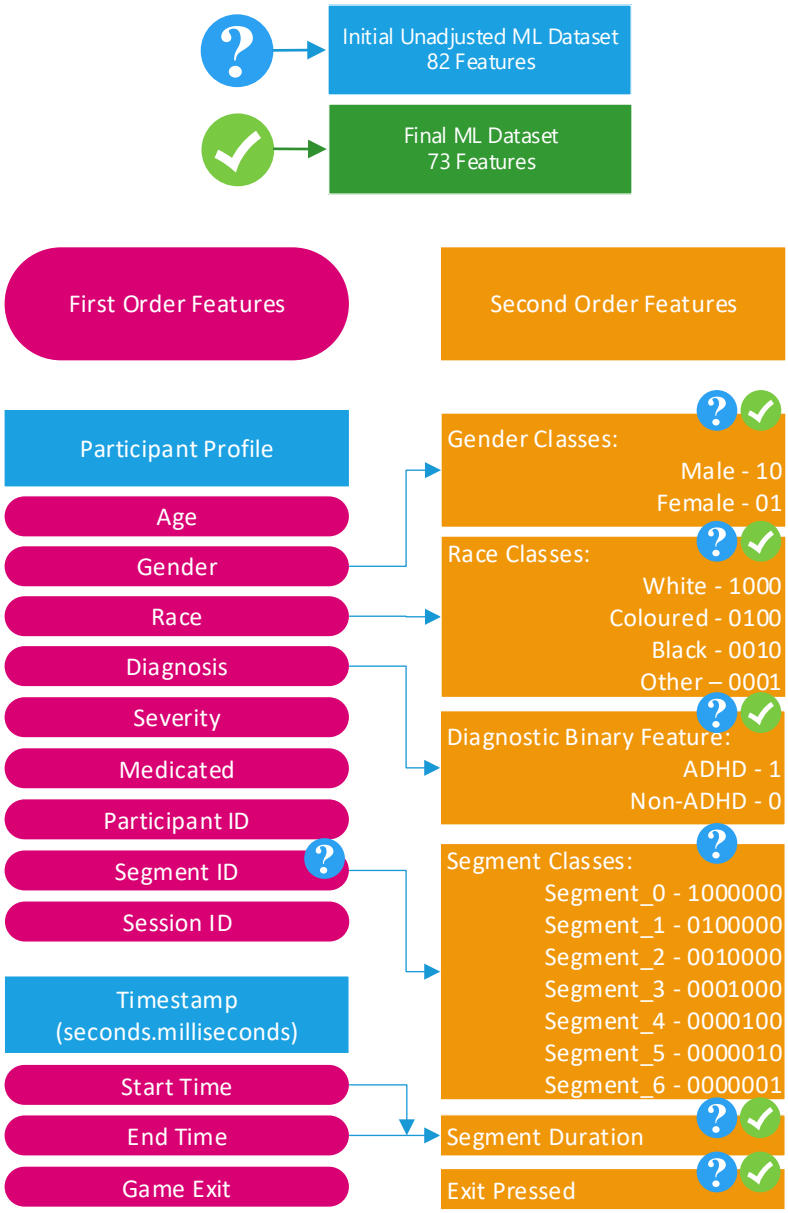
$$if\ C_f < 0; Classification = Normal$$

7.6 Conclusion

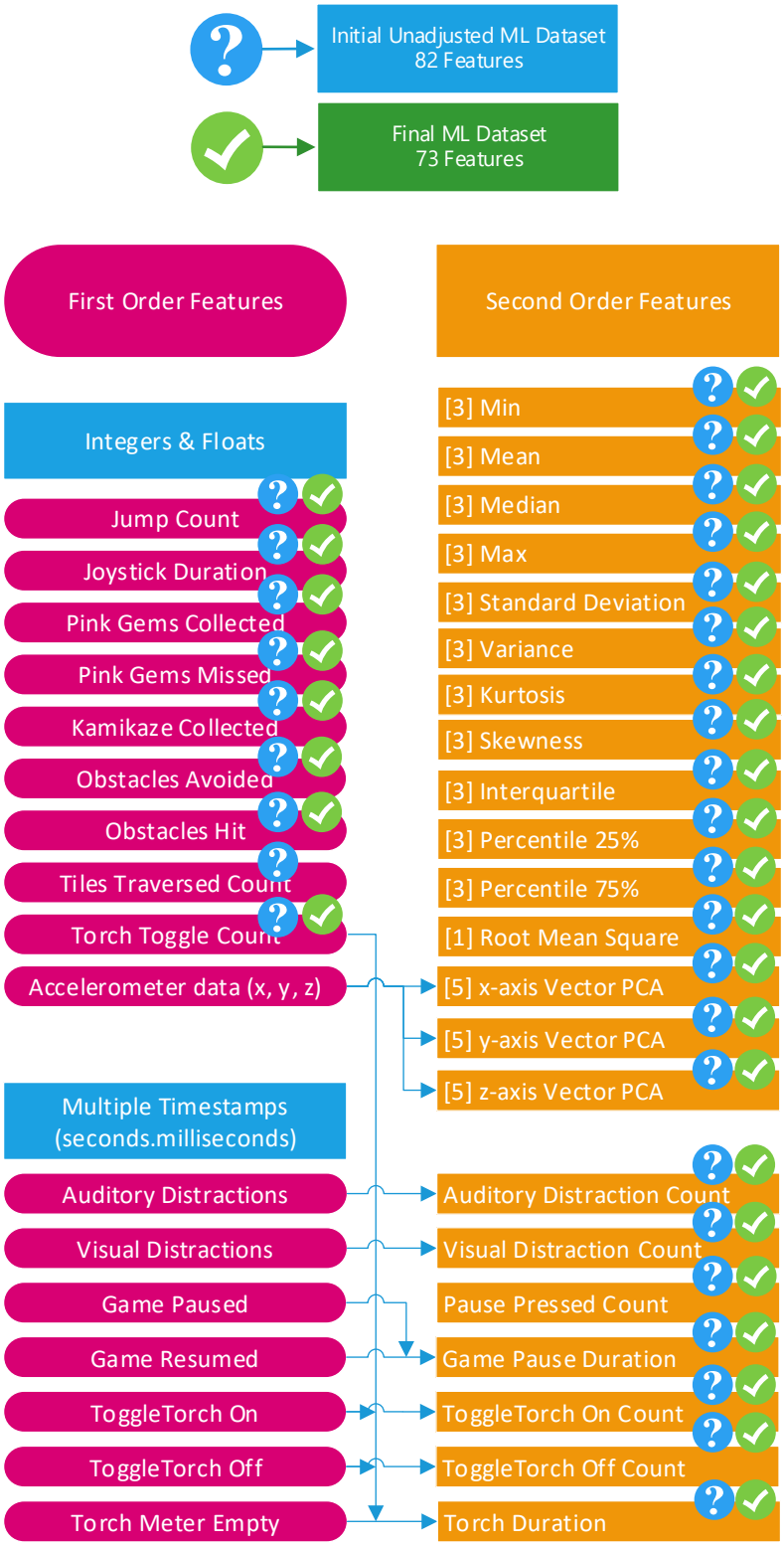
A portable ADHD screening tool was developed which is capable of recording ADHD-I features through a tablet-based gaming interface. The developed tool consists of a single tablet device which is cheaper than products currently on the market and is capable of working in any location with a wireless internet connection. The screening tool was successfully tested in a clinical study where results exceeded that of existing technology. The study successfully met the aims and objectives set out in Chapter One. The null hypothesis was not rejected, providing sufficient evidence to support further development and roll-out of the tool to cares, parents and teachers for ADHD-I screening. The use of a portable ADHD screening tool to detect early onset of ADHD symptoms has the potential to transform ADHD care and change the lives of millions.

APPENDIX A: ADHD SCREENING TOOL FEATURE MATRIX

A.1 First and second-order feature matrix (Part 1).

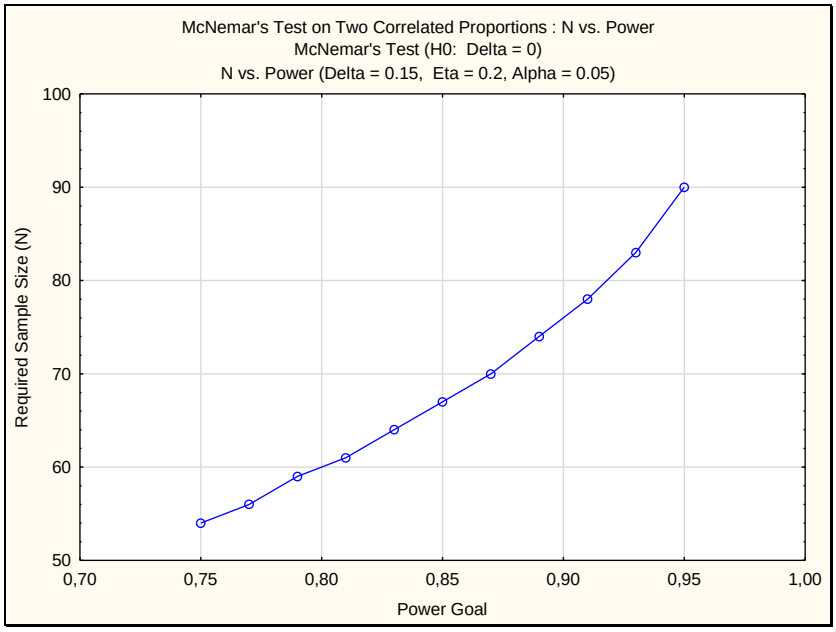


A.2 First and second-order feature matrix (Part 2).

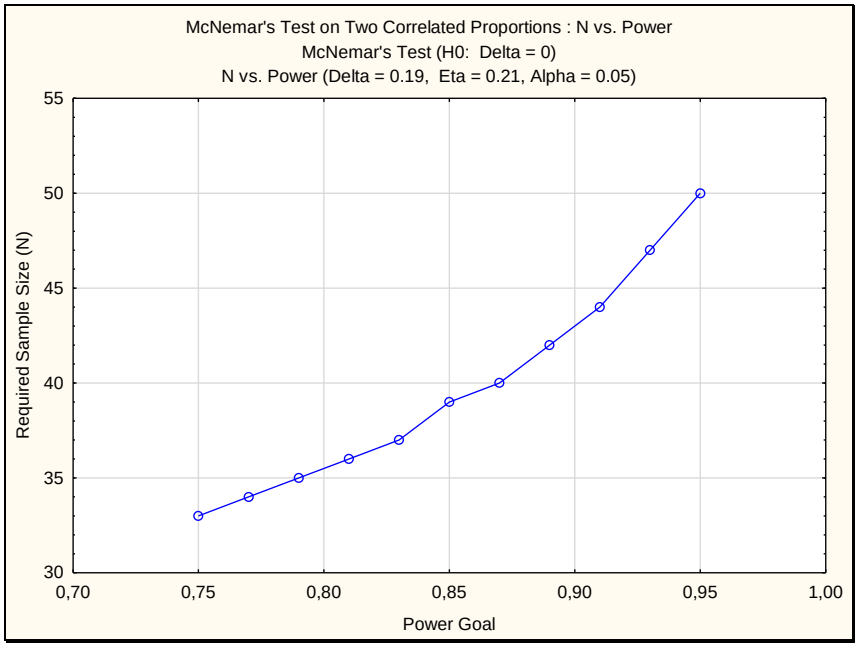


APPENDIX B: CLINICAL STUDY

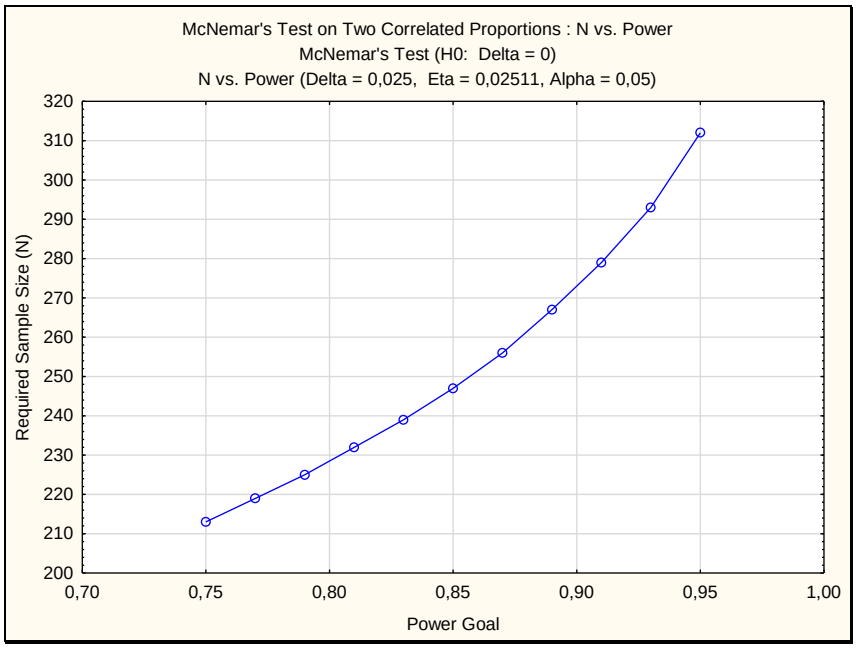
B.1 McNemar's Test for the desired sample size for the desired power goal.



B.2 McNemar's Test for the included sample size.



B.3 McNemar's Test for the projected sample size.



B.4 Ethical clearance documentation.

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Progress Report Approval Letter

05/07/2018

Project ID: 4680

Ethics Reference #: M17/05/019

Title: PANDAS: Paediatric Attention-Deficit/Hyperactivity Disorder Application Software

Dear Mr Hervé Mwamba,

Your request for extension/annual renewal of ethics approval dated 03/07/2018 15:42 refers.

The Health Research Ethics Committee reviewed and approved the annual progress report you submitted through an expedited review process.

The approval of this project is extended for a further year.

Approval date: 05 July 2018

Expiry date: 04 July 2019

Kindly be reminded to submit progress reports two (2) months before expiry date.

Where to submit any documentation

Kindly note that the HREC uses an electronic ethics review management system, *Infonetica*, to manage ethics applications and ethics review process. To submit any documentation to HREC, please click on the following link: https://applyethics.sun.ac.za.

Please remember to use your **Project ID [4680]** and Ethics Reference Number on any documents or correspondence with the HREC concerning your research protocol.

National Health Research Ethics Council (NHREC) Registration Numbers: REC-130408-012 for HREC1 and REC-230208-010 for HREC2

Federal Wide Assurance Number: 00001372

Institutional Review Board (IRB) Number: IRB0005240 for HREC1

Institutional Review Board (IRB) Number: IRB0005239 for HREC2

The Health Research Ethics Committee complies with the SA National Health Act No. 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 Part 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki and the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles, Structures and Processes 2015 (Department of Health).

Yours sincerely,

Francis Masiye,

Coordinator,

Health Research Ethics Committee 2.

APPENDIX C: SCHOOL INVITATION

DATE

BERG Biomedical Engineering Research Group

Department of Mechanical and Mechatronic Engineering, Stellenbosch University

Dear Parent(s)

This is an invitation for your child(ren) between the ages of 6 and 12 to help us with research regarding concentration levels. Desired participants are **normal developmental children** - children who score 6 or less for items 1-9 and items 10-18 on the attached SWAN* Rating Scale.

We are currently developing a tablet-based game with integrated machine learning, to assist in the diagnosis of ADHD. Children will simply be asked to play this game for ±15 min. This research has the potential to help families with early identification of ADHD symptoms, treatment and effective management of medication.

Testing takes place on Saturdays (date will be confirmed) at Cape Gate Therapy Centre (51 Tanner St, Cape Gate, Cape Town – map below) under the supervision of **Rose-Hannah Brown** (021 987 0111).

MAP FROM SCHOOL GOES HERE

Should you have any queries, please call **Romano Swarts** at 084 041 1209. Your participation is dearly appreciated. We look forward to hearing from you.

Kind regards,

Rose-Hannah Brown
Counselling Psychologist: Cape Gate Therapy Centre
E-mail: admin@emotionalwellbeing.co.za

BA, HONS (Couns. Psych.), MA (Couns. Psych.) (Stell)

Romano Swarts
Masters Student: Biomedical Engineering Research Group
E-mail: rswarts@sun.ac.za

BEng (Mechatronic) (Stell)

This research project has ethical approval from the Health Research Ethics Committee (**HREC**) of Stellenbosch University.

Ethics Reference number: M17/05/019
Project Title: PANDAS
(Paediatric Attention-Deficit/Hyperactivity Disorder Application Software)

Ethics Contact Person
Mrs Ashleen Fortuin
Administrator: Health Research Ethics
Tel: (021) 938 9819

BERG Biomedical Engineering Research Group
Department of Mechanical and Mechatronic Engineering, Stellenbosch University

	No	Yes
Please indicate whether you're willing to participate in the study? ☑	☐	☐

NB: If **Yes**, more information will be provided to you via your contact details below. Additionally, the **15 min** testing timeslot will be arranged.

Parent Information

Name & Surname: _____

Contact number: _____

E-mail: _____

Signature: _____ Date: _____

Please only return this page (2) and the rating scale

The SWAN* Rating Scale for ADHD

Child's name: _____ Gender: _____ Age: _____
Completed by: _____ (circle one) Parent Teacher Physician
Date Completed: _____

For each item, check the column that best describes this child over the past six months.

	Not at all	Just a little	Quite a bit	Very much
1. Gives close attention to detail and avoids careless mistakes	____	____	____	____
2. Sustains attention on tasks or play activities	____	____	____	____
3. Listens when spoken to directly	____	____	____	____
4. Follows through on instructions and finishes school work and chores	____	____	____	____
5. Organizes tasks and activities	____	____	____	____
6. Engages in tasks that require sustained mental effort	____	____	____	____
7. Keeps track of things necessary for activities (doesn't lose them)	____	____	____	____
8. Ignores extraneous stimuli	____	____	____	____
9. Remembers daily activities	____	____	____	____
10. Sits still (controls movement of hands or feet or controls squirming)	____	____	____	____
11. Stays seated (when required by class rules or social conventions)	____	____	____	____
12. Modulates motor activity (inhibits inappropriate running or climbing)	____	____	____	____
13. Plays quietly (keeps noise level reasonable)	____	____	____	____
14. Settles down and rests (controls constant activity)	____	____	____	____
15. Modulates verbal activity (controls excessive talking)	____	____	____	____
16. Reflects on questions (controls blurting out answers)	____	____	____	____
17. Awaits turn (stands in line and takes turns)	____	____	____	____
18. Enters into conversation and games without interrupting or intruding	____	____	____	____

Scoring Section: For each question, place a 1 next to the question number below if the response was "not at all" or "just a little" and a 0 if the response was "quite a bit" or "very much".

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____

Sum #'s 1-9 _____ #'s 10-18 _____

Results:

1. If the sum of 1-9 is 6 or greater, the child is likely ADHD- Inattentive type. Consider mental health evaluation.
2. If the sum of 10-18 is 6 or greater, the child is likely ADHD-Hyperactive/Impulsive type. Consider mental health evaluation.
3. If both the sums of 1-9 and 10-18 are 6 or greater, the child is likely ADHD-Combined type. Consider mental health evaluation.
4. If neither sums are 6 or greater, the child likely does not have ADHD or the symptoms are being controlled with current treatment.

*Adapted from James M. Swanson, Ph.D., University of California, Irvine

APPENDIX D: EXTRA RESULTS

D.1 Segment 2 individual classifier results

	Performance Metrics			
	%			
Two-Class Classifier	*Accuracy*	*Sensitivity*	*Specificity*	*AUC*
Averaged Perceptron	76.9	93.5	12.5	0.379
Support Vector Machine	82.1	96.8	25	0.609
Decision Jungle	76.9	96.8	0	0.240
Boosted Decision Tree	79.5	96.8	12.5	0.476
Decision Forest	76.9	96.8	0	0.198
Neural Network	82.1	100	12.5	0.383
Logistic Regression	82.1	100	12.5	0.349
Bayes Point Machine	82.1	96.8	25	0.407
LDSVM	74.4	90.3	12.5	0.395

D.2 Segment 3 individual classifier results

	Performance Metrics			
	%			
Two-Class Classifier	*Accuracy*	*Sensitivity*	*Specificity*	*AUC*
Averaged Perceptron	74.4	93.5	0	0.319
Support Vector Machine	66.7	83.9	0	0.315
Decision Jungle	74.4	93.5	0	0.266
Boosted Decision Tree	84.6	96.8	37.5	0.714
Decision Forest	76.9	96.8	0	0.355
Neural Network	79.5	100	0	0.101
Logistic Regression	74.4	93.5	0	0.220
Bayes Point Machine	74.4	93.5	0	0.133
LDSVM	71.8	90.3	0	0.200

D.3 Segment 4 individual classifier results

	Performance Metrics			
	%			
Two-Class Classifier	*Accuracy*	*Sensitivity*	*Specificity*	*AUC*
Averaged Perceptron	79.5	96.8	12.5	0.407
Support Vector Machine	76.9	96.8	0	0.331
Decision Jungle	76.9	96.8	0	0.258
Boosted Decision Tree	76.9	96.8	0	0.222
Decision Forest	76.9	96.8	0	0.290
Neural Network	74.4	93.5	0	0.512
Logistic Regression	74.4	93.5	0	0.460

Bayes Point Machine	74.4	93.5	0	0.472
LDSVM	69.2	87.1	0	0.504

APPENDIX E: FEATURE SETS

Feature set iteration with Pearson's Correlation Coefficient for each feature.

Stepwise Feature Sets		
Iteration 1 (82 Features)	**Pearson's Correlation**	**Iteration 2 (73 Features)**
z_pc_3	0,216	z_pc_3
ObstaclesHit	0,186	ObstaclesHit
TotalRunTime	0,184	TotalRunTime
ObstaclesAvoided	0,182	ObstaclesAvoided
Age	0,155	Age
x_min	0,141	x_min
z_min	0,139	z_min
JoystickCount	0,138	JoystickCount
ToggleTorchOnCount	0,130	ToggleTorchOnCount
ToggleTorchOffCount	0,121	ToggleTorchOffCount
x_stdD	0,121	x_stdD
x_pc_3	0,119	x_pc_3
y_pc_3	0,113	y_pc_3
x_skew	0,110	x_skew
MeterEmptyCount	0,109	MeterEmptyCount
x_variance	0,107	x_variance
z_max	0,105	z_max
z_stdD	0,104	z_stdD
y_pc_4	0,099	y_pc_4
z_kurtosis	0,099	z_kurtosis
rms_all	0,098	rms_all
z_variance	0,087	z_variance
z_pc_4	0,083	z_pc_4
x_pc_5	0,082	x_pc_5
y_percentile75	0,082	y_percentile75
z_pc_2	0,081	z_pc_2
y_interquartile	0,076	y_interquartile
y_min	0,070	y_min
y_percentile25	0,069	y_percentile25
x_mean	0,069	x_mean
TokenCollected	0,068	TokenCollected
y_pc_1	0,064	y_pc_1
TokenMissed	0,063	TokenMissed
x_max	0,059	x_max
x_pc_1	0,059	x_pc_1

x_pc_2	0,054	x_pc_2
PausePressedCount	0,053	PausePressedCount
x_kurtosis	0,051	x_kurtosis
y_median	0,051	y_median
y_pc_2	0,050	y_pc_2
x_percentile75	0,047	x_percentile75
y_max	0,047	y_max
y_mean	0,047	y_mean
x_interquartile	0,045	x_interquartile
PauseDuration	0,045	PauseDuration
x_pc_4	0,043	x_pc_4
x_percentile25	0,042	x_percentile25
z_skew	0,041	z_skew
y_skew	0,041	y_skew
z_pc_5	0,040	z_pc_5
JumpCount	0,038	JumpCount
z_median	0,032	z_median
y_kurtosis	0,030	y_kurtosis
y_stdD	0,028	y_stdD
y_variance	0,026	y_variance
y_pc_5	0,024	y_pc_5
z_percentile75	0,024	z_percentile75
z_interquartile	0,023	z_interquartile
x_median	0,023	x_median
z_percentile25	0,023	z_percentile25
Gender_Male	0,016	Gender_Male
Gender_Female	0,016	Gender_Female
z_mean	0,015	z_mean
Race_Coloured	0,010	Race_Coloured
Race_White	0,010	Race_White
z_pc_1	0,009	z_pc_1
TorchDuration	0,005	TorchDuration
VD_Count	0,000	VD_Count
AD_Count	0,000	AD_Count
Segment_0	0,000	
Segment_6	0,000	
Segment_2	0,000	
Segment_3	0,000	
Segment_4	0,000	
SegmentId	-	
Segment_1	-	

Segment_5	-	
Race_Black	-	Race_Black
Race_Other	-	Race_Other
TilesTraversed	-	
KamikazeCollected	-	KamikazeCollected
ExitPressed	-	ExitPressed

8 REFERENCES

[1] G. V Polanczyk, E. G. Willcutt, G. A. Salum, C. Kieling, and L. A. Rohde, “ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis,” *Int. J. Epidemiol.*, vol. 43, no. 2, pp. 434–442, Apr. 2014.

[2] S. I. Sharp, A. Mcquillin, H. M. D. Gurling, and H. M. D. Gurling, “Genetics of attention-deficit hyperactivity disorder (ADHD),” *Neuropharmacology*, vol. 57, pp. 590–600, 2009.

[3] U. Preuss *et al.*, “Study design, baseline patient characteristics and intervention in a cross-cultural framework: Results from the ADORE study,” *Eur. Child Adolesc. Psychiatry*, vol. 15, no. Suppl 1, pp. I4–I14, 2006.

[4] U. Díaz-Orueta, G.-L. Cristina, C.-E. Nerea, R. Sánchez-Carpintero, G. Climent, and N. Juan, “AULA virtual reality test as an attention measure: Convergent validity with Conners’ Continuous Performance Test,” *Child Neuropsychol.*, vol. 20, no. 3, pp. 328–342, 2017.

[5] C. D. I. LLC, “Know Your Child Better By Learning The Ages; Stages Of Child Development,” 2015. [Online]. Available: https://childdevelopmentinfo.com/ages-stages/. [Accessed: 22-Jul-2016].

[6] R. C. Kessler *et al.*, “The Prevalence and Correlates of Adult ADHD in the United States: Results From the National Comorbidity Survey Replication,” *Am J Psychiatry*, vol. 163, pp. 716–723, 2006.

[7] R. Thomas, S. Sanders, J. Doust, E. Beller, and P. Glasziou, “Prevalence of Attention-Deficit/ Hyperactivity Disorder: A Systematic Review and Meta-analysis.”

[8] M. D. Rosenberg *et al.*, “A neuromarker of sustained attention from whole-brain functional connectivity,” *Nat. Neurosci.*, vol. 19, no. 1, pp. 165–171, Jan. 2016.

[9] R. Rettner, “Autism’s Rise: Researchers Look at Why Cases Are Increasing,” 2014. [Online]. Available: http://www.livescience.com/48586-autism-prevalence-increase-reporting.html. [Accessed: 23-Jul-2016].

[10] C. and A. with A.-D. D. (CHADD), “CHADD The National Resource on ADHD.” [Online]. Available: http://www.chadd.org/Understanding-ADHD/About-ADHD/Data-and-Statistics/General-Prevalence.aspx. [Accessed: 26-Jul-2016].

[11] L. Burrows, “ADHD pandemic causes researchers to fidget | BrandeisNOW,” 2014. [Online]. Available: http://www.brandeis.edu/now/2014/november/peter-conrad-adhd-

study.html. [Accessed: 26-Jul-2016].

[12] C. T. Gualtieri and L. G. Johnson, “ADHD: Is Objective Diagnosis Possible?,” *Psychiatry (Edgmont).*, vol. 2, no. 11, pp. 44–53, Nov. 2005.

[13] M. Perold, C. Louw, and S. Kleynhans, “Primary school teachers’ knowledge and misperceptions of attention deficit hyperactivity disorder (ADHD).,” *South African J. Educ.*, vol. 30, no. 2002, pp. 457–473, 2010.

[14] L. S. Goldman, M. Genel, R. J. Bezman, and P. J. Slanetz, “Diagnosis and treatment of attention-deficit/hyperactivity disorder in children and adolescents. Council on Scientific Affairs, American Medical Association.,” *JAMA*, vol. 279, no. 14, pp. 1100–7, Apr. 1998.

[15] A. G. Harrison, M. J. Edwards, and K. C. H. Parker, “Identifying students faking ADHD: Preliminary findings and strategies for detection,” *Arch. Clin. Neuropsychol.*, vol. 22, pp. 577–588, 2007.

[16] M. Goetz, M. Vesela, and R. Ptacek, “Notes on the Role of the Cerebellum in ADHD,” *Austin J Psychiatry Behav Sci. Austin J Psychiatry Behav Sci*, vol. 1, no. 1, pp. 1013–3, 2014.

[17] A. F. T. Arnsten, “The Emerging Neurobiology of Attention Deficit Hyperactivity Disorder: The Key Role of the Prefrontal Association Cortex.,” *J. Pediatr.*, vol. 154, no. 5, p. I-S43, May 2009.

[18] D. Purper-Ouakil, N. Ramoz, A.-M. Lepagnol-Bestel, P. Gorwood, and M. Simonneau, “Neurobiology of Attention Deficit/Hyperactivity Disorder,” 2011.

[19] “What is ADHD? (article) | Mental health | Khan Academy.” [Online]. Available: https://www.khanacademy.org/science/health-and-medicine/mental-health/neurodevelopmental-disorders/a/what-is-adhd. [Accessed: 22-Oct-2018].

[20] American Psychiatric Association, “Diagnostic and statistical manual of mental disorders: DSM-5 - Fifth Edition,” 5th ed., Arlington, VA: American Psychiatric Publishing, 2013, pp. 59–66.

[21] N. Grizenko, M. Paci, and R. Joober, “Is the Inattentive Subtype of ADHD Different From the Combined/Hyperactive Subtype?”

[22] S. V. FARAONE, J. BIEDERMAN, W. WEBER, and R. L. RUSSELL, “Psychiatric, Neuropsychological, and Psychosocial Features of DSM-IV Subtypes of Attention-Deficit/Hyperactivity Disorder: Results From a Clinically Referred Sample,” *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 37, no. 2, pp. 185–193, Feb. 1998.

[23] S. Park *et al.*, “Differential perinatal risk factors in children with attention-deficit/hyperactivity disorder by subtype,” *Psychiatry Res.*, vol. 219, no. 3, pp. 609–616, Nov. 2014.

[24] J. Lemiere, H. Wouters, C. Sterken, L. Lagae, E. Sonuga-Barke, and M. Danckaerts, “Are children with ADHD predominantly inattentive and combined subtypes different in terms of aspects of everyday attention?,” *Eur. Child Adolesc. Psychiatry*, vol. 19, no. 8, pp. 679–685, Aug. 2010.

[25] G. Polanczyk, M. Silva de Lima, B. Lessa Horta, J. Biederman, and L. Augusto Rohde, “The Worldwide Prevalence of ADHD: A Systematic Review and Metaregression Analysis,” *Am J Psychiatry*, vol. 1646, 2007.

[26] American Academy of Pediatrics, “ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/ Hyperactivity Disorder in Children and Adolescents.”

[27] H. C. Steinhausen *et al.*, “Co-existing psychiatric problems in ADHD in the ADORE cohort,” *Eur. Child Adolesc. Psychiatry*, vol. 15, no. Suppl 1, pp. I25–I29, 2006.

[28] T. L. Psychiatry, “ADHD in children and young people: prevalence, care pathways, and service provision,” *The Lancet Psychiatry*, Oct. 2017.

[29] J. Fayyad *et al.*, “Cross–national prevalence and correlates of adult attention–deficit hyperactivity disorder,” *Br. J. Psychiatry*, vol. 190, no. 5, pp. 402–409, 2007.

[30] K. Langley *et al.*, “Association of the Dopamine D4 Receptor Gene 7-Repeat Allele With Neuropsychological Test Performance of Children With ADHD,” *Am. J. Psychiatry*, vol. 161, no. 1, pp. 133–138, Jan. 2004.

[31] Shire, “ADHD Diagnosis and Screening in Adults Clinical assessment using DSM-5 TM * criteria and the ASRS checklist,” pp. 1–8.

[32] C. for D. C. and P. CDC, “Symptoms and Diagnosis | ADHD | NCBDDD | CDC,” 2016. [Online]. Available: https://www.cdc.gov/ncbddd/adhd/diagnosis.html. [Accessed: 29-Apr-2017].

[33] National Institute for Health and Care Excellence, “Attention deficit hyperactivity disorder in children and young people - NICE Pathways,” *Clinical guideline CG72*, 2016. [Online]. Available: https://pathways.nice.org.uk/pathways/attention-deficit-hyperactivity-disorder#path=view%3A/pathways/attention-deficit-hyperactivity-disorder/attention-deficit-hyperactivity-disorder-in-children-and-young-people.xml&content=view-node%3Anodes-diagnosing-ad. [Accessed: 16-Nov-2017].

[34] National Institute for Health and Care Excellence, “Attention deficit hyperactivity disorder overview - NICE Pathways,” *Clinical guideline CG72*, 2016. [Online]. Available: https://pathways.nice.org.uk/pathways/attention-deficit-hyperactivity-disorder. [Accessed: 05-Dec-2016].

[35] I. Berger, O. Slobodin, and H. Cassuto, “Usefulness and Validity of

Continuous Performance Tests in the Diagnosis of Attention-Deficit Hyperactivity Disorder Children," *Arch. Clin. Neuropsychol.*, vol. 32, pp. 81–93, 2017.

[36] C. for D. C. and P. CDC, "Other Concerns and Conditions | ADHD | NCBDDD | CDC," 2016. [Online]. Available: https://www.cdc.gov/ncbddd/adhd/conditions.html. [Accessed: 29-Apr-2017].

[37] P. S. Jensen *et al.*, "ADHD comorbidity findings from the MTA study: comparing comorbid subgroups.," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 40, no. 2, pp. 147–58, 2001.

[38] A. Angold, E. J. Costello, and A. Erkanli, "Comorbidity," vol. 40, no. 1, pp. 57–87, 1999.

[39] B. Kadesjo and C. Gillberg, "The Comorbidity of ADHD in the General Population of Swedish School-age Children," *J. Child Psychol. Psychiatry*, vol. 42, no. 4, pp. 487–492, 2001.

[40] A. J. Orinstein and M. C. Stevens, "Brain activity in predominantly-inattentive subtype attention-deficit/ hyperactivity disorder during an auditory oddball attention task," *Psychiatry Res. Neuroimaging*, vol. 223, pp. 121–128, 2014.

[41] S. J. Kooij *et al.*, "European consensus statement on diagnosis and treatment of adult ADHD: The European Network Adult ADHD," *BMC Psychiatry*, vol. 10, no. 67, 2010.

[42] S. J. Ralston, M. J. M. Lorenzo, and M. J. M. Lorenzo, "ADORE – Attention-Deficit Hyperactivity Disorder Observational Research in Europe Background," *Eur Child Adolesc Psychiatry*, vol. 42, 2004.

[43] M. G. Ph.D. *et al.*, "The Texas Children's Medication Algorithm Project: Report of the Texas Consensus Conference Panel on Medication Treatment of Childhood Attention-Deficit/Hyperactivity Disorder. Part I. Attention-Deficit/Hyperactivity Disorder.," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 39, no. 7, pp. 908–919, 2000.

[44] H. Abikoff *et al.*, "Sequential pharmacotherapy for children with comorbid attention-deficit/hyperactivity and anxiety disorders.," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 44, no. 5, pp. 418–27, May 2005.

[45] World Health Organization, "The ICD-10 Classification of Mental and Behavioural Disorders Clinical descriptions and diagnostic guidelines," pp. 1–267.

[46] G. TRIPP, S. L. LUK, E. A. SCHAUGHENCY, and R. SINGH, "DSM-IV and ICD-10: A Comparison of the Correlates of ADHD and Hyperkinetic Disorder," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 38, no. 2, pp. 156–164, Feb. 1999.

[47] M. K. Demaray, J. Elting, and K. Schaefer, “Assessment of attention-deficit/hyperactivity disorder (ADHD): A comparative evaluation of five, commonly used, published rating scales,” *Psychol. Sch.*, vol. 40, no. 4, pp. 341–361, 2003.

[48] L. M. Angello *et al.*, “Assessment of Attention-Deficit / Hyperactivity Disorder : An Evaluation of Six Published Rating Scales,” *Sch. Psycology Rev.*, vol. 32, no. 2, pp. 241–262, 2003.

[49] DefiniPoint, “Online ADHD Rating Scales - DefiniPoint.” [Online]. Available: https://www.attentionpoint.com/definipoint_adhd_rating_scales. [Accessed: 25-Jun-2017].

[50] S. P. Becker, J. M. Langberg, A. J. Vaughn, and J. N. Epstein, “Clinical Utility of the Vanderbilt ADHD Diagnostic Parent Rating Scale Comorbidity Screening Scales,” *J. Dev. Behav. Pediatr.*, vol. 33, no. 3, pp. 221–228, 2012.

[51] J. M. Swanson *et al.*, “Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales.,” *Int. J. Educ. Psychol. Assess.*, vol. 10, no. 1, pp. 51–70, Apr. 2012.

[52] M. Silverstein *et al.*, “Using Clinical Data to Predict Accurate ADHD Diagnoses Among Urban Children,” *Clin. Pediatr. (Phila).*, vol. 55, no. 4, pp. 326–332, 2016.

[53] J. N. Epstein, J. M. Langberg, P. K. Lichtenstein, B. a Mainwaring, C. P. Luzader, and L. J. Stark, “Community-wide intervention to improve the attention-deficit/hyperactivity disorder assessment and treatment practices of community physicians.,” *Pediatrics*, vol. 122, no. May 2014, pp. 19–27, 2008.

[54] B. R. COLLETT, J. L. OHAN, and K. M. MYERS, “Ten-Year Review of Rating Scales. V: Scales Assessing Attention-Deficit/Hyperactivity Disorder,” *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 42, no. 9, pp. 1015–1037, 2003.

[55] A. B. Arnett *et al.*, “The SWAN Captures Variance at the Negative and Positive Ends of the ADHD Symptom Dimension,” *J. Atten. Disord.*, vol. 17, no. 2, pp. 152–162.

[56] J.-A. Ramos-Quiroga *et al.*, “Genome-wide copy number variation analysis in adult attention-deficit and hyperactivity disorder,” *J. Psychiatr. Res.*, vol. 49, pp. 60–67, 2014.

[57] C. P. Jacob *et al.*, “Acetylcholine-metabolizing butyrylcholinesterase (BCHE) copy number and single nucleotide polymorphisms and their role in attention-deficit/hyperactivity syndrome,” *J. Psychiatr. Res.*, vol. 47, pp. 1902–1908, 2013.

[58] C. Brites, C. A. Salgado-Azoni, T. L. Ferreira, R. F. Lima, and S. M. Ciasca, “Development and applications of the SWAN rating scale for assessment of attention deficit hyperactivity disorder: a literature review.”

[59] J. M. Swanson, T. Wigal, and K. Lakes, “DSM-V and the future diagnosis of attention-deficit/hyperactivity disorder,” *Curr. Psychiatry Rep.*, vol. 11, no. 5, pp. 399–406, Oct. 2009.

[60] T. J. C. Polderman, E. M. Derks, J. J. Hudziak, F. C. Verhulst, D. Posthuma, and D. I. Boomsma, “Across the continuum of attention skills: A twin study of the SWAN ADHD rating scale,” *J. Child Psychol. Psychiatry Allied Discip.*, vol. 48, no. 11, pp. 1080–1087, 2007.

[61] D. A. Hay, K. S. Bennett, F. Levy, J. Sergeant, and J. Swanson, “A Twin Study of Attention-Deficit/Hyperactivity Disorder Dimensions Rated by the Strengths and Weaknesses of ADHD-Symptoms and Normal-Behavior (SWAN) Scale,” *Biol. Psychiatry*, vol. 61, no. 5, pp. 700–705, 2007.

[62] L. C. Hartlage and C. J. Long, “Development of Neuropsychology as a Professional Psychological Specialty: History, Training and Credentialing,” in *Handbook of Clinical Child Neuropsychology*, 3rd ed., C. Reynolds and E. Fletcher-Janzen, Eds. Boston, MA: Springer US, 2009, pp. 3–18.

[63] J. C. Reed and H. B. C. Reed, “The Halstead—Reitan Neuropsychological Battery,” Springer US, 1997, pp. 93–129.

[64] R. M. REITAN, “INVESTIGATION OF THE VALIDITY OF HALSTEAD’S MEASURES OF BIOLOGICAL INTELLIGENCE,” *Arch. Neurol. Psychiatry*, vol. 73, no. 1, p. 28, Jan. 1955.

[65] R. M. REITAN and L. A. DAVISON, *Clinical neuropsychology: current status and applications.* Washington, D.C.: Hemisphere Pub. Corp., 1974.

[66] N. L. Nussbaum and M. R. Bunner, “Halstead-Reitan Neuropsychological Test Batteries for Children,” in *Handbook of Clinical Child Neuropsychology*, 3rd ed., Boston, MA: Springer US, 2009, pp. 247–266.

[67] C. J. Golden, “The Nebraska Neuropsychological Children’s Battery,” in *Handbook of Clinical Child Neuropsychology*, Boston, MA: Springer US, 2009, pp. 267–280.

[68] P. D. Harvey, “Clinical applications of neuropsychological assessment.,” *Dialogues Clin. Neurosci.*, vol. 14, no. 1, pp. 91–9, Mar. 2012.

[69] H. . Lovejoy, D.W. and Oakes, “The behavioral health provider as a participant in the disability determination process: Evaluations, terminology, and systems.,” *Neuropsychol. Assess. Work. Inj.*, p. pp.321-343, 2012.

[70] H. J. Oakes and D. W. Lovejoy, “Independent Neuropsychological Evaluations,” *Psychol. Inj. Law*, vol. 6, no. 1, pp. 51–62, 2013.

[71] C. Schoechlin and R. R. Engel, “Neuropsychological performance in adult attention-deficit hyperactivity disorder: Meta-analysis of empirical data,” *Arch. Clin. Neuropsychol.*, vol. 20, pp. 727–744, 2005.

[72] P. S. JENSEN, D. MARTIN, and D. P. CANTWELL, “Comorbidity in ADHD: Implications for Research, Practice, and DSM-V,” *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 36, no. 8, pp. 1065–1079, 1997.

[73] I. Berger and H. Cassuto, “The effect of environmental distractors incorporation into a CPT on sustained attention and ADHD diagnosis among adolescents,” *J. Neurosci. Methods*, vol. 222, pp. 62–68, 2014.

[74] H. Cassuto, A. Ben-Simon, and I. Berger, “Using environmental distractors in the diagnosis of ADHD,” *Front. Hum. Neurosci.*, vol. 7, p. 805, 2013.

[75] I. Berger *et al.*, “Maturational delay in ADHD: evidence from CPT,” 2013.

[76] W. T. D’souza and R. Kavitha, “Human Activity Recognition Using Accelerometer and Gyroscope Sensors,” *2016 IEEE Int. Conf. Consum. Electron. ICCE-Asia 2016*, no. April, 2017.

[77] K. Altun and B. Barshan, “Human Activity Recognition Using Inertial/Magnetic Sensor Units,” vol. 7559, no. January, 2012.

[78] F. Li, K. Shirahama, M. A. Nisar, L. Köping, and M. Grzegorzek, “Comparison of feature learning methods for human activity recognition using wearable sensors,” *Sensors (Switzerland)*, vol. 18, no. 2, pp. 1–22, 2018.

[79] P. Flach, “The ingredients of machine learning,” in *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, Cambridge University Press, 2012, pp. 13–48.

[80] H. Cassuto, A. Ben-Simon, and I. Berger, “Using environmental distractors in the diagnosis of ADHD,” *Front. Hum. Neurosci.*, vol. 7, 2013.

[81] ISRA PHARMACEUTICALS PTY LTD., “Clinical Library - Moxo ADHD TEST,” 2015. [Online]. Available: http://moxo-adhdtest.co.za/shortcodes/. [Accessed: 30-Apr-2017].

[82] “Support - Moxo ADHD TEST,” 2015. [Online]. Available: http://moxo-adhdtest.co.za/support-test/. [Accessed: 05-Jun-2017].

[83] NEUROTECH SOLUTIONS - PROFESSIONALS CHANNEL, “MOXO ADHD Online Test technical requirements - YouTube.” [Online]. Available: https://www.youtube.com/watch?v=aPYhAG6OIOY. [Accessed: 29-Oct-2018].

[84] The TOVA Company, “Introducing the Test of Variables of Attention (T.O.V.A.®).” [Online]. Available: http://www.tovatest.com/. [Accessed: 28-Aug-2017].

[85] C. L. Hall *et al.*, “The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring ADHD in children: a systematic review,” *Eur. Child Adolesc. Psychiatry*, vol. 25, no. 7, pp. 677–699, Jul. 2016.

[86] L. M. Greenberg, C. Holder, L. L. Carol Kindschi, and M. R. Tammy Dupuy, “T.O.V.A Clinical Manual,” 2018.

[87] J. A. Anguera, A. N. Brandes-Aitken, A. D. Antovich, C. E. Rolle, S. S. Desai, and E. J. Marco, “A pilot study to determine the feasibility of enhancing cognitive abilities in children with sensory processing dysfunction,” *PLoS One*, vol. 12, no. 4, p. e0172616, Apr. 2017.

[88] L. A. Reddy, E. Newman, T. K. Pedigo, and V. Scott, “Concurrent Validity of the Pediatric Attention Disorders Diagnostic Screener for Children with ADHD,” *Child Neuropsychol.*, vol. 16, no. 5, pp. 478–493, 2010.

[89] C. Rodríguez, D. Areces, T. García, M. Cueli, and P. González-Castro, “Comparison between two continuous performance tests for identifying ADHD: Traditional vs. virtual reality,” *Int. J. Clin. Heal. Psychol.*, vol. 18, no. 3, pp. 254–263, Sep. 2018.

[90] C. K. Conners and G. Sitarenios, “Conners’ Continuous Performance Test (CPT),” in *Encyclopedia of Clinical Neuropsychology*, New York, NY: Springer New York, 2011, pp. 681–683.

[91] A. Kamenetz, “Can Your Brain Really Be Retrained?,” 2013. [Online]. Available: https://www.fastcompany.com/3022943/can-your-brain-really-be-retrained. [Accessed: 06-Jun-2017].

[92] C. AMPEL, “Game can track, diagnose ADHD,” 2013. [Online]. Available: http://www.stuff.co.nz/technology/games/9018983/Game-can-track-diagnose-ADHD. [Accessed: 06-Jun-2017].

[93] MEDTECH BOSTON, “New Games Promise Accurate ADHD Diagnoses.” [Online]. Available: https://medtechboston.medstro.com/blog/2014/12/03/new-games-promise-accurate-adhd-diagnoses/. [Accessed: 31-Oct-2018].

[94] M. D. Heller, K. Roots, S. Srivastava, J. Schumann, J. Srivastava, and T. S. Hale, “A Machine Learning-Based Analysis of Game Data for Attention Deficit Hyperactivity Disorder Assessment,” *Games Health J.*, vol. 2, no. 5, pp. 291–298, 2013.

[95] J. A. Anguera *et al.*, “Video game training enhances cognitive control in older adults,” *Nature*, vol. 501, no. 7465, pp. 97–101, 2013.

[96] Akili Interactive Labs, “Akili.” [Online]. Available: http://www.akiliinteractive.com/. [Accessed: 29-Aug-2017].

[97] STARS National ADHD Study, “STAR National ADHD Study | Software

Treatment for Actively Reducing Severity of ADHD." [Online]. Available: http://nationaladhdstudy.com/. [Accessed: 29-Aug-2017].

[98] E. Strickland, "Akili: diagnosing Alzheimer's with a game [Resources_Start-Up]," *IEEE Spectr.*, vol. 51, no. 5, pp. 25–26, May 2014.

[99] P. A. Arean *et al.*, "The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial," *Med Internet Res*, vol. 18, no. 22, 2016.

[100] A. F. McDermott, M. Rose, T. Norris, and E. Gordon, "A Novel Feed-Forward Modeling System Leads to Sustained Improvements in Attention and Academic Performance," *J. Atten. Disord.*, 2016.

[101] C. G. Lim *et al.*, "A Brain-Computer Interface Based Attention Training Program for Treating Attention Deficit Hyperactivity Disorder," *PLoS One*, vol. 7, no. 10, 2012.

[102] X. Qian *et al.*, "Brain-computer-interface-based intervention re-normalizes brain functional network topology in children with attention deficit/hyperactivity disorder," *Transl. Psychiatry*, vol. 8, no. 1, p. 149, Dec. 2018.

[103] ATENTIV, "ATENTIVmynd." [Online]. Available: http://www.atentiv.com/index.html. [Accessed: 31-Oct-2018].

[104] NVIDIA, "THE ULTIMATE TABLET FOR GAMERS - NVIDIA SHIELD Store." [Online]. Available: https://www.nvidia.com/en-us/shield/tablet/. [Accessed: 11-Jul-2018].

[105] Android, "Sensors Overview." [Online]. Available: https://developer.android.com/guide/topics/sensors/sensors_overview.html. [Accessed: 20-Oct-2016].

[106] "Project Jupyter | Home." [Online]. Available: https://jupyter.org/. [Accessed: 18-Oct-2018].

[107] L. Hussain, "Detecting epileptic seizure with different feature extracting strategies using robust machine learning classification techniques by applying advance parameter optimization approach," *Cogn. Neurodyn.*, vol. 12, no. 3, pp. 271–294, Jun. 2018.

[108] F. Li *et al.*, "A preliminary study of movement intensity during a Go/No-Go task and its association with ADHD outcomes and symptom severity," *Child Adolesc. Psychiatry Ment. Health*, vol. 10, no. 1, pp. 1–10, 2016.

[109] E. G. Inc, "Introduction to Blueprints." [Online]. Available: https://docs.unrealengine.com/en-US/Engine/Blueprints/GettingStarted. [Accessed: 13-Jul-2018].

[110] K. Keerthi Vasan and B. Surendiran, "Dimensionality reduction using

Principal Component Analysis for network intrusion detection," *Perspect. Sci.*, vol. 8, pp. 510–512, Sep. 2016.

[111] M. H. Shojaeefard, A. Khalkhali, and S. Yarmohammadisatri, "An efficient sensitivity analysis method for modified geometry of Macpherson suspension based on Pearson correlation coefficient," *Veh. Syst. Dyn.*, vol. 55, no. 6, pp. 827–852, 2017.

[112] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond Accuracy, F-score and ROC: a Family of Discriminant Measures for Performance Evaluation."

[113] mlwhiz, "ROC curve," 2015. [Online]. Available: https://www.kaggle.com/yuansun/lb-0-84-for-starters/code. [Accessed: 09-Nov-2018].

[114] T. T. Company, "The T.O.V.A. Store." [Online]. Available: https://www.tovatest.com/store/. [Accessed: 21-Nov-2018].

[115] "SHIELD Tablet User Guide." [Online]. Available: http://support-shield.nvidia.com/tablet-user-guide/index.htm#t=shield-legal.htm. [Accessed: 19-Nov-2018].

[116] J. Zhang and S. Wang, "A fast leave-one-out cross-validation for SVM-like family," *Neural Comput. Appl.*, vol. 27, no. 6, pp. 1717–1730, Aug. 2016.

[117] W. Mao, X. Mu, Y. Zheng, and G. Yan, "Leave-one-out cross-validation-based model selection for multi-input multi-output support vector machine," *Neural Comput. Appl.*, vol. 24, no. 2, pp. 441–451, Feb. 2014.

[118] A. M. Schatz, A. O. Ballantyne, and D. A. Trauner, "Sensitivity and Specificity of a Computerized Test of Attention in the Diagnosis of Attention-Deficit/Hyperactivity Disorder," *Assessment*, vol. 8, no. 4, pp. 357–365, Dec. 2001.

[119] J. Kim, Y. Lee, D. Han, K. Min, D. Kim, and C. Lee, "The utility of quantitative electroencephalography and Integrated Visual and Auditory Continuous Performance Test as auxiliary tools for the Attention Deficit Hyperactivity Disorder diagnosis," *Clin. Neurophysiol.*, vol. 126, pp. 532–540, 2015.

[120] K. Strimbu and J. A. Tavel, "What are biomarkers?," *Curr. Opin. HIV AIDS*, vol. 5, no. 6, pp. 463–6, Nov. 2010.

[121] I. Merriam-Webster, "Neurotypical | Definition of Neurotypical by Merriam-Webster." [Online]. Available: https://www.merriam-webster.com/dictionary/neurotypical. [Accessed: 18-Nov-2018].

[122] Saunders and Elsevier, "Comorbidities | definition of Comorbidities by Medical dictionary," *Miller-Keane Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health*, 2003. [Online]. Available: http://medical-dictionary.thefreedictionary.com/Comorbidities. [Accessed: 24-Jun-2017].

[123] Farlex Partner Medical Dictionary, “Psychopathology | definition of psychopathology by Medical dictionary,” 2012. [Online]. Available: http://medical-dictionary.thefreedictionary.com/psychopathology.

[124] “Aetiology | definition of aetiology by Medical dictionary,” *Farlex Partner Medical Dictionary*, 2012. [Online]. Available: http://medical-dictionary.thefreedictionary.com/etiology. [Accessed: 15-May-2017].

[125] Dictionary.com, “allele,” *The American Heritage® Science Dictionary*. [Online]. Available: http://www.dictionary.com/browse/allele. [Accessed: 29-Apr-2017].

[126] J. E. Mezzich, “International Surveys on the Use of ICD-10 and Related Diagnostic Systems,” *Psychopathology*, vol. 35, no. 2–3, pp. 72–75, Jul. 2002.

[127] D. Lieberman, “HIPAA Security Assessment – Neurotech Solutions Ltd.,” 2012.

[128] D. M. W. Powers, “EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION,” *J. Mach. Learn. Technol. ISSN*, vol. 2, no. 1, pp. 2229–3981, 2011.

[129] “Moxo Test Centers - Moxo ADHD TEST.” [Online]. Available: http://moxo-adhdtest.co.za/adhd-test-centers/. [Accessed: 19-Nov-2018].