KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Technology

Assessment of Cognitive Performance in Elderly Life via Meaningful Play

Karsten Gielis

Supervisors:
Prof. dr. V. Vanden Abeele
Prof. dr. J. Tournoy
Prof. dr. K. Verbert

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Technology (PhD)

June 2021

Assessment of Cognitive Performance in Elderly Life via Meaningful Play

Karsten GIELIS

Examination committee:
Prof. dr. ir. E. Demeester, chair
Prof. dr. V. Vanden Abeele, supervisor
Prof. dr. J. Tournoy, supervisor
Prof. dr. K. Verbert, supervisor
Prof. dr. ir. J. Aerts (UHasselt, Belgium)
Prof. dr. ir. M. De Vos
P. Willems
Prof. dr. M. Birk
(Eindhoven University of Technology, The Netherlands)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Technology (PhD)

June 2021

Voorwoord

Ik herinner me nog heel goed de blik van de persoon die mij inschreef voor de opleiding industrieel ingenieurswetenschappen wanneer deze zag dat ik omschakelde van de faculteit rechtsgeleerdheid: "Ofwel is hij weg na een jaar, ofwel gaat hij negen jaar doen over zijn opleiding.". Uiteindelijk bleek dat tweede deel correct te zijn, maar niemand had zien aankomen dat dit zou komen door een doctoraat. De afgelopen vier jaar mocht ik onderzoek doen aan het e-Media Research Lab, een collectief onderzoekers dat zich inzet voor ingenieuze oplossingen te bedenken voor problemen in de gezondheidszorg, onderwijs, kunst en entertainment. Mijn onderzoek in dit lab is deel van het Dr. Solitaire project, een samenwerking tussen de faculteit industriele ingenieurswetenschappen en de faculteit geneeskunde waar onderzocht wordt of kaartspelen kunnen gebruikt worden om verschillen in cognitieve performantie te detecteren. Dit doctoraat stond mij toe om te groeien op zowel persoonlijk als technisch vlak, met enkele personen in een sleutelrol.

Eerst en vooral wil ik mijn promotoren bedanken, Vero Vanden Abeele, Jos Tournoy, en Katrien Verbert. Vero, ik wist niet wat voor een avontuur startte voor mij toen jij me, in overleg met Luc Geurts en Stef Desmet, op 4 mei 2016 liet weten dat ik geselecteerd was voor het impulsdossier. Doorheen mijn hele doctoraat wist ik dat ik altijd bij jou terecht kon voor raad en inzichten. Bovendien wist je perfect waar je mijn grenzen kon verleggen en waar je mij moest ondersteunen. Ik ben uitermate dankbaar voor het vertrouwen dat je me die dag en gedurende de rest van mijn doctoraat gaf, iets wat ik nooit zal vergeten. Jos, multidisciplinaire projecten zijn maar zo goed als de ondersteuning van beide disciplines. Dank u om mij onder te dompelen in de wereld van de geriatrie. De dagen dat ik mocht observeren en rekruteren in het UZ Leuven hebben mij meer bijgebracht dan je voor mogelijk houdt. Daarboven stond je altijd klaar om af te spreken om mijn focus en deadlines terug scherp te stellen. Katrien, je hebt me geholpen om het enigma machine learning beter te begrijpen. Hartelijk bedankt voor de meetings met opbouwende kritiek en

suggesties, ze hebben dit doctoraat een niveau hoger getild.

Daarnaast wil ik ook graag mijn examencomissie bedanken. Bedankt Peter Willems, om als vertegenwoordiger uit de industrie mee te stappen in dit verhaal. Jouw inzichten werden geapprecieerd. Bedankt Jan Aerts, een expert in data analytics aan boord hebben heeft me altijd twee keer doen nadenken bij elke grafiek en tekening in dit doctoraat. Bedankt om assessor te zijn en voor de feedback. Ook bedankt aan Max Birk en Maarten De Vos om me te begeleiden bij de laatste stappen van mijn traject.

Een doctoraat kan je niet alleen doen. Onderweg kwam ik mensen tegen die geloofden in dit project en, vaak na hun uren, mij ondersteunden. Bedankt Marie-Elena Vanden Abeele, de tweede dr. Vanden Abeele die mijn doctoraat hard vooruit heeft geholpen. Zonder de steun van u, dokters Delva, Lutin, en Wynants en heel de afdeling geriatrie van het Jessa ziekenhuis, had ik nooit genoeg participanten kunnen rekruteren. Je zet je elke dag 100% in voor jouw patiënten, dat je daarnaast je nogmaals 100% inzette voor mijn studie zal mij altijd bijblijven. Bedankt Lies Van Assche, Kathleen Roten, Kelly Pauwels en Hanne Heymans, de psychologen die mij kennis lieten maken met de wereld van de psychodiagnostiek, een discipline die ik nooit gedacht had van zo dichtbij te leren kennen. Bedankt Paul Dierick en Suzy Delarbre, om mij het vertrouwen te geven toen ik kwam spreken voor de COTESS werkgroep met de vraag om samen te werken. Ik hoop dat het Neuropsychology 2.0 project nog een staartje krijgt, dat verdient het.

Gedurende het gehele traject had ik het geluk om terug te kunnen vallen op de steun van mijn familie en mijn vriendenkring. Ik kan niet hard genoeg benadrukken hoeveel ik heb gehad aan mijn ouders, Benny en Fabie. Jullie hebben altijd mijn zorgen uit handen genomen zodat ik mij volop kon concentreren op mijn academische carrière. Ik hoop dat ik ooit een fractie kan worden van de ouders die jullie voor mij zijn. Make, je bent er altijd voor mij geweest. Jammer genoeg was mijn laatste onderwijsinstelling iets te ver gelegen om na de les binnen te springen, maar je zorgde er steeds voor dat ik niets tekort kwam. Pake, Bomma en Bompa, jullie hebben me jammer genoeg niet kunnen zien afstuderen. Maar ik ben er rotsvast van overtuigd dat jullie uitermate fier zouden zijn. Aan de families Gielis en Beckers, bedankt voor achter mij te staan en voor (on)vrijwillige proefkonijnen te zijn. Aan mijn vrienden van Deurne, Diest, Leuven en daar voorbij, bedankt voor de vele avonturen die we tot nu toe al hebben beleefd. Dat ik na zoveel jaren nog altijd zo'n hechte vriendenschappen heb, zegt meer over jullie dan over mij. Speciale dank gaat uit naar Laura Bortels, Thomas De Groote, Laura De Herdt, Jelle Derbaix, Kobe

Houtmeyers, Berto Pals, Simon Renty en Dries Reynders om dit manuscript na te lezen en mij te behoeden voor gênante spelfouten.

Vrienden maak je niet enkel in je vrije tijd, maar blijkbaar ook op je werk. Jonas, we hebben samen een conferentie bezocht in China en les gegeven in Ethiopië, maar het meest van al koester ik de mopjes en discussies die we hadden tijdens de gewone werkdagen. Ine, jij was de eerste persoon die ik ontmoette toen ik startte en ik weet nog goed hoe warm je mij toen ontvangen hebt. Met jou samenwerken was altijd een plezier, ik ga onze design kritieken missen. Hannelore, bedankt voor de talloze keren dat je me hebt laten lachen. Jouw oprechtheid en eerlijkheid sieren je. Robin, bedankt voor jouw gedetailleerde feedback en advies. Je bent een top post-doc. What started as a little research group quickly expanded to a full-fledged research powerhouse. Alemitu, Aqeel, Benjamin, Melkamu, Esmael, Milica, Chetanya, Katta, Duowei, Kymeng, Bert, Dimitri, Jorgen, Ahmed, Yiyuan, ... I thank you all greatly for the fun we had. Ook zou ik graag alle professoren en medewerkers willen bedanken waarmee ik heb mogen samenwerken.

Ten slotte Eline, ik ben blij dat je nog steeds naast mij staat. Elk doctoraat gaat gepaard met hoogtes en laagtes, altijd stond je klaar om samen te vieren of mij op te beuren. Terwijl ik dit hoofdstuk van mijn leven afsluit, staan we aan het begin van een nieuw. Ik kijk er naar uit om samen een huis op te knappen die we dan samen thuis kunnen noemen.

"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful." - George Box

Abstract

Cognitive decline is the deterioration of one or more cognitive functions such as attention, memory, or processing speed. Inherent to aging, most people will encounter some form of cognitive decline during their lifetime while retaining the ability to perform instrumental activities of daily living. However, beyond this age-related cognitive decline due to aging, some people will experience pathological cognitive decline: an abnormal rate of cognitive impairment due to neurological diseases such as dementia or due to physical trauma. In contrast to age-related cognitive decline, this pathological cognitive decline hinders activities of daily living. In between the stages of age-related and pathological cognitive decline, is the stage of Mild Cognitive Impairment (MCI), which is characterized by a slight yet noticeable decline in cognition. Nevertheless, instrumental activities of daily living remain mostly intact in this stage. While the diagnosis of MCI is not always followed by a diagnosis of dementia, people diagnosed with MCI have a markedly higher chance of progressing to dementia. As such, early case-finding of MCI and timely adjusting the diagnosis is crucial to ensure apt medical support. To aid this cause and to better understand the dwindling of cognition, this dissertation sets out to explore the possibilities of using digital card games to assess differences in cognitive performance due to cognitive aging and MCI.

In particular, the use of *digital biomarkers*, i.e., user-generated physiological and behavioral data collected through digital devices, is investigated. Embedded into day-to-day interactions, these digital biomarkers can be used to support diagnosis without interfering with the person's daily routine. In addition, as they are high-resolution in nature, they allow for making informed inferences of neuropsychological processes previously unavailable to psychologists.

In this doctoral dissertation, two different aspects of cognitive decline are measured using different digital card games. First, digital biomarker caused by cognitive aging are assessed using the card game FreeCell. To this end, a generic image processing toolkit was built to extract digital biomarkers from

the Microsoft Solitaire Collection. Using this toolkit, data was captured from three different age categories. Machine learning models trained on this data showed promise in classifying the younger and older age categories but lacked in classifying games played by the middle-aged category.

Second, digital biomarker differences caused by MCI are assessed using the card game Klondike Solitaire. For this part, an Android application was custom-built to capture digital biomarkers while leaving gameplay untouched. Candidate digital biomarkers were identified in collaboration with 11 experts in cognitive decline. Next, gameplay data was captured from both healthy older adults and older adults diagnosed with MCI. A generalized linear mixed model analysis was conducted to investigate differences between healthy older adults and older adults living with MCI. The results of this analysis suggest it is possible to discriminate healthy participants from participants diagnosed with MCI at a group level. In addition, machine learning models were trained to discern games played by older adults with MCI. These models show promise on an individual level and are successful in discerning healthy older adults from adults living with MCI. While exploratory in nature, the results indicate similar psychometric properties as commonly used screening tests.

In sum, these findings suggest that commercial off-the-shelf card games, not built for the purpose of measuring cognition, can be used to capture digital biomarkers of cognitive performance sensitive to the cognitive decline due to aging and MCI.

Beknopte samenvatting

Cognitieve achteruitgang is de achteruitgang van één of meer cognitieve functies zoals aandacht, geheugen of verwerkingssnelheid. Inherent aan ouder worden, zullen de meeste mensen tijdens hun leven te maken krijgen met enige vorm van cognitieve achteruitgang terwijl zij nog wel in staat zijn instrumentele activiteiten van het dagelijks leven uit te voeren. Naast deze normale cognitieve achteruitgang als gevolg van ouder worden, zullen sommige mensen een pathologische cognitieve achteruitgang ervaren: een abnormale mate van cognitieve achteruitgang als gevolg van neurologische aandoeningen zoals dementie of fysiek trauma. In tegenstelling tot normale cognitieve achteruitgang, belemmert deze pathologische cognitieve achteruitgang de activiteiten van het dagelijks leven wel. Tussen de stadia van normale en pathologische cognitieve achteruitgang bevindt zich het stadium van *Mild Cognitive Impairment* (MCI). MCI wordt gekenmerkt door een lichte maar merkbare achteruitgang in cognitie. Niettemin blijven de instrumentele activiteiten van het dagelijks leven in dit stadium grotendeels intact. Hoewel MCI niet altijd evolueert tot dementie, is de kans dat iemand met MCI dementie krijgt aanzienlijk groter. Het vroegtijdig opsporen van MCI en het tijdig bijstellen van de diagnose is dan ook van cruciaal belang. Om *case-finding* te ondersteunen en de achteruitgang van cognitie beter te begrijpen, onderzoekt deze dissertatie de mogelijkheden van digitale kaartspelen om verschillen in cognitie te evalueren die veroorzaakt zijn door cognitieve veroudering en MCI.

In het bijzonder wordt het gebruik van digitale biomarkers onderzocht. Digitale biomarkers omvatten fysiologische gegevens en gedragsgegevens die gegenereerd worden door gebruikers en verzameld worden via digitale apparaten. Ingebed in dagelijkse interacties met technologie kunnen deze digitale biomarkers gebruikt worden om de diagnose te ondersteunen zonder de dagelijkse routine van de persoon te verstoren. Bovendien maakt de hoge resolutie van deze data het mogelijk om geïnformeerde conclusies te trekken over neuropsychologische processen die voorheen niet beschikbaar waren voor psychologen.

In deze dissertatie worden twee verschillende aspecten van cognitieve achteruitgang gemeten met behulp van verschillende digitale kaartspelen. Ten eerste worden verschillen in digitale biomarkers als gevolg van cognitieve veroudering geëvalueerd met behulp van het kaartspel FreeCell. Hiertoe werd een generieke beeldverwerkingstoolkit gebouwd om digitale biomarkers te extraheren uit de Microsoft Solitaire Collectie. Met behulp van deze toolkit werd data verzameld van drie verschillende leeftijdscategorieën. Machine learning modellen die op deze data getraind werden, bleken veelbelovend in het classificeren van de jongere en oudere leeftijdscategorieën, maar bleken matig te zijn in het classificeren van spellen gespeeld door de middelste leeftijdscategorie.

Ten tweede worden verschillen in digitale biomarkers als gevolg van MCI geëvalueerd met behulp van het kaartspel Klondike Solitaire. Voor dit onderdeel werd een Android applicatie op maat gemaakt om digitale biomarkers te capteren zonder het normale spelverloop te storen. Kandidaat digitale biomarkers werden geïdentificeerd in samenwerking met 11 experten in cognitieve achteruitgang. Vervolgens werd data verzameld van zowel gezonde ouderen als ouderen met MCI. Een *generalized linear mixed model analysis* werd uitgevoerd om verschillen tussen gezonde ouderen en ouderen met MCI te onderzoeken. De resultaten van deze analyse suggereren dat het mogelijk is om op groepsniveau gezonde deelnemers te onderscheiden van deelnemers met MCI. Daarnaast werden machine learning modellen getraind om spellen te onderscheiden die gespeeld werden door ouderen met MCI. Deze modellen zijn beloftevol op individueel niveau en zijn succesvol in het onderscheiden van gezonde ouderen en ouderen met MCI. Hoewel exploratief van aard, wijzen de resultaten op vergelijkbare psychometrische eigenschappen als veelgebruikte screeningtests.

Kortom, deze bevindingen suggereren dat commerciële kaartspelen, die niet gemaakt zijn om cognitie te meten, gebruikt kunnen worden om digitale biomarkers te capteren die gevoelig zijn voor de cognitieve achteruitgang als gevolg van veroudering en MCI.

List of Abbreviations

AARP American Association of Retired Persons. 11, 12

AD Alzheimer's Disease. 8, 144

aMCI Amnestic MCI. 83

AUC Area Under the Curve. 21, 73, 76, 126, 127, 129, 130

BDA Big Data Analytics. 145

CAM Confusion Assessment Method. 60

CDR Clinical Dementia Rating scale. 84, 99, 100, 119

COTS Commercial off-the-shelf. 13, 14, 20, 28, 32, 34, 55, 61, 62, 76–78, 82, 83, 85, 86, 111, 113, 114, 118, 119, 134, 135, 138, 139, 142–144

CVA Cerebrovascular Accidents. 54

EAVISE Embedded and Artificially Intelligent Vision Engineering. 20, 135

ESA Entertainment Software Association. 11

GLMM Generalized Linear Mixed Model. 83, 101, 110, 114, 137

IADL Instrumental Activities of Daily Living. 2, 6

ICC Intraclass Correlation. 91, 92

MCI Mild Cognitive Impairment. 2, 3, 6–8, 13, 17–21, 27, 82–84, 86, 87, 89, 91, 99, 101–103, 111–115, 118–120, 126, 129, 130, 136–142, 144

MMSE Mini-Mental State Examination. 8, 27, 60, 84, 99, 100, 119, 129, 130

MoCA Montreal Cognitive Assessment. 8, 27, 60, 84, 86, 99, 100, 119, 130

naMCI non-Amnestic MCI. 83

OpenCV Open Source Computer Vision Library. 32, 40, 135

ROC Receiver Operating Characteristic. 68, 73, 78, 127, 137

SCD Subjective Cognitive Decline. 7

SHAP Shapely Additive exPlanations. 140, 141

SMOTE Synthetic Minority Oversampling. 73

WHO World Health Organization. 27

Contents

List of Figures

List of Tables

Part I

General Introduction

Chapter 1

Introduction

In an ever-aging society, cognitive decline is becoming an increasing point of concern [1]. Cognitive decline can be described as the impairment of one or more cognitive functions such as attention, executive function, or memory [2]. This phenomenon can be divided into two categories: age-related cognitive decline caused by natural processes like aging, and pathological cognitive decline caused by diseases such as Alzheimer's Disease or physical trauma [2]. Most people will encounter some form of cognitive decline in their life due to aging, as indicated by the green trajectories visualized in Figure 1.1. However, some people progress to Mild Cognitive Impairment (MCI) (indicated by the blue line), a condition where one or more cognitive domains are *significantly* impaired [3]. MCI differs from dementia (indicated by the red line), as in the stage of MCI instrumental activities of daily living (IADL), such as managing finances, preparing meals, or shopping, remain essentially intact. While some older adults living with MCI remain stable or even progress to normal levels of cognition, they have a markedly higher chance of progressing to forms of dementia [4].

While cognitive decline is not a direct a cause of death, symptoms caused by advanced cognitive decline such as lapses in judgement, loss of self-awareness, or even partial loss of the ability to speak or swallow, may decrease the patient's life expectancy [5]. To date, no treatment exists in current-day medicine to repair neuronal damage once transpired [6]–[8]. Despite this absence of treatment, early case-finding and regular follow-up of cognitive decline is imperative to support patients and their family [9] by timely diagnosing disease progression, starting (non-) pharmacological treatment to mitigate symptoms, or slowing down disease progression [7], [8], [10], [11].

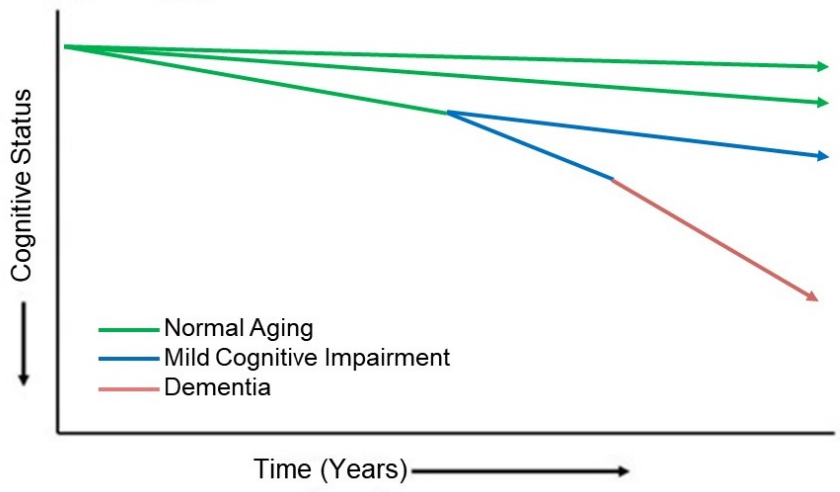


Figure 1.1: Visualisation of cognitive trajectories throughout a human lifespan.

The consequences of dementia, financially and emotionally, are substantial. Around 50 million people worldwide suffer from a form of dementia [12]. This number is predicted to increase to 150 million by 2050, impacting friends, caregivers, and families across the globe. Economically speaking, the cost of dementia is set to grow to 2 trillion dollars by 2030. Spurred by an ever-aging society that wants to maintain older adults' independence and quality of life, governmental organizations urge for further research on assessment of cognitive aging, coining it the *"grand societal challenge"* [13], [14]. In addition, realizing the current impact of dementia and its impending peril, policymakers and (inter)national organizations have called for novel, scalable, and longitudinal tools for improving case-finding and monitoring of dementia [12], [15]–[17].

To provide additional insights into cognitive decline, this dissertation sets out to investigate the possibilities of assessing cognition by means of digital card games. In particular, this dissertation aims to develop and explore game-based digital biomarkers, user-generated and health-related traces collected through connected digital devices [18], to assess differences in cognition due to cognitive aging and MCI.

1.1 Cognitive Decline

1.1.1 Age-related Cognitive Decline: Cognitive Aging

Cognitive aging is described by Blazer [13] as *"the process of gradual, ongoing, yet highly variable changes in cognitive functions that occur as people get older"* and envelops all changes in cognition throughout a person's lifespan. While cognitive aging equally comprises increases in certain cognitive functions, it is in our society most often associated with the decline commonly noted amongst older adults [19]. In the past decades, cognitive aging research has become more important to understand why some older adults are less affected by age than others and how decline could be mitigated.

During early descriptive research in 1955, Wechsler identified certain cognitive functions that appeared to be more resilient to the effects of cognitive aging than others [20]. This led to the theory of *crystallized intelligence* and *fluid intelligence* as first described by Catell [21] in 1963. Fluid intelligence is an umbrella term for all cognitive abilities related to extensive processing of new information, amongst which processing speed, memory, and executive functioning [22], [23]. Crystallized intelligence combines all cognitive abilities linked to knowledge and habitual applications of established information amongst which vocabulary and general knowledge [22], [23]. While crystallized cognitive abilities seem to increase at a steady pace and diminish slightly around 70 years old, fluid cognitive abilities peak around 25 years and decline rapidly afterward (indicated by the full and dotted line in figure 1.2). Even more than 50 years later, the theory of crystallized and fluid intelligence is still widely accepted. Today, current research is focused on refining these categories by further differentiating cognitive functions such as memory (e.g., implicit memory, episodic memory, semantic memory, etc.) [20].

While figure 1.2 accurately depicts the cognitive status of the general population throughout its lifespan, a vast heterogeneity is observed when observing individual cognitive trajectories [22]. Figure 1.3 depicts cross-sectional age trends (N=5315) for fluid reasoning and crystallized knowledge from the Woodcock-Johnson Tests of Cognitive Abilities [25][1]. Every person's trajectory is highly individual: the rate at which it increases and declines afterward, at which age they peak, and how high the peak is. The idiosyncratic cognitive trajectory is shaped by the person's innate intelligence, educational opportunities, lifestyle choices, and other internal and external factors [26], [27]. This heterogeneity

[1] While these cross-sectional studies may suffer from cohort differences, it can still be noted that there are sizeable differences for fluid and crystallized intelligence in the same cohort.

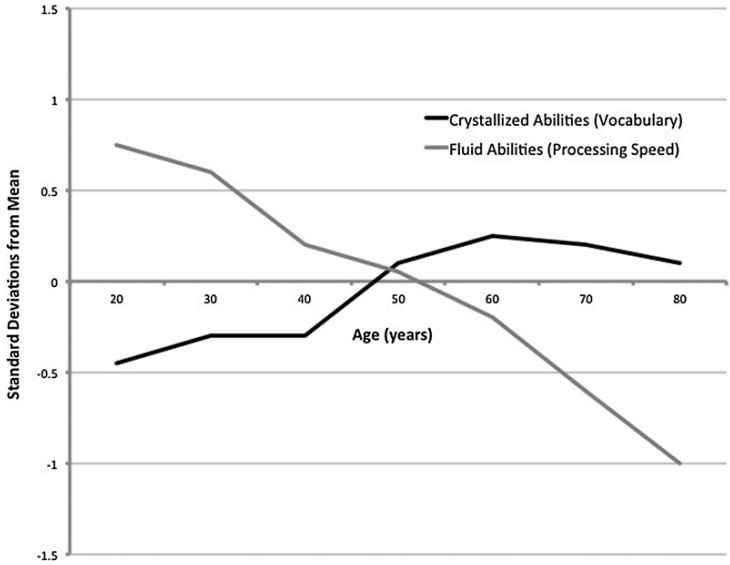
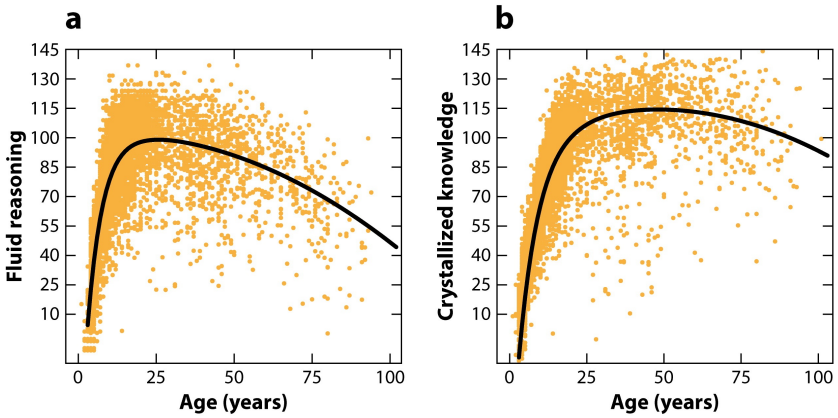


Figure 1.2: Crystallized and fluid intelligence across the lifespan of the average human, adapted from [24].



Tucker-Drob EM. 2019.
Annu. Rev. Dev. Psychol. 1:177–96

Figure 1.3: Cross-sectional age trends for fluid reasoning (a) and crystallized knowledge (b). The average and standard deviation have been rescaled to respectively 100 and 15 units for the ages between 18 and 25. Permission for reprint granted from [22].

complicates accurately distinguishing normal cognitive aging from abnormal pathological cognitive decline [22], [28].

1.1.2 Pathological Cognitive Decline: Dementia and Mild Cognitive Impairment

Pathological cognitive decline is, unlike cognitive aging, not a natural lifelong process but rather the result of disease or trauma. The umbrella term for these conditions that cause symptomatic deficits in cognitive domains is called dementia [29]. Dementia spans over progressive diseases such as Alzheimer's Disease [30] or Lewy-Body dementia [31] and nonprogressive diseases such as Creutzfeld-Jakob [32] or head trauma [29].

Symptoms of dementia are dependant on the type of dementia and the region of the brain where alterations manifest. Amongst others, these symptoms can include loss of short and/or long-term memory, attention span problems, changes in personality, difficulty with language, and more. Consequentially, these alterations in the brain caused by dementia can severely impair mood, behavior, relationships, and hamper essential activities of daily living. Other diseases can mimic symptoms of dementia, such as depression (also known as pseudodementia), thyroid problems, or vitamin deficiencies. However, these symptoms can be reversed by treating the underlying problem. While the neuronal damage of dementia is irreversible with the state of modern medicine, proactive interventions of adjustable risk factors can delay the onset or impede the progression of dementia [12].

However, prodromes of dementia can be detected earlier in the state of Mild Cognitive Impairment [3]. Both MCI and dementia are characterized by a significant impairment in one or more cognitive functions which cannot be attributed to age or other causes. However, they differ in the extent to which activities of daily living (IADL) are preserved. Whereas these IADL are preserved or minimally impaired with MCI, they are severally impaired with dementia. Depending on whether memory is impaired, persons with MCI are categorized as living with *amnestic* or *non-amnestic* MCI (figure 1.4). These two categories of MCI can be further specified as either *single-domain* or *multi-domain*, depending on the amount of cognitive functions that are impaired [33]. A diagnosis of MCI does not always lead to a diagnosis of dementia; some persons remain in the state of MCI and others even recover to normal cognitive levels. However, research has shown that this group has a higher chance of progressing to dementia [4]. Therefore, early case-finding of MCI can assist in timely detecting underlying causes, and allows for taking measures slowing down disease progression [10], such as providing (non-)therapeutical

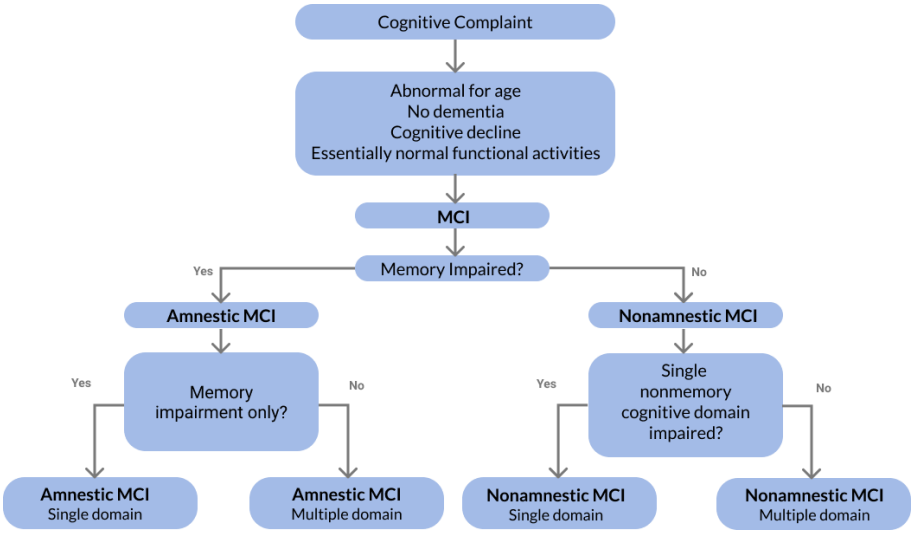


Figure 1.4: Criteria of Mild Cognitive Impairment as described by [33] based on the Key Symposium criteria posed in 2003 [40].

regimens to delay symptoms [7], [8], [10], and providing support for the patient [11]. Despite these benefits of early case-finding, cognitive performance is not regularly monitored amongst community dwellers, care home residents, and hospitalized patients. This can lead to a later diagnosis [5]. Studies point at a general late or underdiagnosis of cognitive impairment amongst older adults [34]–[36].

Recent research pursues further delineation of early stages of cognitive decline. Subjective cognitive decline (SCD) was construed in 2014 to define individuals who have a persistent self-experience of cognitive decline, yet show no objective cognitive impairment in cognitive tests [37]. This construct might prove useful to identify those vulnerable to dementia before pathological decline can be detected. First results support SCD as a late preclinical stage of dementia, with reduced brain volumes and steeper cognitive performance decline being noted by comparing healthy older adults without SCD and SCD [38], [39]. However, the construct of SCD is still being refined and the majority of people with SCD will not progress to pathological forms of cognitive decline. In addition, there is no universally accepted practice to discriminate SCD with MCI and a diagnostic overlap might exist among clinicians [37]. Therefore, for the remainder of this dissertation, the focal point of pathological cognitive decline will be on MCI as it is more mature as a clinical construct.

1.1.3 Detection of Cognitive Decline

Typically, assessment of cognitive decline is initiated by a routine screening at the general practitioner or a cognitive complaint from the patient or their family. The first step in this process is usually to conduct a short cognitive screening test such as the Montreal Cognitive Assessment (MoCA) [41] or Mini-Mental State Examination (MMSE) [42]. When these results indicate a possible cognitive impairment, a complete neuropsychological assessment is conducted as follow-up. These neuropsychological assessments take between 90 minutes to 3 hours and equally assess the patient's history and other potential cognition impairing factors such as language fluency problems or depression [43]. To complement these tests, biomarker analysis (e.g., FDG-PET scans) can be conducted to detect neuronal damage. Recently, special interest in cognitive decline research has gone to biomarker scans. Biomarkers can be defined as "any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease"[44]. As such, they are used to influence diagnosis and can assess the risk of disease progression [44]. For MCI due to AD, four cerebrospinal fluid biomarkers are of main interest: Aβ40, Aβ42, total τ, and phosphorylated τ [45]. An elevated presence of these biomarkers can strengthen the diagnosis of MCI and can give information on the predicted risk of progression to Alzheimer's Disease. Together, this neurological examination leads to the potential diagnosis of MCI.

While our understanding of the intricacies of cognitive decline has come a long way, research pinpoints opportunities for improving modern cognitive assessment. While some physicians tend to solely rely on screening tests for diagnosis, these screening tests may lack in sensitivity or are unable to pinpoint which cognitive functions are impaired [43]. And while thorough neuropsychological assessment remains a golden standard, they require a trained administrator which makes the process vulnerable to interassessor variability and sensitive to white-coat effect [46], [47]. In addition, the mode of administration of the majority of these tests (i.e., pen, paper, and a stopwatch) requires the full attention of the administrator over a lengthy period of time which makes the process prone to errors and limits the number of data points extracted. What requires an extensive cognitive effort from the patient is often reduced to the number of mistakes made and the total time it took to complete said test [48]. Use of biomarkers at the other hand, is invasive and inconvenient for the patient [46], [49]. This whole process is obtrusive for the patient, requiring them to break daily routines and forcing travel for a fragile population [49]. Research on the patient's perspective reveals somber results, with patients describing neuropsychological assessment as bewildering and humiliating [50], [51]. This highly stressful evaluation can make patients self-aware of their condition, leading to feelings of helplessness and distress, which aggravates results and impacts already fragile mental health

[52], [53].

While these disadvantages suggest room for improvement when looking at cognitive assessment as single-point-in-time measurements, deficiencies become more apparent in frequent longitudinal measurements. As pointed out by Piau et al. [46], one of the biggest vulnerabilities in current neuropsychological assessment is that it is performed at discrete points in time at large intervals. Known as the "snapshot problem", this episodic aspect makes measurements vulnerable to temporary lapses in cognition caused by medication, motivation, or even tiredness [48], [49], [54], [55]. In addition, due to the static nature of these pen-and-paper tests, they are susceptible to learning effects, rendering repeated measurements over a short period of time ineffective. Measuring biomarkers are equally unfit for repeated measures as they are costly to conduct and strain limited resources and personnel [48]. Being aware of these challenges, researchers and policymakers promote the development of additional cost-effective and high data resolution tools that support longitudinal cognitive monitoring [3], [48], [56]–[59] which reduce patient-level barriers and are understanding of the patient's experience [52].

1.2 Digital Biomarkers

Cognitive information in the form of digital biomarkers of cognitive performance show promise as complementary tools to the current neuropsychological assessments. Digital biomarkers are defined as "user-generated physiological and behavioral measures collected through connected digital devices to explain, influence and/or predict health-related outcomes" [18]. Sharing the same clinical goals as their biological counterparts, digital biomarkers can be used to support, predict, or influence diagnosis [55]. Embedded in passive and active interactions with wearables, smartphones, or other connected digital devices, digital biomarkers have the potential to unobtrusively capture data related to physical, social, and cognitive health [18]. While clinical encounters offer elaborate information on diseases and their progression, they are often limited in time thus making the data sparse [60], [61]. By providing more continuous measurements of objective health-related information, digital biomarkers can assist in understanding the progress and detect variability of symptoms [60]. In addition, the use of sensors does not require transforming signals to ordinal ranked tasks (e.g., for Parkinson's Disease, the continuous signal of movement is transformed to a numerical quantity of finger taps) [61]. The newfound interest in digital biomarkers reflects in the number of studies published. In the past decade, studies on digital biomarkers indexed in PubMed have increased by 325% [62]. Currently, the use of digital biomarkers has pervaded in a multitude

of domains including, but not limited to, modelling of tremor for movement-related disorders [63], using mammograms to predict breast cancer risk [64], modeling gait changes and sleep disturbances to monitor Alzheimer's Disease [65], [66], detecting arrhythmia [67], using voice analysis to detect signs of depression [68], but also cognitive assessment [46], [49], [69].

The characteristics of digital biomarkers make them suitable to address the challenges caused by traditional cognitive assessment. They are less invasive for the patient, as they are captured from day-to-day interactions with technology [55]. Due to the availability of digital devices in the western world, they are cheaper to acquire, not demanding specialized equipment or trained personnel to administer [18], [55]. In addition, these digital devices allow for more data points to be captured as opposed to manually registering information, leading to a higher granularity and resolution of data. This combination of an increased number of measurements and the higher resolution of data captured during these measurements opens up possibilities for more informed inferences of neuropsychological processes [48]. Intra-individual variability for example, an early indicator of cognitive decline, can be detected [46], [70]–[73]. In addition, the effects of medication can be assessed or events which might impact cognition (e.g., a physical or psychological trauma) can be detected. Finally, these more frequent measurements can be used to create individual cognitive baselines, allowing the patient to be compared with themselves as opposed to norm-referenced data [48].

For Mild Cognitive Impairment, the technologies from which digital biomarkers are captured can be grouped into four distinct categories [46]. First, digital biomarkers can be derived from dedicated embedded or passive sensors. These sensors range from infrared sensors installed in the home to detect movements [74] to passive sensors devices installed in the patient's vehicle [75]. The second category comprises digital biomarkers captured through dedicated wearables. This category uses, amongst others, GPS data from mobile phones [76] and rotation data from inertial sensors [77]. Third, non-dedicated technological solutions can yield digital biomarkers as well. Examples of technology used in this category are mouse pointer movement data [78] and keyboard loggers installed on computers [73]. Finally, digital biomarkers can be captured through dedicated or purposive technologies. This category contains technologies in the form of e.g., personal digital assistants [79], but also ***games*** [80].

Sensor-based technologies are the most mature digital biomarker technology category. However, they show several barriers which hamper adoption [46]. The effort and complexity of installing the sensors makes them difficult to set up by non-technical users. In addition, current-state detection algorithms still show difficulties when differentiating activity between larger households. Dedicated wearables show other complexities, with user acceptance being diminished by

the lower technological skill of the users and little perceived use. While a less mature technology than the other categories, the technologies which can be unobtrusively embedded in day-to-day activities without complex installation show more promise for widespread adoption.

1.3 Digital Games for Assessment of Cognitive Performance

Digital games have stirred the interest of researchers in understanding, measuring, and training cognition for a long time. Suits [81] defined games as "the voluntary attempt to overcome unnecessary obstacles", which hints at the autotelic nature, meaning games are played by the own volition of players. The audiovisual and interactive environments allow for immersive experiences that bring forth an intrinsic motivation to play. This was corroborated by recent research reinforcing that games are more engaging than classical neuropsychological tests [82]–[84]. Moreover, they do not require the presence of a trained administrator, making them more resilient to white-coat effect and administrator bias [48]. In addition to immersive and interactive environments, games provide obstacles bounded by a fixed set of rules, which requires the engagement of the player to overcome them [48]. These obstacles can differ with every play session while keeping the core rules intact. This novelty of the challenges keeps games engaging and also less vulnerable to learning effects [48].

While digital games were already popular amongst young and old, their popularity is increasing rapidly in older populations. In 2019, the American Association of Retired Persons (AARP) [85] conducted a survey of gameplay behavior in North America. Their results show that the 50+ gaming population in North America has grown to 50.6 million, with the average time playing being 5 hours per week. The silver-haired gamer is also a dedicated gamer, with 47% of them gaming daily and 27% multiple days per week. Interestingly enough, this daily gaming frequency increases with age, with 45% of older adults between 50 and 59 gaming daily and 50% of those ages 70 and up. Similar results can be found in the Entertainment Software Association's (ESA) video game industry essential facts of 2020 [86]. Their results indicate that 55% of gamers 65+ are novel gamers, having played video games for ten years or less. This indicates that more older adults are finding their way to digital entertainment. The most favorite game genre in both studies includes casual card games (e.g., Solitaire), a category also found to be popular in independent research by Chesham et al. [87] and De Schutter et al. [88]. All three age groups (50-59, 60-59, and 70+) in the AARP report indicate simply *fun* as their main reason to play [85].

According to the AARP, there has also been a platform shift amongst older adults since 2016, with mobile devices (73%) dethroning computers and laptops (59%). When it comes to digital games for assessment of cognition, it may be preferable to develop applications for large touch devices such as tablets, as argued in a systematic review by Fereira-Brito et al. [89]. The intuitive interface with direct manipulation [90] may contribute to adoption from less familiar users and is more suitable for fragile populations [91], [92].

1.3.1 Serious Games for Assessment of Cognitive Performance

The focal point of game research for assessing cognition has been on serious games, which are “games that do not have entertainment, enjoyment, or fun as their primary purpose” [93]. Instead, they are specifically developed with a serious purpose in mind. One of the earliest examples is "Space Fortress" [94], a "computer-controlled task with game-like qualities", developed in the 80’s under the leadership of the Cognitive Psychological laboratory of the University of Illinois, which was extensively used to train cognition. Up to today, this game-based tool is still used in research labs to unravel the intricacies of cognitive control and to understand learning processes [95]. In the meantime, games to train and measure cognitive performance have boomed, as indicated by the systematic reviews of Ferreira-Brito et al. [89], Lumsden et al. [84], and Valladares-Rodriguez et al. [83].

While these serious games are developed with the intention of capitalizing on the benefits of digital games, as explicated above, pitfalls may prevent them from doing so in reality. In particular, these serious games are at risk of being "chocolate-covered broccoli"[95]–[97], in other words, well-controlled environments to measure cognition with a thin veil of gameplay. In the end, while technically games, they are not motivating players. As they are often based on encapsulating existing neuropsychological tests, they are bound by the rules of said test. Certain gameplay elements might be dropped because they can interfere with measuring the psychological construct [48]. In addition, these games are often developed by academic researchers as opposed to game studios. This lack of expertise and lesser funding can impact the quality of the game, providing lesser immersive and mechanically polished games [95], [98]. A meta-analysis of serious learning games by Wouters et al. [99] corroborates these critiques. The authors compared serious learning games with regular methods. They showed that while serious games might be more effective and even improve retention, they do not improve motivation. This lack of motivation has also been noted in training cognition studies by Boot et al. [95] and Toril et al. [82].

1.3.2 Commercial off-the-shelf Games for Assessment of Cognitive Performance

Commercial off-the-shelf (COTS) games, often developed with the sole purpose of fun in mind, may be a valid alternative to these serious games. As argued by Mandryk and Birk [100], COTS games have the potential to provide digital biomarkers. Using natural and unprompted interactions with these games, information on behavior, cognitive performance, motor performance, social behavior, and affect can be obtained. More specifically, research suggests using casual COTS games for older adults susceptible to MCI as they have a broader appeal, are more accessible, and require a lower cognitive load [87], [101]. Their key advantage, next to the general advantages of digital biomarkers and games described above, is the prospect of *meaningful* play.

Meaningful play is a concept put forward by Salen & Zimmerman [102] in 2003 and can be defined in two separate yet related ways [103]. The first definition of meaningful play reflects how game actions lead to outcomes thus creating meaning: *"Meaningful play in a game emerges from the relationship between player action and system outcome; it is the process by which a player takes action within the designed system of a game and the system responds to the action. The meaning of an action in a game resides in the relationship between action and outcome."*. This descriptive definition illustrates what happens in a game, at a mechanistic level, hence it also highlights what may be captured as raw material for digital biomarkers. The second definition of meaningful play takes a wider viewpoint and defines moments of meaningful play: *"Meaningful play occurs when the relationships between actions and outcomes in a game are both discernible and integrated into the larger context of the game. Creating meaningful play is the goal of successful game design."* The prime focus in this evaluative definition underscores the importance of effectively communicating the results of actions and integrating game outcomes in the grander scheme of the game and the player's life.

Specifically for (digital) meaningful play in elderly life, De Schutter, Vanden Abeele, and colleagues [88], [104]–[108] have done extensive research. Their work states that meaningful play for older adults is often derived from the perception of 'fostering connectedness', 'cultivating oneself and others', and 'contribution to society' [88], [105], [109]. Their user studies paint a picture of gamers that has a fondness for casual games, with social interaction and challenge being the main predictors for investment in playing digital games [88], [105]. A qualitative study (N=35) applying the Uses&Gratifications paradigm showed five categories of older adults playing digital games: "time wasters", "freedom fighters", "compensators", "value seekers" and "ludophiles" [110]. Later research of De Schutter et al. [108] defines three interpretations of game enjoyment:

telic enjoyment or the joy of improving mental and physical health; hedonic enjoyment or the joy of experiencing positive emotions while playing; and eudaimonic enjoyment or the joy of contributing to personal growth. In sum, their research shows that play is enjoyed throughout one's lifespan and that digital games can be a source of meaningful play for older adults [107].

However, using a COTS game is not a silver bullet to solve all problems of serious games. While COTS games are not bound to the demands of psychological tests, they might introduce complexities that may interfere with accurately measuring cognition [48]. For example, adapting difficulty to the player's skill or providing a hint functionality might complicate accurately measuring in-game performance. In addition, these games might change over time, with developers pushing updates, which might complicate comparing previous game performance [98]. In addition, it can prove difficult to extract information from COTS games as the source code and interaction logs are inaccessible to the public[2]. Finally, researchers using either COTS games or serious games still need to recognize that not everyone is an avid gamer. Moreover, those who do enjoy gaming do not necessarily enjoy the genre and type of games promoted by the researchers. Abandonment of the game, which could indicate an inability to play due to cognitive impairment, could equally indicate loss of interest as gaming preference might naturally change over time [100]. This complicates the selection of games worth exploring for digital biomarkers [100].

Klondike Solitaire

A likely suitable choice for a COTS game to assess cognition through digital biomarkers is Klondike Solitaire as it is one of the most popular card games [111]. According to Parlett [112], it was founded in the 18$^{\text{th}}$ century when fortune-telling cards rose in popularity. Three centuries later, it is still popular amongst young and, especially, old. According to De Schutter and Vanden Abeele, card games were amongst the most played activities amongst older adults in their user group [105]. Similar results were noted by Allaire et al., where card games were found to be the most popular amongst digital games [113]. More recently, Solitaire was observed as most popular amongst older adult gamers by Boot et al. [114] during the PRISM randomized field trial. In this year-long study on older adults and leisure, participants had access to a computer where eleven games were installed, amongst which Sudoku, Solitaire, and crossword puzzles. Boot et al. [114] noted that "There was a strong, clear

[2]There are notable exceptions, e.g., League of Legends, EVE online, or Guildwars 2 [100] which are programmed to log game information of their users. However, even if data is readily available for research, it is not logged with monitoring cognition in mind which complicates the process of crafting insightful digital biomarkers.

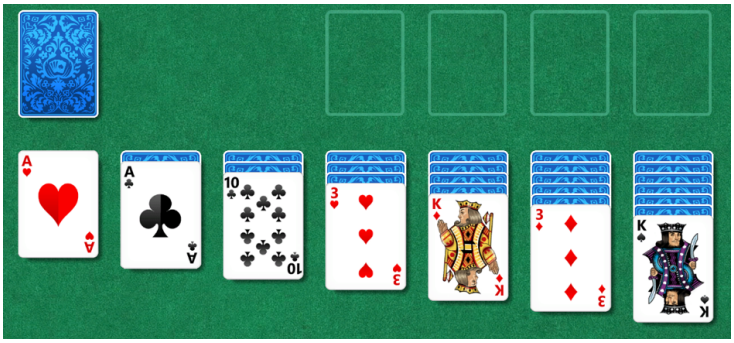


Figure 1.5: Klondike Solitaire as can be seen in the Microsoft Solitaire Collection. The seven build stacks can be seen at the bottom whilst the suit stacks are at the top right. The pile of undealt cards can be seen in the top left.

preference for Solitaire [...]. After Solitaire, there was no clear second choice, and on average participants infrequently played the other games." Additionally, their results showed that Solitaire was being played most consistently.

The rules of Klondike Solitaire are deceitfully simple [115]. In the traditional variant, at the start of the game, 28 of 52 playing cards are divided amongst seven build stacks, as can be seen in figure 1.5. The first build stack contains one card, the second two cards, the third three, all up to the seventh build stack, with each top card being dealt face-up. The remaining 24 cards are placed on the pile. The goal of the game is to sort all cards per rank, starting with the ace and ending with the king, in the four suit stacks. This can be achieved by moving cards between build stacks. Cards can be moved to other build stacks when they differ in color and are one value less (e.g., a black nine of clubs can be placed on a red ten of hearts). As the game progresses, stacks can be cleared. These spots can be solely used to store kings. When stuck, players can always request cards from the pile. Here, cards can be drawn in threes and maintain their first-in-last-out-order (i.e., only the last card drawn can be played).

The allure of Solitaire is surprising as it has one of the lowest success rates of any card game variant [112]. Anecdotal evidence reports an average win rate of 15% [115]. Combining Monte-Carlo methods, Hindsight Optimization, and Sparse sampling, technical win rates up to 35% are noted by Bjarnason et al. [115]. This low win-rate can be explained by the complexity of each initial deal, with the face-down cards having 51 quintillion possible combinations.

FreeCell

FreeCell is another highly popular Solitaire variant. In contrast to Klondike Solitaire, all 52 cards of the deck are dealt at the start and are oriented in the face-up direction, as can be seen in figure 1.6. The increased visible information at the start makes FreeCell, always solveable. The FreeCell configuration consists of three parts. At the bottom, there is the build stack, where all 52 cards are randomly distributed over eight stacks. The top left is called the storage stack, consisting of four free places or FreeCells where the player can temporarily store cards to open up moves. Finally, at the top right, the suit stack can be found. When all cards are sorted per suit in ascending order on these stacks (i.e., all hearts, clubs, diamonds, and spades starting with the ace and ending with the king), the game is won.

To sort all cards on the suit stack, cards have to be moved according to the following rules. Between build stacks, cards can be moved on top of each other when they are of different colors and if the rank of the current top card is one higher. According to the rules, only one card can be moved at the same time; i.e., it is not allowed to move stacks of sorted cards at the same time. However, multiple cards can be moved if there are enough FreeCells and free build stack columns. Mathematically, the maximum number of cards that can be moved in a single move can be calculated as follows:

$$moveablecards = (1 + freestoragespots) * 2^{freecolumns}$$

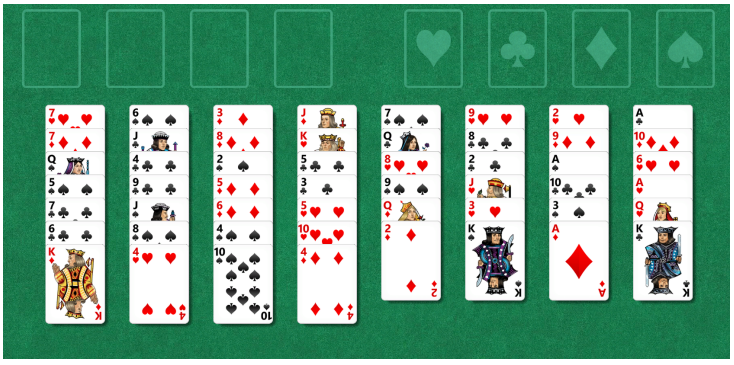


Figure 1.6: FreeCell as can be seen in the Microsoft Solitaire Collection. The 8 cascades can be found at the bottom, the four storage stacks or FreeCells at the top left, and the foundation at the top right.

1.4 Research Objective and Hypotheses

This doctoral dissertation aims to explore the possibilities of game-based digital biomarkers to assess differences in cognitive performance due to cognitive aging and MCI. To assist in resolving this overarching research objective, several research questions have been formulated:

Objective: To assess cognitive performance in elderly life via meaningful play.
RQ1. How can game data be captured from commercial off-the-shelf digital card games?
RQ2. How can insights from game design and cognitive psychology be combined to transform game data into potential digital biomarkers of cognitive performance?
RQ3. To what extent can differences in cognition due to cognitive aging be assessed using digital biomarkers of cognitive performance?
RQ4. To what extent can differences in cognition due to Mild Cognitive Impairment be assessed using digital biomarkers of cognitive performance?

1.5 Methodology

A data science approach is required to address these research questions. Data science can be defined as "the study of the generalizable extraction of knowledge from data" [116]. Driven by the need to handle ever-increasing data flows from today's society, multiple models have been developed to standardize data science processes [117]–[120].

A visualized representation of the six steps of the project can be found in 1.7 [119]. The first phase of this model is *Domain Understanding* which envelops getting to know the perspective from the domain where the data science problem origins. Using the domain knowledge extracted from this phase, the *Data Understanding* phase can start. In this phase, data collection tools can be built and initial data collection studies can start. This first dataset can be used for a first data exploration, to get initial insights from the data, which can be used to solve early issues with the current process, fine-tune hypotheses, or set up plans for further data processing. The third phase, *Data Preparation*, entails every step to create the dataset which will be used to train models like data collection and transformation of data in informative indicators. The fourth phase, *Modeling*, envelops training, optimizing, and reiterating models to extract insights. The results of these models are then critically assessed in the *Evaluation* phase. The whole process and derived insights are evaluated with respect to the objectives

of the study, listing possible next steps. After possibly several iterations, this will lead to the *deployment* phase. This phase is often not the end of the project but can be seen as the organization of all knowledge gathered, written down in a report or deployed in an automated process.

While the process followed in this dissertation adheres to the data science model described above, there is no clear-cut division of each phase between the chapters, with many of them describing several steps of the process. For detecting cognitive aging through FreeCell, the Domain Understanding and Data Understanding part can be mainly found in *chapter 2* as it describes the creation of a computer vision tool used to extract data from FreeCell and the data exploration from the first data collection study. *Chapter 3* comprises the Data Preparation, Modeling, and Evaluation phases of detecting cognitive aging, describing the steps undertaken to create the final dataset, the modeling of machine learning algorithms to detect cognitive aging, and evaluation thereof. For detecting MCI through Klondike Solitaire, *chapter 4* describes the expert consensus study used to gather expert's opinions on the effect of MCI on Klondike Solitaire gameplay. Furthermore, information on Data Understanding, Data Preparation, Modeling, and Evaluation can be found in *chapters 4 and 5*, which contain analyses of digital biomarkers on a group level using statistical interference and on an individual level using machine learning. Finally, this dissertation can be seen as the *Deployment* phase of this project, synthesizing all results gathered.

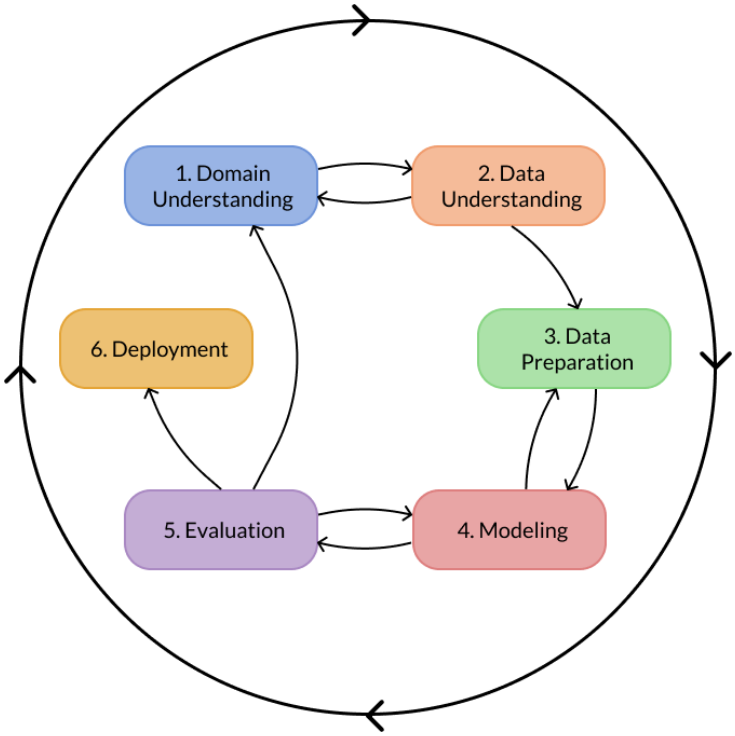


Figure 1.7: Visualisation of all steps taken in this project, adapted from [119].

1.6 Thesis Outline

As this thesis explores the assessment of differences in cognitive performance due to aging and MCI, the remainder of this thesis is divided into three main parts: Part II, which envelopes all research done to detect cognitive decline of cognitive aging through FreeCell; Part III, which details all findings concerning the detection of cognitive decline due to mild cognitive impairment through Klondike Solitaire; and Part IV, which discusses all previous chapters and synthesizes it in general conclusions, recommendations, and future work. A schematic overview can be found in figure 1.8 and an introduction on how the different chapters relate to each other can be found below.

The following two chapters belong to Part II and investigate the impact of cognitive aging through FreeCell. The first step in assessing cognitive aging through FreeCell is to develop tools to capture game data. As argued in section

1.3, extracting information to craft digital biomarkers can prove difficult for COTS games. To answer *RQ1. How can game data be captured from commercial off-the-shelf digital card games?*, a novel method of capturing game metrics without accessing the internal source code or interaction logs was developed. **Part II Chapter 2** describes the development process of a generic image processing toolkit to extract digital biomarkers of cognitive performance from existing casual card games. This toolkit, a collaboration with the Embedded and Artificially intelligent Vision Engineering (EAVISE) research group of KU Leuven, utilizes computer vision to automatically capture, annotate, and process gameplay data from the popular Microsoft Solitaire Collection. Next to detailing this method, results of a data exploration on the effects of aging on FreeCell gameplay were analyzed. The source code of this tool was made publicly available for other researchers to use and adapt.

With Part II Chapter 2 indicating that image processing is a viable way to capture game data with a minimum of stress on the computer, the next chapter builds on the lessons learned from the initial data exploration and expands the analysis. As argued in section 1.1.1, understanding and mapping cognitive aging can help in mitigating decline and might lead to earlier detection of pathological decline. To answer *RQ3. To what extent can differences in cognition due to cognitive aging be assessed using digital biomarkers of cognitive performance?*, a larger number of participants were recruited to explore the viability of machine learning. **Part II Chapter 3** describes the machine learning process used to detect age using digital biomarkers of cognitive performance captured from FreeCell. A Logistic Regression model was trained to classify FreeCell games into one of three age categories (18-25, 40-55, and 65+). Despite the highly variable aspect of cognitive aging, performance metrics showed successful classification of the 18-25 and 65+ categories. However, model performance for the middle 40-55 category showed to be problematic. In sum, the results of this chapter support current theories on fluid and crystallized theory, provide benchmark results for future researchers, and specify pitfalls and opportunities which may inspire future research.

The next two chapters belong to Part III and investigate the impact of MCI through Klondike Solitaire. As argued in section 1.1.2, timely detecting MCI can help with timely changing (non-)therapeutical regimen and providing apt support. Due to the limited related work on casual COTS card game-based digital biomarkers, the collaboration between medical sciences and game research is imperative to crafting digital biomarkers that are theoretically related to cognitive performance. To answer *RQ2. How can insights from game design and cognitive psychology be combined to transform game data into potential digital biomarkers of cognitive performance?*, a methodological method was made to facilitate collaboration and draw insights. **Part III Chapter 4** contains the

expert consensus study used to explore the relation between player actions and cognitive functions in Klondike Solitaire. Agreement between experts was calculated using an intraclass correlation with a two-way fully crossed design with type consistency. A moderate to excellent reliability was achieved for all cognitive functions, which provided insights into the intricacies of Klondike Solitaire gameplay mechanics. Using these insights, digital biomarkers of cognitive performance were defined, leading to the creation of an Android application that can capture necessary information to craft said digital biomarkers. To assess the efficacy of these digital biomarkers on a group level and to partially answer *RQ4. To what extent can differences in cognition due to Mild Cognitive Impairment be assessed using digital biomarkers of cognitive performance?*, digital biomarker differences between older adults living with MCI and a healthy control group were inspected on a group level using Generalized Linear Mixed Models. These models, which took differences in age, tablet experience, and Klondike Solitaire experience into account, showed sizeable and significant effects for 12 out of 23 digital biomarkers tested. Comparing the different categories of digital biomarkers, results indicated biomarkers related to the timing of moves and the outcome of the game to be promising.

In addition to drawing population inferences based on a sample, the dataset of Part III Chapter 4 was equally explored on an individual level to see whether individual cases of MCI can be detected through gameplay. To assess the efficacy of these digital biomarkers on an individual level and to partially answer *RQ4. To what extent can differences in cognition due to Mild Cognitive Impairment be assessed using digital biomarkers of cognitive performance?*, a machine learning analysis was conducted. **Part III Chapter 5** describes the training of nineteen machine learning models with different underlying algorithms to detect games played by older adults with MCI. After training, the three best models were used to predict previously unseen games to test the usability of the models. Results showed F1 scores and AUC's above 0.811 and 0.877 for each of the top-3 models, opening up new research opportunities for longitudinal measurements.

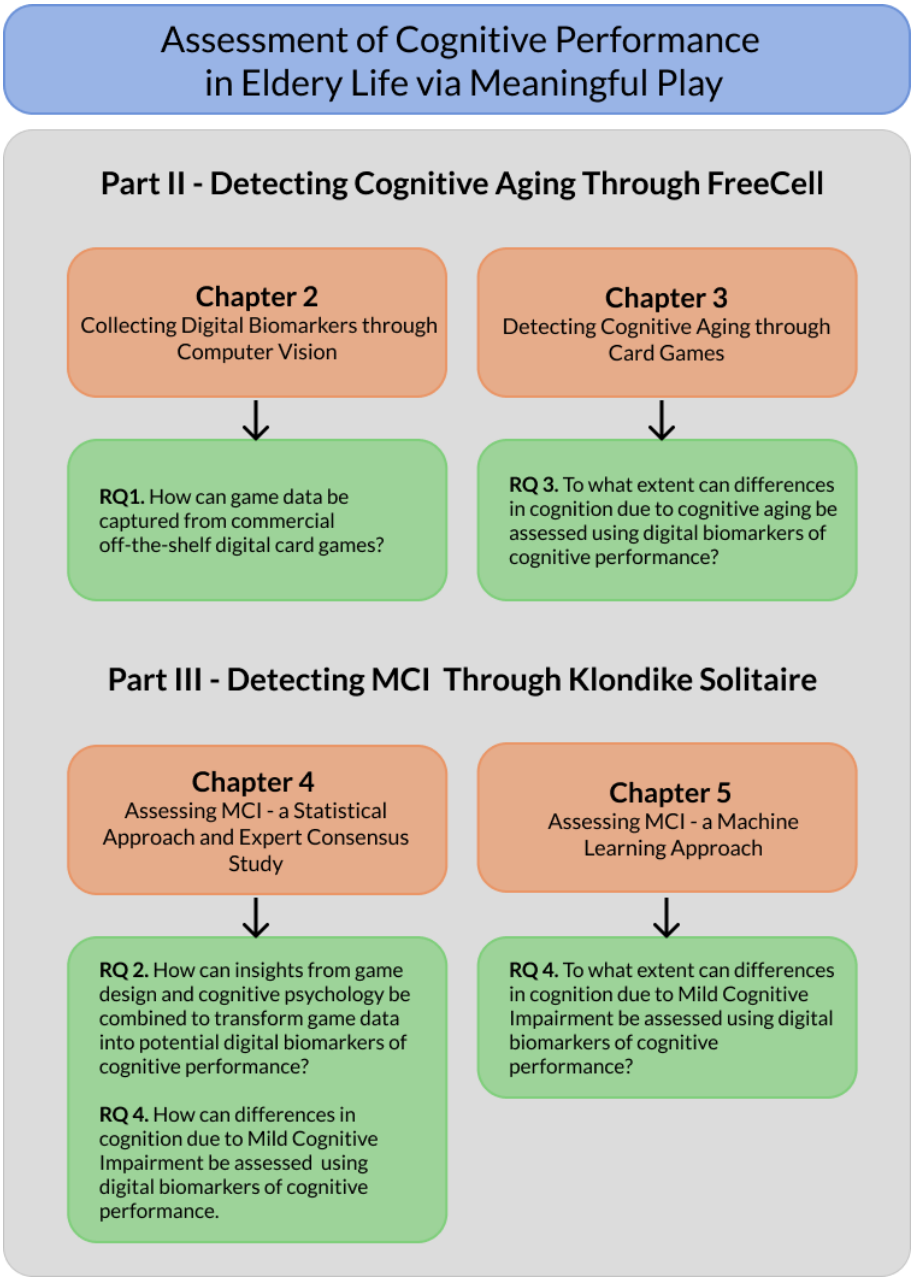


Figure 1.8: Schematic overview of the thesis outline.

1.7 Publications

This dissertation is based on the following scientific publications. A full overview of scientific contributions and outreach can be found in Appendix A.

Papers in Proceedings of International Conferences (blind peer review)

Published K. Gielis, J. Kennes, C. D. Dobbeleer, S. Puttemans, and V. V. Abeele, "Collecting Digital Biomarkers on Cognitive Health Through Computer Vision and Gameplay: An Image Processing Toolkit for Card Games", in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, Xi'an, China: IEEE, Jun. 2019, pp. 1–12, ISBN: 978-1-5386-9138-0. DOI: 10.1109/ICHI.2019.8904511

Published K. Gielis, K. Verbert, J. Tournoy, and V. Vanden Abeele, "Age? It's in the Game: An Exploratory Study on Detection of Cognitive Aging through Card Games", en, in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, Barcelona Spain: ACM, Oct. 2019, pp. 651–664, ISBN: 978-1-4503-6688-5. DOI: 10.1145/3311350.3347193

Papers in International Journal (blind peer review)

In Review K. Gielis *et al.*, "Dissecting Digital Card Games to Yield Digital Biomarkers for the Assessment of Mild Cognitive Impairment: A Methodological Approach and Exploratory Study", *Journal of Medical Internet Research*, Jan. 2021

Published K. Gielis, M.-E. Vanden Abeele, K. Verbert, J. Tournoy, M. De Vos, and V. Vanden Abeele, "Detecting Mild Cognitive Impairment through Digital Biomarkers of Cognitive Performance found in Klondike Solitaire: A Machine Learning Study", *Digital Biomarkers*, Jan. 2021, ISSN: 2504-110X. DOI: 10.1159/000514105

Abstracts in Proceedings of International Conferences

Published K. Gielis, F. Brito, J. Tournoy, and V. Vanden Abeele, "Can Card Games Be Used to Assess Mild Cognitive Impairment?: A Study of Klondike Solitaire and Cognitive Functions", en, in *CHI PLAY '17 Extended Abstracts*, Amsterdam, The Netherlands: ACM Press, 2017, pp. 269–276, ISBN: 978-1-4503-5111-9. DOI: 10.1145/3130859.3131328

Part II

Detecting Cognitive Decline due to Cognitive Aging through FreeCell

Chapter 2

Collecting Digital Biomarkers on Cognitive Health Through Computer Vision and Gameplay: an Image Processing Toolkit for Card Games

This chapter is a copy of the previously published article:
K. Gielis, J. Kennes, C. D. Dobbeleer, S. Puttemans, and V. V. Abeele, "Collecting Digital Biomarkers on Cognitive Health Through Computer Vision and Gameplay: An Image Processing Toolkit for Card Games", in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, Xi'an, China: IEEE, Jun. 2019, pp. 1–12, ISBN: 978-1-5386-9138-0. DOI: 10.1109/ICHI.2019.8904511

Scientific Contribution:
As first author, I lead the writing of the first draft of the manuscript and processed suggestions of co-authors. In addition, I coordinated the data collection and full analysis.

2.1 Introduction

Worldwide millions of people suffer from cognitive disorders such as depression (322 million), anxiety (264 million), or dementia (50 million) [126], [127]. Even milder cognitive impairments such as Mild Cognitive Impairment (MCI) can hamper several cognitive functions such as attention, executive functioning, or social cognition. Depression alone accounts for 4.3% of the Global Burden of Disease, making it the largest cause of disability worldwide [128]. The cost of dementia, not including the emotional stress on families, was estimated at 818 billion dollars in 2015 [126]. Anxiety and depression combined account for a global cost of 1.15 trillion dollars per year [129].

Hence, early diagnosis and frequent follow-up of mental health problems are crucial to managing the disease, allowing for timely treatment and disease progression mitigation. It ensures finding the best sources of support and making informed decisions about the future, even if the disorder is untreatable [30], [130]–[133]. For some cognitive ailments, especially dementia, diagnosis is often non-existent or made in a later stage of the disease. A study in 2015 reported that 58% of all dementia cases in the USA go undiagnosed [134]. This is why, in 2017, the World Health Organization (WHO) endorsed the Global action plan on the public health response to dementia focusing amongst others on diagnosing cognitive impairments in an earlier stage.

A considerable part of diagnosis and follow-up in traditional medicine involves the use of biomarkers. Biomarkers are defined by the WHO as *"any substance, structure, or process that can be measured in the body or its products, and that influences or predicts the incidence of outcome or disease"* [135]. They are an objective way to indicate biological and pathogenic processes or responses to therapeutic interventions, utilized in the fields of disease prediction, diagnosis, and prognosis [135]. A well-known biomarker used in medical diagnosis is e.g., the presence of Amyloid beta in cerebral spinal fluid for Alzheimer's Disease [135], [136]. Asides from biomarkers to aid in the screening, diagnosis, and follow-up of mental health illness, which are often expensive and invasive, neuropsychological tests are common practice. Well-known cognitive tests to aid in the screening for cognitive impairments are e.g., the Mini-Mental State Examination (MMSE) or the Montreal Cognitive Assessment (MoCA) are used to screen for dementia [41], [42]. While these tests are less invasive and less expensive than capturing biomarkers, they are also characterized by lower specificity and sensitivity [42], [137]–[142]. That is why there is an increasing interest in digital biomarkers. Digital biomarkers are defined as *"objective, quantifiable physiological and behavioral data that are collected and measured by means of digital devices such as portables, wearables, implants or digestibles"* [143].

Digital biomarkers have the possibility to give deeper insight to specialists and patients, providing a source of data from text interactions, home data, GPS location, but also games. Mandryk and Birk point out that a variety of activity traces can be gathered from in-the-wild gameplay of commercial off-the-shelf (COTS) games and considered as *digital biomarkers of cognitive health* [100]. The contribution of this paper is to explore gameplay as an additional medium to capture digital biomarkers for mental health. In particular, this paper presents a new method of defining digital biomarkers in games and a toolkit for collecting digital biomarkers of cognitive impairment through gameplay. The toolkit and method are developed to work with existing card games, i.e., the Microsoft Solitaire Suite, that people already play and enjoy. Finally, an exploratory data analysis is shown to demonstrate the feasibility of the toolkit.

2.2 Background

In this section, we first explore the potential digital biomarkers for mental health. Next, we present games as a viable source for capturing these digital biomarkers. Next, we explore challenges when games are used that are developed by research labs. Finally, we present the opportunities of using commercial-of-the-shelves games as a source of digital biomarkers.

2.2.1 Digital Biomarkers

Today, in the cross-domain of computer science, engineering, biomedicine, regulatory science, and informatics, interest is growing in the digital counterparts of biomarkers. The surge in interest has sparked the founding of a journal by Karger in September 2017 [143] dedicated to this topic only. Compared to classical pen-and-paper tests, the use of digital biomarkers has shown advantages such as reduced cost, unobtrusive measurement, and the possibility of continuous data gathering. In contrast to episodic measurements of classical biomarker and pen-and-paper tests, which are often taken bi-annually or yearly, digital biomarkers can be captured on a daily basis. This makes the findings more robust to patients having a momentary lapse, feeling stressed, examined, or being tired because of a bad night's rest. It has also been shown that pen-and-paper tests are vulnerable to practice effects due to the fixed course of the tests [88], [144], [145]. More reliable and comprehensive cognitive test batteries do exist, but these lengthy tests have to be administrated by a trained health professional and require the often frail participant to go to a specialized institution. In these specialized institutions, often one specialist is assigned to follow-up the disease progression as intra-rater reliability is proven to be better

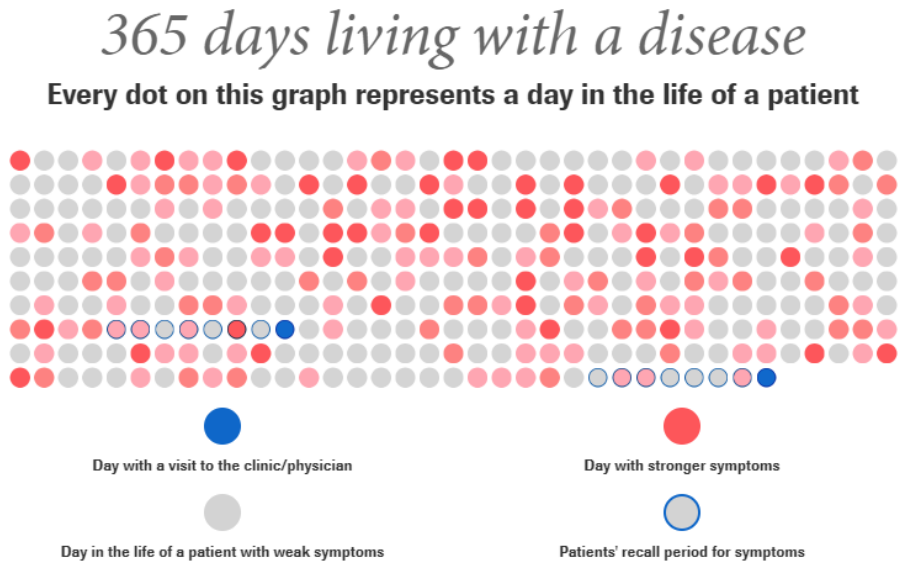


Figure 2.1: Limited contact moments can cause physicians to miss the bigger picture [148].

than inter-rater reliability [146]. However, even these specialists suffer from bias, reducing the overall precision of these follow-up tests [147]. Moreover, as these specialists are overburdened, contact moments can be sparse. As shown by the Digital Biomarkers Department of Roche in Fig 2.1, sparse contact moments can cause neuropsychologist to miss the bigger picture as patients tend to recall symptoms for smaller time periods than the follow-up period actually spans [148]. For people suffering from mental degeneration, remembering and evaluating the severity of symptoms can prove even more difficult.

Digital biomarkers possibly solve many of these problems. They may increase the ecological validity by increasing the temporal and spatial resolution of the captured behavior during activities of daily life [61]. Secondly, they may unlock previously unobtainable sources of behavioral, social, environmental, and physiological data [149] with minimal effort required from physician and patient. Finally, as these digital biomarkers are captured by digital devices, they are less prone to human bias and less susceptible to the white-coat effect [150].

The advantages of digital biomarkers have sparked several initiatives on diverse platforms using various sensors. Redfield et al. [151] used accelerometers of

smartphones to measure gait, finger tapping, voice, and balance as a measure for Parkinson's disease. Saeb et al. [152] found strong evidence that mobile phone sensor data such as GPS and phone usage correlates to depression. Faurholt-Jepsen et al. and Beiwinkel et al. [153], [154] did research on biomarkers and bipolarity, correlating smartphone information such as calls, text messages, and GPS with mental disorders. In a study using the Beiwe app, schizophrenia relapses were correlated with anomalies in patient behavior prior to relapse [155]. Hagler et al. [156] created an in-home monitoring system to assess gait, a predictor of cognitive decline. While in most cases the sample sizes were too small to draw conclusions about the general population, these experiments show promise of continuous monitoring of at-risk populations with minimal effort required from the user.

2.2.2 Digital Biomarkers and Games

Games are a natural source of information on behavior, cognitive performance, motor performance, social behavior and affect, for people of all ages [100]. According to Suits [81], games are interactive systems, where players present "a voluntary attempt to overcome unnecessary obstacles". Through playful design and intuitive rules, players are motivated to push their own boundaries [157]. In certain cases, it can even cause a state of flow [158], a gratifying state where players lose track of time and place because the challenging game activity necessitates all of their attention and skills. This flow experience is found enjoying and makes games autotelic; players are intrinsically motivated to play [81]. As research has shown that a decrease in motivation has an effect on participant performance of classical pen-and-paper cognitive tests [159], the potential of games to maximize motivation makes them a suitable medium for capturing digital cognitive biomarkers.

There are already many examples of gamified collect digital biomarkers that may be indicative for cognitive performance. Episodix, a gamificated California Verbal Learning Test, manages to classify individuals into three categories: healthy, mild cognitively impaired, and Alzheimer Dementia [160]. A mirror game designed by Słowiński et al. captures digital diagnostic biomarkers in the form of non-verbal synchrony and neuromotor functions. Utilizing statistical learning techniques, they could discern users suffering from schizophrenia from their healthy counterparts, with an accuracy of 93% and specificity of 100% [161]. Neuro-World, a collection of 3D mobile games by Jung et al. [162], estimates Mini-Mental State Examination scores from gameplay metrics such as score, time to clear a stage, and the number of cleared levels. Their games test perception, object memory, sequential memory, selective attention, vigilance attention, and visual investigation. Leduc-McNiven et al. developed WarCAT and Lock Picking

[163]. WarCAT, a card game based on War, measures recognition and recall, while Lock Picking measures problem-solving skills by letting the user search for an optimal score. Smartkuber, the augmented reality game for cognitive screening made by Boletsis and McCallum [164] uses five minigames to screen for cognitive impairments, revealing significant correlations and comparable validity to the Montreal Cognitive Assessment, a popular neuropsychological screening test for Mild Cognitive Impairment.

2.2.3 Challenges with serious games

These games fall under the category serious games, defined as "Games that a serious goal, rather than entertainment, enjoyment or fun, as their primary purpose"[165]. In this case, their primary purpose is to provide information on cognitive performance. Unfortunately, research shows that these tailor-made serious games suffer from disadvantages. First, these games made in research labs often miss the funding and development time of commercial games. Developing a serious game that can compare to commercial games in quality is often out of reach due to differences in manpower, budget, and expertise. As research cycles differ from game release schemas, it is likely that the game will be outdated by the time the game is programmed, funding is gathered, and medical ethical clearance is approved [98]. Maintaining the game and shipping updates also prove difficult as this is not the main goal of research labs.

Secondly, despite the efforts to make them as enjoyable as possible, research has shown that custom-made games for cognitive training still fall short in engagement and suffer from attrition in longitudinal studies. The repetitiousness of many gamified assessments and training can lead to participant disengagement, possibly impacting the data quality [84], [166]. Furthermore, it has been reported that it is the affectionate bond with the experimenter and not the cognitive training per se that motivated participants to continue [167]. This suggests that there is a mismatch between the serious games being developed for cognitive functioning and the games people effectively enjoy playing. It may be that serious games, while valid with respect to the mental health purpose, perhaps do not provide 'meaningful play'[103].

2.2.4 Opportunities of Commercial-Of-The-Shelves games

The playing of commercial games is weaved into the fabric of everyday lives; such games are part of the socio-cultural environment[103]. As mentioned above, the power of digital biomarkers lies in its frequent, longitudinal measurement, stressing the importance of the autotelic nature of games. Boot et al. discovered

during post-intervention surveys that the games of the control condition, such as word and puzzle games, were found more enjoyable than those of the gamified test group [168]. This enjoyment of the game led to higher motivation to adhere to the cognitive training, indicating that commercial games may be a better fit for capturing digital biomarkers. This may make COTS games a more valid, suitable medium to gather digital biomarkers for cognitive performance [169]. However, the downside of using these commercial, off-the-shelf games is that the gameplay is less 'controlled', they may demand more complex and variable actions from players, simultaneously addressing multiple cognitive functions from the players. This interplay of different cognitive functions may introduce undesired non-therapeutic effects or add uncertainties in screening and in detecting impairments in cognitive functioning [48].

We are not the first to promote the use of COTS games for assessment of cognitive performance. Jimison et al. showed that there is a correlation between Mild Cognitive Impairment and performance in the game FreeCell. They found that higher variability in scores and more sensitivity to game difficulty are indicative of cognitive impairment[145]. Thompson et al. explored the relation between common puzzle games and standard neuropsychological tests and found that performance on these smartphone-based games is indicative of cognitive ability across several cognitive domains[169]. Furthermore, working memory was correlated to sudoku performance by Grabbe, showing the potential of this popular game for measuring cognitive performance [170].

However, as aforementioned, capturing digital biomarkers in COTS games can be troublesome, as altering the code of the game is impossible without the permission of the game developer. An alternative is recording and annotating gameplay manually, reviewing hours of gameplay, and manually tagging digital biomarkers. Yet, manually annotating game data reintroduces the limitations of the aforementioned classic tests. It reintroduces human error, as manually timing of events is more inaccurate and inconsistent. Secondly, manually annotating is a time-consuming and tedious task. Finally, it limits the number of metrics captured, it refrains from capturing certain digital biomarkers. More fine-grained biomarkers such as speed or certainness of execution are not measurable from manually annotating gameplay alone. Moreover, if the researchers want to explore previously uncaptured biomarkers, the annotating process starts anew.

This paper explores a different method for capturing digital biomarkers, namely computer vision and, more specifically, image processing algorithms. By utilizing machine learning and image processing, gameplay can be analyzed in real-time and digital biomarkers can be extracted and processed in an efficient manner. In order to explore the viability of image processing to capture digital biomarkers on a COTS game, a multithreaded C++ desktop application was developed that utilizes the Open Source Computer Vision Library (OpenCV) [171]. It is

built as a generic toolkit to capture and analyze card gameplay data. It acts as a silent watcher that unobtrusively captures, processes, and analyses gameplay from the standard Microsoft 10 Solitaire Collection. Currently, the code of the game rules is implemented for Klondike Solitaire and FreeCell versions. In the next sections, we will demonstrate how the toolkit operates by using FreeCell as an example.

2.3 Toolkit Concepts

This section contains information necessary to grasp the mechanisms behind the toolkit. First, the rules and board space of FreeCell are illustrated. Secondly, the method of defining digital biomarkers is explained. Finally, the global concept of capturing digital biomarkers in gameplay is described.

2.3.1 The FreeCell Board Space

FreeCell is a well-known and popular Solitaire variant. It is played with all 52 cards in a deck, which are all dealt face-up at the beginning of the game. This transparency of the board makes that almost all FreeCell games are solvable. Of the original 32000 different starts of the FreeCell game (the Microsoft 32K variant), only one is deemed impossible to solve, making approximately 99.99% of all FreeCell deals solvable [172].

As seen in Fig 2.2, the playing board consists of three parts. The large section at the bottom is called the build stack, where all fifty-two cards reside at the start of the game, divided over eight stacks. The part at the top left of the board is called the storage stack. Here, a card can be temporarily stored during the game. The last section of the board, at the top right, is called the suit stack. The goal of the game is to move all the cards here. Playing cards comprise four suits: clubs, diamonds, hearts, and spades. On the corresponding suit stack, the cards need to be placed per suit in ascending order: starting with the ace, then two, three, etc., ending with the king. When all the cards are placed on the suit stack, the game is won.

To accomplish this goal, some rules have to be followed. It is allowed to move cards from one build stack to another if 1) its rank is one lower than the current top card of the pile and 2) of the opposite color. For example, a nine of (red) hearts can be placed on a ten of (black) clubs. The general rule is that only one card is allowed to move at once. However, cards moved on top of each other with alternating colors and descending rank are allowed to move to a

new location, given that there are enough free spots on the build stack and/or storage stack. The maximum number of cards that are allowed to be moved in one single move can be calculated with the following equation:

$$movablecards = (1 + freespots) * 2^{freecolumns}$$

.

Figure 2.2: The FreeCell Board Space.

2.3.2 From Game to Digital Biomarker

As aforementioned, measuring digital biomarkers from COTS games can be troublesome since these games simultaneously require multiple cognitive functions, in contrast to gamified tests, which are custom made to capture a specific cognitive function. Therefore, it is imperative to outline the methods that were used to explore, extract and define specific digital biomarkers from gameplay. In order to translate gameplay into digital biomarkers, we applied a methodical approach existing of three phases. We first started by creating an exhaustive list of game events. In the second phase, we converted them into player mistakes. In the third phase, we quantified these mistakes to transform them into possible digital biomarkers.

For the first phase, two researchers in the field of human-computer interaction (KG and VV, co-authors of this paper) and two master students (JK and CD,

also co-authors of this paper) created a list of all possible game events for the game FreeCell. The literature on the topic of FreeCell and its rules was gathered, studied, and processed [145], [172]–[176]. This literature ranged from optimal solvers to previous cognitive studies to hardness analysis. This gave insight into the common pitfalls, optimal solving strategies, and cognitive studies previously done on the subject. Next to this literature study, the game was played in several sessions. In a series of iterations, a list of game events was drafted and refined until no more game events were found. Through this processed literature, in combination with the information gathered through the extensive gameplay, a thorough, comprehensive list of game events was generated. These game events consisted, among others, of game outcomes (e.g., game won or lost), player moves (e.g., storing a card in the storage), and incorrect player moves (e.g., placing a card on another card with the same color on the build stack).

In the second phase, to reduce this list and prevent duplicate records for the same event (e.g., positively and negatively phrased game outcomes), all game events were converted as player actions that may be indicative of cognitive impairment. For example: 'User makes a correct move' was translated into the player mistake 'User makes an incorrect move'. Next, player actions were further specified. For example, 'User makes an incorrect move' was further detailed into 'User makes a rank error' and 'User makes a suit error'. This resulted in a set of 16 possible player actions indicative of cognitive performance. Next, this set of player actions was reviewed again, and only those actions that can be captured unambiguously via playing behavior were retained. Therefore, player actions that required insight into the current mindset of the player were not captured. Additionally, only player actions that are unquestionably erroneous remained. For example, "Player stores a card with no clear advantage." was omitted as well.

Finally, in the third phase of our systematic approach, these remaining player actions were quantified. In other words, for each player action, the measurable element was determined as well as the type and range of (i.e., the game outcome that is measurable on a quantitative scale). These final elements are considered as potential digital biomarkers as they can be unambiguously captured and are potentially influenced by cognitive status. For FreeCell, 10 digital biomarkers were defined. Next to these biomarkers, metadata concerning the games and moves are captured. Table 2.1 shows all 10 digital biomarkers and metadata as captured in FreeCell.

All digital biomarkers are designed to measure at the lowest level possible. In this manner, they can become the building blocks of more complex composite digital biomarkers. For example, digital biomarkers such as 'Think time before making an erroneous move' or 'longest error streak', can always be extracted

as a combination of these original digital biomarkers. Furthermore, metadata of the x- and y- coordinates can be used to calculate the speed of moves. By capturing this information at the lowest level, there are few limitations on the number of derivatives or combinations of digital biomarkers.

2.3.3 Efficiently Capturing Digital Biomarkers

To efficiently capture digital biomarkers, the toolkit should not process images when the user is not interacting with the game (e.g., thinking of the next move). Therefore, event-driven interrupts are programmed to ensure optimal performance. These events are triggered by the user and consist of a combination of keyboard, controller, or mouse input. As FreeCell is solely played with the left mouse button, the event-driven interrupts consist of left-clicking, double-clicking, and dragging.

The general program flow can be seen in Fig 2.3. First, the program waits for user events. Secondly, when these events arise, the program immediately captures the screen and corresponding user input. This combination is crucial to determine the action and outcome of the user. Thirdly, image processing is utilized to capture any visual cues of the outcome (e.g., the program cards that have been moved). Finally, the user input is combined with the visual cues of the game to evaluate the event. It is crucial that when such a game event happens, the program stops any calculations it is doing at that point (e.g., processing previous moves). To solve this, three threads are started at the beginning of the program. The first thread is the main thread; it processes inputs and screenshots, stores the digital biomarkers, and handles other critical information such as coordinates of important locations. The second thread is purely dedicated to listening and capturing user input. The third thread is triggered by the user input and captures the next stable screenshot. This last thread has the highest priority of all threads, as capturing the screen as soon as possible after the user has made a move is crucial to determine the outcome.

2.4 Implementation

This section contains information concerning the implementation of the program. First, the setup is illustrated, explaining all necessary steps to play the game. Secondly, the card region extraction algorithm is clarified. Thirdly, how single cards are extracted is described. Fourthly, the card classification algorithm is explained. Finally, performance metrics of the program are given.

Table 2.1: Digital Biomarkers and Metadata captured in FreeCell.

Digital Biomarker	Explanation	Value
Suit Error (SE)	This error is prompted when a card is placed on another card with incompatible suits.	total
Rank Error (RE)	This error is prompted when a card is placed on another card with incompatible ranks.	total
Moved Too Many Cards Error (MMCE)	This error is prompted when a card or a group of cards is moved when there is not enough room to execute said move.	total
Unmovable Card Error (UCE)	This error is prompted when the user tries to move a card which is unmovable (i.e. there are still cards above the card that need to be moved before the original card can be moved).	total
Think Time (TT)	Think Time is defined as the time between the last card placed and the first card touched to make a new move.	ms
Move Time (MT)	This is the time necessary for a user to move a card from one place to the other.	ms
Game Result (GR)	The outcome of the game, whether the user was able to place all cards on the four suit stacks and won the game.	WON/LOST
End of Game (EoG)	Whether the user gave up or the game indicated that there were no more moves.	YES/NO
Number of undo's (NU)	The number of undo's requested by the user.	total
Number of hints (NH)	The number of hints requested by the user.	total
Move Details	Metadata of each move is stored such as x- and y-coordinates, the selected card, source location, destination location and the number of cards moved.	x-coordinate, y-coordinate, rank/suit (e.g. 5H for five of hearts), location (0-15)
Game Information	Metadata concerning the game: difficulty of the game, seed to generate the deal, the starting time, and the end time of the game is logged	Easy/Normal/Hard, seed number, UNIX Timestamp

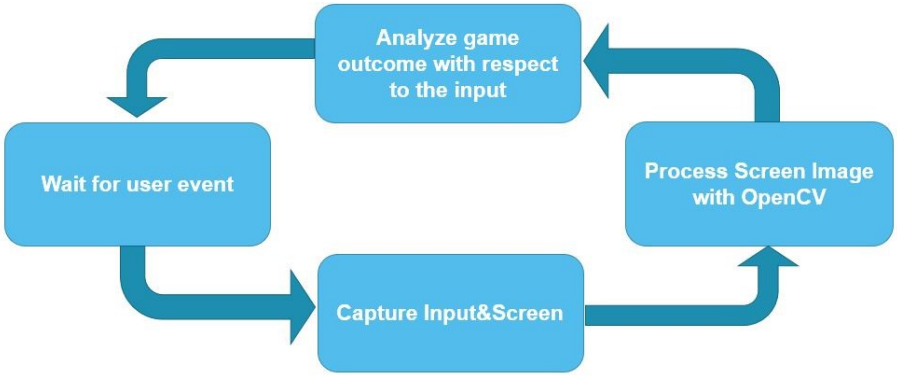


Figure 2.3: General program flow.

2.4.1 Setup

Once the program is activated, it starts the Microsoft Solitaire Collection. From there on, the program continuously monitors the state it is in: playing, choosing a game, selecting a difficulty, starting a game, ending a game, etc. To determine the state, the program follows the state diagram as illustrated in Fig 2.4. Depending on the specific state of the game, interactions of the user will be interpreted in a different way. For example, in the PLAYING state, double-clicking will trigger an event to detect changes in the playing board state. While in the MAINMENU state double-clicking is ignored. To prevent essential board information from being obfuscated by pop-ups or animations and to ensure move stability, hints need to be turned off in the settings, as well as single tap to move, alerts, tutorial, background animations, and end animations.

As some players may have the game open prior to launching the program, the game does an initial check on whether the starting state is MAINMENU or PLAYING. The central state of the program is the PLAYING state. Unless the player accesses the menu, requests hints, or starts a new game, the program will just follow the natural game progress. The program processes these changes in state by detecting button clicks. For these buttons, the dynamic position of the button is extracted through contour detection.

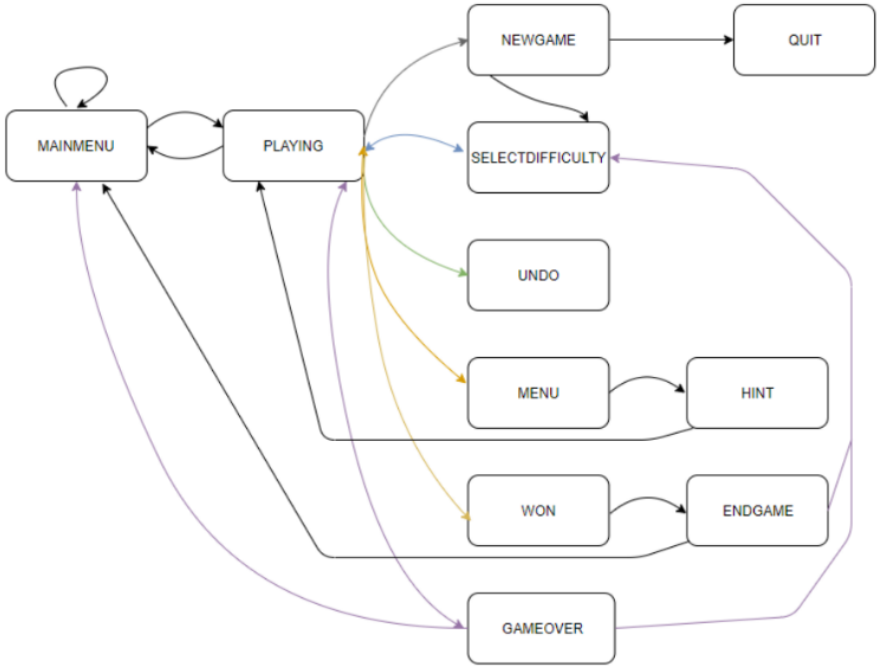


Figure 2.4: State diagram of the program.

2.4.2 Card Region Extraction

At the beginning of a game, each card is clearly visible. However, as the game progresses, some card stacks tend to grow larger and cards tend to overlap (Fig 2.5). Due to this overlap, each card needs to be extracted (i.e., the visible portion of each card needs to be separated) and classified (i.e., the rank and suit need to be determined) at the start of the game. This way, a model of the entire playing field is mapped. During the rest of the game, only the top cards are extracted and classified to monitor the progress of the game.

To extract all cards, the screenshot of the board will be split into different pieces according to the card regions. For FreeCell, as aforementioned, there are eight regions on the build stacks, four on the storage stacks, and four on the suit stacks. The coordinates of these 16 card regions will be used to determine the actual region the user interacts with. To find these regions, the screenshot of the board is first converted to grayscale, and next, by thresholding converted

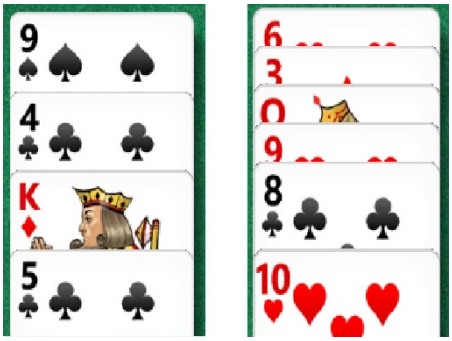


Figure 2.5: At the start of the game, the rank and suit of each card is clearly visible (left). After the game progresses, the rank and suit can be hidden by overlapping cards (right).

to a binary image (Fig 2.6). Thresholding is an image segmentation technique that creates a binary image of a grayscale image based on a manually selected threshold. Then, all the contours of this image are found using the contour detection algorithms of OpenCV. Contours are the curves of continuous points that have the same color, or intensity [171]. After filtering, only the contours that are larger than the size of a card (i.e., the card stacks) remain. We can draw a rectangle around each of these contours. The coordinates of these rectangles are then stored in a vector. This way, we can compare them with the coordinates of each click, thus being able to detect the cards the users interact with.

2.4.3 Unique Card Extraction

With the card regions defined, each card can be extracted and classified. The width of the cards is defined based on the width of the card region contours. As the ratio between the card width and height is resolution independent, the height can be inferred from the width. Then, the screenshot is divided into sixteen card regions as seen in Fig 2.7.

Since cards are stacked, not all are completely visible on the screen. At the start of the game, cards that are partially visible have an aspect ratio of 0.4 (width over length). This way, if we extract card images of 0.4 times the card height, we can extract the card sections with the rank and the suit clearly visible (Fig 2.8). These sections are split in rank and suit using contour detection. They are stored in separate vectors and are ready for classification.

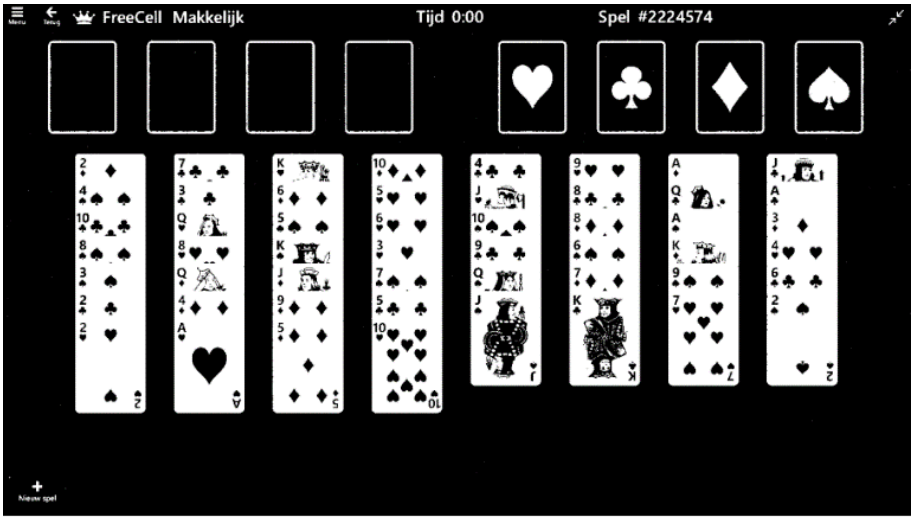


Figure 2.6: A threshold image of FreeCell. All relevant information is shown in white while all noise is eliminated.

Figure 2.7: The 16 Card Regions of FreeCell.

Figure 2.8: An extracted card region with the first card extracted through the aspect ratio.

2.4.4 Card Classification

To classify the rank and suit, three algorithms are applied, as shown in Fig 2.9. First, the contours of the rank are extracted as individual images. Secondly, the images are rescaled to a standard size (40x50 pixels for ranks, 50x50 for suits). Finally, the images are converted into a binary black and white image. These binary images are classified using a k-nearest neighbors classifier trained on different sets of rank and suit images. With every rank and suit classified, a digital representation of the board can be build, and consecutive player actions can be interpreted.

2.4.5 Performance

To test the resource efficiency of our toolkit, a performance test was done on a computer (8GB RAM, i7 Intel Core 2.7 GHz). The extraction and

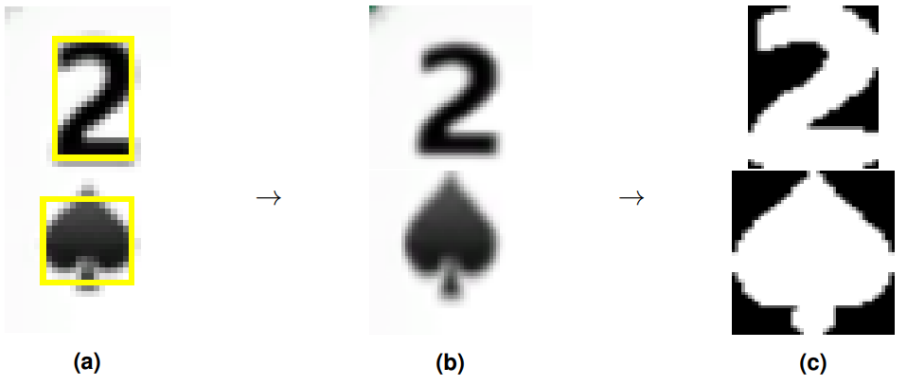


Figure 2.9: (a) detecting the contours of the rank and suit (b) Extracted rank and suit. (c) The binary image of rank and suit ready for classification.

classification performance of the whole board was tested over 10 different boards. Classification of the whole board, necessary at the start of the games, took on average 51ms. The classification of all top cards, necessary for interpreting moves, took on average 32.1ms. Finally, the extraction and classification of the game seed number on the top right took on average 260.3 ms. This performance allows for near real-time evaluation of gameplay, with response times that are lower than the theoretical limits of what can be perceived by players [177]. This was confirmed during the exploratory study (see chapter V). Participants of the exploratory tests did not notice any interference of the program while playing the game.

2.5 Exploratory Data Analysis

To explore the potential of the toolkit for capturing digital biomarkers, a first exploratory study was conducted. Digital biomarkers were captured from three different age groups. According to literature on cognitive performance, age-related cognitive decline is a natural process [178]–[180]. Primarily working memory, motor control, episodic memory, spatial ability, reasoning, and processing speed deteriorate as people grow older [178], [181]. In other words, people need more time to complete tasks and find it harder to keep important information in mind. Hence, the impact of age on player actions through digital biomarkers was explored as it may be indicative of cognitive performance. We aimed to explore whether digital biomarkers could discriminate among age

groups and possibly show a decline for the older groups. The biomarkers were categorized into three groups: Time-related, Error-related, and Outcome-related Digital Biomarkers.

2.5.1 Method

Digital biomarkers were captured from users across three different age groups (18-25,45-55,65+). The first age group, from now on referred to as youth, contained 21 participants. The middle group, referred to as middle-aged, contained 12 participants. The oldest group, referred to as elderly, contained 11 participants. Each of these participants lived independently, had no known cognitive impairments or prior cognitive complaints. In addition, all participants were new to FreeCell.

As they had no previous FreeCell experience, each participant was first briefed about the rules and mechanics of FreeCell via a fixed presentation. After this presentation, each participant got to play a practice game (seed number #25001) [182]. During this practice game, questions were allowed concerning the game rules. After this practice game, each participant played the same identical games (seeds #34898, #2365418, and #8840193). The choice for identical seeds eliminated differences in-game performance due to the chance of having a more 'generous' deal. During these three games, questions were not allowed and players continued playing until they either finished the game, the game ended because of a lack of possible moves or until the user deemed that he/she was stuck and requested to end that game.

2.5.2 Results

This data was visually explored to give insight into age-related playing differences. The goal is to show the possibilities of the toolkit. The information is divided into three distinct categories: time-related digital biomarkers, outcome-related digital biomarkers, and error-related digital biomarkers.

Time-related Digital Biomarkers

We were most interested in the digital biomarkers related to time spent thinking before making a move, as this can possibly correlate to important cognitive functions for daily activities such as attention, executive function, and planning. The 'young' age group has an average think-duration of 6871.84ms (sd: 2467.59 ms). The middle-aged group has an average think-duration of 10383.40ms (sd:

5816.19 ms), while the 'elderly' have an average think-duration of 13423.65ms (sd: 7089.50 ms) 2.10.

We explored the difference in time spent thinking before an erroneous or successful move, as seen in Fig 2.11 and Fig 2.12. For successful moves, players in the youth category thought on average 6805.76 ms (sd: 2402.36 ms), players in the middle-aged category 10755.85 ms (sd: 5840.93 ms); and the oldest group 13241.61 ms (6750.04 ms). For erroneous moves, players in the youth category thought on average 7337.47 ms (sd: 4881.20 ms), players in the middle-aged category 8487.77 ms (sd: 5462.14 ms); and the oldest group 15448.09 ms (13395.10 ms). For all age groups, except for the middle-aged group, time spent thinking before making a successful move was shorter than for an erroneous move. The average think-duration of each move in time is shown in Fig. 16. The x-axis indicates the move number of the game, meaning, the first value on the x-axis corresponds to the first move of the game. The y-axis corresponds to the average think-duration of that specific move. Concerning Move Time, as shown in Fig 2.14, people in the young category took on average 1578.82 ms (sd: 809.52) to move a card. For the middle-aged category, this was 1661.28 ms (sd:791.49 ms). The oldest category took 2103.04 ms (sd: 1298.41 ms) on average to make a move.

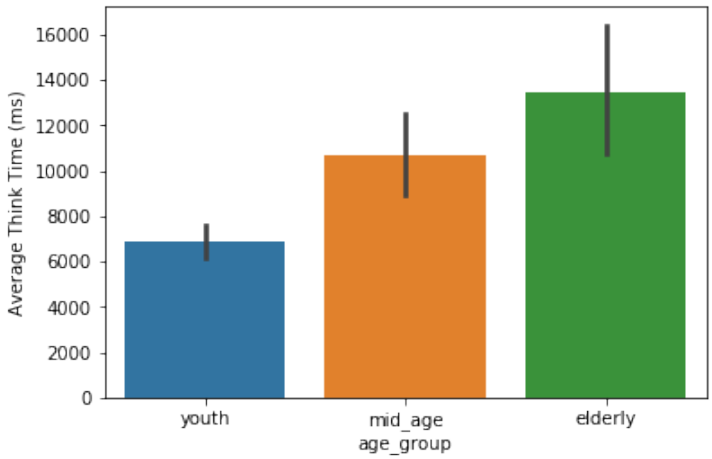


Figure 2.10: Average Think Time. The vertical line resembles the 95% confidence interval.

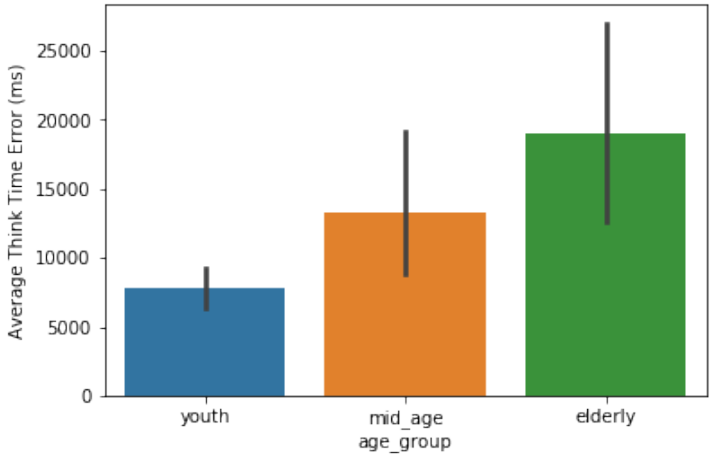


Figure 2.11: Average Think Time for an error.

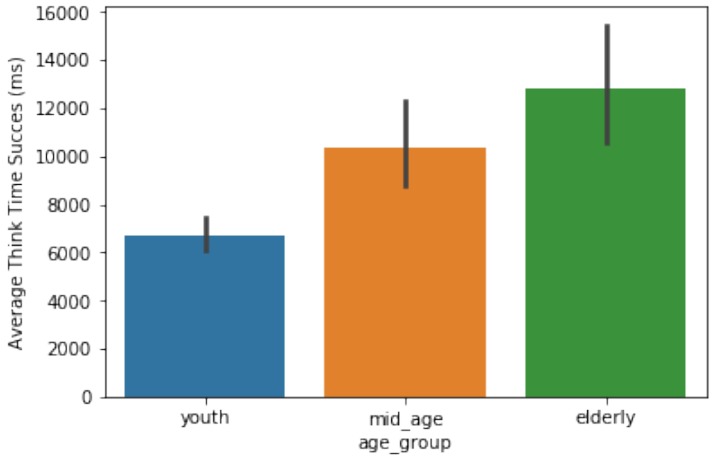


Figure 2.12: Average Think Time for a successful move.

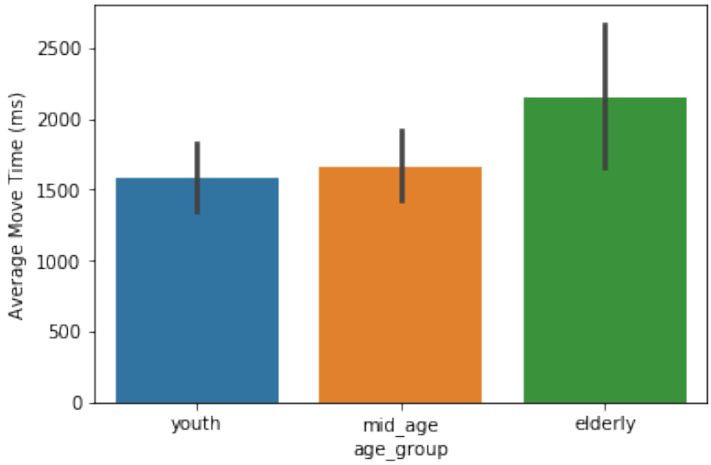


Figure 2.13: Average Move Time.

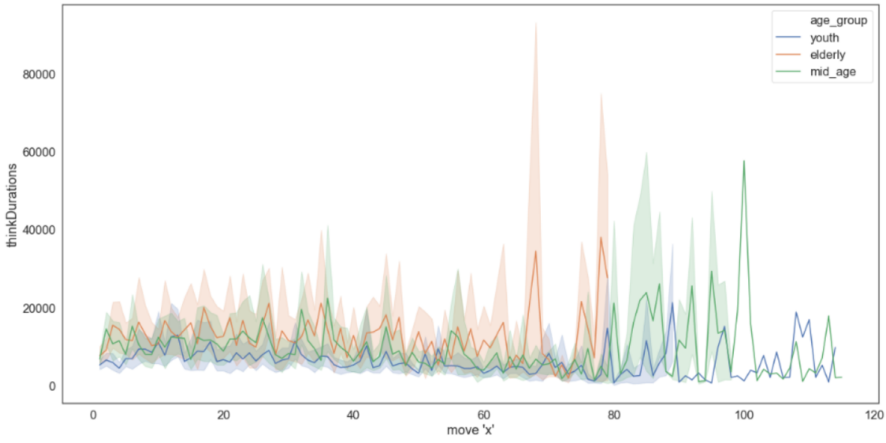


Figure 2.14: Average Think Time on the xth move.

Error-related Digital Biomarkers

Errors made during gameplay may be indicative of planning, executive functioning, and attention as players are required to think ahead, processing the next couple of moves. The average total amount of errors made by each age group can be seen in Fig 2.15. On average, the youth group made 12.2 mistakes, the middle-aged group 12.4 mistakes, and the elderly made 7.1 mistakes. Concerning Rank Errors, the youngest group made 5.3 errors on average, the middle-aged group 5.0 errors, and the oldest group 3.7 errors. For Suit Errors, 3.5, 3.6, and 1.6 errors were made on average for the youngest, middle, and oldest category. Regarding Unmovable Card Errors, the youngest, middle-aged, and oldest categories made 1.2, 1.3, and 0.5 errors respectively. Regarding the Too Many Cards Moved errors, the youngest age group made 2.3 errors on average, the middle-aged group 4.3 errors, and the eldest group 2.5 errors. For requesting hints, an average of 0.04 was found for the youngest group, 0.25 for the middle-aged group, and 1.00 for the oldest group. For correcting unwanted moves, on average, 0.86 was found for the youngest group, 0.31 for the middle-aged group, and 0.39 for the oldest group.

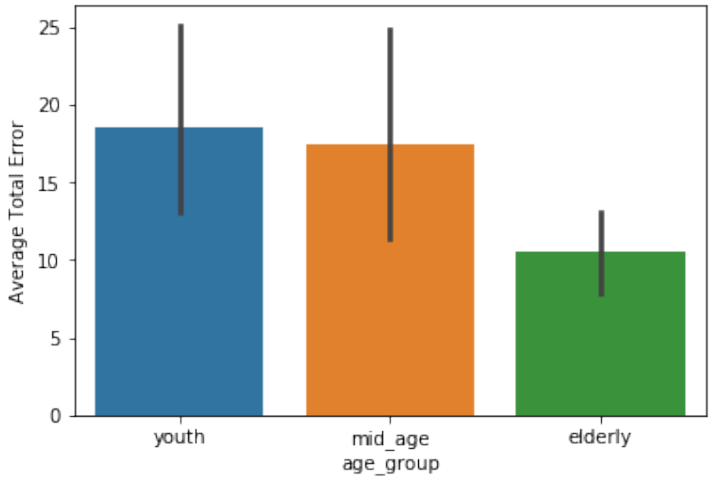


Figure 2.15: Average Total Errors.

Outcome-related Digital Biomarkers

Fig 2.16 shows the percentages of the game won and lost. The results display that the win rate decreases with the increase of age group. The percent of games won by youth, middle-aged, and elderly is 91.3%, 77.1% and 60.9% respectively.

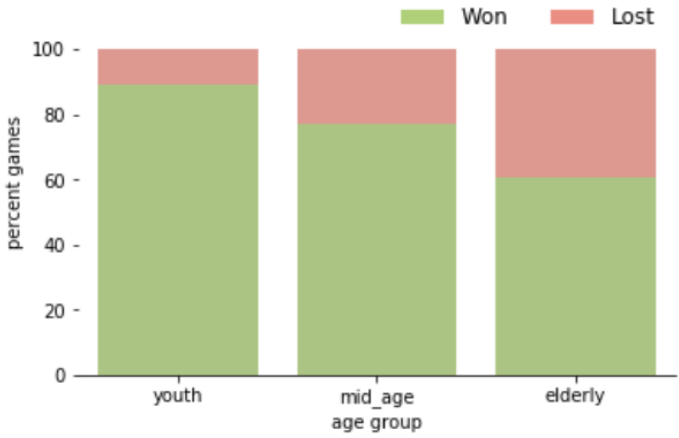


Figure 2.16: Total Percentage of games won.

2.6 Discussion

Today, many cognitive impairments go undiagnosed, and those patients who have been diagnosed have sparse follow-up moments with neuropsychologists due to restricted time and funding. This problem may be mitigated by adding digital biomarkers to the toolbox of neuropsychological assessment. Previous research has shown that daily interactions with technology can provide a trail of information on cognitive performance. This may be an efficient method of gathering digital biomarkers whilst reducing the effort from the user and the healthcare system. Such digital biomarkers can help fill in the gap between consultations and can potentially help in screening, diagnosis, and prognosis of cognitive health.

In this paper, we explored a novel way of capturing digital biomarkers via gameplay of card games, by utilizing image processing. The contribution of our work lies in the presentation of the generic toolkit using image processing,

and a first exploration of player actions that can be indicative of cognitive performance.

We built a generic image processing toolkit for card games to collect digital biomarkers on cognitive health. We chose the Microsoft Solitaire Collection, coded general image processing algorithms to detect cards, and implemented game rules for two games: Klondike Solitaire and FreeCell. As these card games are popular amongst young and old, and weaved in the daily lives of people, it may be a good fit to capture digital biomarkers. Modularity of the program was kept in mind while programming, such that this toolkit can easily be extended to other card games such as TriPeaks and Pyramid. The card detection algorithm is generalizable for all card games in the Microsoft Solitaire set, only the game rules need to be implemented. Performance tests showed that this toolkit is able to capture digital biomarkers at real-time with minimum stress on the computer. All participants of the exploratory tests did not notice any interference of the program while playing the game. No visual or auditory queues came up. The program did not stress the performance of the computer, gameplay remained as smooth as if the toolkit's software program was not there. However, technical improvements can still be made to the toolkit. Up to date, threshold values are manually selected. This can be set automatically using techniques such as Otsu Thresholding, making the program more resilient to changes [183]. Furthermore, currently, all animations need to be turned off to ensure image stability during processing. Hence, updates could be made to make the toolkit more durable and less susceptible to animations. Moreover, more advanced machine learning techniques, such as deep learning, can be explored to further improve the robustness of classification [184]–[186]. The toolkit can also be adapted to other 2D-games with minor adjustments. New game rules need to be implemented and the machine learning models need to be retrained to detect new targets. Furthermore, new digital biomarkers need to be defined as not all digital biomarkers from this study are generalizable for all games.

As a first exploratory study, data from 44 participants from three different age groups was captured. In this paper, we limited ourselves to visualizing the data and descriptive statistics. Results from this exploratory study suggest that data gathered via the generic toolkit can discriminate among different age groups of cognitively healthy participants and possibly provide information on cognitive performance. At a group level, all time-related digital biomarkers show a steady decline the older the age group. This can be expected as cognitive functions critical to cognitive aging such as processing speed and working memory tend to decline [187]. As expected, the older the age group, the fewer games were won on average. However, for Error-related digital biomarkers, the reverse was true. Older adults made fewer errors than their younger and middle-aged

counterparts. This could indicate that older adults need more time to think of a move but make their moves with more caution.

However, as this is a first, exploratory investigation, this study also has its shortcomings and any interpretations need to be done in a conditional manner. First, the groups were small and unbalanced, making results not generalizable to a wider population. Secondly, we compared cognitive healthy age groups as opposed to groups with cognitive impairments. Thirdly, differences were found at the group level only, no investigation was carried out at the individual level. If daily interactions are going to be predictors of cognitive performance, results should be obtained at the individual level. To this end, data should be captured over a longer period of time. Results should be compared inter-group and intra-individually. In this manner, a more accurate analysis of digital biomarkers as bearers of cognitive information can be performed (i.e., improving sensitivity and specificity). Hence, in the future, data should be captured over a longer period of time from larger populations and populations with cognitive impairments. Ultimately, with the help of machine learning models, cognitively healthy participants could be discerned from their impaired counterparts on the basis of multiple combined digital biomarkers. Finally, more complex composite digital biomarkers should be explored.

2.7 Conclusion

Early diagnosis and frequent follow-up of cognitive health problems are crucial to managing disease progression, allowing for timely treatment. Digital biomarkers obtained via gameplay have the potential to aid in early diagnosis of cognitive health issues. To this end, we developed a generic toolkit for card games using image processing to capture digital biomarkers indicative of cognitive performance. First, we applied a methodical approach to define 10 digital biomarkers indicative of cognitive performance. Next, we implemented the toolkit, on the top of the Microsoft 10 Solitaire Collection, as a multithreaded C++ desktop application, utilizing the Open Source Computer Vision Library to unobtrusively monitor games. Performance tests showed that this toolkit is able to capture digital biomarkers in real-time with minimum stress on the CPU. Finally, we conducted an exploratory user study to verify whether we can discriminate amongst different age groups, characterized by different cognitive performance due to normal cognitive aging. The results of the exploratory study suggest, at a group level, that age groups differ. Time-based digital biomarkers and outcome-related measures show a steady decline the older the age group. Although this is only a first exploratory study, the results suggest promise of

the use of games, weaved in the daily life of players, for the capturing of digital biomarkers for cognitive health.

2.8 Access to the source code

We would like to invite all researchers to build on, repurpose, and utilize this tool. All source code can be found on `https://github.com/kgielis/ImageProcessingMicrosoftSolitaireCollection`. All work based on this code should be referenced correctly. Fair use and modification is allowed, as described by The GNU General Public License v3.0.

Chapter 3

Age? It's in the Game: An Exploratory Study on Detection of Cognitive Aging through Card Games

This chapter is a copy of the previously published article:
K. Gielis, K. Verbert, J. Tournoy, and V. Vanden Abeele, "Age? It's in the Game: An Exploratory Study on Detection of Cognitive Aging through Card Games", en, in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, Barcelona Spain: ACM, Oct. 2019, pp. 651–664, ISBN: 978-1-4503-6688-5. DOI: 10.1145/3311350.3347193

Scientific Contribution:
As first author, I lead the writing of the first draft of the manuscript and processed suggestions of co-authors. In addition, I coordinated the data collection and full analysis.

3.1 Introduction

The western world population is aging rapidly; the proportion of people over 60 years old will grow from 12% to 22% between 2015 and 2050 [188]. As overall life expectancy increases, so increases the importance of scrutinizing the complex relationship between cognition and the aging brain. Even though most people will not experience dementia, when growing older, they will experience at least some subtle cognitive changes associated with aging. Cognitive aging is defined by Blazer as *"the process of gradual, ongoing, yet highly variable changes in cognitive functions that occur as people get older"* [13]. Cognitive performance in older adults may improve in areas building on accumulated knowledge and experience. Yet, it may decline in other areas related to memory, attention, delayed recall, processing speed, and executive function [189].

Cognitive aging is a natural, lifelong process and not a disease. Hence, cognitive aging is different from cognitive decline due to neurological diseases such as Alzheimer's disease, cerebrovascular accidents (CVA), or Parkinson's disease. Still, the gradual dwindling of certain cognitive functions will eventually lead to reduced performance on more complex, instrumental activities of daily living, such as driving a car, optimizing financial decisions, or complying with therapeutic regimens [13]. Hence, a better understanding of the different changes in cognition and functioning that accompany aging contributes to taking timely measures and seeking support for healthy and independent living. Moreover, a better understanding also helps setting normal cognitive aging apart from pathological decline such as manifested in Mild Cognitive Impairment, the precursor to Alzheimer's [190].

To address the "grand societal challenge" of cognitive, yet healthy aging [13], [14], governmental organizations have called for an increase in the research and development of tools for the longitudinal assessment of cognitive aging and the charting of cognitive aging trajectories. More specifically, there is a call for the identification and validation of novel tools that capture cognitive performance on real-world tasks over a longer period of time, and that are sensitive to early and subtle changes.

Digital games have long been promoted to understand, measure, and improve cognition[95]. Already in the mid-eighties, Space Fortress was developed as a serious game to measure and train memory, attention, dual-tasking ability, and psychomotor control [94]. To date, it is still used in research labs and perhaps the most notable and systematic game-based tool to understand the relationship between fundamental cognitive abilities and skill development [95].

However, it has also been reported that playing Space Fortress can be a frustrating experience, as it was developed by psychologists, not professional

game developers [95]. It provides primitive graphics, lacks an engaging story, and its level of difficulty does not adapt to the player's skill. Unfortunately, this reported lack of appeal is not unique to Space Fortress. This may equally apply to other serious games developed to assess cognitive functioning. Competing with commercial games is often not feasible, as serious game researchers do not possess similar budgets compared to AAA games developers. As a consequence, serious games may appear feeble next to their commercial counterparts that provide, among others, a superior aesthetic design, various worlds to invoke curiosity, and complex algorithms to tailor the game's difficulty to the player.

If video games are to measure cognitive aging in a natural setting and over a longer period, they must be well designed, and they must consider the preferences of players. Otherwise, engagement will suffer and adherence to longitudinal game-based assessment may be low [191]. Therefore, it has been argued to turn to commercial off-the-shelf (COTS) video games [96], [98]. COTS games, and in particular casual games, are already played often across different ages and gender [113], [192]. Playing them is perceived as an enjoyable activity in and of itself. Often, such games are part of the social fabric of the player's life [193]. Casual games may equally lend themselves to cognitive assessment and may even provide a set of 'digital biomarkers' for cognitive performance [100].

Biomarkers can be defined as *"objectively measured and evaluated indicators of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention"* [194]. Two examples of these biomarkers in the field of dementia are β-amyloid or τ in the human cerebrospinal fluid. The combined presence of these biomarkers could indicate Alzheimer's Disease [195]. Digital biomarkers, then, can be understood as "user-generated physiological and behavioral measures collected through connected digital devices that can be used to explain, influence and/or predict health-related outcomes" [18]. Digital biomarkers on cognitive performance are those data sources that help to measure and assess cognition-related functions such as executive function, attention, memory, performance, etc. [49]. Early results hint at their use in various conditions such as bipolar disorder, schizophrenia, and cognitive decline [152], [153], [155], [156].

To summarize, casual games may provide a means of monitoring cognitive performance over time and follow-up on an individual's cognitive aging trajectory. To this end, this study explores the extent to which cognitive aging can be assessed via casual games. In particular, we investigate whether we can predict age group on the basis of game data of the popular card game FreeCell, one of the most popular games of the Microsoft Solitaire collection [196]. Digital biomarkers of cognitive performance (i.e., game measures indicative of cognitive aging) were captured using an image processing toolkit, detailed and made available in previous work [121]. FreeCell gameplay data was gathered from 52

players, across three different age categories (18-25, 40-55, 65+), playing 130 games. First, game metrics (features) most indicative of cognitive aging were identified. Next, using machine learning, a model was trained to categorize participants according to age categories. The results also inform us of which game metrics are most indicative of cognitive aging. The results also suggest it is possible to distinguish younger from older players. However, accurate prediction of middle-aged players was found problematic. We discuss these findings in relation to known models of cognitive aging and conclude with the limitations of our machine learning approach.

3.2 Background

In this section, we first provide an introduction to cognitive aging. Next, the role of games for brain training and assessment is discussed. Finally, opportunities for casual games as tools to assess cognitive performance are given.

3.2.1 Understanding Cognitive Aging

As aforementioned, cognitive aging is a natural process, selectively affecting cognitive processes; some cognitive functions are more perceptible to aging than others. Most of the cognitive abilities robust to aging are classified as crystallized intelligence, whilst the most sensitive are classified as fluid intelligence [187]. Crystallized intelligence pertains to those skills that are well-practiced or to the type of knowledge that accumulates over the life span through development, educational experience, and culture [197]. Examples of crystallized intelligence are language skills, vocabulary (see Figure 3.1, vocabulary), general knowledge, or learned practices. Crystallized abilities gradually improve at a slow yet steady pace to taper off slightly at circa 70 years (see Figure 3.2, full line). Fluid intelligence, in contrast, points to those abilities involving problem-solving and reasoning on the basis of new information. Examples of cognitive domains underlying fluid intelligence are executive functioning, processing speed, and memory. Many fluid cognitive abilities peak around the age of 25 and then decline at a steady rate (see Figure 3.2, dotted line) [23], [198].

Age Sensitive Cognitive Abilities

While the above distinction between crystallized and fluid abilities is well-known, it is limited in detailing the specific changes in cognitive functioning. Therefore,

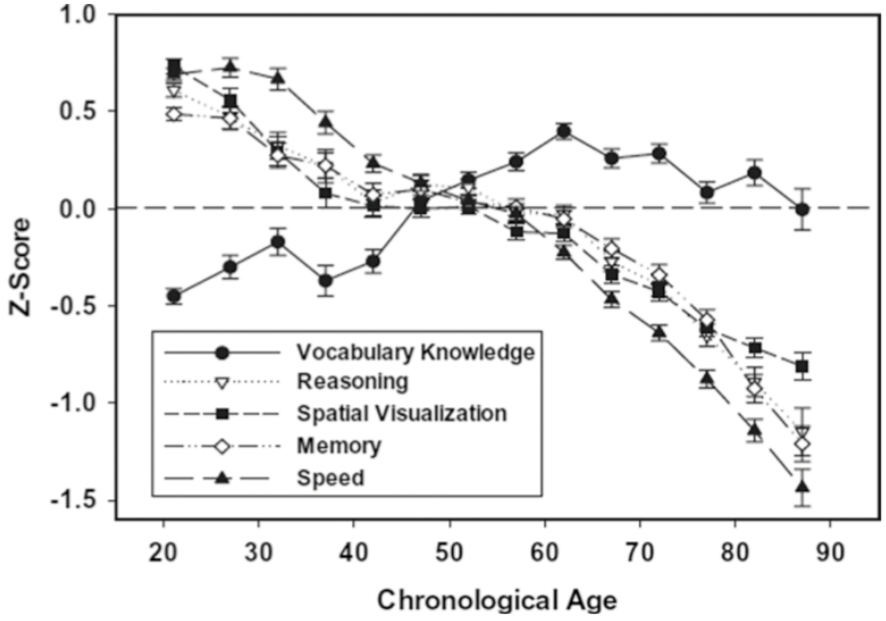


Figure 3.1: Means and standard errors for composite scores of diverse cognitive abilities as a function of age (permission for reprint granted)[23].

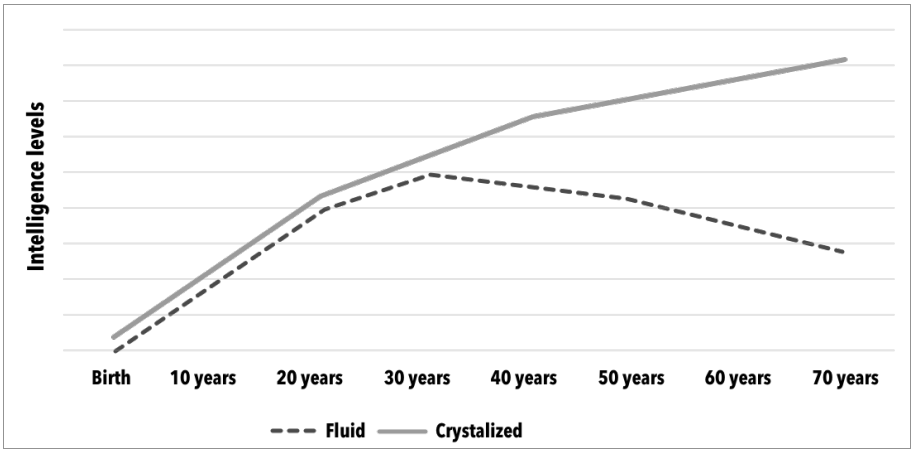


Figure 3.2: Fluid and crystallized intelligence across the age span, image adapted from [199].

we highlight the underlying cognitive domains most sensitive to aging below, based on [13], [187], [200].

Processing speed refers to the speed at which cognitive and motor activities are executed [187]. It can be seen as a measure of the efficiency of cognitive functioning and is highly sensitive to aging. Moreover, it is a building stone for other cognitive functions, as decreased processing speed hinders the efficiency of other cognitive functions such as learning, attention, speech processing, and memory (see Figure 3.1, Speed).

Attention refers to the ability to concentrate and focus on specific stimuli. It can also be seen as a measure of the capacity for processing information. Attention also deteriorates with age. More complex attention tasks are even more sensitive to aging, such as selective and divided attention. Selective attention is the ability to focus on specific information in the environment while ignoring irrelevant information. Divided attention is the ability to focus on multiple tasks simultaneously, such as talking on the phone while preparing a meal. Both divided and selective attention are particularly prone to aging.

Working memory refers to the ability to momentarily hold information in memory while simultaneously manipulating such information. Although it is often used as a synonym for short-term memory, the term working memory conveys an additional emphasis on the active manipulation of buffered information. Older adults perform significantly worse on tasks involving working memory, such as performing simple calculations.

Declarative (explicit) memory refers to the conscious, long-term process of storing and recalling information for a longer period. A decline in declarative memory is perhaps one of the most commonly reported complaints by older adults. Two types of declarative memory are particularly influenced by aging (see Figure 3.1, Memory): semantic memory and episodic memory. Semantic memory involves the capacity to store and recall concepts, numbers, or words and, in particular, knowing the meaning of the concepts. Semantic memory shows a decline only in late life, in particular in the ability to store and recall new words. Episodic memory (also known as autobiographical memory) involves the conscious recollection of previous autobiographic experiences situated in time and space and their associated emotions. Episodic memory shows a steady decline throughout aging.

Executive functioning refers to the ability to self-monitor, plan, organize, reason, be mentally flexible, and solve problems. Research has shown that such abilities to self-steer, particularly mental flexibility, decline with age, especially after the age of 70 years old (see Figure 3.1, Reasoning). Research has also shown that aging negatively affects response inhibition, which is the ability to inhibit

automatic responses in favor of producing novel responses.

Visuospatial construction refers to the ability to visually put together individual parts of an object to make a coherent whole (for example, visually assembling furniture from a box of parts). Visuospatial construction, closely related to spatial visualization, also declines over time.

While ample research has demonstrated the effects of cognitive aging on the aforementioned cognitive functions, it has to be noted that the exact moment when cognitive functions start to deteriorate and the rate at which this happens is highly variable from individual to individual [187], [200]. This variability is due to factors such as life experience, health, educational level, socioeconomic status, genetics, etc. [13]. Moreover, participation in certain activities, building cognitive reserve, and engaging in cognitive retraining are all potential approaches to achieving successful cognitive aging [201].

3.2.2 Video Games for an Aging Population

Video games for brain training

One of the activities to achieve successful aging may be video gameplay. In the past decade, we have witnessed the proliferation of commercial game-based programs, claiming to "train the brain" or "mitigate the risk of dementia" (e.g. Brain Age [202], Lumosity [203], Brain Fitness Program [204]). Many of such claims lack scientific validation and companies have been fined for deceptive advertising practices [205]. A meta-analytic video game training study published in 2018 concluded that there is, currently, no evidence of a causal relationship between video game training and enhanced cognitive ability [206]. On the other hand, a systematic review by Shah et al., reviewing the empirical evidence for commercially available game-based brain training products, lends support to findings that some commercial brain training games may contribute to healthy brain aging [207]. Other recent studies suggest that certain game-based training may improve certain cognitive functions such as cognitive control [208] or visual attention [209]. Of particular interest are the outcomes of the ACTIVE research program, the largest study on cognitive training using standardized outcome measures [210]. Spanning over 20 years and involving 2,832 participants, the findings suggest that cognitive training may have an impact on processing speed. Even though the ACTIVE trial was not really a game-based intervention, the results can support the notion of computerized brain training. These conflicting results make video game brain training an interesting yet controversial topic.

Videogames to measure cognitive aging

In this paper, we use games as tools to measure and achieve insight into cognitive aging, as opposed to the abovementioned studies that focus on using video games for the training of cognitive functions. Video games have been and are increasingly being to aid in the understanding and measuring of cognitive capacity, brain plasticity, development, and aging, and individual differences(Boot, 2015). As aforementioned, Space Fortress [94], an outcome of the Learning Strategies Program, funded by the Defense Advanced Research Projects Agency [211], was perhaps the first game to be explicitly designed to measure cognitive functions such as memory, attention, dual-tasking ability, speed, and psychomotor control in a standardized manner. Up to this day, Space Fortress is still being used by many researchers [212]. Additionally, in the past years, many game researchers have designed various serious games to understand and measure specific cognitive functions, such as virtual wayfinding navigation [213], inhibitory control [214], visual attention [215], executive function [216], episodic memory [160], etc.

Of particular interest is the recent work of Silva Neto et al. and Tong et al. [217], [218]. Silva Neto et al. created three serious games and compared the performance of these games against the various cognitive domains of the Montreal Cognitive Assessment (MoCA). They found correlations with every domain of the MoCA except abstraction, supporting the notion of serious games as a means of cognitive evaluation. Tong et al. carried out a validation study for their whack-a-mole type of game in a hospital emergency environment. Significant correlations were found between game performance and multiple cognitive assessments, including but not limited to the Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), and the Confusion Assessment Method (CAM).

Unfortunately, these serious games risk being accused of 'chocolate-covered broccoli' [219], i.e. neuropsychological tests superficially repackaged as games [95], [96]. It has been reported that participants of game-based interventions, designed to measure and train cognitive functions, lack motivation [82], [220], [221]. This lack of meaningful play becomes increasingly problematic for those games that aim to measure and understand cognitive aging not as a single-point-in-time measurement in a research lab, but rather as a tool for longitudinal measurement in the context of the home.

Casual games to provide meaningful play

To ensure meaningful play, it may be useful to turn to those games that are already played often and perceived as an enjoyable, meaningful activity in and of itself, i.e., Commercial-Of-The-Shelf (COTS) games. Characteristic for these games is that they are part of the social fabric of the gamer's life [193]. They have been appropriated and are not part of the reductionist discourse that focuses on gameplay solely to mitigate decline [107]. While players do not play these games for serious cognitive health purposes, they may still be used to measure cognitive performance. Moreover, playing COTS games is not reserved for a young audience. Contradictory to common prejudice, studies show that digital games are played across different age spans and gender[192]. Research findings consistently report that although only 10% of those who play games identify themselves as "gamers" [222], approximately 50% of all ages and genders play games. Across different demographic groups, the most common game genre are puzzle games and strategy games [223]. When analyzing game genre preferences specifically for older players and female players, again the importance of "casual" games (e.g., card games, puzzle games)[113] is emphasized.

COTS games to assess normal&pathological cognitive ageing

Today, casual games are already being used to measure and understand cognitive aging. Yet most studies rely on statistical techniques to correlate performance outcomes of games with outcomes on classic neuropsychological tests [169]. For example, Baniqued et al. [96] examined the degree to which commercially available video games tap into certain cognitive abilities, and found performance outcomes of casual games categorized as tapping into working memory and reasoning to correlate with neuropsychological test outcomes of working memory and fluid intelligence tasks.

To date, few studies attempt to model and present algorithms for inferring users' cognitive performance. Notably, Jimison et al. present techniques for monitoring computer interactions with FreeCell to detect sustained changes in cognitive performance [145]. However, the monitored user interactions were limited to the appropriateness of a move (as compared to a solver). Hagler et al. developed a computational model of executive function based on a decomposition of the gameplay data of a scavenger hunt type of game [216]. Leduc-McNiven equally explored models for cognitive performance [163]. However, it remains unclear to what extent these models have been empirically validated. Moreover, gameplay data seems to be limited to more optimal versus less optimal player moves.

Mandryk and Birk point out that a variety of activity traces gathered from

in-the-wild gameplay of COTS games can be considered as digital biomarkers of cognitive health [100]. However, when dealing with a multitude of biomarkers, this necessitates the use of computational approaches for big data. To this end, they provide a machine learning pipeline for digital game-based biomarkers of cognitive health (see Figure 3.3).

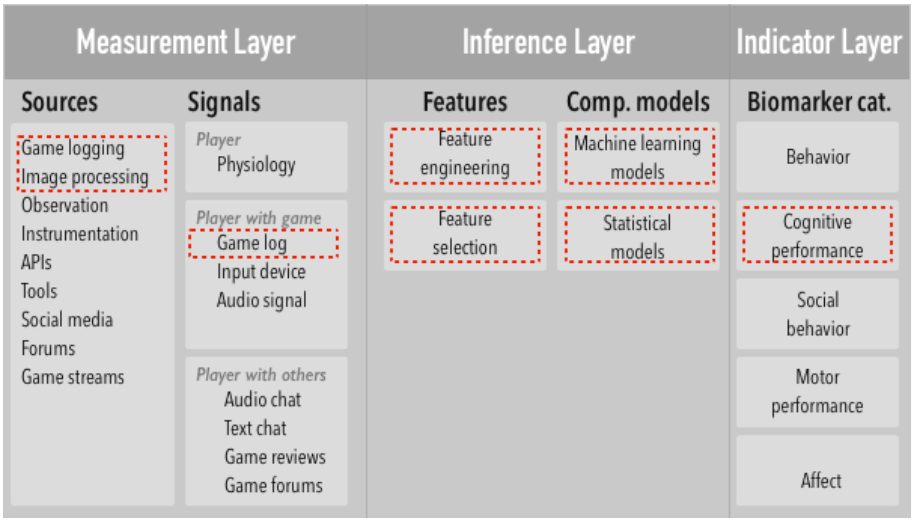


Figure 3.3: The machine learning pipeline as put forward by Mandryk and Birk [100], dotted rectangles visualize those elements used in this study.

3.2.3 FreeCell

FreeCell is a popular Solitaire variant that can be found on every Windows 10 computer in the Microsoft Solitaire Suite [182]. The game board is composed of three parts: storage, suit, and build stacks (see Figure 3.4). At the start of the game, all cards are dealt face-up.

Similar to other Solitaire games, FreeCell is won when all cards are moved to the respective suit stacks, starting with the aces and ending with the kings. To realize this, a player has to move cards. Cards can be moved from one build stack to another, only if its rank is one lower than the current top card of the pile and of the opposite color. The storage stack can be used as a place to temporarily store cards to a maximum of four cards, one per spot. The general rule is that only one card can be moved at a time. However, when there are sufficient empty places at the build stack and/or storage stack, groups of cards

Figure 3.4: The FreeCell board space.

with alternating colors and descending rank are allowed to be moved in one go to a new pile on the build stack.

Specific to FreeCell is that 99.99% of all deals are solvable [172]. Moreover, since all cards are dealt face-up at the start of the game, this makes it a game of strategy, not of sheer luck. As a result, the game is popular, particularly among older players [193]. Moreover, since the release of Windows 95, Microsoft has shipped its operating system with the Microsoft Solitaire Collection. Given the popularity of Microsoft's version of FreeCell [182], its omnipresence, and its focus on strategic planning, it was deemed a good candidate for the longitudinal collection of digital biomarkers on cognitive performance.

Therefore, this study explores the extent to which cognitive aging can be predicted via digital biomarkers captured via FreeCell play. To this end, we followed the pipeline as put forward by Mandryk and Birk [100], illustrated in Figure 3. First, digital biomarkers of cognitive performance were captured using a toolkit for image processing of card gameplay [121], from three distinctive age groups (18-25,40-55,65+). Next, features were engineered and selected. Finally, a machine learning model was trained. If successful in predicting age group, this may indicate that differences in cognitive performance due to cognitive aging can be detected through casual gameplay.

3.3 Method

3.3.1 Participant Details

The participants were recruited using a snowball sampling method from three age groups. The youngest category (18-25) will be referred to as 'younger adults, the middle category (40-55) as 'middle-aged adults', and the oldest category (65+) as 'older adults'. Inclusion criteria were that each participant lived independently, had no prior cognitive complaints or a history of mental illnesses, no medical conditions that could influence the measurement (i.e., motor and visual disorders), and, important, no prior experience in playing FreeCell, to assure we were capturing fluid rather than crystallized intelligence. At the start of each game session, demographic information concerning the participant's gender, use of the computer, and education level was inquired (see Table 3.1).

Table 3.1: Demographic information of the participants per age group.

	18-25	40-45	65+
Age	Average:21.87 SD:1.92	Average:48.69 SD:3.83	Average:70.95 SD:4.64
Gender	Male:12 Female:8	Male:7 Female:9	Male:10 Female:6
Highest Obtained Diploma	Elementary:2 High school:9 Bachelor's:3; Master's:6	Elementary:0 High school:2 Bachelor's:6 Master's:8	Elementary:1 High school:6 Bachelor's:5 Master's:4
Computer Use	Daily:18 Weekly:2 Yearly:0 Never:0	Daily:14 Weekly:2 Yearly:0 Never:0	Daily:13 Weekly:0 Yearly:1 Never:2

3.3.2 Capturing Digital Biomarkers in COTS Games

Capturing digital biomarkers was done via an image processing toolkit for card games [121], built to work for the standard Microsoft Solitaire Collection, and allowing for unobtrusive measurements of digital biomarkers. The program acts as a silent watcher, identifying every card on the game board and analyzing

every move of the player. It utilizes machine learning and image processing to analyze gameplay in real-time, extracting and processing digital biomarkers in an efficient manner. The toolkit is generalized for the entire Microsoft Solitaire Collection, meaning that the same card detection algorithms can be used for every Solitaire game on the platform. Currently, the game logic is implemented for Klondike and FreeCell, with TriPeaks, Spider, and Pyramid being developed. More information on the toolkit and its inner workings can be found in [121].

3.3.3 Digital Biomarkers in FreeCell

A methodical approach consisting of three phases was used to explore, define, and extract digital biomarkers found in FreeCell. For the first phase, an exhaustive list of game events was defined. Two researchers (KG and VDA) and two master students studied the literature on the topic of FreeCell and its rules [145], [172]–[176]. Next, the game was played for several sessions. The insights from the literature and the game sessions were distilled into a list of game events. This list was drafted and refined until no more game events were found. These game events consisted, among others, of game outcomes (e.g., game won or lost), player moves (e.g., storing a card in the storage stack), and incorrect player moves (e.g., placing a card on another card with the same color in the build stack).

For the second phase, the list was curated to prevent duplicate records for the same game event. This resulted in a set of 16 events indicative of cognitive performance. Next, this set of player actions was reviewed again, and only those actions that could be unambiguously captured via playing behavior were retained. Player actions that required insight into the current mindset of the player were not captured.

In the end, seven potential candidates to form digital biomarkers remained. Next to these, metadata concerning the games and moves was captured as well, such as the game seed, start and end time, etc. (See Table 3.2).

3.3.4 Procedure

Each game session was played in a distraction-free environment. As the participants were unfamiliar with the game, each participant was first briefed about the rules and mechanics of FreeCell via a fixed presentation. After this presentation, each participant got to play a practice game (seed number #25001) [182]. After this practice game, each participant played up to three identical games (seeds #34898, #2365418, and #8840193), data was captured

Table 3.2: Digital biomarkers and metadata captured during FreeCell.

Digital Biomarker	Explanation	Form
Suit Error (SE)	This error is prompted when a card is placed on another card with incompatible suits.	total
Rank Error (RE)	This error is prompted when a card is placed on another card with incompatible ranks.	total
Moved Too Many Cards Error (MMCE)	This error is prompted when a card or a group of cards is moved when there is not enough room to execute said move.	total
Unmovable Card Error (UCE)	This error is prompted when the user tries to move a card which is unmovable (i.e. there are still cards above the card that need to be moved before the original card can be moved).	total
Think Time (TT)	Think Time is defined as the time between the last card placed and the first card touched to make a new move.	milliseconds
Move Time (MT)	This is the time necessary for a user to move a card from one place to the other.	milliseconds
Game Result (GR)	The outcome of the game, whether the user was able to place all cards on the four suit stacks and won the game.	WON/ LOST
Move Details	Metadata of each move is stored such as x- and y-coordinates, the selected card, source location, destination location and the number of cards moved.	x-coordinate, y-coordinate, rank/suit (e.g. 5H for five of hearts), location(0-15)
Game Information	Metadata concerning the game: the difficulty of the game, seed to generate the deal, the starting time, and the end time of the game is logged	Easy/Normal/ Hard, seed number, UNIX Timestamp

using the toolkit. The choice for identical seeds minimized differences in game performance due to the chance of having a more 'generous' deal. During these three games, questions were not allowed, and players continued playing until either they finished the game, the game ended because of a lack of possible moves, or until the users deemed that they were stuck and requested to end the game. As not all participants were willing or able to complete all three games, a total of 130 games were captured for 52 persons.

3.3.5 Tools and Machine Learning Process

Data coming from the toolkit underwent several phases to optimize the machine learning model: preprocessing, outlier detection, feature engineering, and feature selection. To achieve this, Jupyter Notebooks was used with as main libraries: pandas and NumPy for data manipulation and analysis, seaborn for data visualization, and scikit-learn for machine learning [224]–[228].

Once the data was cleaned, the dataset was randomly split into test data and training data with stratified age group sampling. As is good practice, this test dataset was not used until the final evaluation, where the retrained model was put to the test by classifying this new data.

We experimented with different classifiers; multiple machine learning algorithms were trained and compared. However, a discussion of the performance of the different models is beyond the scope of this paper. Logistic Regression was the best model for our data set. This is in line with the findings that this is a classifier that works well with a smaller data set and a smaller number of features [229].

The training data was first used to tune the hyperparameters of the model and to get insight into the general performance of the model [229]. Generally, a training set is further split into a training and validation set. However, for smaller datasets (as ours), this can be problematic, as this would render the training or validation set too small. To solve this, 10-fold validation was applied, this method is visualized in Figure 3.5 [229].

Once the hyperparameters were optimized, the learning curves of the model were plotted to detect the level of over- and underfitting in the model (see Figure 3.6). Overfitting can be caused by various factors, such as too many features, too little data, too many iterations, etc. A good learning curve should have similar scores for the training as well as the cross-validation data. A high training score and a low cross-validation score can be a sign of overfitting, while both a low training and cross-validation score can be a sign of underfitting [230]. Once it was confirmed that both over- and underfitting were as low as possible,

the model was put to a final test. Scoring measures for multiclassification such as ROC-plots, confusion matrices, and performance metrics such as accuracy and F1-score were calculated.

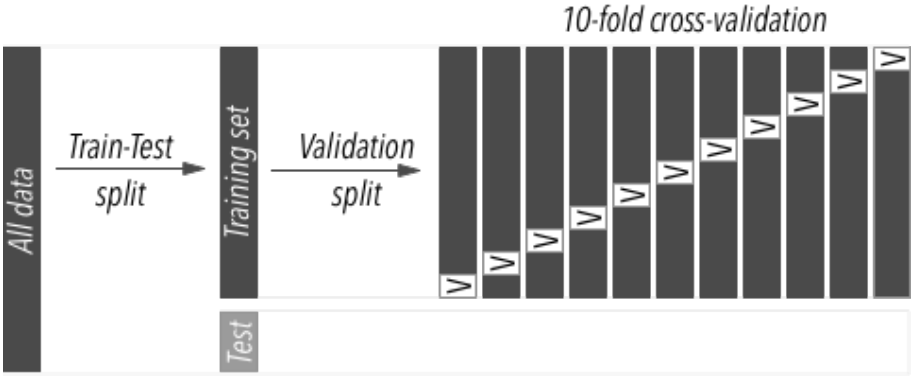


Figure 3.5: Illustration of the 10-fold validation, used in our study as recommended for smaller data sets [231].

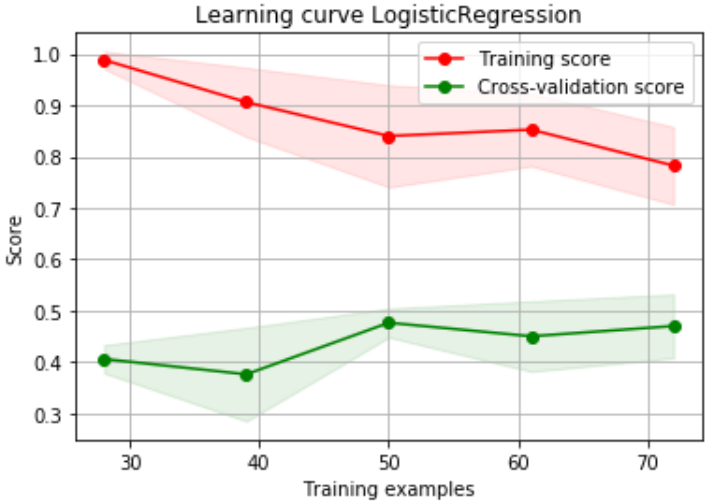


Figure 3.6: Learning curve of the logistic regression algorithm with 24 features. The discrepancy between training and cross-validation score is a strong indication of overfitting.

3.4 Results

3.4.1 Preprocessing

To prepare our data for classification, unnecessary meta-data such as usernames or the exact starting time of the game were deleted as this brings no cognitive information to the machine learning model. For every user, the correct age category was assigned.

3.4.2 Outlier Detection

Outliers were removed according to Tukey's fences [232], games with feature values that were larger than 1.5 times the interquartile range were removed. This caused 25 games to be removed, leaving 105 games up for classification.

3.4.3 Feature Engineering

Next, additional features were generated, combining the candidates for digital biomarkers of Table 3.2. These composite or derivative digital biomarker candidates were established based on insights gathered from observing the players during data acquisition. Table 3.3 provides all features used to train the machine learning models.

3.4.4 Feature Selection

As we have 24 features for a limited dataset, feature selection has to be done with care. Training machine learning models with too many features can lead to overfitting. This phenomenon happens when the model captures the noise of the data. In the end, this results in a model that performs well on the training set but performs significantly worse on the validation, such a model is not generalizable for unseen data. To inspect for correlation between features, a correlation plot was drawn (see Figure 3.7). This correlation plot shows that many errors and time related measurements are heavily correlated. Adjacent to the full correlation plot, a strip with the highest correlations of features with age group is shown. Moreover, the learning curves of the first logistic regression model on the basis of 24 features showed strong discrepancies between the training and cross-validation score (Figure 3.6). This indicates that the model is heavily overfitted and will possibly not perform well on newer observations.

Table 3.3: Overview of the 24 digital biomarkers, both basic and composite, used as features for machine learning.

Digital Biomarker	Explanation	Features
Game Result	The outcome of the game.	WON/Lost
Error Percentage	The number of errors divided by the total amount of moves	percentage
Suit Error Percentage	The number of suit errors divided by the total amount of moves	percentage
Rank Error Percentage	The number of rank errors divided by the total amount of moves	percentage
Moved Too Many Cards Error	The number of moved too many card errors divided by the total amount of moves	percentage
Unmovable Card Errors	The amount of moved unmovable card errors divided by the total amount of moves	percentage
Maximum Error Streak	The longest streak of error moves of that game	total
Number of Moves	The total number of moves made in the game	total
Number of Storage Moves	The total number of moves spent manipulating the storage	total
Total Time	Total time spent for that game	total
Think Time	The time between the last card placed and the first card touched to make a new move.	Average, standard deviation
Think Time Success	Time spent thinking before executing a successful move	Average, standard deviation
Think Time Error	Time spent thinking before executing an erroneous move	Average, standard deviation
Move Time	Time spent moving a card	Average, standard deviation
Move Time Success	Time spent moving a card before executing a successful move	Average, standard deviation
Move Time Error	Time spent moving a card before executing an erroneous move	Average, standard deviation
Cards Moved	The number of cards moved each move.	Average, standard deviation

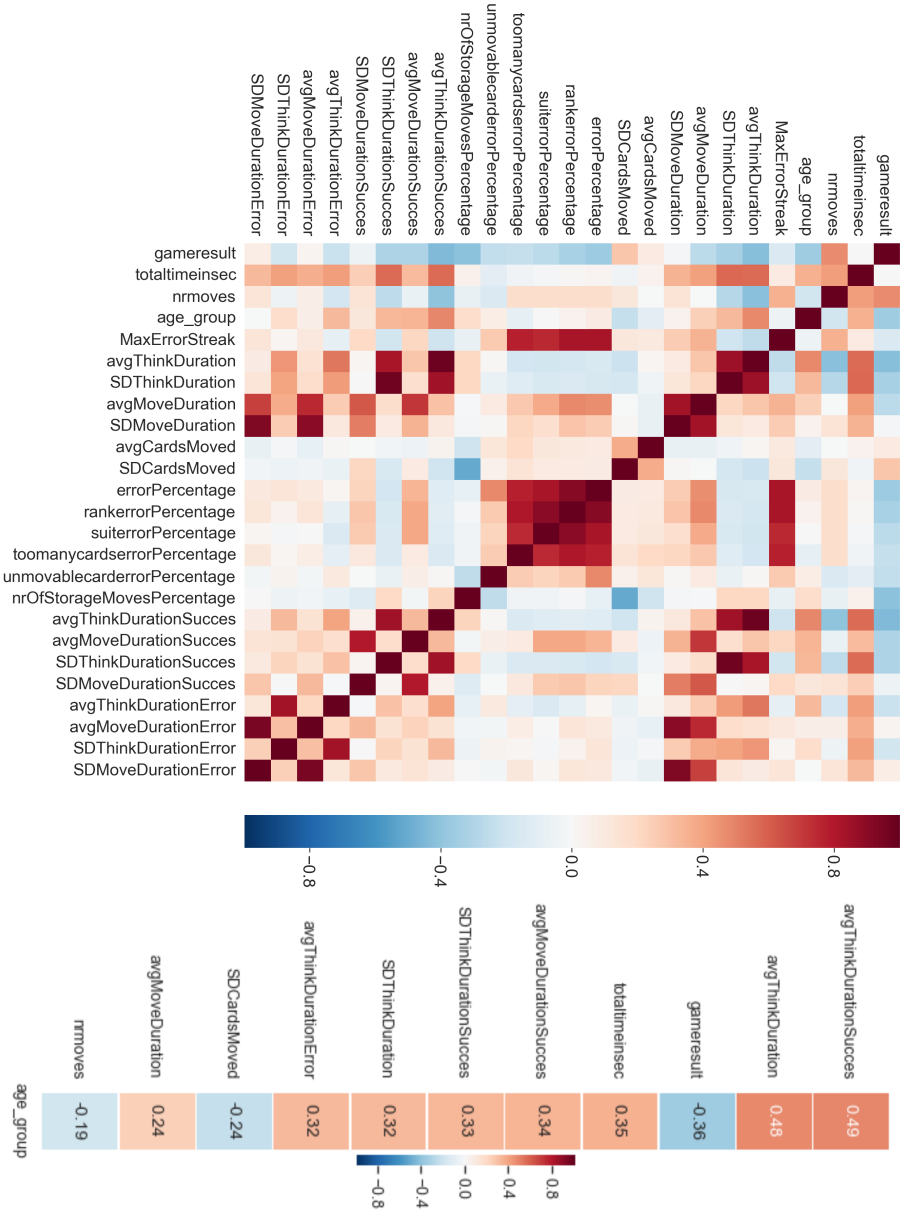


Figure 3.7: Correlation plot for the full feature set.

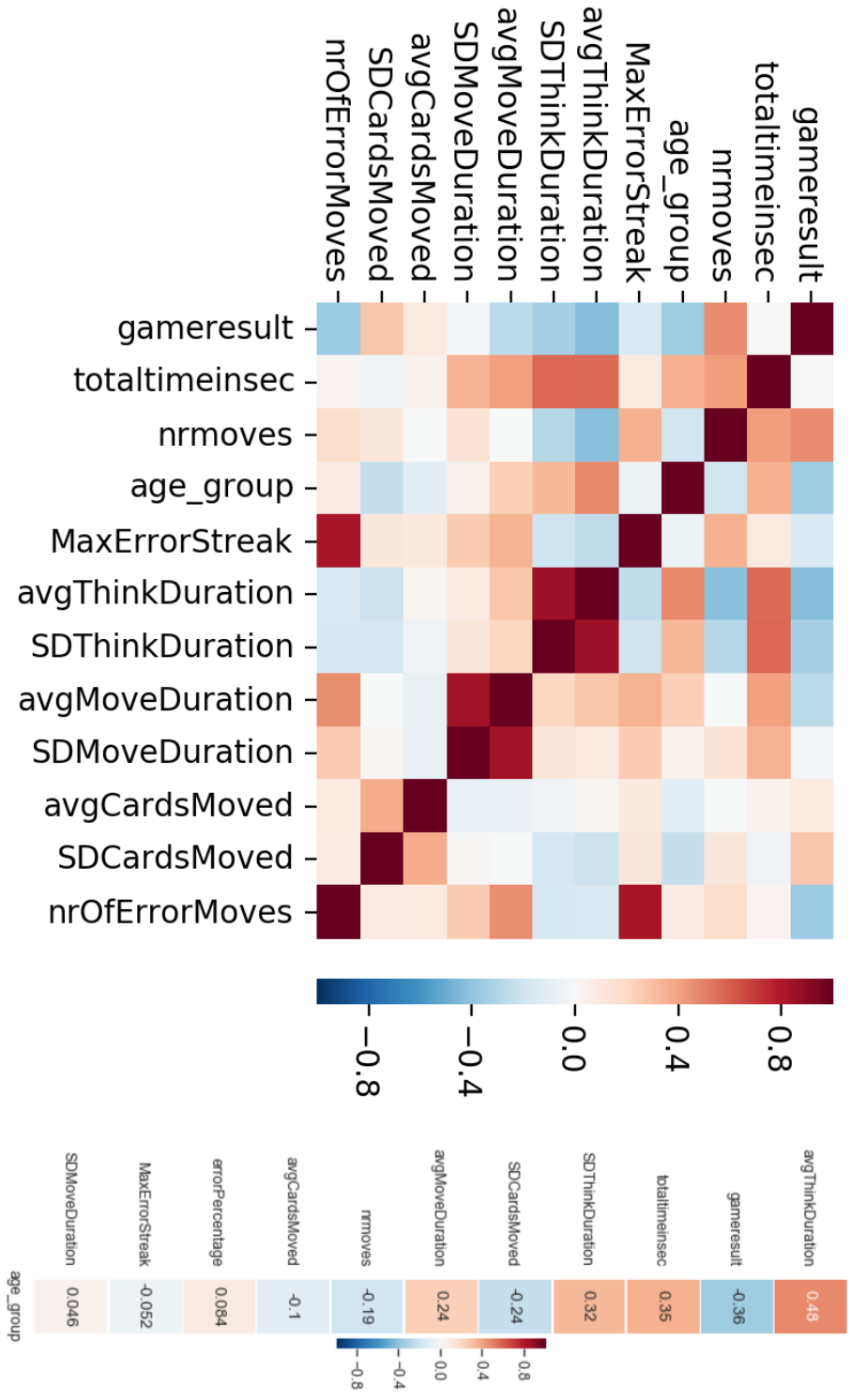


Figure 3.8: Correlation plot for the reduced feature set.

To counter this overfitting, the number of features was reduced. Upon experimentation, the following 11 features were selected based on a series of iterations: Game Result, Total Time, Number of Moves, Maximum Error Streak, Think Time (average and standard deviation), Move Time (average and standard deviation), Cards Moved (average and standard deviation), and Error Percentage. The motivation of this selection is that these features contain maximum information (e.g., error percentage contains information of all errors made) or that they contain information based on observations during game sessions and/or theories. The correlation plot for the reduced feature set can be seen in Figure 3.8.

3.4.5 Training, Validating, and Testing of the Model with 11 Features

All games were again split into train and test data with a 75% to 25% distribution. Stratified sampling was used to ensure that every age category was equally represented in both sets. To make sure every class was balanced, Synthetic Minority Oversampling (SMOTE) was applied to the training set [233]. This is a robust oversampling method that minimizes the impact on the model. A standard scalar was applied to prevent larger ranging features from influencing the model [234]. Models were trained and hyperparameters were tuned using a 10-fold cross-validation random search with accuracy as scoring metric.

Furthermore, learning curves (Figure 3.10) were plotted and inspected to detect under- and overfitting. This time the discrepancy between training and cross-validation scores was less severe. Finally, the confusion matrix (Figure 3.9), Receiving Operating Characteristic (ROC) curve with Area Under Curve (AUC) (Figure 3.11), and performance metrics (Table 3.4) were calculated to evaluate the final model.

The performance metrics of the model to discriminate the three classes range from 0.615 (Precision) to 0.734 (AUC). The confusion matrix shows that young adults can be well discriminated (nine correctly classified and two misclassified), whilst middle-aged adults (4 correctly classified and 5 misclassified) and older adults (4 correctly classified and 3 misclassified) are more often misclassified. The ROC curves further detail that younger adults and older adults can be well discriminated in one-versus-all situations, as their respective AUC's are 0.88 and 0.81. However, for the middle-aged adults versus-all situation, an AUC of 0.45 can be noted, performing below what is to be expected by chance. This lowers the overall performance of the model.

Table 3.4: Performance metrics of the test set.

Accuracy	Precision	Recall	F1	AUC
0.630	0.615	0.621	0.629	0.734

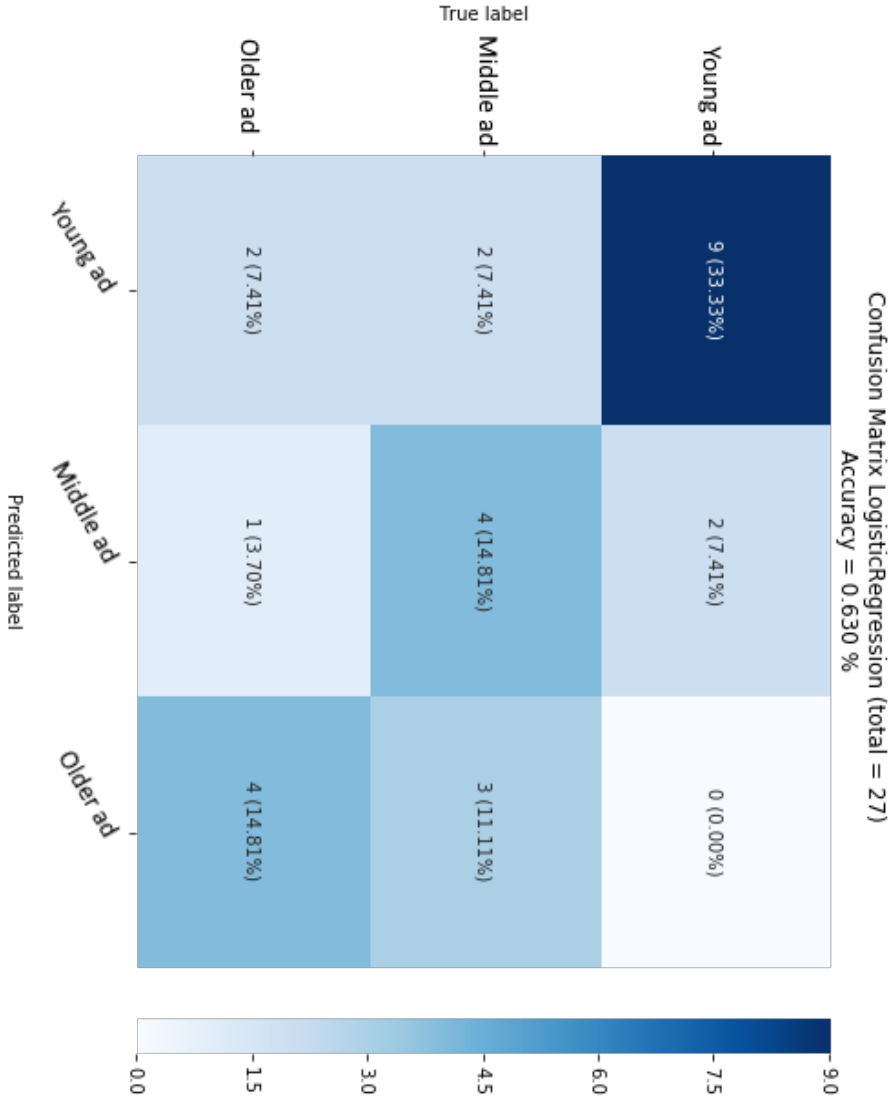


Figure 3.9: Confusion matrix of the test set.

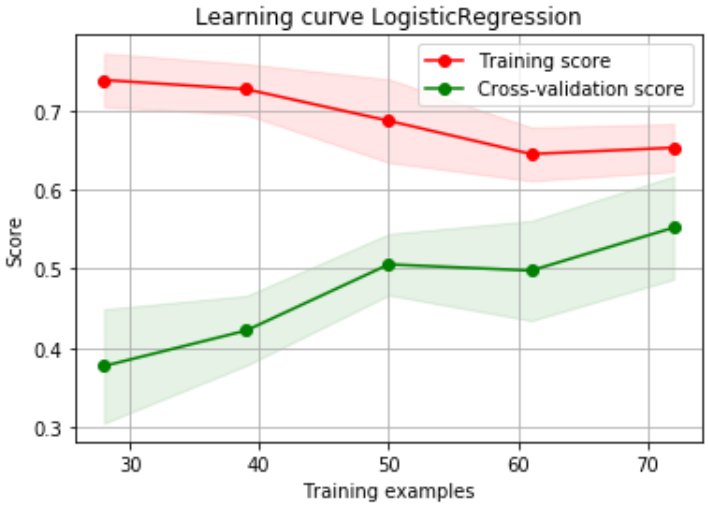


Figure 3.10: Learning curves for the optimized model.

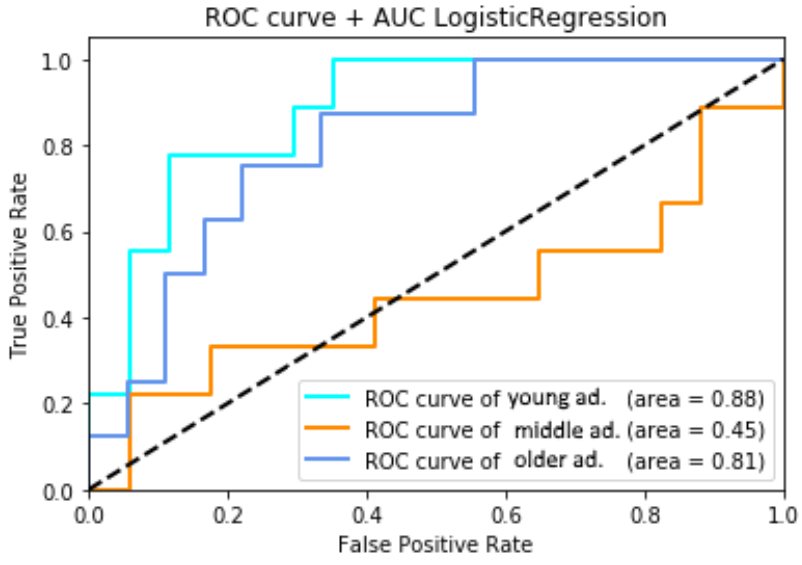


Figure 3.11: ROC and AUC for each class (one-versus-all).

3.5 Discussion

In this paper, we explored to what extent we can predict age groups based on differences in cognitive performance due to cognitive aging. For this classification, we relied on digital biomarkers detected and captured via FreeCell gameplay. The contribution of our work lies in the investigation of digital biomarkers most indicative for cognitive performance, using COTS games, and the exploration to which extent machine learning models can be used to classify games into their respective age groups. To the best of our knowledge, this study is the first study to attempt to predict the age group on the basis of digital biomarkers obtained through casual gameplay.

3.5.1 Adequate Performance Metrics

Interpreting the performance metrics of our model is not straightforward, given the lack of other comparative studies that use COTS games to predict cognitive aging. Beyond the realm of player-computer interaction, in medical diagnosis, high AUC's (>.95) are sought. However, in applied psychology, AUC values of .70 and higher are already considered as strong effects [235].

Scrutinizing our model further, our results suggest that we can predict young participants with good performance, but we are less successful in classifying middle-aged participants and distinguishing them from older adults. This lack of performance for this middle-aged group might be found in the limited data set. The learning curve of our model has not converged completely, and a rising trend can be noted. This suggests that adding more data points is beneficial and that performance would likely improve with more participants. However, it may also be that the model reflects the variability that is inherent in cognitive aging trajectories. As emphasized by [187], [200], the age at which cognitive functions start to deteriorate and the rate at which this happens is highly variable from individual to individual. Given the lack of other comparative studies, it remains unclear what performance measures can actually be expected. To this end, this paper can set a first benchmark and stimulate other researchers to outperform this model.

3.5.2 Relating Features to Cognitive Aging

As can be seen from the correlations between features and age group (Figure 3.8), the features with the highest correlations were time-based. Older adults spent more time thinking before executing either a successful or erroneous move.

They also spent more time moving the card itself. This confirms the sensitivity of 'processing speed' to cognitive aging [187], [210]. Our study also demonstrated that older adults moved fewer cards to the storage stacks, a necessary part to progress in the game. This suggests they did not play as strategically as young and middle-aged participants, aligning with a lesser performance of executive functioning [187].

3.5.3 Suitable for Longitudinal Measurement

Ultimately, for the charting of cognitive aging trajectories, COTS games like FreeCell are to be played on a continuous basis. These continuous measurements, as opposed to episodic measurements, are likely to be more resistant to unwanted factors such as stress, dehydration, or tiredness, as opposed to episodic measurements in memory clinics [54], [199], [236]. Our observations confirmed that participants were eager to play FreeCell, even after the game session ended. Anecdotal remarks from these game sessions support the hypothesis that such games would be played frequently and over a longer time period.

3.6 Limitations

As aforementioned, this is a first exploratory study. Our inference from the data is based on theoretical backgrounds of cognitive aging and calendar age. No validated instruments were used to measure cognitive performance or motor proficiency of our sample. Individual dispositions, socio-economic status, and lifestyle choices may significantly influence cognitive aging. Therefore, follow-up research should investigate the addition of validated instruments to measure cognitive functioning as this would improve the validity of measuring cognitive aging. Predictions relying on these measures instead of calendar age could improve the performance of the models as these are more directly linked to cognitive performance.

Moreover, we relied on a limited sample and learning curves suggest it may be beneficial to add data points. Finally, we only used the tool for a single point in time measurement. Ultimately, such a game-based measurement would span several months and be done frequently to equally capture intra-individual differences, giving deeper insight into the personal cognitive profile of the player.

3.7 Conclusion

In this study, we investigated to what extent we can predict cognitive age from player data from the COTS game FreeCell, as part of the Microsoft Solitaire Collection. Fifty-two participants, belonging to three age groups (young, middle-aged, and older adults) played 130 games. Candidates for digital biomarkers were captured using an image processing toolkit tailored to capture digital biomarkers from card gameplay. Features were engineered, and a machine learning model, based on Logistic regression, was trained. Performance metrics range between 0.615 (Precision) and 0.734 (AUC). Upon inspecting the ROC curves and confusion matrix, it becomes apparent that the model is successful in classifying the youngest and oldest age group, but is less successful for the middle age group. Given the lack of other comparative studies, it remains unclear to what extent the low performance metrics for the middle-aged group are the result of a limited sample or rather because of the inherent variability in cognitive aging between individuals. Nevertheless, the results show that COTS games lend themselves to capturing biomarkers for cognitive performance and that these biomarkers support current theories on fluid versus crystallized intelligence.

Part III

Detecting Cognitive Decline due to Mild Cognitive Impairment through Klondike Solitaire

Chapter 4

Dissecting Digital Card Games to Yield Digital Biomarkers for the Assessment of Mild Cognitive Impairment: a Methodological Approach and Exploratory Study

This chapter is a copy of the article currently in review:
K. Gielis *et al.*, "Dissecting Digital Card Games to Yield Digital Biomarkers for the Assessment of Mild Cognitive Impairment: A Methodological Approach and Exploratory Study", *Journal of Medical Internet Research*, Jan. 2021

Parts of this chapter have been previously published as Work In Progress:
K. Gielis, F. Brito, J. Tournoy, and V. Vanden Abeele, "Can Card Games Be Used to Assess Mild Cognitive Impairment?: A Study of Klondike Solitaire and Cognitive Functions", en, in *CHI PLAY '17 Extended Abstracts*, Amsterdam,

The Netherlands: ACM Press, 2017, pp. 269–276, ISBN: 978-1-4503-5111-9. DOI: 10.1145/3130859.3131328

Scientific Contribution:
As first author, I lead the writing of the first draft of the manuscript and processed suggestions of co-authors. In addition, I coordinated the data collection and full analysis.

4.1 Introduction

Mild Cognitive Impairment (MCI) is a clinical entity defined as a transitional state between normal and pathological aging, where one or more cognitive domains are significantly impaired yet activities of daily living are still preserved [3]. Early detection of MCI is important for signaling possible prodromes of dementia, monitoring the progression of possible decline, taking supportive measures, and detecting any possible underlying causes. Unfortunately, cognitive impairment is still underdiagnosed [34]–[36]. In response, governmental bodies have called for novel, scalable, and longitudinal tools to assist in the early screening and monitoring of dementia [15]–[17]. To answer this call, researchers have explored the use of digital games as a suitable medium for assessing cognitive impairment [145], [160], [163], [237]. Games are autotelic in nature, tapping into the intrinsic motivation to play [81], [238], hence captivating a player's attention. Furthermore, digital games are a natural source of information on player behavior, cognitive performance, motor skills, social conduct, and affective experiences [100].

As such, digital games may help by providing digital biomarkers of cognitive performance. Biomarkers, defined as "objectively measured and evaluated indicators of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [194] have a longstanding tradition in dementia research [137], [239]. Complementary to their biological counterparts, digital biomarkers are "user-generated physiological and behavioral measures collected through connected digital devices to explain, influence and/or predict health-related outcomes" [18]. User interaction with digital games produces dense and detailed behavioral traces that may inform on the users' social health, praxis, and cognition.

Today, the focal point of research assessing cognitive performance has been serious games, i.e., games intentionally designed and built for a serious purpose, and not solely to entertain [93]. While serious gaming interventions show potential, they are typified by lesser funding, shorter development cycles, and missing know-how of the video game industry, which affect in-game quality such as graphics, music, and storytelling [98], [240]. This may lead to frustrating player experiences, a lack of engagement and lesser attention during gameplay, which may lower the reliability and validity of any findings, and possibly cause attrition in longitudinal studies [82], [99], [168]. Therefore, most recently, Mandryk&Birk [100] have argued turning to commercial off-the-shelf (COTS) video games instead. Instead of spending limited resources on building a serious game, researchers can devote themselves to investigating existing games already enjoyed by the target population. While not designed to measure cognition, COTS games are woven into the fabric of everyday life and may be

able to provide digital biomarkers of cognitive performance that are reflective of cognitive status [100], [122].

This study aims to explore the possibilities of COTS card games to screen cognition amongst MCI patients. First, it describes a study involving 11 experts in the domain of MCI to craft 23 candidate digital biomarkers from the digital card game Klondike Solitaire. Subsequently, a data acquisition campaign was set up involving 46 participants: 23 healthy older adults and 23 older adults with MCI. Participants were asked to play three games on a tablet. We examined the game data on differences at a group level for the candidate digital biomarkers using a Generalized Linear Mixed Model (GLMM) Analysis. The results show that 12 out of 23 candidate digital biomarkers differ significantly between both groups, taking age, tablet experience, and Klondike experience into account. By providing a methodological approach and an exploratory study for crafting digital biomarkers, articulating the rationale and the different steps taken, we hope to inform future researchers who aim to leverage the use of commercial off-the-shelf video games to yield digital biomarkers

4.2 Background

4.2.1 Mild Cognitive Impairment

Persons diagnosed with MCI show a deficit in cognition in at least one cognitive domain that cannot be attributed to age or any other disease, yet they do not fulfill the diagnosis of dementia [101]. Persons with MCI however, have a higher risk of progressing to a form of dementia such as Lewy body dementia [31], vascular dementia [241], or, the most common form of dementia, Alzheimer's Disease [242]. Depending on the early symptoms, persons with MCI can be classified into two groups: amnestic MCI (aMCI) and non-Amnestic MCI (naMCI). The aMCI group shows a significant memory deficit, whereas for naMCI mainly a non-memory impairment (e.g. language) is present [243]. For both aMCI and naMCI, a further distinction can be made between persons with one cognitive domain impaired (single domain MCI), and multiple cognitive domains impaired (multiple domain MCI). Even though no treatments exist with the current state of modern medicine to cure the neuronal damage of these progressive forms of dementia [7], [8], early diagnosis matters [9], as there are several measures that can be taken to slow down disease progression [10], starting (non-) pharmacological treatment for delaying symptoms [7], [8], [10], or support patient and family with the appropriate counseling [11].

4.2.2 Detecting Mild Cognitive Impairment

Typically, the process leading to a diagnosis of MCI is set into motion by a cognitive complaint from the older adult, relative, or (informal) caregiver, followed by a presumptive identification through a screening test. The most used cognitive screening tests for MCI are the Montreal Cognitive Assessment (MoCA) [41] and the Mini-Mental State Examination (MMSE) [42]. These cognitive screening tests primarily focus on evaluating language, visual skills, memory, orientation, attention, and executive functions [244]. Despite their widespread use, the psychometric properties of the screening tests alone are insufficient to draw firm conclusions regarding MCI diagnosis [245].

Therefore, this presumptive identification is in turn followed by an elaborate neuropsychological assessment (i.e., a cognitive test battery) and possibly a biomarker scan or a neuroimaging scan [101], [243]. This neuropsychological assessment assesses cognitive skills and level of impairment more thoroughly. In addition, they may include a semi-guided interview with a relative or caregiver to evaluate the change in symptoms over time such as in the Clinical Dementia Rating scale (CDR) [246]. However, this neuropsychological assessment is laborious and time-intensive, requiring skilled test administrators, who despite their training, are still subject to interassessor variability [46]. In addition, from a patient perspective, the process has been described as bewildering, highly stressful, and even humiliating [50], [51], contributing to malperformance. This in turn can make patients self-aware of impairment, leading to feelings of distress or helplessness, possibly spiraling into even worse performance [52], [53]. While biological and imaging biomarkers are becoming more common to support diagnosis, they remain expensive and invasive which makes them equally unfit for high-frequency measurements [46]. As a result, health professionals and policymakers welcome additional tools supporting monitoring of cognition [3], [56]–[59] which reduce patient-level barriers and are more considerate of patients' experiences [52].

4.2.3 Serious Games for the Assessment of Cognitive Functions

Serious (digital) games are "games that do not have entertainment, enjoyment, or fun as their primary purpose" [93]. One early and longstanding tradition [94] is the use of serious games for cognitive evaluation [83]. Space Fortress[93], [94] is perhaps the first research game to measure and train cognitive control and related cognitive functions. Ever since, the popularity of creating serious games and game-based interventions to measure, detect, and train cognition has only

increased, as indicated by systematic reviews on this topic by Ferreira-Brito et al. [89], Lumsden et al. [84], and Valladares-Rodriguez et al. [83].

Serious games may provide certain advantages for the assessment of cognitive performance *compared to* standard cognitive assessment. Firstly, by offering an interactive and immersive audiovisual experience, serious games can be considered to be more engaging than classical tests [82]–[84]. As ensuring the full attention of the participant is paramount in neuropsychological testing, such increased engagement may also result in more reliable research results; previous research has linked effort to cognitive test performance in healthy undergraduate students[159]. Second, games allow embedding cognitive tasks in a (virtual) audiovisual world that more closely mimics the actual lived-in world, allowing for better transfer of task results and providing higher ecological validity [247]. However, it has to be noted that skills learned through serious games might still be difficult to generalize to skills needed in a real life context [247]. Third, serious games can be designed in such a manner that they minimize the need for the presence of a trained administrator. Setting a pace, reading out loud, or cueing can be integrated into the game itself. In this manner, test administer bias is reduced and white-coat effects can be minimized [47], [150]. If assessments are possible with less supervision and manual effort, this is also more scalable, as testing becomes less resource-intensive [247]. However, this lack of supervision while measuring has an important caveat. Measurements made in a personal environment make it more difficult to prevent distractions that influence gameplay behavior.

Even though serious games show promising results and have merit for patient and physician, serious games are at risk of being accused of being 'chocolate-cov[95]–[97]; neuropsychological tests disguised by a thin layer of gameplay. This can lead to games that are suboptimal in terms of aesthetic quality and game mechanics [98] and negatively impact gameplay [95]. A meta-analysis of serious games [99] shows that while serious games can be more effective and improve retention compared to conventional methods, they are not found more motivating. Similar signs of lack of motivation have been noted in game-based interventions designed to train cognitive functions [82], [168].

This lack of sustained engagement contrasts with surveys on gameplay among older adults. A large scale (n=3737) survey of older adults' attitude towards video games, conducted in 2019 by the American Association of Retired Persons [85], highlights that older adults enjoy playing digital games. Out of nine reasons to play, "to have fun" was indicated to be the top reason (78%) to play video games, "to help stay mentally sharp" comes in second (69%). In the 70+ age category, this difference becomes marginal with 73% indicating "to have fun" and 72% indicating "to stay mentally sharp". Therefore, to increase engagement and to tap into intrinsic motivation, popular COTS video games may present

an interesting alternative. These games are already woven into the daily life of the older adult, providing meaningful play [103], [193].

4.2.4 Commercial Off-The-Shelf Video Games for Mental Health

COTS games may have the ability to retain players over a longer period and to support continuous measurements of cognitive performance. As frequent measurements are more sensitive to detecting small deviations in cognitive performance of older adults [48], this could lead to a better interpretation of the patient's cognitive trajectory. Furthermore, fluctuations in cognitive performance [248], a common feature of dementia, may be more easily detected. Additionally, this continuous monitoring enables establishing an intraindividual cognitive baseline [61]. This cognitive baseline can be used to compare patients with themselves, as opposed to comparing to normed references. In turn, this can lead to improved management and care [3]. Nevertheless, a prominent disadvantage of COTS games is that researchers have less control over which cognitive functions are measured in the game [96].

Recent research studies on using COTS games to measure cognitive impairment have generated promising results. Jimison et al. [145] used FreeCell to compare cognitive performance amongst an MCI group and a healthy control group, by means of an 'optimized solver' . Their results indicated that based on gameplay, the group with MCI could be discerned from the healthy control group. Regarding Sudoku, another popular game amongst older adults, Grabbe [170] showed that performance in the game was significantly related to measures of working memory. Using a set of smartphone-based puzzle games, which also contained Sudoku, Thompson et al. [169] explored smartphone-based games as a means of portable cognitive assessment and monitoring. Performance of these games correlated to several measures of cognition, among which visual memory, verbal learning, and reasoning. Finally, Wallace et al. [237] developed a word search and Sudoku game that incorporated hints to reduce frustration amongst MCI patients. Their first study with two patients indicated that cognitive performance could be measured with COTS gameplay, comparing game performance with the MoCA and the Repeatable Battery for the Assessment of Neuropsychological Status [249]. Synthesizing these results, these studies suggest that COTS games yield promise for the assessment of cognitive impairment but that further research is warranted.

Across the previously mentioned studies, different lines of reasoning are given to justify the game of choice as suitable for neuropsychological evaluation. Grabbe [170] analyzed components of Sudoku and linked them to working memory

based on a subjective analysis. Jimison et al. [145] chose FreeCell because it was the most popular game in their focus group. Wallace et al. [237] chose a word search game and sudoku above other games due to certain properties such as the percentage of successful deals. Lastly, Thompson et al. [169] chose games based on face validity with regards to target cognitive functions. While these reasons are valid arguments for choosing a game, it can be noted that these studies have no arguments rooted in empirical evidence for their game of choice.

4.2.5 Klondike Solitaire

One of the most popular card games among older adults is Klondike Solitaire, also known as Patience, Fascination, or even just Solitaire [111]. The popularity of Klondike Solitaire amongst older adult gamers was recently noted by Boot et al. [114]. For one year, participants had access to a computer where eleven games were installed, amongst which Sudoku, Solitaire, and crossword puzzles. They noted that "There was a strong, clear preference for Solitaire [...]. After Solitaire, there was no clear second choice, and on average participants infrequently played the other games." Additionally, their results showed that of all games, Solitaire was being played most consistently of those 11 games.

This popular card game is played with a standard 52-card deck with 28 cards dealt in seven build stacks and the other 24 cards put in a pile, as can be seen in Figure 4.1. The goal of the game is to order all cards from ace to king on the four corresponding suit stacks. Cards can be moved on top of other build stacks if their rank is one lower than the current top card and of the opposite color. Cards can be requested from the pile to be put on the talon.

4.2.6 Study Objective

Given the popularity of Klondike Solitaire among the older population, and given the need for engaging, ecologically valid, scalable tools to assist in the screening and monitoring of MCI , this paper set out to investigate the feasibility of Klondike Solitaire to yield digital biomarkers of MCI . To this end, the study is comprised of the following investigations: 1) an exploration of digital biomarkers of cognitive performance, based on player actions of Klondike Solitaire and 2) an evaluation of candidate digital biomarkers captured in Klondike Solitaire to measure differences between healthy older adults and older adults living with Mild Cognitive Impairment.

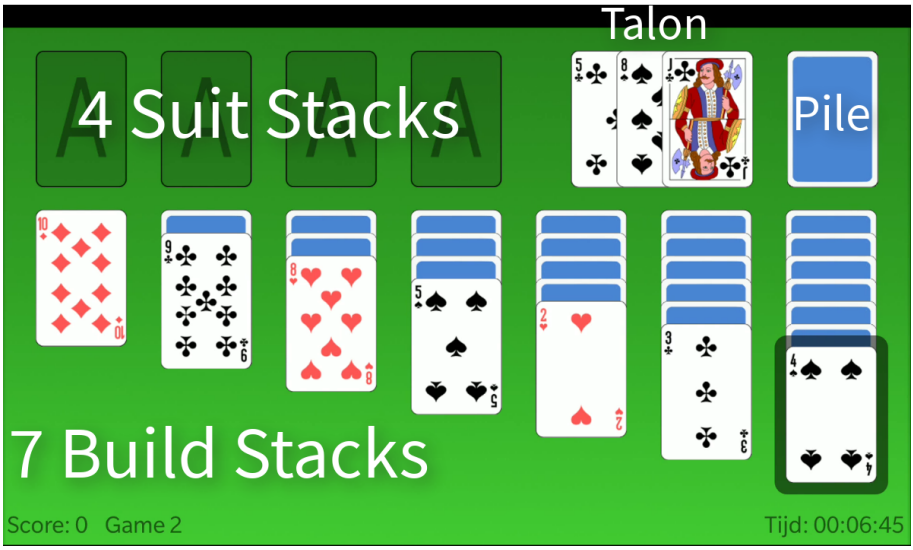


Figure 4.1: Klondike Solitaire. The seven build stacks can be seen at the bottom, the suit stacks are at the top left. The pile of undealt cards can be seen in the top right.

4.3 Crafting Candidate Digital Biomarkers in Klondike Solitaire

To explore the potential digital biomarkers of cognitive performance present in Klondike Solitaire, we first conducted an expert consensus study, involving 11 experts. In this first part of the paper, we discuss the three steps taken to come to a final list of 23 candidate digital biomarkers.

4.3.1 Step 1: Defining Player Actions

To transform gameplay into player actions (PA's), a methodical approach was applied. Four researchers in the field of human-computer interaction carried out the following tasks. First, the literature on Klondike Solitaire was studied, ranging from scientific work [115], [176], [250]–[254] to more general sources [255]–[257]. Afterward, the game was played for a minimum of ten sessions of thirty minutes by each of the researchers. Combining this theoretical background with practical experience, a list of game events was drafted, first independently,

then reviewed in team. This list was iterated three times until no more game events were found. Game events included, but were not limited to, game outcomes (e.g., the game was won or lost), correct player moves (e.g., the player moves a card between build stacks), and erroneous player moves (e.g., player moves cards on each other which are not in descending order on the build stack).

This set of game events were then converted to player actions; they were described as an action of the player rather than an event of the game. Next, all these player actions were transformed into their negative equivalents, e.g., "The Player takes little time to think of a move" was reworded as "The Player takes a lot of time to think of a move." The reason for this is twofold. It causes duplicate PA's (the positive and negative equivalent, e.g., moving cards fast or moving card slowly) to be combined, reducing rating complexity for the professionals. Furthermore, the negative equivalent aimed to facilitate the rating process as impaired cognition will lead to reduced performance in gameplay. After this step, 21 player actions (see Table 4.1) were defined for evaluation.

4.3.2 Step 2: Defining Cognitive Functions

A set of cognitive functions was drafted in five phases (see Figure 4.2). A first draft was made starting from the cognitive functions tested in the most used MCI screening tests [41], [42], [258]. Next, during a trial with one psychologist, we replaced abstraction with object recognition to more clearly indicate problems with finding cards based on key articles on cognitive aging and cognition [187], [259]–[262]. In addition, to better delineate attention it was specified as selective attention. In phase 4, a pilot study was done with an expert on memory and age-related disorders (with 23 years of clinical and research experience). Based on this pilot testing, it was decided to split executive functioning into inhibitory control, cognitive planning, and mental flexibility. Memory was further specified as working memory, and lack of motor skills as apraxia. In the final iteration, cognitive functions ostensibly not present in Solitaire, i.e. orientation in time and space, and language, were removed to reduce the rating complexity. This resulted in a set of nine cognitive functions which can be found in Figure 4.2 phase 5.

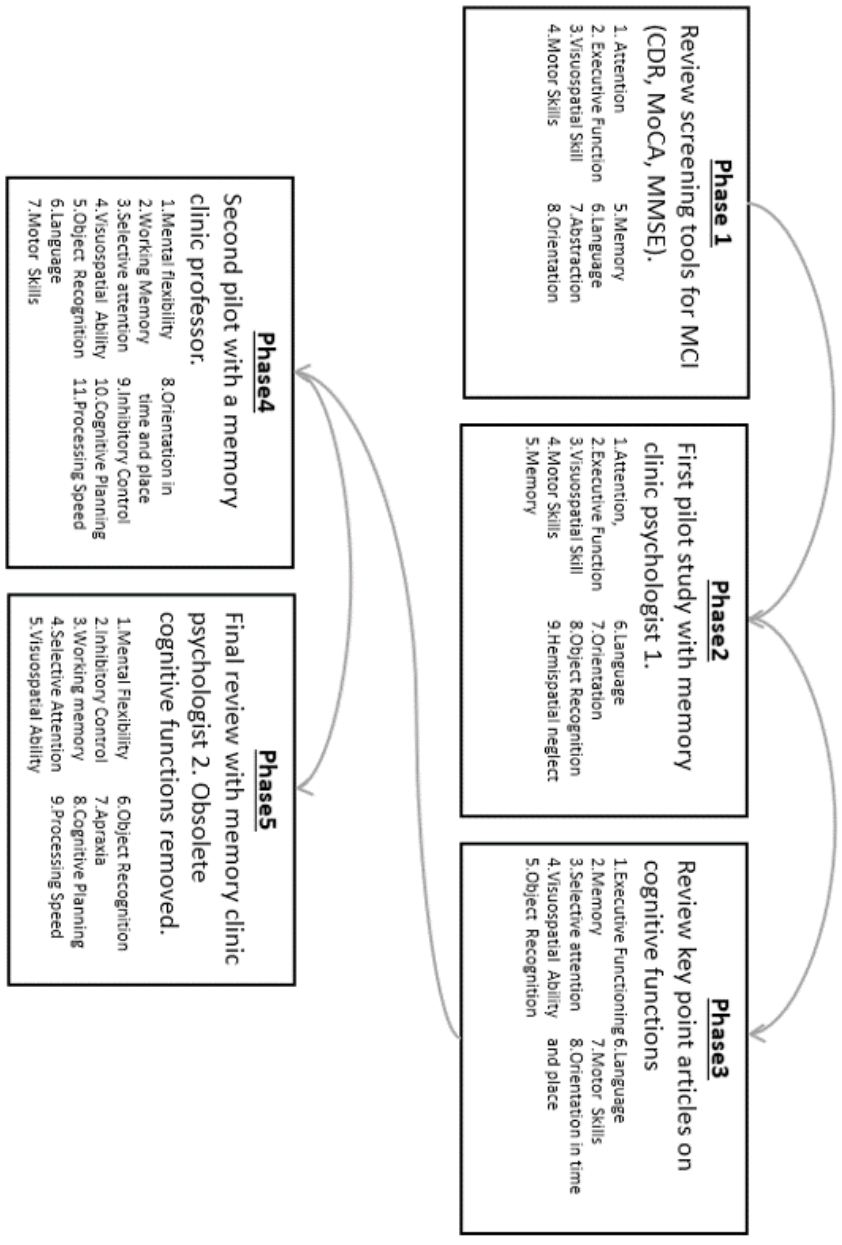


Figure 4.2: The five phases through which the cognitive functions present in Klondike Solitaire were identified.

Protocol for Rating Functions and Actions

As a next step, experts were asked to rate the extent to which each player action was related to a specific cognitive function. These experts were recruited using a snowball sampling method starting from two leading memory clinics in Belgium. Three experts were a Doctor of Medicine experienced in cognitive decline and eight (neuro)psychologist. Seven participants identify as female, four identify as male. The average age of all participants is 45 years (SD= 13.3) with an average amount of working experience of 20years (SD=14). Three co-authors of this paper (LVA, PD, and FFB) also participated as an expert. None of the experts were compensated for participating in the study.

Before rating, every expert received a standardized introduction comprising a video that explains all concepts of the game [263], a video that visualizes all 21 Player Actions [264], and a document that explains all nine cognitive functions. This introduction aimed to prevent confusion concerning the game terminology, interpretation of player actions, and cognitive functions. It also included a delineation of the target group to amnestic multiple domain MCI . Experts could revisit these videos and documents at any time.

After this introduction, every expert received a coding sheet where they could map the 21 player actions to the nine cognitive functions. Each cell had to be filled in according to the following four-point scale:

0: This cognitive function has no significant correlation to the player action.
1: This cognitive function correlates weakly to the player action.
2: This cognitive function correlates moderately to the player action.
3: This cognitive function correlates strongly to the player action.

Finally, they were also given the choice to explain their train of thought in the optional further clarification column.

Expert Agreement on Player Actions and Cognitive Functions

The intraclass correlation (ICC) for each player action as variables of interest with cognitive functions were computed. Additionally, we computed ICC's for each of the cognitive functions as variables of interest with all player actions considered as observations [265]. All calculations were executed using IBM SPSS Statistics 23 [266]. The ICC was calculated to verify the rater agreement [267] on player actions and cognitive functions, based on a two-way random fully crossed design with type consistency [268]. According to the criteria of Koo et al. [269], ICC's lower than 0.5 are indicative of low reliability, ICC's

between 0.5 and 0.75 are indicative of moderate reliability, ICC's between 0.75 and 0.9 are indicative good reliability, and ICC's greater than 0.9 are indicative of excellent reliability.

We found intraclass correlations for all player actions to score above 0.75, suggesting good to excellent reliability according to Koo et al. [269]. Except for four cognitive functions (i.e., Mental Flexibility, Visuospatial Ability, Object Recognition, and Apraxia which score 0.68, 0.42, 0.66, and 0.71, respectively), all intraclass correlations of cognitive functions scored above 0.75, suggesting good to excellent reliability.

Cognitive Functions present in Klondike Solitaire

An overview of the associations between individual player actions and cognitive functions, according to the expert mapping can be found in Table 4.1. In addition, a full overview of all Intraclass Correlations with 95% confidence intervals can be found in section 4.4. We found that all player actions were related by the experts to one or more cognitive functions with an average association above two, which indicates a moderate to strong relation to the cognitive function. Similarly, we found that for each cognitive function, at least one player action has an average association above two.

Player actions	**Mental Flexibility**	**Inhibitory Control**	**Working memory**	**Selective Attention**	**Visuospatial Ability**	**Object Recognition**	**Apraxia**	**Cognitive Planning**	**Processing Speed**
PA1. Player takes a lot of time to think of a move.	1.64 ±1.12	0.73 ±1.01	1.82 ±0.4	1.55 ±0.82	1.18 ±0.98	1.27 ±0.79	0.27 ±0.47	2.45 ±0.52	2.55 ±0.82
PA2. Player takes a lot of time to move the card.	0.73 ±1.01	0.73 ±1.1	0.64 ±0.5	0.64 ±0.67	1.45 ±1.04	0.64 ±0.67	1.64 ±0.92	0.91 ±0.83	2.09 ±1.04
PA3. Player does not move a suitable card from the talon to the suit stack.	2.27 ±0.79	0.73 ±0.9	2 ±0.77	2.55 ±0.69	1.45 ±0.93	1.64 ±0.67	0.73 ±1.01	2.18 ±0.75	0.91 ±1.04
PA4. Player does not move a suitable card from the build stack to the suit stack.	1.82 ±0.98	0.91 ±0.83	1.82 ±0.75	2.36 ±0.67	1.36 ±0.92	1.36 ±0.92	0.27 ±0.47	1.73 ±0.79	1±1

Player actions	**Mental Flexibility**	**Inhibitory Control**	**Working memory**	**Selective Attention**	**Visuospatial Ability**	**Object Recognition**	**Apraxia**	**Cognitive Planning**	**Processing Speed**
PA5. Player does not move a suitable card from the talon to the build stack.	2.18 ±0.75	1.27 ±0.9	1.91 ±0.7	2.64 ±0.67	1.55 ±1.04	1.73 ±0.79	0.18 ±0.4	2.09 ±0.94	0.82 ±1.08
PA6. Player does not move a suitable card from one build stack to another build stack.	2.36 ±0.81	1.09 ±0.94	2 ±0.77	2.45 ±0.69	1.45 ±0.93	1.64 ±0.92	0.18 ±0.4	2.18 ±0.98	0.73 ±1.01
PA7. Player does not place an ace immediately on an empty suit stack.	1.27 ±1.01	0.73 ±1.01	2.18 ±0.4	2.36 ±0.92	1 ±0.77	1.18 ±1.17	0.45 ±0.69	2.09 ±0.7	1.09 ±1.04
PA8. Player does not put a king on an empty build stack.	1.45 ±1.13	0.73 ±0.9	2 ±0.77	2.27 ±0.9	1 ±0.77	1.55 ±1.13	0.36 ±0.67	2.09 ±0.7	1 ±0.89
PA9. Player moves cards without benefit (e.g. putting a jack from one lady to another).	1.45 ±1.04	1.82 ±0.98	2.18 ±1.17	1.64 ±1.03	0.82 ±0.75	1.45 ±1.21	0.18 ±0.4	2.27 ±1.01	0.45 ±0.52
PA10. Player flips a lot through the pile.	2 ±0.89	2.55 ±0.69	1.73 ±1.01	1.82 ±1.08	1 ±0.89	1.45 ±1.13	1 ±1	2.09 ±1.04	0.91 ±0.7
PA11. Player moves a card onto a card with the same color.	1.73 ±1.1	2.55 ±0.52	2.18 ±0.98	2.18 ±0.98	1 ±1	1.82 ±1.17	0.27 ±0.65	1.36 ±0.81	0.45 ±0.69
PA12. Player moves a card to another card with the wrong number (E.g., placing a five on a ten).	1.18 ±1.08	2 ±1	2.27 ±0.9	1.91 ±0.94	1.09 ±1.04	2.09 ±0.94	0.45 ±0.69	1.45 ±0.93	0.36 ±0.5
PA13. Player selects the cards with a very bad precision (taps on edge or next to the card).	0.45 ±0.69	0.73 ±0.79	0.27 ±0.47	0.64 ±0.81	2.27 ±0.9	0.82 ±0.75	2.27 ±0.79	0.45 ±0.82	0.45 ±0.69

Player actions	**Mental Flexibility**	**Inhibitory Control**	**Working memory**	**Selective Attention**	**Visuospatial Ability**	**Object Recognition**	**Apraxia**	**Cognitive Planning**	**Processing Speed**
PA14. Player starts tapping on the playfield with no apparent target (with short interval, fidget tapping).	0.73 ±0.79	2.27 ±1.01	0.27 ±0.47	0.82 ±0.87	0.73 ±0.9	0.45 ±0.52	1.55 ±1.29	0.91 ±1.04	0.73 ±0.9
PA15. Player presses the undo button a lot.	1.82 ±0.6	2.45 ±0.69	1.73 ±1.1	1.36 ±1.12	0.64 ±0.67	0.64 ±0.67	0.73 ±1.01	2.27 ±1.01	1.27 ±1.1
PA16. Player requests a lot of hints.	1.91 ±1.04	1.73 ±1.01	2 ±1	1.45 ±0.93	0.64 ±0.81	1 ±0.77	0.45 ±0.69	2.27 ±1.01	1.18 ±0.75
PA17. Player takes a very long time to finish games.	2.18 ±1.25	1 ±1.34	2.18 ±0.75	1.64 ±1.21	1.09 ±0.83	1.18 ±0.98	0.91 ±0.83	2.64 ±0.5	2.91 ±0.3
PA18. Player does not have a high score in the game.	2.18 ±0.98	2 ±1	2.36 ±1.03	1.91 ±1.04	1.45 ±0.93	1.36 ±0.92	0.91 ±0.94	2.27 ±0.9	1.55 ±1.04
PA19. Player does not win a lot of games (low win ratio).	2.36 ±0.67	1.82 ±1.08	2.64 ±0.5	2 ±1	1.36 ±0.92	1.18 ±0.87	1 ±0.89	2.82 ±0.4	1.64 ±0.81
PA20. Player's scores of different games vary greatly.	2.27 ±1.1	1.64 ±1.12	2.27 ±0.79	2.36 ±1.12	0.73 ±0.9	0.73 ±0.9	0.64 ±0.92	2.18 ±1.08	1.82 ±1.08
PA21. Player's win ratio decreases rapidly as the difficulty of the game increases.	2.36 ±0.67	1.91 ±0.94	2.64 ±0.67	2.18 ±0.87	1.18 ±0.87	1.09 ±0.94	0.82 ±0.87	2.64 ±0.81	1.64 ±1.03

Table 4.1: Average of the experts' ratings for each Player Action and Cognitive Function.

4.3.3 Step 3: Defining Candidate Digital Biomarkers

These player actions were captured via the game as potential digital biomarkers, i.e., measurable factors of the game such as score duration of the game, detailed moves etc. These candidate digital biomarkers were enriched with additional information about the game. This contextualization is important to ensure an unambiguous interpretation of the cognitive information derived from the gameplay. For example, whereas a game played with a lot of moves on the pile may indicate that a player progressed in the game, it may equally indicate that the player does not realize that they are stuck. By calculating the percentage of pile moves by dividing it by the total amount of moves made, a more informative metric can be obtained. In this manner, 23 potential digital biomarkers were defined which we further classified in one of five categories: time-based, performance-based, error-based, execution-based, auxiliary-based, and result-based. Time-based digital biomarkers are biomarkers related to the speed of player actions. Performance-based digital biomarkers are biomarkers related to optimal gameplay (i.e., is the game played according to strategies that ensure optimal performance). Error-based digital biomarkers relate to making incorrect moves according to the Solitaire rules. Auxiliary-based digital biomarkers are interactions that are not part of the core gameplay, i.e., requesting undo's and hints. Finally, result-based digital biomarkers are biomarkers that evaluate the final outcome of the game (e.g., how far did the participant get in the game). A full overview of all digital biomarkers and their contextualizations can be found in Table 4.3.3.

Related PA	Digital Biomarker	Description	Contextualization	Value
Time-based				
PA1	Think Time	Time spent thinking of a move. Defined as the time necessary to find and touch a suitable card.	Average, Standard Deviation	Number (ms)
PA2	Move Time	Time spent moving card(s). Defined as the time necessary to move a suitable card to the destination.	Average, Standard Deviation	Number (ms)
PA1, PA2	Total Time	Total time to make a move. Defined as the combination of Think Time and Move Time.	Average, Standard Deviation	Number (ms)
Performance -based				
PA3	Final Beta Error	Whether there were still moves possible when quitting a game.	None	Boolean
PA3, PA4, PA5, PA6	Beta Error	Amount of pile moves made with moves remaining on the board, divided by the total amount of Pile Moves.	Percentage	0-100%
PA7	Ace Beta Error	Amount of missed opportunities to place a king on the suit stacks, divided by the total of game moves.	Percentage	0-100%
PA8	King Beta Error	Amount of missed opportunities to place a king on an empty spot, divided by the total of game moves.	Percentage	0-100%
PA10	Pile Move	Amount of pile moves, divided by the total amount of board moves.	Percentage	0-100%

Related PA	Digital Biomarker	Description	Contextualization	Value
Error-based				
PA11, PA12	Successful Move	Amount of successful moves, divided by the total amount of board moves.	Percentage	0-100%
PA11, PA12	Erroneous Move	Amount of erroneous moves, divided by the total amount of board moves.	Percentage	0-100%
Execution -based				
PA13	Accuracy	Accurateness of selecting a card, defined by how close a card was touched to the center.	Average, Standard Deviation	0-100%
PA14	Taps	Amount of actuations on non- UI elements.	None	Number
Auxiliary -based				
PA15	Undo Move	Amount of undos requested.	Percentage	0-100%
PA16	Hint Move	Amount of hints requested.	Percentage	0-100%
Result -based				
PA17	Gametime	Total time spent playing a game.	None	Number (ms)
PA18	Score	Final score of a game.	None	Number
PA19	Solved	Whether the game was completed or not. Indicator of how successfully the game is played.	None	Boolean
PA19	Cards Moved	Amount of cards selected for each move. An additional indicator of how successfully the game is played; as games of Solitaire progress, longer stacks of cards are moved.	Average, Standard Deviation	Number

Table 4.2: Digital Biomarkers related to the player actions (PA) in Klondike Solitaire. [1]

4.4 ICC Results

Variable of interest	Intraclass Correlation	95% Confidence Interval	
		Lower Bound	Upper Bound
Cognitive Function			
Mental Flexibility	0.68	0.43	0.85
Inhibitory Control	0.83	0.69	0.92
Working memory	0.82	0.67	0.91
Selective Attention	0.81	0.66	0.91
Visuospatial Ability	0.42	-0.03	0.73
Object Recognition	0.66	0.39	0.84
Apraxia	0.71	0.48	0.86
Cognitive Planning	0.79	0.62	0.90
Processing Speed	0.87	0.78	0.94
Player Action			
PA1	0.92	0.81	0.98
PA2	0.84	0.63	0.96
PA3	0.87	0.70	0.97
PA4	0.85	0.64	0.96
PA5	0.90	0.76	0.97
PA6	0.91	0.78	0.98
PA7	0.91	0.78	0.97
PA8	0.87	0.69	0.96
PA9	0.88	0.71	0.97
PA10	0.87	0.69	0.97
PA11	0.87	0.70	0.97
PA12	0.82	0.56	0.95
PA13	0.94	0.85	0.98
PA14	0.83	0.59	0.95
PA15	0.90	0.75	0.97
PA16	0.86	0.68	0.96
PA17	0.91	0.78	0.97
PA18	0.86	0.68	0.96
PA19	0.91	0.78	0.98
PA20	0.91	0.79	0.98
PA21	0.93	0.84	0.98

4.5 Evaluating Digital Biomarkers

The aim of this second study was to explore the potential of these candidate digital biomarkers of cognitive performance. Relying on 46 participants, we captured data and performed a Generalized Mixed Model analysis to examine differences between healthy participants and participants diagnosed with MCI .

4.5.1 Participants

In total, 23 healthy participants and 23 participants with MCI) participants were enrolled. Older adults with MCI were recruited by two of the leading memory clinics in Belgium. Healthy participants were recruited using a snowball sample starting from multiple senior citizen organizations. All healthy participants had a minimum age of 65 years, were fluent in written and verbal Dutch, had 20/20 (corrected) vision, no motor impairments, and lived independently or semi-independently at home, service flat, or care home. Exclusion criteria for healthy participants were subjective memory concerns from participant, caretaker, or clinician. Additionally, they were screened using the MMSE, MoCA, and CDR. To minimize the risk of including potential individuals with MCI among healthy participants, a cut-off score of 27 on the MMSE, 26 on the MoCA, and a score of 0 on the CDR were enforced. Participants living with MCI were formally diagnosed with multiple-domain amnestic MCI based on Petersen's diagnostic criteria [33] by one of the two collaborating memory clinics. Participants with MCI were excluded when scoring less than 23 on the MMSE to avoid including participants which are on the border between the diagnosis of MCI and dementia. In addition, all participants recruited had prior experience with Klondike Solitaire. This familiarity with the rules was imperative as participants with MCI may have problems with memorizing and recalling new game rules in their short-term memory. Moreover, the rationale underlying this study is to draw from games already played and enjoyed by participants and where the rules are crystallized in memory. Demographic and basic neuropsychological data of both groups can be found in Table 4.3.

[1]PA's 20, and 21 were not captured as the single-point-in-time setup would not allow comparing scores and win ratio's with ranging difficulty over time. In addition, PA9 was not tested as the current software would not allow for detecting these moves.

Table 4.3: Demographic and Neuropsychological Data.

	Healthy (n=23)	**MCI (n=23)**
Age	70 (SD=5.4)	80 (SD=5.2)
Education[a]	22% 30% 48%	17% 57% 26%
Sex (F/M/X)	47% 53% 0%	57% 43% 0%
Tablet Proficiency[b]	52% 9% 0% 9% 30%	13% 9% 9% 4% 65%
Klondike Proficiency[b]	13% 26% 13% 47% 0%	30% 35% 9% 26% 0%
MMSE Score	29.61 (SD=0.65)	26.17 (SD=1.75)
MoCA Score	28.09 (SD=1.28)	NA
CDR Score	0 (SD=0)	NA

[a] Participants were categorized into three education groups based on the 1997 International Standard Classification of Education (UNESCO United Nations Educational, Scientific and Cultural Organization 2003): a. ISCED 1/2 b. ISCED 3/4 c. ISCED 5/6.

[b] Participants were categorized into five proficiency groups based on frequency of use: a. Daily b. Weekly c. Monthly d. Yearly or less e. Never

4.5.2 Data Collection Tools

All game sessions were completed on a Lenovo Tab 3.10 Business tablet running Android 6.0. A Solitaire application created by Bielefeld [270] under the LGPL 3 license was modified to capture and store game metrics which served as building blocks for the digital biomarkers of cognitive performance.

4.5.3 Data Collection Procedure

Each observation was carried out between 9 AM and 5 PM in the home environment of the participant to ensure a familiar and comfortable environment. An observation took between two to three hours and consisted of two main parts: 1) a game session where game-based digital biomarkers of Klondike Solitaire were collected on a tablet and 2) a neuropsychological examination where cognitive information was collected.

Each game session started with a standardized five-minute introduction where the tablet, the game mechanics, and possible touch interactions were explained. This was followed by a practice game, identical for all participants, where questions to the researcher were allowed and the participant could get used to the touch controls. Data from this practice game was not used for analysis. After this practice game, the participant played three games in succession. The order and games were equally identical across all participants. All games were purposefully chosen through prior playtesting, in that they were solvable, and

varied in difficulty level. During these three games, no questions were allowed and gameplay continued until the participants either finished the game or indicated that they deemed further progress impossible. All game sessions and cognitive evaluations were conducted by the same researcher to avoid differences due to researcher bias.

4.5.4 Ethical Statement

This study is in accordance with the declaration of Helsinki and GDPR compliant. Ethical approval was given by the Ethics Committee of UZ/KU Leuven, Belgium, CTC S59650. Due to the fragile nature of our participants, utmost care was given to inform them. Tests were conducted solely after written informed consent.

4.5.5 Statistical Analysis

To assess the difference between healthy participants and participants diagnosed with MCI , a GLMM analysis was performed using R [271] with the lme4 library [272]. Concerning the design of our GLMM, the fixed effects comprised of MCI, age, tablet proficiency, and Klondike proficiency. Random effects were modeled as random intercepts for gameseed and participant. In addition, by-participant random slopes for the effect of MCI were modeled.

Continuous digital biomarkers (e.g., Think Time Average) were modeled using a GLMM with the identity link function. Binary outcomes (e.g. Solved or not solved) were modeled using a GLMM with the logit link function. The significance of the effect of MCI was determined using the Likelihood Ratio Test which compares the model with a model without the effect of MCI, both estimated without Restricted Maximum Likelihood [273], [274]. Assumptions of homoscedasticity and normality were visually inspected using residual plots. To provide supplemental information on the fit of the models, the marginal R^2, and the conditional R^2 are given, as specified in [275]. Given the exploratory nature, we did not correct for family-wise inflation error [276].

4.6 Results

The results of the GLMM summary on the effect of MCI can be found below. A visualization of digital biomarker performance for all groups across all games

can be found in Figures 4.3,4.4,4.7,4.5,4.8. A summary can be found in Table 4.6.6.

4.6.1 Time-based Digital Biomarkers

For time-based digital biomarkers (Figure 4.3), MCI significantly affected Think Time Average ($chi^2(1)= 7.658$, $p= 0.006$), increasing it by 1119.947 ms ± 405.81 (SD). MCI equally significantly affected Think Time Standard Deviation ($chi^2(1)= 5.173$, $p= 0.023$), increasing it by 1112.533 ms ± 490.53 (SD). However, MCI did not significantly affect Move Time Average ($chi^2(1)= 2.737$, $p= 0.098$) or Move Time Standard Deviation ($chi^2(1)= 2.651$, $p= 0.103$). MCI significantly affected Total Time Average ($chi^2(1)= 5.286$, $p= 0.021$), increasing it by 1278.263 ms ± 573.84 (SD), and Total Time Standard Deviation ($chi^2(1)= 4.16$, $p= 0.041$), increasing it by 1315.598 ms ± 673.67 (SD).

4.6.2 Performance-based Digital Biomarkers

For performance-based digital biomarkers (Figure 4.4), MCI did not significantly affect Final Beta Error Percentage ($chi^2(1)= 0.213$, $p= 0.645$). MCI did equally not significantly affect Beta Error Percentage ($chi^2(1)= 0.836$, $p= 0.36$), Ace Beta Error Percentage ($chi^2(1)= 0.117$, $p= 0.733$), or King Beta Error Percentage ($chi^2(1)= 0.506$, $p= 0.477$). MCI significantly affect Pile Move Percentage ($chi^2(1)= 7.544$, $p= 0.006$), increasing it by 13.333% ± 4.88 (SD).

4.6.3 Error-based Digital Biomarkers

For error-based digital biomarkers (Figure 4.7), MCI significantly affected Successful Move Percentage, ($chi^2(1)= 5.949$, $p= 0.015$), lowering it by 8.913% ± 3.6 (SD). MCI did also significantly affect Erroneous Move Percentage, ($chi^2(1)= 4.892$, $p= 0.027$), increasing it by 3.624% ± 1.65 (SD).

4.6.4 Execution-based Digital Biomarkers

For execution-based digital biomarkers (Figure 4.5), MCI significantly affected Accuracy Average ($chi^2(1)= 4.085$, $p= 0.043$), lowering it by 3.817 % ± 1.9 (SD). MCI did not significantly affect Accuracy Standard Deviation ($chi^2(1)= 0.036$, $p= 0.849$) or Taps, ($chi^2(1)= 3.82$, $p= 0.051$).

4.6.5 Result-based Digital Biomarkers

For result-based digital biomarkers (Figure 4.6), MCI did not significantly affect Gametime ($chi^2(1)= 3.071$, $p= 0.08$). MCI significantly affected Solved ($chi^2(1)= 6.93$, $p= 0.008$), lowering it by 2.63 ± 1.01 (SD). MCI also significantly Cards Moved Average ($chi^2(1)= 4.928$, $p= 0.026$), lowering it by 0.119 cards $\pm$ 0.05 (SD), and Cards Moved Standard Deviation ($chi^2(1)= 6.733$, $p= 0.009$), lowering it by 0.38 cards $\pm$ 0.15 (SD).

4.6.6 Auxiliary-based Digital Biomarkers

For auxiliary-based digital biomarkers (Figure 4.8), none of these candidate biomarkers reached significance: Undo Move Percentage ($chi^2(1)= 0.467$, $p= 0.494$), Hint Move Percentage ($chi^2(1)= 2.402$, $p= 0.121$).

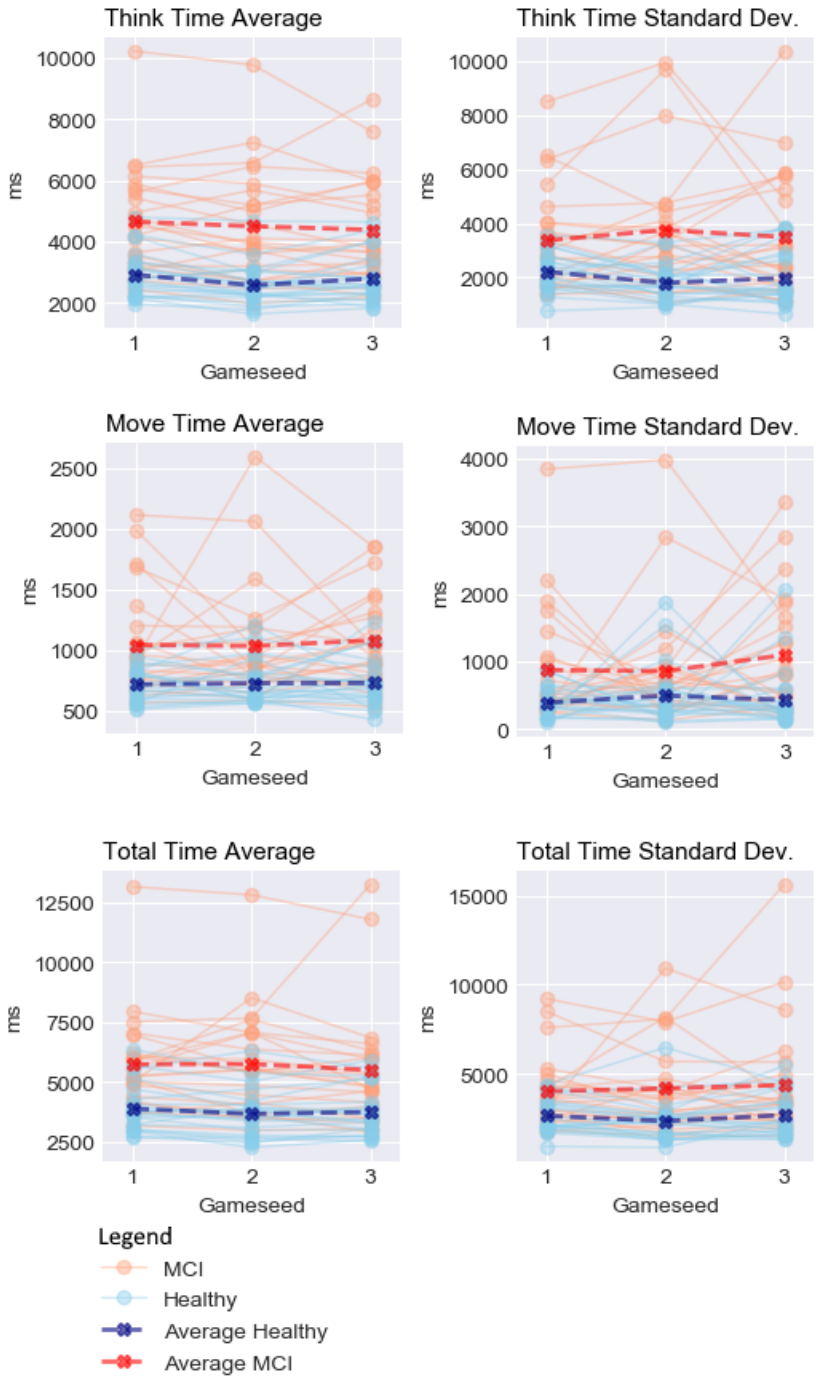


Figure 4.3: Performance on time-based digital biomarkers for both groups.

Figure 4.4: Performance on performance-based digital biomarkers for both groups.

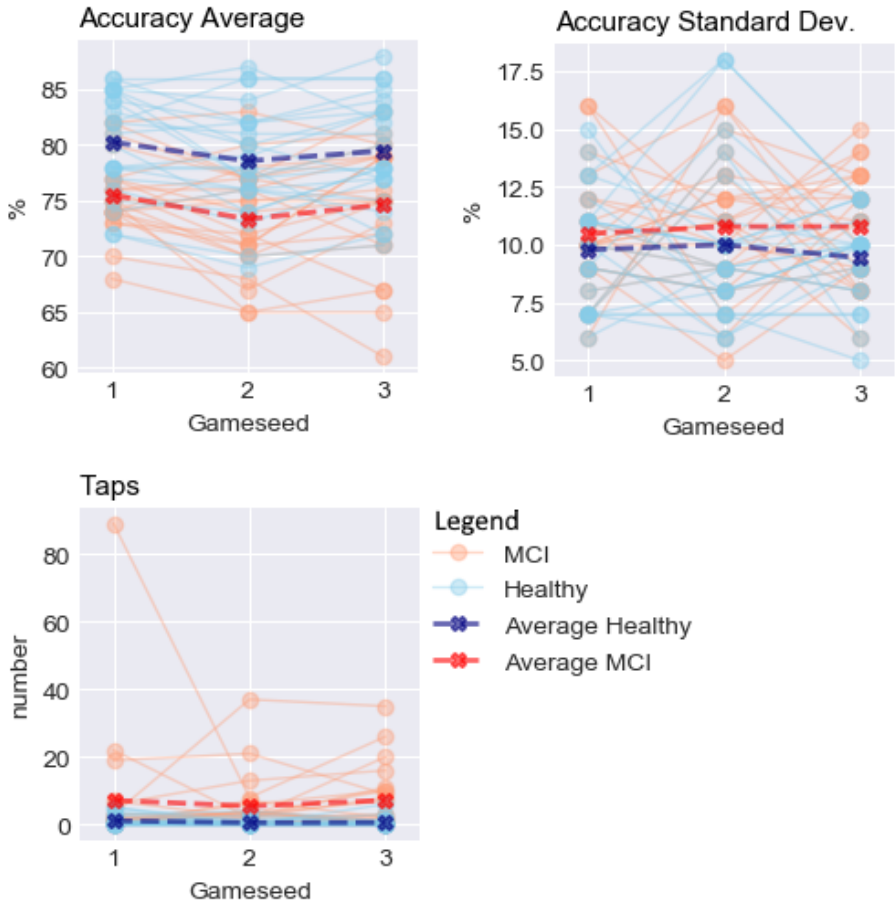


Figure 4.5: Performance on execution-based digital biomarkers for both groups.

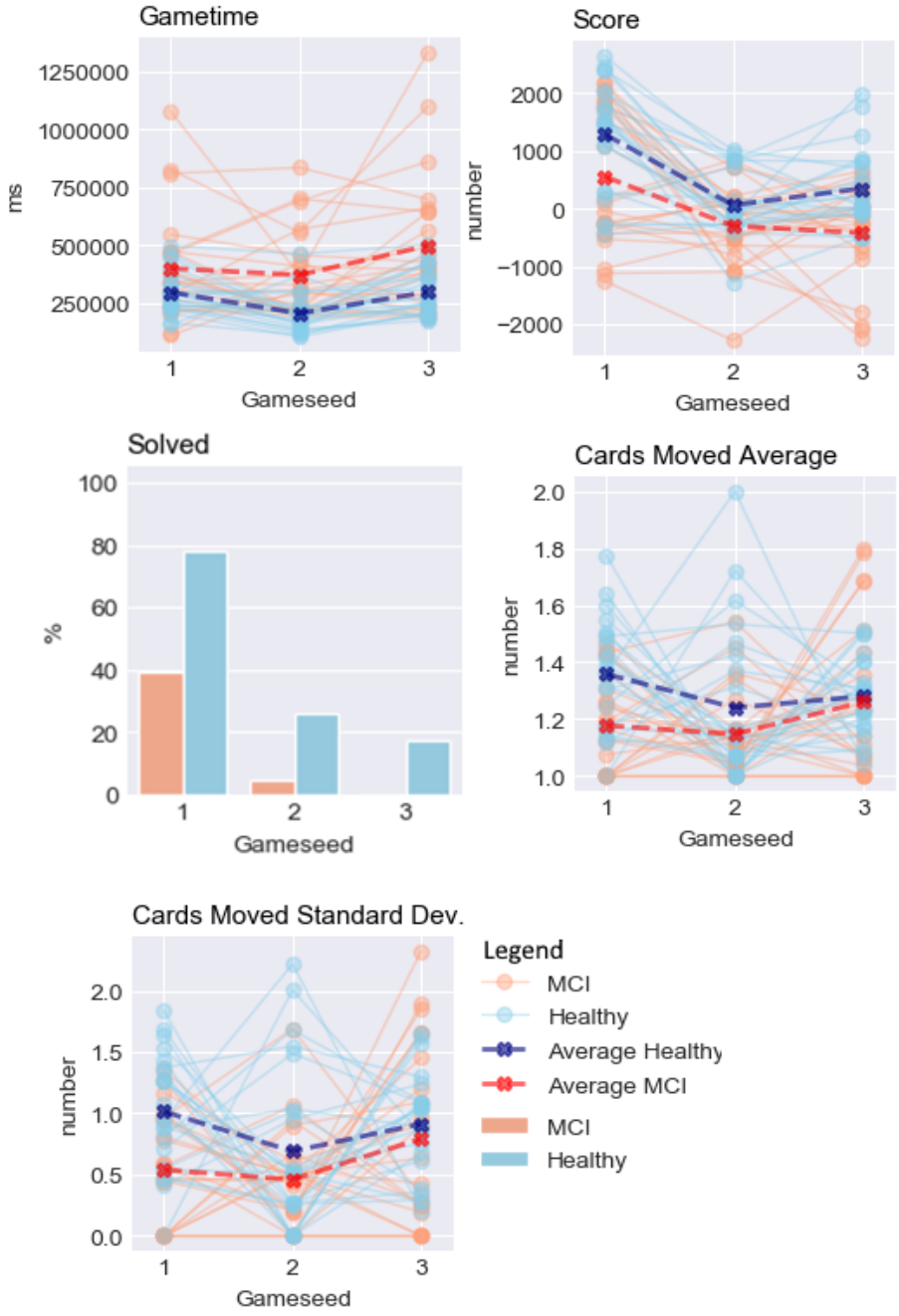


Figure 4.6: Performance on result-based digital biomarkers for both groups.

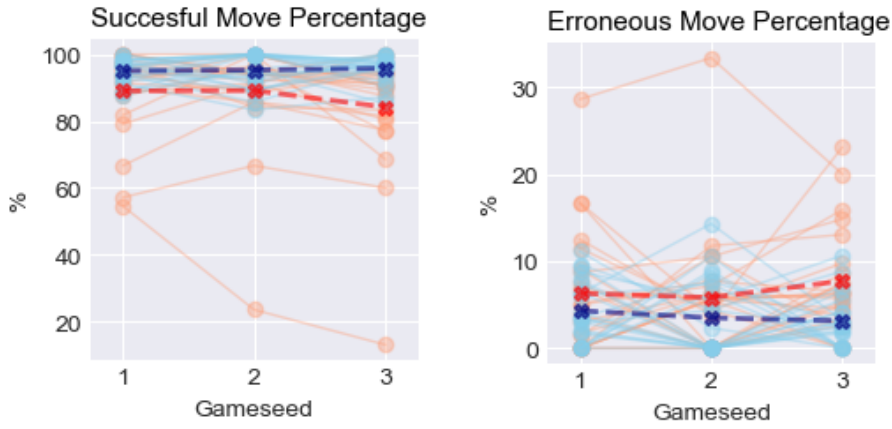


Figure 4.7: Performance on error-based digital biomarkers for both groups.

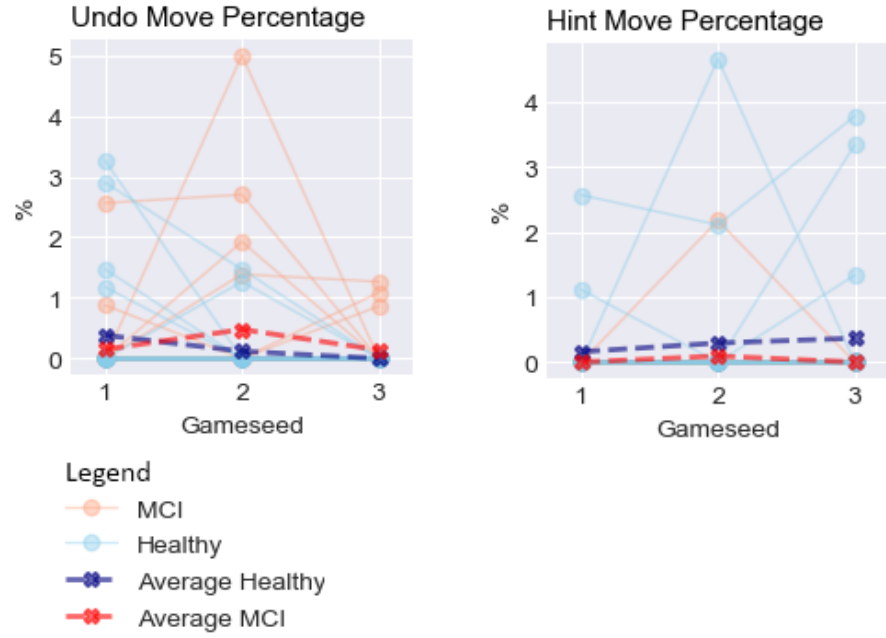


Figure 4.8: Performance on auxiliary-based digital biomarkers for both groups.

Digital Biomarker	**Constant (SD)**	**β (SD)**	**p Chi²**	**R²m (R²c)**
Time-based				
Think Time Average	-1371.778 (1415.444)	1119.947 (405.815)	0.006 **	0.416 (0.904)
Think Time Standard Deviation	-814.527 (1720.073)	1112.533 (490.53)	0.023 *	0.211 (0.655)
Move Time Average	-508.575 (373.89)	156 (95.547)	0.098	0.257 (0.579)
Move Time Standard Deviation	-856.605 (847.852)	323.599 (202.032)	0.103	0.137 (0.419)
Total Time Average	-912.419 (2149.177)	1278.263 (573.839)	0.021 *	0.318 (0.870)
Total Time Standard Deviation	206.569 (2676.062)	1315.598 (673.665)	0.041 *	0.176 (0.715)
Performance-based				
Final Beta Error	-7.233 (4.131)	0.435 (0.922)	0.645	0.096 (0.068)
Beta Error Percentage	-7.203 (33.849)	6.108 (6.879)	0.36	0.089 (0.371)
Ace Beta Error Percentage	-0.132 (0.629)	0.051 (0.137)	0.733	0.023 (0.209)
King Beta Error Percentage	-3.682 (5.918)	0.907 (1.323)	0.477	0.028 (0.230)
Pile Move Percentage	71.759 (24.052)	13.333 (4.88)	0.006 **	0.097 (0.513)

**p <.01; * p <.05

Digital Biomarker	**Constant (SD)**	**β (SD)**	**p Chi²**	**R²m (R²c)**
Error-based				
Successful Move Percentage	87.486 (9.443)	-8.913 (3.595)	0.015 *	0.104 (0.795)
Erroneous Move Percentage	9.486 (6.529)	3.624 (1.651)	0.027 *	0.081 (0.466)
Execution-based				
Accuracy Average	92.134 (9.167)	-3.817 (1.903)	0.043 *	0.246 (0.805)
Accuracy Standard Deviation	4.519 (3.746)	0.137 (0.772)	0.849	0.056 (0.196)
Taps	-5.113 (10.704)	5.334 (2.762)	0.051	0.098 (0.500)
Auxiliary-based				
Undo Move Percentage	0.228 (0.955)	0.135 (0.205)	0.494	0.008 (0.151)
Hint Move Percentage	-0.58 (0.87)	-0.311 (0.204)	0.121	0.046 (0.491)
Result-based				
Gametime	-167427.7 (187325.5)	93211.27 (53699.25)	0.08	0.198 (0.690)
Score	29.03 (1389.752)	-744.433 (286.576)	0.009 **	0.105 (0.612)
Solved	-2.954 (4.578)	-2.63 (1.007)	0.008 **	0.186 (0.152)
Cards Moved Average	1.111 (0.262)	-0.119 (0.054)	0.026 *	0.061 (0.093)
Cards Moved Standard Deviation	0.135 (0.705)	-0.38 (0.147)	0.009 **	0.072 (0.152)

**p <.01; * p <.05

Table 4.4: GLMM results for each digital biomarker.

4.7 Discussion

Mild Cognitive Impairment is a neurological disorder that is linked to an increased risk of developing dementia. As such, early detection of cognitive deterioration is essential for timely diagnosis and for allowing tailored care and treatment. Collecting digital biomarkers via COTS games may help by providing cognitive information through behavior traces of activities already integrated into the daily life of older adults. In this study, we investigated in particular whether Klondike Solitaire can yield digital biomarkers. In the paragraphs below, we discuss our findings and reflect on the different potential digital biomarkers, their relation to cognitive functions, and the ethical implications of their use for cognitive assessment purposes.

4.7.1 Dissecting Digital Biomarkers

Out of 23 candidate digital biomarkers, we found 12 to differ significantly between older adults with MCI and a healthy control group. This supports the use of digital card games for monitoring cognitive performance and possibly detecting differences in cognitive performance caused by MCI.

While overall findings are promising, not all candidate biomarkers performed equally. We saw for time-based digital biomarkers that biomarkers related to coming up with a move, Think Time Average and Standard Deviation, were significantly affected by MCI. In contrast, biomarkers related to the actual physical movement of cards, Move Time Average and Standard Deviation, were not significantly affected. Total Time Average (p-value = 0.021), which contains Move Time as well as Think Time, was significantly affected yet was less significant than Think Time Average (p-value = 0.006). These results indicate that segmenting in-game actions can be beneficial as they can more accurately isolate cognitive functions such as praxis and cognitive planning.

For performance-based digital biomarkers, in contrast with expectations, none of the biomarkers related to beta errors were proven to differ significantly. Upon rewatching gameplay, it became clear that there were two different types of beta errors, strategic versus unintentional. Unfortunately, due to the current configuration of the application, it was impossible to discriminate between both types. This is further discussed in the Limitations section. On the other hand, Pile Move Percentage was proven to differ significantly. This may indicate that older adults with MCI may not recognize the same cards being returned as fast as their healthy counterparts.

Results equally indicated that participants with MCI made more mistakes, as both error-based digital biomarkers (i.e., successful move percentage and erroneous move percentage) were significant. In contrast, none of the auxiliary-based digital biomarkers showed to differ significantly. Upon inspecting the data, it was noted that none of both groups consistently used these functionalities, which may have contributed to the lack of significance.

Finally, four out of five digital biomarkers in the result-based category were significant, three of them with $p<0.01$ (i.e., Score, Solved, Cards Moved Standard Deviation). The outcome of these measures is the result of a series of consequent moves, each of them being potentially crucial to complete the game. For example, one lapse in attention or executive functioning can cause important moves to be overlooked, in turn making the game unsolvable. While overall Gametime was not significant, this can be explained due to the fact that time spent in the game on itself does not indicate lesser performance. Time-based digital biomarkers, which are equally measures of time but contextualized with the amounts of moves made, show more significant results (i.e., Think Time Average, Think Time Standard Deviation, Total Time Average, Total Time Standard Deviation), stressing the importance of contextualization.

In sum, our findings are in accordance with the earlier work of Jimison et al. [145] where FreeCell, another Solitaire variant. Using card gameplay, we can discern older adults with MCI from a healthy control group. Moreover, the results gathered from this study are in line with previous research by Bankiqued et al. [96] and Ángeles Quiroga et al. [277]. Bankiqued et al. [96] found that casual games that tap working memory and reasoning can be robustly related to performance on working memory and fluid intelligence. Similar research on commercial video games by Ángeles Quiroga et al. [277] found high relationships between video games and general intelligence test performance. Our findings confirm these findings at a finer granularity and show that when scrutinizing player actions time-based, error-based, and result-based biomarkers yield promise in particular.

4.7.2 Future Work

In this study, participants with MCI were diagnosed with multiple-domain amnestic MCI, based on Petersen's diagnostic criteria [33]. As MCI is a multidimensional clinical entity, it would be interesting to explore whether Klondike Solitaire is suitable for monitoring the cognitive status of participants with non-amnestic MCI as well. The focus on executive functioning can be useful for both subtypes, as it has been shown that both MCI subtypes have a similar decrease in executive functioning [278]. Although we acknowledge

that the evaluation of other cognitive functions such as anterograde memory, retrograde memory, orientation, and language is paramount to get a complete overview of the patient's cognitive profile, these cognitive functions were not identified by the experts and were thus not included in our analysis.

4.7.3 Reflections on the Use of COTS games to assess Cognitive Performance

COTS games also have their limitations. First, neuropsychological assessments are typically designed to assess a broad yet targeted spectrum of cognitive functions. Moreover, different tests are devised to measure one cognitive function in particular. COTS games, and more particularly digital card games, were found more limited in the cognitive functions they can specifically assess. When using COTS games, it may be hard to separate the evaluations of specific cognitive functions. In this study, experts judged every single player action to be related to at least two cognitive functions.

Second, using COTS games as an instrument to measure cognitive performance and possibly flag MCI necessitates ethical reflection. We envisioned the use of COTS games to be used only in accordance with the patient, with the positive aspiration that this could aid in the longitudinal monitoring of cognitive deterioration, more accurately measuring cognitive performance and variance. This project grew out of an ambition to escape the limitations of serious games and providing meaningful play to older adults. Yet, we have to acknowledge that we may have transformed an activity previously considered enjoyable, yet innocent, into an instrumental activity that may even trigger a sense of health surveillance [279]. Observational notes taken during this study did not reveal any verbal remarks of stress from the participants diagnosed with MCI. However, such remarks were made by several of the healthy participants, as some felt pressure to outperform participants living with MCI. Further research is needed to understand how the instrumentalization of COTS games impacts the player experience of patients.

Third, it has to be noted that deriving digital biomarkers from digital games may not be relevant for all older adults. Not everyone is an avid gamer, and even those who are may have preferences for different game genres. In addition, these preferences might change over time [100]. While digital card games, such as Klondike Solitaire, are in general a popular pastime for the population susceptible to MCI [85], [86], [105], [113], [114], they might not be for coming generations. Therefore, it is important to identify other accessible games suitable for cognitive monitoring with a broad appeal.

Finally, the interaction between healthcare professional and patient, oftentimes found stimulating and motivating in and of itself, is crucial for full assessment. Hence, we argue that COTS games for screening and monitoring of cognitive impairment should not be used as a replacement of current neuropsychological examination but rather as a source of additional information.

4.7.4 Limitations

Fine-tuning Beta Errors

In contrast with expectations, Beta Error related digital biomarkers proved to be insignificant. Upon inspecting games of both groups, it became clear that there are two types of beta errors: Build Stack Beta Errors and Suit Stack Beta Errors. The former represent missed moves between build stacks. These errors were rarely on purpose and occurred less in the healthy participants' group, based on observation. In contrast, the latter represent missed moves between build stacks and suit stacks. We observed that this latter category is utilized strategically to prevent the inability of placing future cards. Observations suggest these occurred more often in the healthy participants' group. Unfortunately, due to the current configuration of the application, it was impossible to discriminate between these two types of Beta errors. Hence, this points to the importance of further contextualization and refinement of measurement of Beta errors, and biomarkers in general which should be addressed in future work.

Limited sample size

An a priori power analysis [280] estimated the adequate sample size to lie between 32 to 88 participants (assuming comparable effect sizes as cognitive screening instruments to detect MCI [281]). Due to strict inclusion criteria, only 46 participants were eligible. While this strict protocol kept data quality in mind, the sample size may have impacted the effects estimated in this study. It could be that our study is underpowered, leading to some digital biomarkers to be wrongfully found insignificant. Future studies should therefore still critically inspect the different digital biomarkers and results obtained.

Additionally, because of the average age difference between both groups, we chose a GLMM for our statistical analysis, as it can factor in confounding effects. A side exploration included trained machine learning models to predict age instead of MCI on the same dataset. These models were found less performant than the ones modeling MCI, underscoring that the effect of MCI is greater

than the effect age in our dataset. Nevertheless, it is a limitation we have to acknowledge and take into account while interpreting the results.

4.8 Conclusion

This study gives insight into the cognitive functions addressed while playing digital card games, and assesses its potential of screening for MCI. To this end, eleven experts in neuropsychology or geriatrics mapped associations of player actions in Klondike Solitaire and cognitive functions. Upon this exercise, that showed experts agreed player actions were related to cognitive functions a, 23 potential digital biomarkers of cognitive performance were crafted. A Generalized Linear Mixed Effects analysis, taking effects of age, tablet experience, and Solitaire experience into account, compared digital biomarker performance between an MCI group and a healthy control group. We found a significant and sizeable effect for 12 of 23 digital biomarkers, despite strict inclusion criteria and natural variations in human cognition. These exploratory results support the notion of detecting Mild Cognitive Impairment through Klondike Solitaire.

Chapter 5

Detecting Mild Cognitive Impairment through Digital Biomarkers of Cognitive Performance found in Klondike Solitaire: a Machine Learning Study

This chapter is a copy of the previously published article:
K. Gielis, M.-E. Vanden Abeele, K. Verbert, J. Tournoy, M. De Vos, and V. Vanden Abeele, "Detecting Mild Cognitive Impairment through Digital Biomarkers of Cognitive Performance found in Klondike Solitaire: A Machine Learning Study", *Digital Biomarkers*, Jan. 2021, ISSN: 2504-110X. DOI: 10.1159/000514105

Scientific Contribution:
As first author, I lead the writing of the first draft of the manuscript and processed suggestions of co-authors. In addition, I coordinated the data collection and full analysis.

5.1 Introduction

Mild Cognitive Impairment (MCI) is a condition where one or more cognitive domains are slightly impaired, yet instrumental activities of daily living are still intact [3], [43]. People with MCI have a higher chance of progressing to a form of dementia, moreover, it can also signal other neurologic or psychiatric diseases such as vascular disease or depression [3], [43], [56]. Therefore, timely detection of patients with MCI is necessary to provide support and devise a (non)pharmaceutical management approach [3], [101]. While clinically valid, modern cognitive assessment is limited by the mode of administration, often pen, paper, and stopwatch [48]. These modes of administration require continuous attention from a trained administrator, limiting the type and amount of data points captured, and make measurements vulnerable to administrator bias and white coat effect [47], [282]. As a consequence, this lack of accurate high-resolution data can make it difficult to make informed inferences of neuropsychological processes [48]. Cognitive assessment through digital biomarkers of cognitive performance could be an addition to the current cognitive toolset by contributing to a more complete cognitive profile [48]. Digital biomarkers [18], [61] are user-generated physiological and behavioral measures, captured through connected digital devices, which can provide high-resolution, objective, and quantifiable cognitive data [46].

For MCI, the systems measuring digital biomarkers of cognitive performance can be categorized into four groups [46]: systems using dedicated or passive sensors, systems with wearable sensors, non-dedicated technological solutions (e.g. software that captures text input), and dedicated or purposive technologies such as games. Games are in an unique position to yield digital biomarkers as they are autotelic in nature, meaning they are played for the enjoyment they offer, without the need or request from a third person. Hence, they are intrinsically motivating and do not necessitate an administrator, thereby avoiding white coat effect and related biases. Moreover, they can provide different challenges with every playthrough while leaving the fundamental game rules intact [48]. This possibility of supplying novel challenges contrasts with the static property of classical cognitive testing, which makes administering them over a short period of time more prone to learning effects [48].

Whereas prior research in games and cognition focused primarily on games specifically made for the purpose to measure cognition (i.e. serious games), current research is investigating commercial off-the-shelf (COTS) video games as a medium for digital biomarkers of cognitive performance [100]. While both serious and COTS games may provide more interactive, immersive, and engaging experiences, when compared to traditional cognitive screening [82]–[84], COTS games have the important advantage of already being woven into the

daily life of older adults. Previous research indicates that serious games for training and measuring cognition still lack engagement, and suffer from attrition in longitudinal studies [82], [96], [99]. As such, this study explores whether Klondike Solitaire, an existing popular Solitaire card variant [114], can be used to detect differences in cognitive performance amongst healthy older adults and those with MCI.

To this end, Klondike Solitaire data from 23 healthy older adults and 23 older adults with MCI were captured. Derived digital biomarkers of cognitive performance were used to train machine learning models to classify individuals belonging to either group. Successful classification of MCI through machine learning supports the efficacy of COTS games to detect differences in cognitive performance on an individual level.

5.2 Materials and Methods

5.2.1 Participants

Participants with MCI were recruited from two leading memory clinics in Belgium and were clinically diagnosed with multiple domain amnestic MCI according to Petersen's diagnostic criteria [283]. Healthy participants were recruited using a snowball sample starting from multiple senior citizen organizations and were screened using two commonly used cognitive screening tests and a structured interview: the Montreal Cognitive Assessment (MoCA), the Mini-Mental State examination (MMSE), and the Clinical Dementia Rating (CDR) scale [41], [246], [284]. The inclusion and exclusion criteria of both groups can be found in Table 5.1. Out of 64 enrolled participants, 23 healthy older adults and 23 older adults with MCI fulfilled all inclusion criteria. These 46 participants all played the same three games, resulting in a total of 138 games captured.

5.2.2 Study Overview

This study is part of an overarching study that assesses cognitive performance through meaningful play (Clinical Trial ID NCT02971124). Every observation was conducted in the home of the participant between 9am and 5pm to ensure a familiar and distraction-free environment. All sessions were completed on a Lenovo Tab 3 10 Business tablet running Android 6.0. All Klondike Solitaire games were played on a custom-build Solitaire application which captured several game metrics, originally created by Bielefeld [270] under the LGPL 3

Table 5.1: Study inclusion and exclusion criteria.

Inclusion Criteria
Minimum 65 years old
Lives independently or semi-independently at home, service flat, or care home.
Prior Solitaire Experience
Fluent in written and verbal Dutch
No visual or motoric deficits
Stable medical condition
Exclusion Criteria Healthy Group
MMSE<27, MoCA<26, or CDR>0
Exclusion Criteria MCIGroup
Non-amnestic or single-domain MCI
MMSE<23

license. In this application, cards requested from the pile came in three, with unlimited passes through the pile. Points could be earned or lost by making the following moves: cards put from build to suit stack added 60 points, cards put from pile to suit stack added 45 points, revealing cards on the build stack added 25 points, retrieving cards from suit to build stack subtracted 75 points, and going through the whole pile subtracted 200 points. Before playing Solitaire, a standardized five-minute introduction of the tablet and game was given. In addition, a practice game was played where questions to the researcher were allowed. Afterwards, three rounds of Klondike Solitaire, each with a different shuffle were played in succession. To prevent unfair shuffles (dis)advantages, deck shuffles were identical for all participants for each round. These three shuffles were chosen beforehand by the researchers so that they were solvable and varied in difficulty. While playing these three rounds, no questions were allowed, and gameplay continued until the rounds were finished or until the participant indicated that they deemed no further moves were possible.

5.2.3 Data Analysis

While playing Klondike Solitaire, general game data such as the total time, score, and outcome were captured. In addition, for every single move, the timestamp, touch coordinates, origin card information, destination card information, and the possibility of other moves on the board was logged. This game data was used to calculate the digital biomarkers of cognitive performance (see Table 5.2). These digital biomarkers can be seen as basic game metrics enriched with game information. This contextualization is important to aid the interpretation

of the cognitive information from the game. For example: a larger number of pile moves made can be interpreted as progression in the game, but can equally be interpreted as the player not realizing that they are stuck. By dividing the amount of pile moves by the number of total moves, a more informative candidate digital biomarker can be obtained. This contextualization resulted in 61 candidate digital biomarkers of Klondike Solitaire (Table 5.2) to be classified in one of five categories: result-based, which contains biomarkers which are related to performance at the end of a game; performance-based, which contains biomarkers which are related to performance during the game; time-based, which contains biomarkers related to time; execution-based, which contain biomarkers related to physical execution of moves; and auxiliary-based, which contain biomarkers related to help features.

For model training, a machine learning procedure was adapted from Sebastian Raschka [285] (Figure 5.1), using Scikit-learn [227] as the main machine learning library. All data were split using a randomized stratified sampling method (102 games from 34 participants in the training set, 36 games from 12 participants in the test set). To prevent data leakage due to identity confounding [286], rounds were split subject-wise instead of record-wise (i.e. all rounds of a participant were either all in the test set or the training set). Heavily correlated features ($p>0.9$) were removed to prevent multicollinearity [287]. In total, 26 features remained after selection which are indicated in bold in Table 5.2. Afterwards, features were scaled using a Standard Scaler. As each algorithm has its inherent biases with none being superior over the rest, nineteen classification models were trained, ranging from linear models like Logistic Regression up to non-linear models like Gaussian Naïve Bayes [285]. The selection of our models was based on their maturity, popularity, and support available in the sci-kit learn machine learning library. To evaluate them during the training phase, 5-fold cross-validated F1 scores were compared. The hyperparameters of the three most performant models were further optimized. Ultimately, these three best performing models were evaluated on the test dataset.

Table 5.2: Potential Digital Biomarkers of Cognitive Performance in Klondike Solitaire, divided into five categories. Remaining features, after multicollinearity and zero value checks, used to train the models, are in bold and indicated with an asterix.

Digital Biomarker	**Description**	**Aggregation**	**Data Type**
Result-Based			
Score	Final score of a game.	**Value***	Integer [-∞,+∞]
Solved	Whether the game was completed or not.	**Value***	Boolean
Gametime	Total time spent playing a game, expressed in ms.	**Value***	Integer [0,+ ∞]
Total Moves	Total amount of moves made during the game.	**Sum***	Integer [0,+ ∞]
Performance-Based			
Successful Move	Amount of successful moves.	**Percentage***	Double [0.00%-100.00%]
Erroneous Move	Amount of erroneous moves.	**Percentage***	Double [0.00%-100.00%]
Rank Error	Amount of rank errors.	**Percentage***	Double [0.00%-100.00%]
Suit Error	Amount of suit errors.	**Percentage***	Double [0.00%-100.00%]
King Error	Amount of kings misplaced	Percentage	Double [0.00%-100.00%]
Ace Error	Amount of aces misplaced	Percentage	Double [0.00%-100.00%]
Pile Move	Amount of pile moves.	**Percentage***	Double [0.00%-100.00%]
Cards Moved	Amount of cards selected for each move.	**Average***, Median, Standard Deviation	Double [0.00,+ ∞]

Digital Biomarker	Description	Aggregation	Data Type
Beta Error	Amount of pile moves with moves remaining on the board.	**Percentage***	Double [0.00%-100.00%]
King Beta Error	Amount of missed opportunities to place a king on an empty spot.	Percentage	Double [0.00%-100.00%]
Ace Beta Error	Amount of missed opportunities to place a king on the suit stacks.	Percentage	Double [0.00%-100.00%]
Final Beta Error	Whether there was a missed move when quitting a game.	**Value***	Boolean
Time-based			
Think Time	Time spent thinking of a move, expressed in ms.	**Average***, **Standard Deviation***, **Min***, Max, Median	Integer [0,+ ∞]
Think Time Successful	Time spent thinking of a successful move, expressed in ms.	Average, Median, Standard Deviation, Min, Max	Integer [0,+ ∞]
Think Time Erroneous	Time spent thinking of an erroneous move, expressed in ms.	Average, Median, Standard Deviation, Min, Max	Integer [0,+ ∞]
Move Time	Time spent moving card(s), expressed in ms.	**Average***, **Standard Deviation***, **Min***, Max, Median	Integer [0,+ ∞]
Move Time Successful	Time spent moving card(s) for a successful move, expressed in ms.	Average, Median, Standard Deviation, Min, Max	Integer [0,+ ∞]

Digital Biomarker	**Description**	**Aggregation**	**Data Type**
Move Time Erroneous	Time spent moving card(s) for an erroneous move, expressed in ms.	Average, Median, Standard Deviation, Min, Max	Integer [0,+ ∞]
Total Time	Total time to make a move, expressed in ms.	**Average***, **Standard Deviation***, **Min***, Max, Median	Integer [0,+ ∞]
Execution-based			
Accuracy	Accurateness of selecting a card, defined by how close a card was touched to the center.	**Average***, **Standard Deviation***, **Min***, **Max***, Median	Double [0.00%-100.00%]
Taps	Actuations on non-game or UI elements.	**Sum***	Integer [0,+ ∞]
Auxiliary-based			
Undo Move	Amount of undo's requested.	Percentage	Double [0.00%-100.00%]
Hint Move	Amount of hints requested	Percentage	Double [0.00%-100.00%]

* Indicates remaining features after multicollinearity and zero value checks

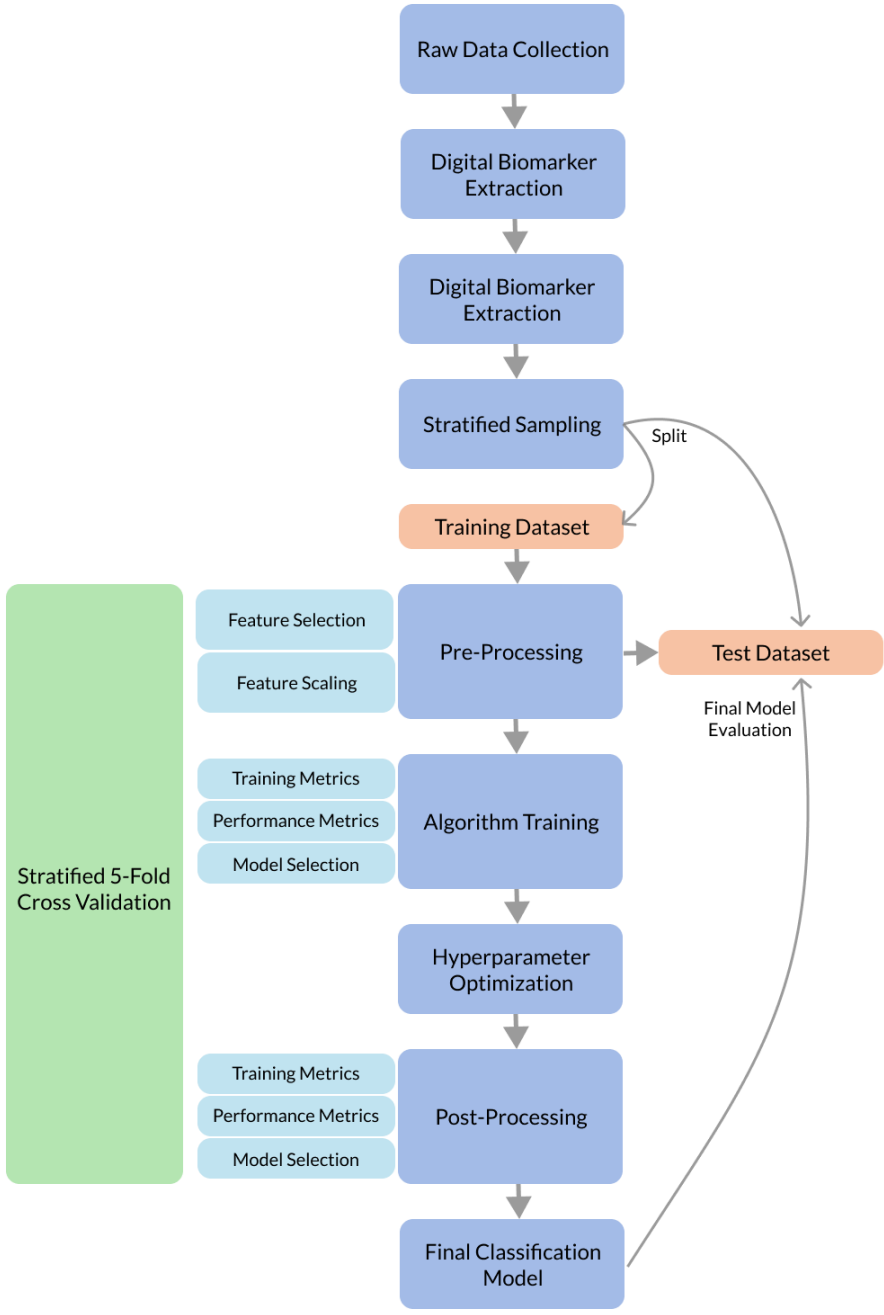


Figure 5.1: Machine Learning process based on the work of Raschka [285].

5.3 Results

5.3.1 Study Population

In total, 46 participants (23 MCI and 23 Healthy) were enrolled, resulting in 138 rounds of Klondike Solitaire captured. Demographic and basic neuropsychological data of both groups can be found in Table 5.3.

Table 5.3: Demographic and neuropsychological data for both groups.

DEMOGRAPHIC INFORMATION		
	Healthy (n=23)	**MCI (n=23)**
Age	70 (SD=5.4)	80 (SD=5.2)
Education[1]	22%/30%/48%	17%/57%/26%
Sex (F/M/X)	47%/53%/0%	57%/43%/0%
Tablet Proficiency[2]	52%/9%/0%/9%/30%	13%/9%/9%/4%/65%
Klondike Proficiency[2]	13%/26%/13%/47%/0%	30%/35%/9%/26%/0%
MMSE Score	29.61 (SD=0.65)	26.17 (SD=1.75)
MoCA Score	28.09 (SD=1.28)	NA
CDR Score	0 (SD=0)	NA

[1]Participants were categorized into three education groups according to the 1997 International Standard Classification of Education [288]: a. ISCED 1/2 b. ISCED 3/4 c. ISCED 5/6.
[2]Participants were categorized into five proficiency groups based on frequency of use: a. Daily b. Weekly c. Monthly d. Yearly or less e. Never

5.3.2 Model Performance

The average results of all selected digital biomarkers of cognitive performance across the three rounds for both groups can be found in Table 5.4. The 5-fold cross-validated F1 validation score of the nineteen initial base models was on average 0.738 [1]. The validation performance metrics of the three best finetuned models obtained an F1 score of 0.812 (SD=0.058) for the Gradient Boosting classifier, 0.797 (SD=0.074) for the Nu-Support Vector classifier, and 0.792 (SD=0.102) for the Extra Trees classifier. Test performance metrics of these models for the 36 rounds in the test set achieved an F1 score of 0.821 with an AUC of 0.892 for the Gradient Boosting classifier, an F1 Score of 0.824 with an AUC of 0.901 for the Nu-Support Vector classifier, and an F1 score of 0.811

[1]An overview of all performance metrics of each model can be found in appendix A

with an AUC of 0.877 for the Extra Trees classifier. Confusion matrixes and ROC curves for these three models can be found in Figure 5.2.

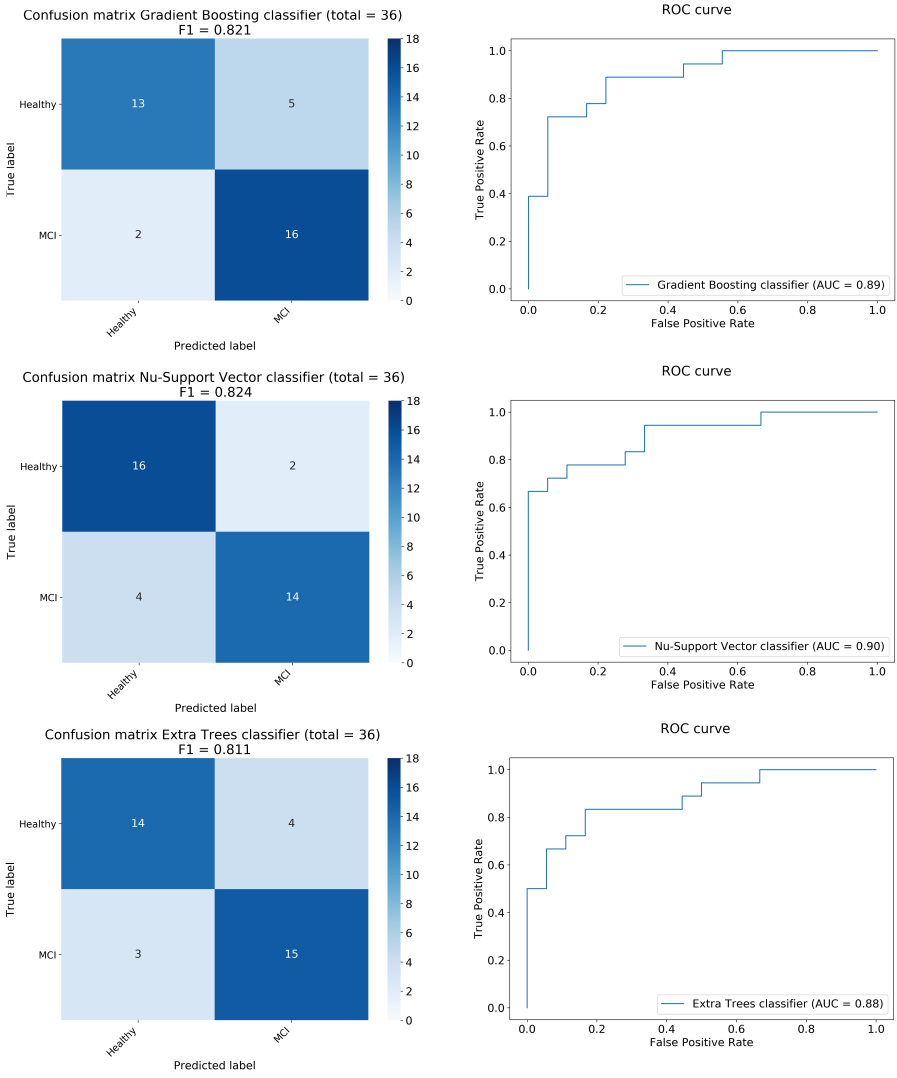


Figure 5.2: Test performance metrics on a per game basis.

Table 5.4: Average performance scores of both groups across all rounds.

Candidate Biomarker	**Healthy**	**MCI**
Result-Based		
Score	565.22 (SD=896.92)	-56.3 (SD=1032.16)
Solved	28 out of 69 games solved	10 out of 69 games solved
Gametime	266107.33 (SD=100546.06)	422283.35 (SD=243018.32)
Total Moves	68.49 (SD=17.45)	72.59 (SD=28.54)
Performance-Based		
Succesful Move Percentage	95.37 (SD=4.28)	87.45 (SD=15.86)
Erroneous Move Percentage	3.65 (SD=3.62)	6.62 (SD=6.7)
Rank Error Percentage	1.85 (SD=2.34)	4.51 (SD=6.18)
Suit Error Percentage	2.33 (SD=2.74)	3.59 (SD=4.83)
Pile Move Percentage	47.36 (SD=16.93)	56.66 (SD=16.34)
Average Cards Moved	1.29 (SD=0.21)	1.19 (SD=0.2)
Beta Error Percentage	45.25 (SD=27.83)	57.37 (SD=29.98)
Final Beta Error	0.13 (SD=0.34)	0.33 (SD=0.47)
Time-based		
Average Think Time	2765.71 (SD=734.83)	4514.78 (SD=1749.75)
Standard Deviation Think Time	1999.72 (SD=812.16)	3544.32 (SD=2181.62)
Minimum Think Time	957.04 (SD=223.42)	1289.55 (SD=573.65)
Average Move Time	722.16 (SD=169.82)	1050.45 (SD=426.31)
Standard Deviation Move Time	440.04 (SD=383.42)	943.64 (SD=872.37)
Minimum Move Time	376.35 (SD=97.09)	458.03 (SD=140.38)
Average Total Time	3767.28 (SD=992.82)	5666.61 (SD=2221.33)
Standard Deviation Total Time	2560.54 (SD=1123.73)	4191.06 (SD=2576.13)
Minimum Total Time	741.12 (SD=234.66)	842.41 (SD=414.25)
Execution-based		
Average Accuracy	79.43 (SD=4.73)	74.51 (SD=4.68)
Standard Deviation Accuracy	9.74 (SD=2.67)	10.68 (SD=2.63)
Minimum Accuracy	51.88 (SD=18.47)	49.06 (SD=13.72)
Maximum Accuracy	96.07 (SD=2.33)	92.58 (SD=4.48)
Taps	0.77 (SD=1.41)	6.61 (SD=12.84)

5.4 Discussion

Digital biomarkers of cognitive performance, embedded into casual gameplay, can be used for cognitive monitoring. By evaluating the efficacy of these candidate digital biomarkers of cognitive performance to discriminate healthy older adults from older adults with MCI, new research opportunities may be opened for monitoring the cognitive trajectories of older adults.

In total, 136 rounds were collected from 46 participants (23 healthy and 23 persons diagnosed with MCI). Derived digital biomarkers were used to train 19 diverse machine learning models which were optimized for the F1 score (the harmonic mean of precision and recall). The choice for optimizing for the F1 score is twofold. First, the possible damage of False Negatives, as well as False Positives, is significant. False negatives, in this study older adults with MCIbeing classified as healthy, could postpone diagnosis, leading to longer undetected disease progression. False Positives, in this study healthy older adults being classified as MCI, could have an equally detrimental impact. Misdiagnosis of cognitive impairment could further spiral the depression of the healthy older adult. Second, F1 score is a robust parameter for unbalanced datasets. Should these studies be expanded to real-life settings where MCIand populations are not equal, this scoring parameter will likely still be of relevance to other researchers.

After hyperparameter finetuning, the 5-fold cross-validated F1 training score on the validation set was above 0.792 for each of the three selected models. When evaluated on the test set, each of these models had an F1 score above 0.811 and an AUC above 0.877. The ROC curves of each model also reveal promising decision thresholds to maximize Sensitivity (True Positive Rate) and Specificity (1-False Positive Rate). It can also be noted that the three selected models come from different machine learning model techniques: a bagged decision tree ensemble (Extra Trees), a boosted decision tree ensemble (Gradient Boosting), and a Support Vector model (Nu-Support Vector) [227]. These high performances on validation and test, combined with the variety of techniques used, indicate that the digital biomarkers contain cognitive information and that successful classification is not hinging on the intricacies of a certain model. In contrast, these robust results indicate that digital biomarkers of cognitive performance, measured while playing Klondike Solitaire, are impacted by MCI. When combined, these digital biomarkers may even be used to train machine learning models to discern older adults with MCIfrom their healthy counterparts, lending support for use for detecting cognitive decline.

The performance metrics of our models appear to be in line to those of current day neuropsychological screening tests. Two of the most common screening tests for discriminating MCI from healthy are the MoCA [41] and the MMSE[284]. In

a systematic review by Pinto et al. [244], a mean AUC of 0.883 was found for the MoCA and a mean AUC of 0.780 for the MMSE. While this study is not meant as a validation study of Klondike Solitaire, our results indicate possible comparative psychometric properties. However, the performance metrics appear to be below the findings of prior studies using serious games. Valladares-Rodríguez et al. [289] investigated the use of machine learning models to discriminate amongst healthy older adults, older adults with MCI, and older adults with Alzheimer's Disease. Their serious game set Panoramix consists of seven games based on seven pre-existing neuropsychological tests such as the California Verbal Test. Their Random Forest classifier obtained a global training accuracy of 1.00, a global F1-score of 0.99, a Sensitivity score of 1.00 for MCI, and a Specificity score of 0.7 for MCI. Direct comparison with this study is however problematic due to different inclusion criteria, absence of a hold-out test set, and ternary classification.

Although this study focused on discerning healthy older adults from older adults with MCI using a single point in time measurement, these findings may well have a bearing on their use of frequent cognitive monitoring. As pointed out by Piau et al. [46], perhaps the biggest shortcoming of today's neuropsychological examination is that it is taken at discrete points in time at large intervals. This makes results vulnerable to temporary alterations in motivation or cognition (e.g. stress or tiredness). As argued by Pavel et al. [48], general principles of measurement may be extended to psychological processes. By increasing the amount of measurements, measurement uncertainty caused by imperfections of the tool can be reduced and natural variations in cognition caused by characteristics of the phenomenon can be detected. The spatial and temporal richness of data derived from longitudinal gameplay may allow for a more detailed cognitive profile and could signal events where cognition was altered (e.g. impact of changes in medication regimen or traumas) [48]. In addition, personal cognitive baselines can be created which allow the individual to be compared with themselves as opposed to normative data [61]. These cognitive baselines could be used to detect subtle cognitive fluctuations, an early indicator of cognitive change [46], [48], [290].

Finally, there are limitations to this study that should be addressed in future work. In particular, the small sample size refrains us from drawing any absolute conclusions. This might lead to potential bias in the test set, explain the performance discrepancy between the test and validation set. In addition, discrepancies in age, tablet experience, and Klondike Solitaire experience between both groups may equally confound results. Confirmatory studies with larger and more balanced sample sizes are needed to further investigate the psychometric properties of using casual card games for screening.

5.5 Conclusion

This study set out to investigate the suitability of the card game Klondike Solitaire to detect Mild Cognitive Impairment through machine learning. The major finding of this study is that casual card games, not built for the purpose of measuring cognition, can be used to capture digital biomarkers of cognitive performance which are sensitive to cognitive impairment caused by MCI. Hence, the popularity of casual games amongst today's older generations may prove useful for supplying cognitive information between consultations. Notwithstanding the relatively small sample size, this work offers valuable insights into the use of casual games to detect cognitive impairments.

5.6 Statements

5.6.1 Acknowledgement

The authors would like to thank all participants who volunteered for this study. The authors also thank the staff of the memory clinics of University Hospital Leuven and Jessa Hospital for making recruitment possible.

5.6.2 Statement of Ethics

This study has been conducted in compliance with the declaration of Helsinki and all applicable national laws and rulings concerning privacy. Approval was granted by the Ethics Committee Research UZ/KU Leuven, Belgium, CTC S59650. All tests were conducted after written informed consent from the participants. Collected data related to the cognitive status during the observations was made anonymous and stored in a secure database. All participants were informed that no information would be used for diagnostic or clinical purposes.

Part IV

General Discussion

Chapter 6

Discussion

In this dissertations, we investigated the possibilities of assessing cognitive performance in older populations through meaningful play, in particular via digital card games. Therefore, a series of studies were conducted which assessed either the impact of cognitive aging through FreeCell or the impact of Mild Cognitive Impairment through Klondike Solitaire. This chapter first recapitulates the research questions posed in the introduction and reflects on the scientific contributions made during this doctoral study. Next, we discuss directions for future work and how COTS games can further evolve to tools used for frequent longitudinal monitoring. Finally, we end this chapter with a general evaluation of the outcomes, a discussion of the limitations, and a reflection on the ethical implications of using commercial off-the-shelf video games for cognitive assessment.

6.1 Revisiting Research Questions

Objective: To assess cognitive performance in elderly life via meaningful play

RQ 1. How can game data be captured from commercial of-the-shelf digital card games?

In this doctoral thesis, we have shown that card game data can be captured from digital card games. One of the biggest technical hurdles in working with

COTS games is extracting cognitive information from said games, as there is no access to the source code nor to interaction logs, complicating the process of creating insightful digital biomarkers. To this end, we built an Android application of Klondike Solitaire. This allowed us to capture raw data used for analysis in chapters 4 and 5.

However, one could argue that such a custom built variant of Klondike Solitaire, while providing meaningful play, still does not represent a truly *commercial* off-the-shelve game. Therefore, we explored capturing data from the commercial Microsoft Solitaire Collection [291], currently shipped with Windows 10. A novel method of extracting game data from COTS games was explored by means of computer vision in collaboration with the EAVISE lab of KU Leuven. Chapter 2 details the development of this image processing toolkit for card games. The toolkit, a multithreaded C++ desktop application built with OpenCV, captures images from the Microsoft 10 Solitaire Collection, analyzes them, and calculates digital biomarkers of cognitive performance. In total, two master theses [292], [293] have been building on and improving this toolkit. With the current toolkit, it is possible to extract potential digital biomarkers from the whole Microsoft 10 Solitaire Collection. Results showed that it was possible to accurately analyze gameplay in real-time without interference and with minimal stress on the computer. On the other hand, this study also showed the caveats for this approach. We found it was necessary to turn down in-game animations, which interfered with analyzing the images. Additionally, it required manually setting the correct image threshold values. In the last iteration of this toolkit [294], we explored integrating deep learning models in the toolkit which may alleviate some of the current limitations.

To the best of the author's knowledge, this study presents the first attempt to extract digital biomarkers from card game play through computer vision. Therefore, the study can function as a proof-of-concept, paving the way forward for using COTS games as a medium to collect digital biomarkers. By detailing the methods and releasing the source code, we hope to inspire future research to apply this process to other games of interest.

RQ 2. How can insights from game design and cognitive psychology be combined to transform game data into potential digital biomarkers of cognitive performance?

This research question tackles one of the biggest methodological challenges of this dissertation. Multidisciplinary in its nature, this project hinges on a successful collaboration between engineering and medical sciences. To bridge this gap, we conducted many observations, guided interviews, and casual conversations with

physicians and patients. Results of this extensive research were first published in a Work-In-progress in 2017 [125] and are brought together in Chapter 4, which is currently under review.

In this chapter, we asked eleven experts in the domain of MCI to rank 21 player actions of Klondike Solitaire to nine cognitive functions. The results showed that all player actions were at least moderately to strongly correlated to one other cognitive function. Similarly, each cognitive function had at least one player action to which it was on average moderately to strongly correlated. These results were validated using intraclass correlations, which primarily indicated good to excellent reliability for both player actions and cognitive functions. Using these insights, the potential digital biomarkers were crafted.

One of the most important findings to emerge from this study is the importance of *contextualisation* (i.e., the enrichment of game metrics with additional information from the game, to be able to function as biomarker). As indicated by Pavel et al. [48], merely using end-results such as the score or total time played is often insufficient to get a thorough overview of a cognitive profile. While motivating and fun-inducing, these end-results are not directly related to a single cognitive function nor do they take difficulty into account. By dissecting the game into more granular digital biomarkers, more direct links to cognitive functions can be made. This effect of contextualization can be seen in e.g., the analysis of the average total time of a move and of gametime. While gametime was found to differ insignificantly, the average total move time differed significantly between both groups.

In sum, the empirical findings in this study support our methodology for creating digital biomarkers. To fully optimize the use of digital biomarkers, contextualization is imperative to make accurate inferences of underlying cognitive processes.

RQ 3. To what extent can differences in cognition due to cognitive aging be assessed using digital biomarkers of cognitive performance?

To assess whether traces of cognitive aging could be detected via digital card games, a data acquisition study was set up, gathering potential digital biomarkers found in FreeCell from three different age groups. Chapter 3 details the machine learning process used to train models to classify games in one of the three age groups. Important to note is that none of the participants had ever played FreeCell before taking part in this study. Each participant was only given a standardized presentation of the game rules and a practice game. Consequently, each participant had to rely predominantly on fluid intelligence [295] while playing the game. Despite the natural high heterogeneity of cognition in humans,

sufficient global performance metrics were obtained on the test set. However, individual ROC curves showed discrepancies amongst age groups, with the model being suboptimal in distinguishing games played by the middle-aged adults from younger and older counterparts.

While the effects of cognitive aging are more subtle than those of Mild Cognitive Impairment, these results show that traces of aging can be captured through game-based digital biomarkers, which can be used to train machine learning models to distinguish older and younger age groups.

RQ 4. To what extent can differences in cognition due to Mild Cognitive Impairment be assessed using digital biomarkers of cognitive performance?

To evaluate this research question, potential digital biomarkers from Klondike Solitaire were gathered from 23 older adults living with MCI and 23 healthy older adults. Two analyses were conducted on this dataset to evaluate differences on a group level using statistical models (chapter 4) and on an individual level using machine learning models (chapter 5). The GLMM analyses showed that 12 out of 23 digital biomarkers differed significantly across groups while controlling for factors such as Klondike Solitaire experience, tablet experience, and age. In addition, the machine learning process followed in chapter 5 showed that multiple models coming from different underlying algorithms showed comparable psychometric properties as common cognitive screening tests [244].

Combined, these two studies show that digital biomarker performance differed both at the group level, through statistical analysis, and at the individual level, through machine learning analysis. This lend support to the use of COTS games for discriminating older adults with MCI from healthy older adults, vouching for further research and validation studies.

6.2 Overall Scientific Contribution

The overall scientific contribution of this dissertation can be divided into four segments, following the research contribution categories for Human-Computer Interaction of Wobbrock and Kientz [296]. An overview of all the contributions in this dissertation can be found in figure 6.1.

First, we have **one methodological contribution**. Developing game-based digital biomarkers is a process that necessitates successful collaboration between data science and medical science. As such, a multi-step process was developed to deconstruct the game into digital biomarkers. By detailing the process

used to craft the potential digital biomarkers in collaboration with experts, we hope to inform future researchers on how to craft novel digital biomarkers. Second, **two artifacts** were created to capture potential digital biomarkers of cognitive performance. For Klondike Solitaire, an existing Android app of Bielefeld [270] was adapted to yield digital biomarkers. Next, for FreeCell, a tool was developed that could unobtrusively analyze gameplay in real-time using image processing techniques. To the best knowledge of the authors, this is one of the first studies exploring this method of extracting game data. By making the code of both applications publicly available, we hope to inform future research and inspire them to build other digital biomarker yielding COTS games. Third, **two datasets** were gathered over the past four years: a dataset containing digital biomarkers of FreeCell from three different age groups and a dataset containing digital biomarkers of Klondike Solitaire from healthy older adults and older adults living with MCI. By equally making these publicly available, we hope to incite other researchers to scrutinize these datasets or use them as benchmarks. Finally, **three empirical contributions** were made in the form of quantitative analyses of these datasets. These results provide support for the hypothesis that COTS games may be used for the assessment of cognitive performance. By clearly and transparently documenting every step leading to these analyses, from methodology to artifacts and datasets, we aim to strengthen the findings of these empirical contributions.

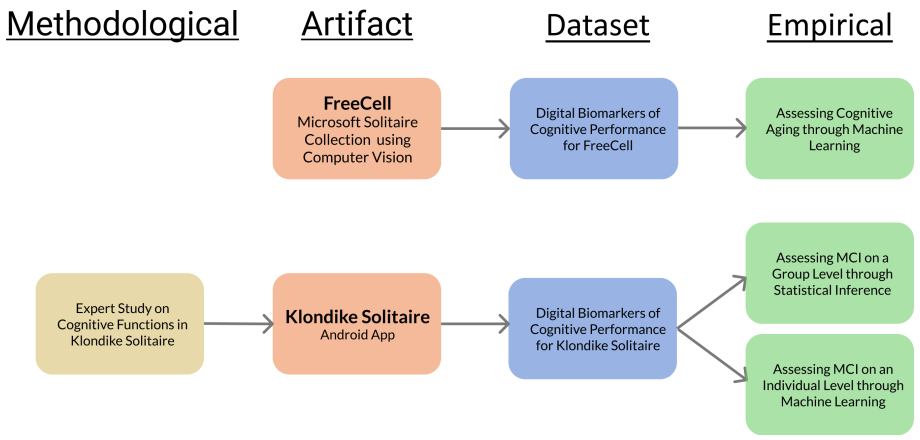


Figure 6.1: Schematic overview of the scientific contributions of this dissertation.

6.3 Towards Frequent Longitudinal Assessment of Cognitive Performance

A natural progression of this work is to investigate how the unique benefits of games can be utilized for frequent (i.e., daily or multiple times per week) longitudinal cognitive monitoring. The results gathered in this dissertation show the value of game-based digital biomarkers for single-point-in-time measurements. However, research stresses the importance of longitudinally monitoring changes in cognition to timely change diagnosis or adjust disease management approach [3]. Unfortunately, a critical review on current MCI management poses that "the optimal timing, choice, and cost-effectiveness of longitudinal cognitive assessments are unclear", indicating that the amount of cognitive assessments is constrained by the inherent cost and time necessitated from patient and physician [297]. In addition, as pointed out by Pavel et al. [48], classical neuropsychological tests are not designed to be enjoyed and are fixed in their format. This makes them suboptimal for frequent longitudinal measurements due to a lack of adherence and vulnerable to learning effects. COTS games, with the important sidenote that they are meaningful for the player, could provide a cost-efficient tool with changing challenges, reducing learning effects, while keeping the player engaged. To develop frequent longitudinal game-based cognitive monitoring, additional studies and improvements to existing work will be necessary.

6.3.1 Improving Current Models

After scrutinizing our existing models, two main opportunities for improvement become clear. First, a more general improvement is to further explore the contextualization of the digital biomarkers. In this dissertation, only the surface of contextualization was scratched by segmenting digital biomarkers and relating them to other digital biomarkers. From a cognitive perspective, it can be interesting to also explore patient reactions to game events. For example, nobody makes an error on purpose. This surprises the player when the game does not validate their move and throws them off their game. As such, an interesting biomarker to explore is the time necessary to come up with a move after an erroneous move. Another example is investigating series of timings necessary to think of a move. Should these be consequently low and successful, this might indicate that the player is capable of planning ahead, devising multiple steps, and remembering them. Second, the way beta errors are captured in Klondike Solitaire can be refined. These beta errors aimed to detect missed opportunities of moving cards by scanning the whole board for possible

moves when the player requested new cards from the pile. Observational notes made while visiting participants regularly indicated that this was one of the most notable differences in playstyle between groups. When results showed that these beta errors did not differ significantly, a selection of games was manually replayed to see whether these beta errors were correctly captured. After verifying the accurateness of the data, it stood out that the beta errors made by both groups were different in nature. The healthy group appeared to make fewer beta errors revolving around accidentally missing moves between build stacks. Their beta errors were more intentional in nature and happened because the user did not want to place cards on the suit stacks in case that would prevent future cards from the pile to be placed on the build stacks. Therefore, the hypothesis that beta errors happen more in the MCI group needs to be refined.

6.3.2 Unlocking the Black Box for Interpretable Machine Learning

In contrast to humans, machine models do not articulate their rationale. When new data is presented, it puts out a regression or classification, substantiated by complex interactions between feature weights. The medical world is, rightly so, skeptical when adopting complex black-box approaches where data is fed into the system and a prediction is *automagically* given. To resolve this distrust, making models more transparent might aid in building successful data-driven medical models that are adopted by the medical world.

First, dimensionality reduction techniques that transform features, such as principal component analysis, should be abstained from. These techniques aim to reduce the complexity of models by reducing the number of features while keeping a maximum amount of information by making (non-)linear combinations. While proven to be effective, these techniques make it more difficult to understand how decisions are made by the model as they sever the direct link to the original features. Second, the decision-making process of models needs to be investigated by calculating feature importance [285]. Feature importance quantifies how useful a feature is to predict the target variable and gives insight into how decisions are made in general. Third, in addition to this general understanding of models, more advanced machine learning interpretability techniques, such as Shapely Additive exPlanations (SHAP) [298], [299], can be used to make models more transparent. SHAP can be used to calculate feature importance in an algorithm-agnostic way, facilitating the process of comparing models built using different algorithms. Most importantly, SHAP can be used to dissect the decision-making process of *individual* predictions, making it possible to understand the rationale of the model on a case-by-case level. This rationale

behind a classification can be more valuable to the physician than the outcome itself as it can provide insight into what behavior led to the classification. Finally, the information gathered using the techniques above must by effectively relayed to the physician with utmost care. Using a dashboard application [300], cognitive information can be visually communicated at a glance. In this dashboard, special attention should be given to the change in biomarker performance over time (as monitoring changes in cognition is important for MCI [3]) and how this affects the certainty of the prediction made by the model.

6.3.3 Detecting Abnormal Change

Frequent longitudinal monitoring opens up new possibilities to quantify cognitive impairment. Pavel et al. [48] have argued that the general benefits of increasing measurements equally apply to measuring cognition. The advantage of increasing measurements is twofold. First, measurement uncertainty caused by random instrument errors can be reduced. Second, it allows for creating individual cognitive baselines that contain personal cognitive information. This baseline can be used to compare the person with itself instead of normative data [61] and make it possible to detect and evaluate alterations in cognition caused by changes in medication or trauma. Furthermore, this cognitive baseline allows for detecting specific temporal characteristics of the disease [48]. For MCI, it could be interesting to detect cognitive fluctuations [248]. Cognitive fluctuations are temporary alterations in cognition commonly found in MCI but also in Lewy Body Dementia or Alzheimer's Disease [248], [290], [301]. Tracking these cognitive fluctuations and how they evolve, can help us better understand the layers of transition leading to dementia. In addition, they can assist in making changes to diagnosis and disease management by timely signaling changes in cognition.

Longitudinal cognitive monitoring can build on the techniques of assessing cognition used in this dissertation. Refining the machine learning models trained for these studies, it can be interesting to investigate the certainty of the model's prediction over time. Most binary classification algorithms output a raw value, which is transformed into a probability value [285]. Comparing this value with a learned cut-off value leads to the eventual classification. Monitoring how these probability values change over time can inform the physician that the game behavior of the patient is gradually moving further from or towards typical MCI behavior learned by the model. Combining the trends of these values with SHAP, corroborating this probability number, can lead to insights into the person's cognitive trajectory. A specific implementation, as suggested by Pavel et al. [48], is to use sliding window techniques to aggregate data to detect these trends.

6.3.4 Understanding the Gamer behind the Impairment

To get a better grasp on the person playing the game, research should investigate the reasons behind playing Klondike Solitaire. Research from Greenberg et al. [302] and Sherry et al. [303] touch on the different gratifications which games have to offer, ranging from challenge to diversion. It is intriguing that Klondike Solitaire, a game with such a low win rate and highly depending on random shuffles, can be such a popular pastime. In the context of game-based cognitive monitoring, this natural low win rate can be seen as an advantage. For our research on the impact of MCI on Klondike Solitaire gameplay, many healthy older adults and older adults with MCI were visited in their home environment, each of them playing several rounds of Klondike Solitaire and undergoing a thorough neuropsychological test battery. Observational notes made during these visits indicated that both groups showed signs of frustration and distress during the neuropsychological tests. This emotional effect of neuropsychological examination has also been studied by Wong et al. [52] and Lai et al. [53].

In contrast, while playing Klondike Solitaire, none of the participants showed signs of frustration. Verbal utterances made by the participants while playing indicated that blame often shifted from their own performance to "a bad shuffle of the deck"[1]. This random aspect of shuffles in the game can make Klondike Solitaire more suitable for longitudinally monitoring, especially in populations with MCI where cognitive impairment will often hinder game performance over time. Investigating the motivation behind play and the effect of random factors on frustration might assist in pinpointing other games suitable for longitudinal cognitive monitoring.

6.3.5 Overcoming Longitudinal Barriers

Finally, to better grasp the implications of frequent longitudinal monitoring, current and possible future limitations have to be investigated. First, future researchers should refine existing studies. As indicated in Chapter 4 and 5, current experiments are underpowered and should be revisited with larger sample sizes and with the possible improvements described in 6.3.1 to more accurately assess the impact of MCI to Solitaire gameplay. Confirming these results and investigating suggested improvements is essential to train more robust models that can be evaluated in clinical trials. Second, the concerns on computerized neuropsychological assessment devices voiced by Bauer et al. [304] equally apply on digital biomarkers derived from COTS games. Before

[1]During the experiments, none of the participants were informed that each round of Solitaire was completable. As such, their expectations to actually solve a round were kept low, reducing frustration when suboptimal play prevented them from completing a round.

implementing them in practice, these games must follow the same standards as neuropsychological tests. In addition, the effects of self-administration on the psychometric properties have to be investigated. Furthermore, the effect of differing specifications has to be taken into account. For these studies, all participants did an identical experiment on an identical device. Certain specifications of digital devices such as the CPU, RAM, or even resolution of the screen might impact the measurements. Third, as indicated by Bent et al. [62], digital biomarkers currently have lacking standards and validation methods. Piau et al. [305] specifically note that while the first large-scale monitoring initiatives were conducted ten years ago, little to no cross-referencing of digital biomarkers with biological or imaging biomarkers has been done. To mature as a research domain, a clear framework must be built that can be used to critically evaluate and validate potential digital biomarkers. After taking these concerns into consideration, studies can be set up to investigate the effects of measuring digital biomarkers longitudinally.

Finally, it has to be recognized that longitudinally measuring cognition may complicate measurements. For example, practice effects, a common problem among cognitive tests where participants improve their results when assessed [306], could arise when users play the game more frequently. Investigating this effect and taking frequency of play into account when training machine learning models can mitigate inaccurate interpretations of results. In addition, while the data from this dissertation was captured in the home environment of the participant, it was still supervised. Data captured from unsupervised longitudinal measurements will inadvertently be influenced by interruptions, causing outliers in the data. Future research should be aware of these effects and mitigate them by using aggregations that are resilient to outliers (e.g. medians opposed to minimums and maximums) or by removing outliers. Moreover, it has to be noted that the motivation or enjoyment to play the game might fleet over time. While abandonment of the game might indicate that the person is unable to play the game due to their cognitive impairment, it might equally indicate a shift in interest. Being aware of these different causes that have the same outcome might contribute to a more accurate longitudinal assessment. Critically assessing these complications in future research is crucial to prevent negative outcomes.

6.4 Ethical reflections

The premise of this dissertation, i.e., using COTS games to assess cognitive performance, prompted much debate during the course of the doctoral trajectory with other researchers, reviewers, and participants. This topic necessitates

ethical reflection and touches on larger ethical discussions such as digital health [307], [308], patient engagement [309], and the use of machine learning [310], [311]. In this section we will touch on the most important topics.

The intention of this dissertation is to enable people to play games that are meaningful to them, in a familiar environment, with the possibility of unobtrusively assessing cognitive performance. While the technical possibilities of game-based digital biomarkers are promising, it has to be recognized that we are transforming an innocent pastime into an instrumental activity that might trigger feelings of surveillance [279], extending the assessing gaze of health care providers into the living room of older adults [312] and contributing to a discourse where such practices are normalized. To ensure capturing a sufficient amount of data, it may even be tempting to stimulate people to play a minimum amount of games per week through notifications. This form of "prescription gaming" may confront the patient with their condition and increase their burden [312]. As such, it has to be investigated that the assumption of meaningful play [103] is not violated by adding digital biomarker measurements. The pressure of performing well, on what was previously a leisure activity, could lead to the abandonment of the game. Moreover, passive monitoring in fragile populations, like those of MCI, could inadvertently increase stigma and lead to unnecessary intensified concern from the family and surroundings. Especially during the conversion process from MCI to AD, patients might suffer from decisional impairment, stressing the importance of being meticulous in the informed consent process [313]. *Therefore, to ensure maintaining the* ***meaningful*** *aspect of the game, research has to investigate the impact of adding digital biomarkers to COTS games. In addition, utmost care must be taken to assist the patients in their medical decision-making.*

Even with proper consent, care must be taken with how and to whom information is relayed. Tools that monitor cognitive performance fit in a trend towards *the quantified self* and the digitally engaged patient. This current trend in digital healthcare promotes individuals to self-monitor their daily activities (e.g., sleep, steps, or mood) and stimulates self-reflection [307]. This trend is driven by the assumption that the individual is empowered by showing meaningful data-driven insights tailored to the individual. However, research [312] has shown that if healthcare data contradicts the patient's own subjective interpretation of their medical status, feelings of anxiety, helplessness, and fear could be induced. Equally for MCI, as indicated by Petersen et al. [3], how information is shared has to be weighed against the risk of inducing anxiety for a disease that might not even progress. *Therefore, it is advised to solely relay information to the physician and rely on their judgment to share what is beneficial to the patient.*

Moreover, even when information is exclusively shared with the physician, researchers must be conscientious when developing algorithms. Algorithms

excel in finding patterns in data yet are not able to discriminate between correlation and causality [310]. This may cause models that are performant on test data to fail in real-life scenarios, a phenomenon better known as "shortcut learning" [311]. Biases can be induced by supplying wrong data to the algorithm, a famous example being a classifier of pneumonia that learned to detect a hospital-specific metal token instead of learning to detect pneumonia. This form of Big Data Analytics (BDA) favors data-driven approaches that are atheoretical in nature. These techniques revolve around capturing as much data as possible and using machine learning techniques to gather inductive knowledge and find correlations in the data. In the long run, BDA approaches equally need epistemological reflection to be relevant, to have a scientific contribution they cannot abstain from relating to theory [314]. *Therefore, in the context of leveraging complex interactive technologies such as games for yielding digital biomarkers, we advocate for the systematic inclusion of domain experts and patients, using a multi-step process that prioritizes comprehensibility and patient trust over accuracy and cost-effectiveness.*

6.5 Conclusion

Today, assessment of cognitive performance is an agglomeration of cognitive screening tests, biomarker examinations, and thorough neuropsychological assessment. Individually, none of these instruments are considered accurate enough to consistently support a reliable diagnosis, screening, or case-finding. Game-based digital biomarkers may function as an additional instrument to the aforementioned neuropsychological assessment ecosystem, empowering clinicians in case finding and diagnosis while providing patients with a less stressful alternative. In essence, Solitaire games are a constant stream of stimuli in the form of cards, prompted by the player, which require comparison with other stimuli based on two factors (color and rank). Their easy to grasp rules allow for an unlimited amount of challenges without changing core gameplay. However, care will need to be given to the use and implementation of such games in health care systems. Considering the results in this dissertation on the efficacy of game-based digital biomarkers for single-point-in-time measurements, I look forward to the prospect of meaningful play, elucidating the current blind spot between consultations and contributing to a more complete cognitive profile.

Appendix A

Model Performances

This section details the 19 classification models trained for the study described in chapter 5. In addition to the 19 models, a dummy classifier was trained to act as a baseline and to ascertain that no data was added or removed during the preprocessing stage. Models were selected based on their popularity, performance, and support in the sci-kit learn machine learning library.

All models were trained in four different scenario's: without preprocessing, with transformed features, with scaled features, and with transformed and scaled features. Heavily skewed features were transformed using natural log and square root transformation depending on their level of skewness. Transforming features is advised as some machine learning algorithms (e.g., Logistic Regression) assume normally distributed data. After each scenario, the average 5-fold cross-validated F1 score of all models without the dummy classifier was calculated.

An overview of all Training set F1 scores for each scenario can be found in table A. Results showed that the preprocessing scenario without preprocessing had an average 5-fold cross-validated F1 score of 0.650, which was the lowest of all scenarios. The scenario with only transforming features had an average of 0.665, transforming and scaling features had an average of 0.699, and only scaling features had an average of 0.741. The three most performant models of the most performant pre-processing scenario were selected for hyperparameter optimization. The results of these fine-tuned models are described in section 5.3. While the scenario with scaled features had on average the best results, it can be noted that the Random Forest Classifier had an F1 score of 0.8035 in the transformed and scaled scenario, the Bagging Classifier an F1 score of 0.7909 in the transformed scenario, and the XGBoost Decision Tree Classifier an F1 score of 0.7702 in the no preprocessing scenario.

Model	No Preprocessing	Transformed Features	Transformed and Scaled Features	Scaled Features
Nu-Support Vector Classifier	0.7068	0.5414	0.6853	0.7912
Gradient Boosting Classifier	0.7428	0.693	0.759	0.7719
Extra Trees Classifier	0.7579	0.7525	0.7405	0.7693
Logistic Regression CV Classifier	0.7439	0.7247	0.7307	0.7665
Support Vector Classifier	0.5315	0.5637	0.5637	0.7625
XGBoost Decision Tree Classifier	0.7702	0.6862	0.7358	0.7608
Bagging Classifier	0.6903	0.7909	0.7450	0.7491
XGBoost Random Forest Classifier	0.7469	0.7390	0.7219	0.7462
Multi-layer Perceptron Classifier	0.3454	0.6611	0.7042	0.7448
Random Forest	0.7889	0.7592	0.8035	0.7435
AdaBoost Classifier	0.7307	0.6658	0.7391	0.7391
Decision Tree Classifier	0.7035	0.7137	0.7188	0.7324
Gaussian Naïve Bayes Classifier	0.7334	0.742	0.7328	0.7225
Logistic Regression	0.6314	0.6484	0.7547	0.7161
Linear Stochastic Gradient Descent Classifier	0.2708	0.5048	0.4624	0.7154
K-Nearest Neighbors Classifier	0.5771	0.5681	0.5616	0.7104
Quadratic Discriminant Analysis Classifier	0.7281	0.6217	0.761	0.7046
Linear Support Vector Classifier	0.5072	0.591	0.6452	0.7005
Gaussian Process Classifier	0	0.7284	0.7200	0.6761
Dummy Classifier	0.1290	0.1290	0.1290	0.129
Average	0.650	0.665	0.699	0.741

Table A.1: Training F1 scores of the 19 models for all four preprocessing scenarios.

Bibliography

[1] A. L. Norman, J. L. Woodard, J. E. Calamari, E. Z. Gross, N. Pontarelli, J. Socha, B. DeJong, and K. Armstrong, "The fear of Alzheimer's disease: Mediating effects of anxiety on subjective memory complaints", en, *Aging & Mental Health*, vol. 24, no. 2, pp. 308–314, Feb. 2020, ISSN: 1360-7863, 1364-6915. DOI: 10.1080/13607863.2018.1534081. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/13607863.2018.1534081 (visited on 10/19/2020).

[2] M. D. Binder, N. Hirokawa, and U. Windhorst, Eds., *Encyclopedia of Neuroscience*, en. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. DOI: 10.1007/978-3-540-29678-2. [Online]. Available: http://link.springer.com/10.1007/978-3-540-29678-2 (visited on 10/16/2020).

[3] R. C. Petersen *et al.*, "Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology", en, *Neurology*, vol. 90, no. 3, pp. 126–135, Jan. 2018, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.0000000000004826. [Online]. Available: http://www.neurology.org/lookup/doi/10.1212/WNL.0000000000004826 (visited on 09/10/2019).

[4] R. O. Roberts *et al.*, "Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal", en, *Neurology*, vol. 82, no. 4, pp. 317–325, Jan. 2014, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.0000000000000055. [Online]. Available: http://www.neurology.org/cgi/doi/10.1212/WNL.0000000000000055 (visited on 03/25/2020).

[5] M. C. Polidori, G. Nelles, U. Senin, and P. Mecocci, "Cognitive Decline", in *Learning Geriatric Medicine*, R. Roller-Wirnsberger, K. Singler, and M. C. Polidori, Eds., Series Title: Practical Issues in Geriatrics, Cham: Springer International Publishing, 2018, pp. 67–80. DOI: 10.1007/978-

3-319-61997-2_7. [Online]. Available: http://link.springer.com/10.1007/978-3-319-61997-2_7 (visited on 10/16/2020).

[6] W. H. Organisation, *Global action plan on the public health response to dementia 2017-2025*. Geneva, Switzerland, 2017, ISBN: 978-92-4-151348-7.

[7] R. Knight, M. Khondoker, N. Magill, R. Stewart, and S. Landau, "A Systematic Review and Meta-Analysis of the Effectiveness of Acetylcholinesterase Inhibitors and Memantine in Treating the Cognitive Symptoms of Dementia", en, *Dementia and Geriatric Cognitive Disorders*, vol. 45, no. 3-4, pp. 131–151, 2018, ISSN: 1420-8008, 1421-9824. DOI: 10.1159/000486546. [Online]. Available: https://www.karger.com/Article/FullText/486546 (visited on 03/30/2020).

[8] L. Webster *et al.*, "Core outcome measures for interventions to prevent or slow the progress of dementia for people living with mild to moderate dementia: Systematic review and consensus recommendations", en, *PLOS ONE*, vol. 12, no. 6, T. J. Quinn, Ed., e0179521, Jun. 2017, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0179521. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0179521 (visited on 03/30/2020).

[9] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. Kokmen, and E. G. Tangelos, "Aging, Memory, and Mild Cognitive Impairment", *International Psychogeriatrics*, vol. 9, no. S1, pp. 65–69, 1997. DOI: 10.1017/S1041610297004717.

[10] J. Rodakowski, E. Saghafi, M. A. Butters, and E. R. Skidmore, "Non-pharmacological interventions for adults with mild cognitive impairment and early stage dementia: An updated scoping review", en, *Molecular Aspects of Medicine*, vol. 43-44, pp. 38–53, Jun. 2015, ISSN: 00982997. DOI: 10.1016/j.mam.2015.06.003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0098299715000382 (visited on 07/03/2019).

[11] G. Livingston *et al.*, "START (STrAtegies for RelaTives) study: A pragmatic randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of a manual-based coping strategy programme in promoting the mental health of carers of people with dementia", en, *Health Technology Assessment*, vol. 18, no. 61, pp. 1–242, Oct. 2014, ISSN: 1366-5278, 2046-4924. DOI: 10.3310/hta18610. [Online]. Available: https://www.journalslibrary.nihr.ac.uk/hta/hta18610/ (visited on 07/03/2019).

[12] W. H. Organization, "Risk reduction of cognitive decline and dementia: WHO guidelines", World Health Organization, Geneva, Switzerland, Tech. Rep., 2019.

[13] Committee on the Public Health Dimensions of Cognitive Aging, Board on Health Sciences Policy, and Institute of Medicine, *Cognitive Aging: Progress in Understanding and Opportunities for Action*, eng, ser. The National Academies Collection: Reports funded by National Institutes of Health. Washington (DC): National Academies Press (US), 2015. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK285319/ (visited on 03/18/2019).

[14] E.
bibinitperiod I. S. a. T. R. H. G. C. M. Department for Business, *Government announces 300 million pound for landmark ageing society grand challenge*, en, Mar. 2018. [Online]. Available: https://www.gov.uk/government/news/government-announces-300-million-for-landmark-ageing-society-grand-challenge (visited on 04/09/2019).

[15] "National Dementia Strategy 2014-2019", Federal Office of Public Health, Switserland, Tech. Rep., 2016.

[16] "National Action Plan on Dementia 2025", Healthcare Denmark, Tech. Rep., 2018. [Online]. Available: https://www.healthcaredenmark.dk/media/1626294/HCD-Dementia-white-paper-v1-single-0318.pdf (visited on 12/05/2019).

[17] "A Dementia Strategy for Canada", Public Health Agency of Canada, Tech. Rep., 2019, p. 110. [Online]. Available: https://www.canada.ca/content/dam/phac-aspc/images/services/publications/diseases-conditions/dementia-strategy/National%20Dementia%20Strategy_ENG.pdf (visited on 07/04/2019).

[18] J. Torous, J. Rodriguez, and A. Powell, "The New Digital Divide For Digital Biomarkers", en, *Digital Biomarkers*, Jun. 2017, ISSN: 2504-110X. DOI: 10.1159/000477382. [Online]. Available: https://www.karger.com/Article/FullText/477382 (visited on 07/09/2019).

[19] F. Martínez-Santiago, M. R. García-Viedma, J. A. Williams, L. T. Slater, and G. V. Gkoutos, "Aging Neuro-Behavior Ontology", *Applied Ontology*, vol. 15, no. 2, pp. 219–239, May 2020, ISSN: 18758533, 15705838. DOI: 10.3233/AO-200229. [Online]. Available: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/AO-200229 (visited on 11/12/2020).

[20] N. D. Anderson and F. I. M. Craik, "50 Years of Cognitive Aging Theory", en, *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 72, no. 1, pp. 1–6, Jan. 2017, ISSN: 1079-5014, 1758-5368. DOI: 10.1093/geronb/gbw108. [Online]. Available: https://academic.oup.com/psychsocgerontology/article-lookup/doi/10.1093/geronb/gbw108 (visited on 11/12/2020).

[21] R. B. Cattell, "Theory of fluid and crystallized intelligence: A critical experiment.", en, *Journal of Educational Psychology*, vol. 54, no. 1, pp. 1–22, Feb. 1963, ISSN: 1939-2176, 0022-0663. DOI: 10.1037/h0046743. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0046743 (visited on 11/12/2020).

[22] E. M. Tucker-Drob, "Cognitive Aging and Dementia: A Life-Span Perspective", en, *Annual Review of Developmental Psychology*, vol. 1, no. 1, pp. 177–196, Dec. 2019, ISSN: 2640-7922, 2640-7922. DOI: 10.1146/annurev-devpsych-121318-085204. [Online]. Available: https://www.annualreviews.org/doi/10.1146/annurev-devpsych-121318-085204 (visited on 11/12/2020).

[23] T. A. Salthouse, "Selective review of cognitive aging", en, *Journal of the International Neuropsychological Society*, vol. 16, no. 05, pp. 754–760, Sep. 2010, ISSN: 1355-6177, 1469-7661. DOI: 10.1017/S1355617710000706. [Online]. Available: http://www.journals.cambridge.org/abstract_S1355617710000706 (visited on 03/27/2019).

[24] M. Lally and S. Valentine-French, *Lifespan Development: A Psychological Perspective*. College of Lake County, 2019.

[25] F. Navarro, "The woodcock-johnson tests of cognitive ability, third edition", Mar. 2010.

[26] J. J. McArdle, E. Ferrer-Caja, F. Hamagami, and R. W. Woodcock, "Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span", eng, *Developmental Psychology*, vol. 38, no. 1, pp. 115–142, Jan. 2002, ISSN: 0012-1649.

[27] J. J. McArdle and L. Wang, "Modeling age-based turning points in longitudinal life-span growth curves of cognition.", in *Applied data analytic techniques for turning points research.* Ser. Multivariate applications series. New York, NY, US: Routledge/Taylor & Francis Group, 2008, pp. 105–127.

[28] T. A. Salthouse, "Trajectories of normal cognitive aging.", en, *Psychology and Aging*, vol. 34, no. 1, pp. 17–24, Feb. 2019, ISSN: 1939-1498, 0882-7974. DOI: 10.1037/pag0000288. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/pag0000288 (visited on 11/12/2020).

[29] K. V. Papp, "Dementia", en, in *Encyclopedia of Clinical Neuropsychology*, J. S. Kreutzer, J. DeLuca, and B. Caplan, Eds., New York, NY: Springer New York, 2011, pp. 804–809. DOI: 10.1007/978-0-387-79948-3_1118. [Online]. Available: http://link.springer.com/10.1007/978-0-387-79948-3_1118 (visited on 11/17/2020).

[30] A. Association, "2016 Alzheimer's Disease Facts and Figures", Alzheimer's & Dementia 2016, Tech. Rep., 2016.

[31] J. P. M. Kane *et al.*, "Clinical prevalence of Lewy body dementia", en, *Alzheimer's Research & Therapy*, vol. 10, no. 1, p. 19, Dec. 2018, ISSN: 1758-9193. DOI: 10.1186/s13195-018-0350-6. [Online]. Available: https://alzres.biomedcentral.com/articles/10.1186/s13195-018-0350-6 (visited on 03/30/2020).

[32] G. Mackenzie and R. Will, "Creutzfeldt-Jakob disease: Recent developments", en, *F1000Research*, vol. 6, p. 2053, Nov. 2017, ISSN: 2046-1402. DOI: 10.12688/f1000research.12681.1. [Online]. Available: https://f1000research.com/articles/6-2053/v1 (visited on 12/31/2020).

[33] R. C. Petersen, "Mild cognitive impairment as a diagnostic entity", en, *Journal of Internal Medicine*, vol. 256, no. 3, pp. 183–194, Sep. 2004, ISSN: 0954-6820, 1365-2796. DOI: 10.1111/j.1365-2796.2004.01388.x. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-2796.2004.01388.x (visited on 04/01/2020).

[34] E. Bartfay, W. J. Bartfay, and K. M. Gorey, "Prevalence and correlates of potentially undetected dementia among residents of institutional care facilities in Ontario, Canada, 2009-2011: Prevalence of potentially undetected dementia", en, *International Journal of Geriatric Psychiatry*, vol. 28, no. 10, pp. 1086–1094, Oct. 2013, ISSN: 08856230. DOI: 10.1002/gps.3934. [Online]. Available: http://doi.wiley.com/10.1002/gps.3934 (visited on 06/28/2017).

[35] A. Cherubini *et al.*, "Underrecognition and Undertreatment of Dementia in Italian Nursing Homes", en, *Journal of the American Medical Directors Association*, vol. 13, no. 8, 759.e7–759.e13, Oct. 2012, ISSN: 15258610. DOI: 10.1016/j.jamda.2012.05.015. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1525861012001661 (visited on 06/28/2017).

[36] V. G. Valcour, K. H. Masaki, J. D. Curb, and P. L. Blanchette, "The Detection of Dementia in the Primary Care Setting", en, *Archives of Internal Medicine*, vol. 160, no. 19, p. 2964, Oct. 2000, ISSN: 0003-9926. DOI: 10.1001/archinte.160.19.2964. [Online]. Available: http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.160.19.2964 (visited on 07/03/2019).

[37] F. Jessen *et al.*, "The characterisation of subjective cognitive decline", en, *The Lancet Neurology*, vol. 19, no. 3, pp. 271–278, Mar. 2020, ISSN: 14744422. DOI: 10.1016/S1474-4422(19)30368-0. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1474442219303680 (visited on 02/19/2021).

[38] P. Vannini *et al.*, "Hippocampal hypometabolism in older adults with memory complaints and increased amyloid burden", en, *Neurology*, vol. 88, no. 18, pp. 1759–1767, May 2017, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.0000000000003889. [Online]. Available: http://www.neurology.org/lookup/doi/10.1212/WNL.0000000000003889 (visited on 02/19/2021).

[39] A. Koppara *et al.*, "Cognitive performance before and after the onset of subjective cognitive decline in old age", en, *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 2, pp. 194–205, Jun. 2015, ISSN: 2352-8729, 2352-8729. DOI: 10.1016/j.dadm.2015.02.005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1016/j.dadm.2015.02.005 (visited on 02/19/2021).

[40] B. Winblad *et al.*, "Mild cognitive impairment - beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment", en, *Journal of Internal Medicine*, vol. 256, no. 3, pp. 240–246, Sep. 2004, ISSN: 0954-6820, 1365-2796. DOI: 10.1111/j.1365-2796.2004.01380.x. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-2796.2004.01380.x (visited on 12/31/2020).

[41] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment: MOCA: A BRIEF SCREENING TOOL FOR MCI", en, *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, Apr. 2005, ISSN: 00028614, 15325415. DOI: 10.1111/j.1532-5415.2005.53221.x. [Online]. Available: http://doi.wiley.com/10.1111/j.1532-5415.2005.53221.x (visited on 09/14/2016).

[42] T. N. Tombaugh and N. J. McIntyre, "The Mini-Mental State Examination: A Comprehensive Review", en, *Journal of the American Geriatrics Society*, vol. 40, no. 9, pp. 922–935, Sep. 1992, ISSN: 00028614. DOI: 10.1111/j.1532-5415.1992.tb01992.x. [Online]. Available: http://doi.wiley.com/10.1111/j.1532-5415.1992.tb01992.x (visited on 09/14/2016).

[43] N. D. Anderson, "State of the science on mild cognitive impairment (MCI)", en, *CNS Spectrums*, vol. 24, no. 1, pp. 78–87, Feb. 2019, ISSN: 1092-8529, 2165-6509. DOI: 10.1017/S1092852918001347. [Online]. Available: https://www.cambridge.org/core/product/identifier/S1092852918001347/type/journal_article (visited on 12/17/2020).

[44] K. Strimbu and J. A. Tavel, "What are Biomarkers?", *Current opinion in HIV and AIDS*, vol. 5, no. 6, pp. 463–466, Nov. 2010, ISSN: 1746-630X. DOI: 10.1097/COH.0b013e32833ed177. [Online]. Available: https:

//www.ncbi.nlm.nih.gov/pmc/articles/PMC3078627/ (visited on 02/01/2019).

[45] V. V. Giau, E. Bagyinszky, and S. S. A. An, "Potential Fluid Biomarkers for the Diagnosis of Mild Cognitive Impairment", en, *International Journal of Molecular Sciences*, vol. 20, no. 17, p. 4149, Aug. 2019, ISSN: 1422-0067. DOI: 10.3390/ijms20174149. [Online]. Available: https://www.mdpi.com/1422-0067/20/17/4149 (visited on 12/21/2020).

[46] A. Piau, K. Wild, N. Mattek, and J. Kaye, "Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review", en, *Journal of Medical Internet Research*, vol. 21, no. 8, e12785, Aug. 2019, ISSN: 1438-8871. DOI: 10.2196/12785. [Online]. Available: http://www.jmir.org/2019/8/e12785/ (visited on 03/24/2020).

[47] M. Schlemmer and O. Desrichard, "Is Medical Environment Detrimental to Memory? A Test of A White Coat Effect on Older People's Memory Performance", en, *Clinical Gerontologist*, vol. 41, no. 1, pp. 77–81, Jan. 2018, ISSN: 0731-7115, 1545-2301. DOI: 10.1080/07317115.2017.1307891. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/07317115.2017.1307891 (visited on 09/11/2019).

[48] M. Pavel, H. Jimison, S. Hagler, and J. McKanna, "Using Behavior Measurement to Estimate Cognitive Function Based on Computational Models", in *Cognitive Informatics in Health and Biomedicine: Understanding and Modeling Health Behaviors*, V. L. Patel, J. F. Arocha, and J. S. Ancker, Eds., Cham: Springer International Publishing, 2017, pp. 137–163, ISBN: 978-3-319-51732-2. DOI: 10.1007/978-3-319-51732-2_7. [Online]. Available: https://doi.org/10.1007/978-3-319-51732-2_7.

[49] P. Dagum, "Digital biomarkers of cognitive function", en, *npj Digital Medicine*, vol. 1, no. 1, p. 10, Dec. 2018, ISSN: 2398-6352. DOI: 10.1038/s41746-018-0018-4. [Online]. Available: http://www.nature.com/articles/s41746-018-0018-4 (visited on 06/07/2019).

[50] S. Martin *et al.*, "Attitudes and preferences towards screening for dementia: A systematic review of the literature", en, *BMC Geriatrics*, vol. 15, no. 1, p. 66, Dec. 2015, ISSN: 1471-2318. DOI: 10.1186/s12877-015-0064-6. [Online]. Available: http://bmcgeriatr.biomedcentral.com/articles/10.1186/s12877-015-0064-6 (visited on 09/03/2020).

[51] K. Krohne, Å. Slettebø, and A. Bergland, "Cognitive screening tests as experienced by older hospitalised patients: A qualitative study: Cognitive screening tests as experienced by older hospitalised patients",

en, *Scandinavian Journal of Caring Sciences*, vol. 25, no. 4, pp. 679–687, Dec. 2011, ISSN: 02839318. DOI: 10.1111/j.1471-6712.2011.00878.x. [Online]. Available: http://doi.wiley.com/10.1111/j.1471-6712.2011.00878.x (visited on 09/03/2020).

[52] S. Wong and C. Jacova, "Older Adults' Attitudes towards Cognitive Testing: Moving towards Person-Centeredness", en, *Dementia and Geriatric Cognitive Disorders Extra*, vol. 8, no. 3, pp. 348–359, Oct. 2018, ISSN: 1664-5464. DOI: 10.1159/000493464. [Online]. Available: https://www.karger.com/Article/FullText/493464 (visited on 09/03/2020).

[53] J. M. Lai, K. A. Hawkins, C. P. Gross, and J. H. Karlawish, "Self-Reported Distress After Cognitive Testing in Patients With Alzheimer's Disease", en, *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 63, no. 8, pp. 855–859, Aug. 2008, ISSN: 1079-5006, 1758-535X. DOI: 10.1093/gerona/63.8.855. [Online]. Available: https://academic.oup.com/biomedgerontology/article-lookup/doi/10.1093/gerona/63.8.855 (visited on 09/03/2020).

[54] P. Alhola and P. Polo-Kantola, "Sleep deprivation: Impact on cognitive performance", *Neuropsychiatric Disease and Treatment*, vol. 3, no. 5, pp. 553–567, Oct. 2007, ISSN: 1176-6328. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2656292/.

[55] L. M. Babrak *et al.*, "Traditional and Digital Biomarkers: Two Worlds Apart?", en, *Digital Biomarkers*, vol. 3, no. 2, pp. 92–102, Aug. 2019, ISSN: 2504-110X. DOI: 10.1159/000502000. [Online]. Available: https://www.karger.com/Article/FullText/502000 (visited on 12/22/2020).

[56] M. J. Summers and N. L. J. Saunders, "Neuropsychological measures predict decline to Alzheimer's dementia from mild cognitive impairment.", en, *Neuropsychology*, vol. 26, no. 4, pp. 498–508, Jul. 2012, ISSN: 1931-1559, 0894-4105. DOI: 10.1037/a0028576. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0028576 (visited on 07/22/2020).

[57] B. Dubois *et al.*, "Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria", en, *The Lancet Neurology*, vol. 6, no. 8, pp. 734–746, Aug. 2007, ISSN: 14744422. DOI: 10.1016/S1474-4422(07)70178-3. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1474442207701783 (visited on 07/22/2020).

[58] M. S. Albert *et al.*, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines

for Alzheimer's disease", en, *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 270–279, May 2011, ISSN: 15525260. DOI: 10.1016/j.jalz.2011.03.008. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S155252601100104X (visited on 08/13/2019).

[59] S. Weintraub *et al.*, "Measuring cognition and function in the preclinical stage of Alzheimer's disease", en, *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 4, no. 1, pp. 64–75, Jan. 2018, ISSN: 2352-8737, 2352-8737. DOI: 10.1016/j.trci.2018.01.003. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1016/j.trci.2018.01.003 (visited on 07/22/2020).

[60] C. Rodarte, "Pharmaceutical Perspective: How Digital Biomarkers and Contextual Data Will Enable Therapeutic Environments", en, *Digital Biomarkers*, Aug. 2017, ISSN: 2504-110X. DOI: 10.1159/000479951. [Online]. Available: https://www.karger.com/Article/FullText/479951 (visited on 04/08/2021).

[61] E. R. Dorsey, S. Papapetropoulos, M. Xiong, and K. Kieburtz, "The First Frontier: Digital Biomarkers for Neurodegenerative Disorders", en, *Digital Biomarkers*, Jul. 2017, ISSN: 2504-110X. DOI: 10.1159/000477383. [Online]. Available: https://www.karger.com/Article/FullText/477383 (visited on 02/02/2019).

[62] B. Bent *et al.*, "The digital biomarker discovery pipeline: An open-source software platform for the development of digital biomarkers using mHealth and wearables data", en, *Journal of Clinical and Translational Science*, vol. 5, no. 1, e19, 2021, ISSN: 2059-8661. DOI: 10.1017/cts.2020.511. [Online]. Available: https://www.cambridge.org/core/product/identifier/S2059866120005117/type/journal_article (visited on 04/08/2021).

[63] N. Mahadevan, C. Demanuele, H. Zhang, D. Volfson, B. Ho, M. K. Erb, and S. Patel, "Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device", en, *npj Digital Medicine*, vol. 3, no. 1, p. 5, Dec. 2020, ISSN: 2398-6352. DOI: 10.1038/s41746-019-0217-7. [Online]. Available: http://www.nature.com/articles/s41746-019-0217-7 (visited on 04/08/2021).

[64] D. Arefan, A. A. Mohamed, W. A. Berg, M. L. Zuley, J. H. Sumkin, and S. Wu, "Deep learning modeling using normal mammograms for predicting breast cancer risk", en, *Medical Physics*, vol. 47, no. 1, pp. 110–118, Jan. 2020, ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.13886. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13886 (visited on 04/08/2021).

[65] M. Mc Carthy and P. Schueler, "CAN DIGITAL TECHNOLOGY ADVANCE THE DEVELOPMENT OF TREATMENTS FOR ALZHEIMER'S DISEASE?", *The Journal of Prevention of Alzheimer's Disease*, pp. 1–4, 2019, ISSN: 22745807. DOI: 10.14283/jpad.2019.32. [Online]. Available: https://link.springer.com/article/10.14283/jpad.2019.32 (visited on 04/08/2021).

[66] L. C. Kourtis, O. B. Regele, J. M. Wright, and G. B. Jones, "Digital biomarkers for Alzheimer's disease: The mobile/wearable devices opportunity", en, *npj Digital Medicine*, vol. 2, no. 1, p. 9, Dec. 2019, ISSN: 2398-6352. DOI: 10.1038/s41746-019-0084-2. [Online]. Available: http://www.nature.com/articles/s41746-019-0084-2 (visited on 04/08/2021).

[67] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network", en, *Nature Medicine*, vol. 25, no. 1, pp. 65–69, Jan. 2019, ISSN: 1078-8956, 1546-170X. DOI: 10.1038/s41591-018-0268-3. [Online]. Available: http://www.nature.com/articles/s41591-018-0268-3 (visited on 04/08/2021).

[68] S. S. Ghosh, G. Ciccarelli, T. F. Quatieri, and A. Klein, "Speaking one's mind: Vocal biomarkers of depression and Parkinson disease", en, *The Journal of the Acoustical Society of America*, vol. 139, no. 4, pp. 2193–2193, Apr. 2016, ISSN: 0001-4966. DOI: 10.1121/1.4950530. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.4950530 (visited on 04/08/2021).

[69] K. Leduc-McNiven, R. T. Dion, M. Aljumaili, R. D. McLeod, M. R. Friesen, and K. Gutenschwager, "Serious Games and Machine Learning for Detecting Mild Cognitive Impairment", English, in *Proceedings on the International Conference on Artificial Intelligence*, Athens, 2019. [Online]. Available: https://search.proquest.com/docview/2362908930?accountid=17215.

[70] J. Kaye, "Home-based technologies: A new paradigm for conducting dementia prevention trials", en, *Alzheimer's & Dementia*, vol. 4, no. 1S1, S60–S66, Jan. 2008, ISSN: 15525260. DOI: 10.1016/j.jalz.2007.10.003. [Online]. Available: http://doi.wiley.com/10.1016/j.jalz.2007.10.003 (visited on 12/22/2020).

[71] J. Kaye *et al.*, "Unobtrusive measurement of daily computer use to detect mild cognitive impairment", en, *Alzheimer's & Dementia*, vol. 10, no. 1, pp. 10–17, Jan. 2014, ISSN: 15525260. DOI: 10.1016/j.jalz.2013.01.011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1552526013000435 (visited on 12/22/2020).

[72] H. H. Dodge, N. C. Mattek, D. Austin, T. L. Hayes, and J. A. Kaye, "In-home walking speeds and variability trajectories associated with mild cognitive impairment", en, *Neurology*, vol. 78, no. 24, pp. 1946–1952, Jun. 2012, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.0b013e318259e1de. [Online]. Available: http://www.neurology.org/cgi/doi/10.1212/WNL.0b013e318259e1de (visited on 12/22/2020).

[73] J. Austin, K. Hollingshead, and J. Kaye, "Internet Searches and Their Relationship to Cognitive Function in Older Adults: Cross-Sectional Analysis", en, *Journal of Medical Internet Research*, vol. 19, no. 9, e307, Sep. 2017, ISSN: 1438-8871. DOI: 10.2196/jmir.7671. [Online]. Available: http://www.jmir.org/2017/9/e307/ (visited on 12/22/2020).

[74] T. Suzuki and S. Murase, "Influence of Outdoor Activity and Indoor Activity on Cognition Decline: Use of an Infrared Sensor to Measure Activity", en, *Telemedicine and e-Health*, vol. 16, no. 6, pp. 686–690, Jul. 2010, ISSN: 1530-5627, 1556-3669. DOI: 10.1089/tmj.2009.0175. [Online]. Available: https://www.liebertpub.com/doi/10.1089/tmj.2009.0175 (visited on 12/22/2020).

[75] A. Seelye, N. Mattek, N. Sharma, P. Witter, A. Brenner, K. Wild, H. Dodge, and J. Kaye, "Passive Assessment of Routine Driving with Unobtrusive Sensors: A New Approach for Identifying and Monitoring Functional Level in Normal Aging and Mild Cognitive Impairment", *Journal of Alzheimer's Disease*, vol. 59, no. 4, pp. 1427–1437, Aug. 2017, ISSN: 13872877, 18758908. DOI: 10.3233/JAD-170116. [Online]. Available: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JAD-170116 (visited on 12/22/2020).

[76] J. Y. Tung, R. V. Rose, E. Gammada, I. Lam, E. A. Roy, S. E. Black, and P. Poupart, "Measuring Life Space in Older Adults with Mild-to-Moderate Alzheimer's Disease Using Mobile Phone GPS", en, *Gerontology*, vol. 60, no. 2, pp. 154–162, 2014, ISSN: 1423-0003, 0304-324X. DOI: 10.1159/000355669. [Online]. Available: https://www.karger.com/Article/FullText/355669 (visited on 12/22/2020).

[77] M. Mancini, H. Schlueter, M. El-Gohary, N. Mattek, C. Duncan, J. Kaye, and F. B. Horak, "Continuous Monitoring of Turning Mobility and Its Association to Falls and Cognitive Function: A Pilot Study", en, *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 71, no. 8, pp. 1102–1108, Aug. 2016, ISSN: 1079-5006, 1758-535X. DOI: 10.1093/gerona/glw019. [Online]. Available: https://academic.oup.com/biomedgerontology/article-lookup/doi/10.1093/gerona/glw019 (visited on 12/22/2020).

[78] A. Seelye, S. Hagler, N. Mattek, D. B. Howieson, K. Wild, H. H. Dodge, and J. A. Kaye, "Computer mouse movement patterns: A potential marker of mild cognitive impairment", en, *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 4, pp. 472–480, Dec. 2015, ISSN: 2352-8729, 2352-8729. DOI: 10.1016/j.dadm.2015.09.006. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1016/j.dadm.2015.09.006 (visited on 12/22/2020).

[79] M. Allard *et al.*, "Mobile Technologies in the Early Detection of Cognitive Decline", en, *PLoS ONE*, vol. 9, no. 12, K. Herholz, Ed., e112197, Dec. 2014, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0112197. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0112197 (visited on 12/22/2020).

[80] M. Hyde *et al.*, "Sea hero quest: The world's first mobile game where anyone can help scientists fight dementia", 2016.

[81] B. Suits, T. Hurka, and F. Newfeld, *The grasshopper: games, life, and utopia*, eng, 3. ed. Peterborought, Ont: Broadview Press, 2014, OCLC: 891787747, ISBN: 978-1-55481-215-8.

[82] P. Toril, J. M. Reales, and S. Ballesteros, "Video game training enhances cognition of older adults: A meta-analytic study", eng, *Psychology and Aging*, vol. 29, no. 3, pp. 706–716, Sep. 2014, ISSN: 1939-1498. DOI: 10.1037/a0037507.

[83] S. Valladares-Rodríguez, R. Pérez-Rodríguez, L. Anido-Rifón, and M. Fernández-Iglesias, "Trends on the application of serious games to neuropsychological evaluation: A scoping review", en, *Journal of Biomedical Informatics*, vol. 64, pp. 296–319, Dec. 2016, ISSN: 15320464. DOI: 10.1016/j.jbi.2016.10.019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1532046416301563 (visited on 12/03/2019).

[84] J. Lumsden, E. A. Edwards, N. S. Lawrence, D. Coyle, and M. R. Munafò, "Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy", en, *JMIR Serious Games*, vol. 4, no. 2, e11, Jul. 2016, ISSN: 2291-9279. DOI: 10.2196/games.5888. [Online]. Available: http://games.jmir.org/2016/2/e11/ (visited on 02/03/2019).

[85] B. Nelson Kakulla, "Gaming Attitudes and Habits of Adults Ages 50-Plus", en, AARP Research, Tech. Rep., Dec. 2019. DOI: 10.26419/res.00328.001. [Online]. Available: https://www.aarp.org/research/topics/technology/info-2019/2020-gaming-trends-older-americans.html (visited on 09/02/2020).

[86] E. S. Association, "Essential Facts About the Video Game Industry", Tech. Rep., 2020. [Online]. Available: http://www.theesa.com/wp-content/uploads/2018/05/EF2018_FINAL.pdf (visited on 04/08/2019).

[87] A. Chesham, P. Wyss, R. M. Müri, U. P. Mosimann, and T. Nef, "What Older People Like to Play: Genre Preferences and Acceptance of Casual Games", en, *JMIR Serious Games*, vol. 5, no. 2, e8, Apr. 2017, ISSN: 2291-9279. DOI: 10.2196/games.7025. [Online]. Available: http://games.jmir.org/2017/2/e8/ (visited on 07/24/2020).

[88] B. De Schutter, "Never Too Old to Play: The Appeal of Digital Games to an Older Audience", en, *Games and Culture*, vol. 6, no. 2, pp. 155–170, Mar. 2011, ISSN: 1555-4120, 1555-4139. DOI: 10.1177/1555412010364978. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1555412010364978 (visited on 08/03/2017).

[89] F. Ferreira-Brito *et al.*, "Game-based interventions for neuropsychological assessment, training and rehabilitation: Which game-elements to use? A systematic review", en, *Journal of Biomedical Informatics*, vol. 98, p. 103 287, Oct. 2019, ISSN: 15320464. DOI: 10.1016/j.jbi.2019.103287. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1532046419302060 (visited on 06/15/2020).

[90] B. Shneiderman, "The future of interactive systems and the emergence of direct manipulation†", en, *Behaviour & Information Technology*, vol. 1, no. 3, pp. 237–256, Jul. 1982, ISSN: 0144-929X, 1362-3001. DOI: 10.1080/01449298208914450. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01449298208914450 (visited on 12/29/2020).

[91] W. IJsselsteijn, H. H. Nap, Y. A. de Kort, and K. Poels, "Digital game design for elderly users", in *Proceedings of the 2007 conference on Future Play*, Toronto, Canada: ACM, 2007, pp. 17–22, ISBN: 978-1-59593-943-2. DOI: 10.1145/1328202.1328206. [Online]. Available: http://portal.acm.org/citation.cfm?id=1328202.1328206&coll=&dl=ACM (visited on 03/01/2009).

[92] J. Wittland, P. Brauner, and M. Ziefle, "Serious Games for Cognitive Training in Ambient Assisted Living Environments – A Technology Acceptance Perspective", in *Human-Computer Interaction – INTERACT 2015*, J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, and M. Winckler, Eds., Cham: Springer International Publishing, 2015, pp. 453–471, ISBN: 978-3-319-22701-6.

[93] D. Michael, *Serious games: games that educate, train and inform*. Boston, Mass: Thomson Course Technology, 2006, OCLC: ocm62780159, ISBN: 978-1-59200-622-9.

[94] A. Mané and E. Donchin, "The Space Fortress game", *Acta Psychologica, 71*, vol. 71, no. 1, pp. 17–22, Aug. 1989. [Online]. Available: https://scholar-google-com.kuleuven.ezproxy.kuleuven.be/scholar_lookup?title=The%20Space%20Fortress%20game&publication_year=1989&author=A.M.%20Man%C3%A9&author=E.%20Donchin (visited on 04/09/2019).

[95] W. R. Boot, "Video games as tools to achieve insight into cognitive processes", *Frontiers in Psychology*, vol. 6, Jan. 2015, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2015.00003. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00003/abstract (visited on 09/03/2020).

[96] P. L. Baniqued *et al.*, "Selling points: What cognitive abilities are tapped by casual video games?", en, *Acta Psychologica*, vol. 142, no. 1, pp. 74–86, Jan. 2013, ISSN: 00016918. DOI: 10.1016/j.actpsy.2012.11.009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0001691812001916 (visited on 05/02/2019).

[97] S. Valladares-Rodriguez, R. Pérez-Rodríguez, L. Anido-Rifón, and M. J. Fernández-Iglesias, "Cognitive Assessment through Casual Video Games and Machine Learning - Doctoral Consortium Contributions", in *Doctoral Consortium - DCICT4AW, (ICT4AgeingWell 2015)*, SciTePress, 2015, pp. 3–11.

[98] M. Colder Carras, A. J. Van Rooij, D. Spruijt-Metz, J. Kvedar, M. D. Griffiths, Y. Carabas, and A. Labrique, "Commercial Video Games As Therapy: A New Research Agenda to Unlock the Potential of a Global Pastime", English, *Frontiers in Psychiatry*, vol. 8, 2018, ISSN: 1664-0640. DOI: 10.3389/fpsyt.2017.00300. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyt.2017.00300/full#B54 (visited on 04/09/2019).

[99] P. Wouters, C. van Nimwegen, H. van Oostendorp, and E. D. van der Spek, "A meta-analysis of the cognitive and motivational effects of serious games.", en, *Journal of Educational Psychology*, vol. 105, no. 2, pp. 249–265, 2013, ISSN: 1939-2176, 0022-0663. DOI: 10.1037/a0031311. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0031311 (visited on 08/03/2017).

[100] R. L. Mandryk and M. V. Birk, "The Potential of Game-Based Digital Biomarkers for Modeling Mental Health", en, *JMIR Mental Health*, vol. 6, no. 4, e13485, Apr. 2019, ISSN: 2368-7959. DOI: 10.2196/13485. [Online]. Available: http://mental.jmir.org/2019/4/e13485/ (visited on 04/29/2019).

[101] E. G. Tangalos and R. C. Petersen, "Mild Cognitive Impairment in Geriatrics", en, *Clinics in Geriatric Medicine*, vol. 34, no. 4, pp. 563–589, Nov. 2018, ISSN: 07490690. DOI: 10.1016/j.cger.2018.06.005. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0749069018309984 (visited on 11/04/2019).

[102] K. S. Tekinbaş and E. Zimmerman, *Rules of play: game design fundamentals*. Cambridge, Mass: MIT Press, 2003, ISBN: 978-0-262-24045-1.

[103] K. Salen and E. Zimmerman, "Game design and meaningful play", *Handbook of Computer Game Studies*, pp. 59–79, 2005.

[104] B. De Schutter and V. Vanden Abeele, "Meaningful Play in Elderly Life", eng, Montreal, Quebec, Canada, 2008. [Online]. Available: $$Uhttps://lirias.kuleuven.be/retrieve/114734$$DMain%20Article%20[freely%20available].

[105] ——, "Designing meaningful play within the psycho-social context of older adults", en, ACM Press, 2010, pp. 84–93, ISBN: 978-1-60558-907-7. DOI: 10.1145/1823818.1823827. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1823818.1823827 (visited on 08/14/2016).

[106] B. De Schutter and S. Malliet, "The older player of digital games: A classification based on perceived need satisfaction", *Communications*, vol. 39, no. 1, Jan. 2014, ISSN: 1613-4087, 0341-2059. DOI: 10.1515/commun-2014-0005. [Online]. Available: https://www.degruyter.com/view/j/comm.2014.39.issue-1/commun-2014-0005/commun-2014-0005.xml (visited on 08/03/2017).

[107] B. De Schutter and V. V. Abeele, "Towards a gerontoludic manifesto", *Anthropology & Aging*, vol. 36, no. 2, pp. 112–120, 2015.

[108] B. De Schutter and J. A. Brown, "Digital Games as a Source of Enjoyment in Later Life", en, *Games and Culture*, vol. 11, no. 1-2, pp. 28–52, Jan. 2016, ISSN: 1555-4120, 1555-4139. DOI: 10.1177/1555412015594273. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1555412015594273 (visited on 04/07/2021).

[109] V. A. Vanden Abeele and V. Van Rompaey, "Introducing human-centered research to game design: Designing game concepts for and with senior citizens", in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '06, Montréal, Québec, Canada: Association for Computing Machinery, 2006, pp. 1469–1474, ISBN: 1595932984. DOI: 10.1145/1125451.1125721. [Online]. Available: https://doi.org/10.1145/1125451.1125721.

[110] E. Katz, J. G. Blumler, and M. Gurevitch, "Uses and Gratifications Research", *Public Opinion Quarterly*, vol. 37, no. 4, p. 509, 1973, ISSN: 0033362X. DOI: 10.1086/268109. [Online]. Available: https://academic.oup.com/poq/article-lookup/doi/10.1086/268109 (visited on 04/07/2021).

[111] A. H. Morehead and G. Mott-Smith, *The Complete Book of Solitaire and Patience Games: The Most Comprehensive Book of Its Kind: Over 225 Games*, English. New York: Bantam, Aug. 1983, ISBN: 978-0-553-26240-7.

[112] D. S. Parlett, *A history of card games*. Oxford ; New York: Oxford University Press, 1991, ISBN: 978-0-19-282905-4.

[113] J. C. Allaire, A. C. McLaughlin, A. Trujillo, L. A. Whitlock, L. LaPorte, and M. Gandy, "Successful aging through digital games: Socioemotional differences between older adult gamers and Non-gamers", en, *Computers in Human Behavior*, vol. 29, no. 4, pp. 1302–1306, Jul. 2013, ISSN: 07475632. DOI: 10.1016/j.chb.2013.01.014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0747563213000174 (visited on 07/26/2019).

[114] W. R. Boot *et al.*, "Exploring Older Adults' Video Game Use in the PRISM Computer System", *Innovation in Aging*, vol. 2, no. 1, 2018, _eprint: https://academic.oup.com/innovateage/article-pdf/2/1/igy009/33326049/igy009.pdf, ISSN: 2399-5300. DOI: 10.1093/geroni/igy009. [Online]. Available: https://doi.org/10.1093/geroni/igy009.

[115] R. Bjarnason, A. Fern, and P. Tadepalli, "Lower Bounding Klondike Solitaire with Monte-Carlo Planning", in *Proceedings of the Nineteenth International Conference on International Conference on Automated Planning and Scheduling*, ser. ICAPS'09, event-place: Thessaloniki, Greece, AAAI Press, 2009, pp. 26–33, ISBN: 978-1-57735-406-2.

[116] V. Dhar, "Data science and prediction", en, *Communications of the ACM*, vol. 56, no. 12, pp. 64–73, Dec. 2013, ISSN: 0001-0782, 1557-7317. DOI: 10.1145/2500499. [Online]. Available: https://dl.acm.org/doi/10.1145/2500499 (visited on 10/21/2020).

[117] S. Angée, S. I. Lozano-Argel, E. N. Montoya-Munera, J.-D. Ospina-Arango, and M. S. Tabares-Betancur, "Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data & Analytics Projects", en, in *Knowledge Management in Organizations*, L. Uden, B. Hadzima, and I.-H. Ting, Eds., ser. Communications in Computer and Information Science, Cham: Springer International Publishing, 2018, pp. 613–624, ISBN: 978-3-319-95204-8. DOI: 10.1007/978-3-319-95204-8_51.

[118] S. Sabrina Jahan, "Human Resources Information System (HRIS): A Theoretical Perspective", *Journal of Human Resource and Sustainability Studies*, vol. 02, no. 02, pp. 33–39, 2014, ISSN: 2328-4862, 2328-4870. DOI: 10.4236/jhrss.2014.22004. [Online]. Available: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/jhrss.2014.22004 (visited on 10/21/2020).

[119] R. Wirth, "Crisp-dm: Towards a standard process model for data mining", in *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 2000, pp. 29–39.

[120] marktab, *What is the Team Data Science Process?*, en-us. [Online]. Available: https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview (visited on 10/21/2020).

[121] K. Gielis, J. Kennes, C. D. Dobbeleer, S. Puttemans, and V. V. Abeele, "Collecting Digital Biomarkers on Cognitive Health Through Computer Vision and Gameplay: An Image Processing Toolkit for Card Games", in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, Xi'an, China: IEEE, Jun. 2019, pp. 1–12, ISBN: 978-1-5386-9138-0. DOI: 10.1109/ICHI.2019.8904511.

[122] K. Gielis, K. Verbert, J. Tournoy, and V. Vanden Abeele, "Age? It's in the Game: An Exploratory Study on Detection of Cognitive Aging through Card Games", en, in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, Barcelona Spain: ACM, Oct. 2019, pp. 651–664, ISBN: 978-1-4503-6688-5. DOI: 10.1145/3311350.3347193.

[123] K. Gielis *et al.*, "Dissecting Digital Card Games to Yield Digital Biomarkers for the Assessment of Mild Cognitive Impairment: A Methodological Approach and Exploratory Study", *Journal of Medical Internet Research*, Jan. 2021.

[124] K. Gielis, M.-E. Vanden Abeele, K. Verbert, J. Tournoy, M. De Vos, and V. Vanden Abeele, "Detecting Mild Cognitive Impairment through Digital Biomarkers of Cognitive Performance found in Klondike Solitaire: A Machine Learning Study", *Digital Biomarkers*, Jan. 2021, ISSN: 2504-110X. DOI: 10.1159/000514105.

[125] K. Gielis, F. Brito, J. Tournoy, and V. Vanden Abeele, "Can Card Games Be Used to Assess Mild Cognitive Impairment?: A Study of Klondike Solitaire and Cognitive Functions", en, in *CHI PLAY '17 Extended Abstracts*, Amsterdam, The Netherlands: ACM Press, 2017, pp. 269–276, ISBN: 978-1-4503-5111-9. DOI: 10.1145/3130859.3131328.

[126] W. H. Organisation, *WHO | Dementia: A public health priority*, 2019. [Online]. Available: http://www.who.int/mental_health/neurology/dementia/en/ (visited on 01/31/2019).

[127] W. H. Organization, "Depression and Other Common Mental Disorders: Global Health Estimates", World Health Organization, Geneva, Tech. Rep., 2017.

[128] W. H. Organisation, "Comprehensive mental health action plan 2013-2020", World Health Organisation, Tech. Rep. sixty-sixth world health assembly, May 2013, p. 27. [Online]. Available: http://apps.who.int/gb/ebwha/pdf_files/WHA66/A66_R8-en.pdf.

[129] D. Chisholm, K. Sweeny, P. Sheehan, B. Rasmussen, F. Smit, P. Cuijpers, and S. Saxena, "Scaling-up treatment of depression and anxiety: A global return on investment analysis", en, *The Lancet Psychiatry*, vol. 3, no. 5, pp. 415–424, May 2016, ISSN: 22150366. DOI: 10.1016/S2215-0366(16)30024-4. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2215036616300244 (visited on 01/31/2019).

[130] J. W. Feightner and G. Worrall, "Early detection of depression by primary care physicians", eng, *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, vol. 142, no. 11, pp. 1215–1220, Jun. 1990, ISSN: 0820-3946.

[131] C. Laske *et al.*, "Innovative diagnostic tools for early detection of Alzheimer's disease", en, *Alzheimer's & Dementia*, vol. 11, no. 5, pp. 561–578, May 2015, ISSN: 15525260. DOI: 10.1016/j.jalz.2014.06.004. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1552526014024637 (visited on 07/31/2017).

[132] G. Mushtaq, N. H. Greig, J. A. Khan, and M. A. Kamal, "Status of acetylcholinesterase and butyrylcholinesterase in Alzheimer's disease and type 2 diabetes mellitus", eng, *CNS & neurological disorders drug targets*, vol. 13, no. 8, pp. 1432–1439, 2014, ISSN: 1996-3181.

[133] R. C. Petersen *et al.*, "Current concepts in mild cognitive impairment", eng, *Archives of Neurology*, vol. 58, no. 12, pp. 1985–1992, Dec. 2001, ISSN: 0003-9942.

[134] G. M. Savva and A. Arthur, "Who has undiagnosed dementia? A cross-sectional analysis of participants of the Aging, Demographics and Memory Study", en, *Age and Ageing*, vol. 44, no. 4, pp. 642–647, Jul. 2015, ISSN: 0002-0729. DOI: 10.1093/ageing/afv020. [Online]. Available: https://academic.oup.com/ageing/article/44/4/642/66339 (visited on 02/01/2019).

[135] B. D. W. Group, "Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework", en, *Clinical Pharmacology & Therapeutics*, vol. 69, no. 3, pp. 89–95, 2001, ISSN: 1532-6535. DOI: 10.1067/mcp.2001.113989. [Online]. Available: https://ascpt.onlinelibrary.wiley.com/doi/abs/10.1067/mcp.2001.113989 (visited on 02/01/2019).

[136] N. Sharma and A. N. Singh, "Exploring Biomarkers for Alzheimer's Disease", *Journal of Clinical and Diagnostic Research : JCDR*, vol. 10, no. 7, KE01–KE06, Jul. 2016, ISSN: 2249-782X. DOI: 10.7860/JCDR/2016/18828.8166. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5020308/ (visited on 02/18/2019).

[137] R. M. Ahmed, R. W. Paterson, J. D. Warren, H. Zetterberg, J. T. O'Brien, N. C. Fox, G. M. Halliday, and J. M. Schott, "Biomarkers in dementia: Clinical utility and new directions", en, *J Neurol Neurosurg Psychiatry*, vol. 85, no. 12, pp. 1426–1434, Dec. 2014, ISSN: 0022-3050, 1468-330X. DOI: 10.1136/jnnp-2014-307662. [Online]. Available: https://jnnp.bmj.com/content/85/12/1426 (visited on 02/22/2019).

[138] K. Blennow, H. Hampel, M. Weiner, and H. Zetterberg, "Cerebrospinal fluid and plasma biomarkers in Alzheimer disease", en, *Nature Reviews Neurology*, vol. 6, no. 3, pp. 131–144, Mar. 2010, ISSN: 1759-4766. DOI: 10.1038/nrneurol.2010.4. [Online]. Available: https://www.nature.com/articles/nrneurol.2010.4 (visited on 02/22/2019).

[139] J. D. Doecke *et al.*, "Blood-Based Protein Biomarkers for Diagnosis of Alzheimer Disease", en, *Archives of Neurology*, vol. 69, no. 10, pp. 1318–1325, Oct. 2012, ISSN: 0003-9942. DOI: 10.1001/archneurol.2012.1282. [Online]. Available: https://jamanetwork.com/journals/jamaneurology/fullarticle/1217314 (visited on 02/22/2019).

[140] C. Reitz and R. Mayeux, "Alzheimer disease: Epidemiology, diagnostic criteria, risk factors and biomarkers", *Biochemical Pharmacology*, Alzheimer's Disease - Amyloid, Tau and Beyond, vol. 88, no. 4, pp. 640–651, Apr. 2014, ISSN: 0006-2952. DOI: 10.1016/j.bcp.2013.12.024. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0006295213008083 (visited on 02/22/2019).

[141] T. Smith, N. Gildeh, and C. Holmes, "The Montreal Cognitive Assessment: Validity and Utility in a Memory Clinic Setting", en, *The Canadian Journal of Psychiatry*, vol. 52, no. 5, pp. 329–332, May 2007, ISSN: 0706-7437. DOI: 10.1177/070674370705200508. [Online]. Available: https://doi.org/10.1177/070674370705200508 (visited on 02/22/2019).

[142] T. Tapiola, I. Alafuzoff, S.-K. Herukka, L. Parkkinen, P. Hartikainen, H. Soininen, and T. Pirttilä, "Cerebrospinal Fluid beta-Amyloid 42 and Tau Proteins as Biomarkers of Alzheimer-Type Pathologic Changes in the Brain", en, *Archives of Neurology*, vol. 66, no. 3, pp. 382–389, Mar. 2009, ISSN: 0003-9942. DOI: 10.1001/archneurol.2008.596. [Online]. Available: https://jamanetwork.com/journals/jamaneurology/fullarticle/796677 (visited on 02/22/2019).

[143] K. Journal, "Digital Biomarkers | Karger Journal", en, *Digital Biomarkers*, 2018. [Online]. Available: https://www.karger.com/Journal/Home/271954 (visited on 02/15/2019).

[144] S. A. Cooley, J. M. Heaps, J. D. Bolzenius, L. E. Salminen, L. M. Baker, S. E. Scott, and R. H. Paul, "Longitudinal Change in Performance on the Montreal Cognitive Assessment in Older Adults", en, *The Clinical Neuropsychologist*, vol. 29, no. 6, pp. 824–835, Aug. 2015, ISSN: 1385-4046, 1744-4144. DOI: 10.1080/13854046.2015.1087596. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/13854046.2015.1087596 (visited on 06/28/2017).

[145] H. Jimison, M. Pavel, J. McKanna, and J. Pavel, "Unobtrusive monitoring of computer interactions to detect cognitive status in elders", eng, *IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 8, no. 3, pp. 248–252, Sep. 2004, ISSN: 1089-7771.

[146] J. L. Palmer, M. A. Coats, C. M. Roe, S. M. Hanko, C. Xiong, and J. C. Morris, "Unified Parkinson's Disease Rating Scale-Motor Exam: Inter-rater reliability of advanced practice nurse and neurologist assessments: Neurological assessment inter-rater reliability", en, *Journal of Advanced Nursing*, vol. 66, no. 6, pp. 1382–1387, Apr. 2010, ISSN: 03092402, 13652648. DOI: 10.1111/j.1365-2648.2010.05313.x. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-2648.2010.05313.x (visited on 02/02/2019).

[147] S. K. Inouye, C. H. van Dyck, C. A. Alessi, S. Balkin, A. P. Siegal, and R. I. Horwitz, "Clarifying confusion: The confusion assessment method. A new method for detection of delirium", eng, *Annals of Internal Medicine*, vol. 113, no. 12, pp. 941–948, Dec. 1990, ISSN: 0003-4819.

[148] Roche, *The measure of you*, en, 2019. [Online]. Available: https://www.roche.com/about/priorities/personalised_healthcare/digital-biomarkers.htm (visited on 02/14/2019).

[149] J. Torous, J.-P. Onnela, and M. Keshavan, "New dimensions and new tools to realize the potential of RDoC: Digital phenotyping via smartphones and connected devices", eng, *Translational Psychiatry*, vol. 7, no. 3, e1053, 2017, ISSN: 2158-3188. DOI: 10.1038/tp.2017.25.

[150] A. Shehab and A. Abdulle, "Cognitive and autonomic dysfunction measures in normal controls, white coat and borderline hypertension", *BMC Cardiovascular Disorders*, vol. 11, no. 1, p. 3, Jan. 2011, ISSN: 1471-2261. DOI: 10.1186/1471-2261-11-3. [Online]. Available: https://doi.org/10.1186/1471-2261-11-3 (visited on 02/19/2019).

[151] M. M. Redfield *et al.*, "Isosorbide Mononitrate in Heart Failure with Preserved Ejection Fraction", en, *New England Journal of Medicine*, vol. 373, no. 24, pp. 2314–2324, Dec. 2015, ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMoa1510774. [Online]. Available: http://www.nejm.org/doi/10.1056/NEJMoa1510774 (visited on 02/02/2019).

[152] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, and D. C. Mohr, "Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study", en, *Journal of Medical Internet Research*, vol. 17, no. 7, e175, Jul. 2015, ISSN: 1438-8871. DOI: 10.2196/jmir.4273. [Online]. Available: http://www.jmir.org/2015/7/e175/ (visited on 02/02/2019).

[153] M. Faurholt-Jepsen, M. Vinberg, M. Frost, E. M. Christensen, J. E. Bardram, and L. V. Kessing, "Smartphone data as an electronic biomarker of illness activity in bipolar disorder", en, *Bipolar Disorders*, vol. 17, no. 7, pp. 715–728, Nov. 2015, ISSN: 13985647. DOI: 10.1111/bdi.12332. [Online]. Available: http://doi.wiley.com/10.1111/bdi.12332 (visited on 02/02/2019).

[154] T. Beiwinkel, S. Kindermann, A. Maier, C. Kerl, J. Moock, G. Barbian, and W. Rössler, "Using Smartphones to Monitor Bipolar Disorder Symptoms: A Pilot Study", en, *JMIR Mental Health*, vol. 3, no. 1, e2, Jan. 2016, ISSN: 2368-7959. DOI: 10.2196/mental.4560. [Online]. Available: http://mental.jmir.org/2016/1/e2/ (visited on 02/02/2019).

[155] I. Barnett, J. Torous, P. Staples, L. Sandoval, M. Keshavan, and J.-P. Onnela, "151. Relapse Prediction in Schizophrenia through Digital Phenotyping", en, *Biological Psychiatry*, vol. 83, no. 9, S61–S62, May 2018, ISSN: 00063223. DOI: 10.1016/j.biopsych.2018.02.169. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0006322318302713 (visited on 02/03/2019).

[156] S. Hagler, D. Austin, T. Hayes, J. Kaye, and M. Pavel, "Unobtrusive and Ubiquitous In-Home Monitoring: A Methodology for Continuous Assessment of Gait Velocity in Elders", *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 813–820, Apr. 2010, ISSN: 0018-9294. DOI: 10.1109/TBME.2009.2036732. [Online]. Available: http://ieeexplore.ieee.org/document/5339193/ (visited on 02/03/2019).

[157] J. McGonigal, *Reality is broken: why games make us better and how they can change the world*. New York: Penguin Press, 2011, OCLC: ocn646111856, ISBN: 978-1-59420-285-8.

[158] M. Csikszentmihalyi, *Finding flow: The psychology of engagement with everyday life*, ser. Finding flow: The psychology of engagement with everyday life. New York, NY, US: Basic Books, 1997, ISBN: 978-0-465-04513-6.

[159] J. DeRight and R. S. Jorgensen, "I Just Want My Research Credit: Frequency of Suboptimal Effort in a Non-Clinical Healthy Undergraduate Sample", en, *The Clinical Neuropsychologist*, vol. 29, no. 1, pp. 101–117, Jan. 2015, ISSN: 1385-4046, 1744-4144. DOI: 10.1080/13854046.2014.989267. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/13854046.2014.989267 (visited on 02/04/2019).

[160] S. Valladares-Rodriguez, M. J. Fernández-Iglesias, L. Anido-Rifón, D. Facal, and R. Pérez-Rodríguez, "Episodix: A serious game to detect cognitive impairment in senior adults. A psychometric study", en, *PeerJ*, vol. 6, e5478, Sep. 2018, ISSN: 2167-8359. DOI: 10.7717/peerj.5478. [Online]. Available: https://peerj.com/articles/5478 (visited on 01/14/2019).

[161] P. Słowiński *et al.*, "Unravelling socio-motor biomarkers in schizophrenia", en, *npj Schizophrenia*, vol. 3, no. 1, Dec. 2017, ISSN: 2334-265X. DOI: 10.1038/s41537-016-0009-x. [Online]. Available: http://www.nature.com/articles/s41537-016-0009-x (visited on 02/03/2019).

[162] H.-T. Jung, H. Lee, K. Kim, B. Kim, S. Park, T. Ryu, Y. Kim, and S. I. Lee, "Estimating Mini Mental State Examination Scores using Game-Specific Performance Values: A Preliminary Study", in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI: IEEE, Jul. 2018, pp. 1518–1521, ISBN: 978-1-5386-3646-6. DOI: 10.1109/EMBC.2018.8512516. [Online]. Available: https://ieeexplore.ieee.org/document/8512516/ (visited on 01/07/2019).

[163] K. Leduc-McNiven, B. White, H. Zheng, R. D McLeod, and M. R Friesen, "Serious games to assess mild cognitive impairment: 'The game is the assessment'", *Research and Review Insights*, vol. 2, no. 1, 2018, ISSN: 25152637. DOI: 10.15761/RRI.1000128. [Online]. Available: http://www.oatext.com/serious-games-to-assess-mild-cognitive-impairment-the-game-is-the-assessment.php (visited on 09/07/2018).

[164] C. Boletsis and S. McCallum, "Smartkuber: A Serious Game for Cognitive Health Screening of Elderly Players", en, *Games for Health Journal*, vol. 5, no. 4, pp. 241–251, Aug. 2016, ISSN: 2161-783X, 2161-7856. DOI: 10.1089/g4h.2015.0107. [Online]. Available: http://online.liebertpub.com/doi/10.1089/g4h.2015.0107 (visited on 08/03/2017).

[165] P. Felicia and R. Tennyson, Eds., *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches*, en, ser. Advances in Game-Based Learning. IGI Global, 2011. DOI: 10.4018/978-1-60960-495-0. [Online]. Available: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-60960-495-0 (visited on 02/19/2019).

[166] P. L. Ackerman, R. Kanfer, and C. Calderwood, "Use it or lose it? Wii brain exercise practice and reading for domain knowledge.", en, *Psychology and Aging*, vol. 25, no. 4, pp. 753–766, 2010, ISSN: 1939-1498, 0882-7974. DOI: 10.1037/a0019277. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0019277 (visited on 08/04/2017).

[167] A. C. S. Torres, "Cognitive effects of video games on old people", *International Journal on Disability and Human Development*, vol. 10, no. 1, Jan. 2011, ISSN: 2191-0367, 2191-1231. DOI: 10.1515/ijdhd.2011.003. [Online]. Available: https://www.degruyter.com/view/j/ijdhd.2011.10.issue-1/ijdhd.2011.003/ijdhd.2011.003.xml (visited on 09/10/2018).

[168] W. R. Boot, D. Souders, N. Charness, K. Blocker, N. Roque, and T. Vitale, "The Gamification of Cognitive Training: Older Adults' Perceptions of and Attitudes Toward Digital Game-Based Interventions", in *Human Aspects of IT for the Aged Population. Design for Aging*, J. Zhou and G. Salvendy, Eds., vol. 9754, Cham: Springer International Publishing, 2016, pp. 290–300. DOI: 10.1007/978-3-319-39943-0_28. [Online]. Available: http://link.springer.com/10.1007/978-3-319-39943-0_28 (visited on 01/07/2019).

[169] O. Thompson, S. Barrett, C. Patterson, and D. Craig, "Examining the Neurocognitive Validity of Commercially Available, Smartphone-Based Puzzle Games", *Psychology*, vol. 03, no. 07, pp. 525–526, 2012, ISSN: 2152-7180, 2152-7199. DOI: 10.4236/psych.2012.37076. [Online]. Available: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/psych.2012.37076 (visited on 02/03/2019).

[170] J. W. Grabbe, "Sudoku and Working Memory Performance for Older Adults", en, *Activities, Adaptation & Aging*, vol. 35, no. 3, pp. 241–254, Jul. 2011, ISSN: 0192-4788, 1544-4368. DOI: 10.1080/01924788.2011.

596748. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01924788.2011.596748 (visited on 01/04/2019).

[171] G. Bradski, "The OpenCV Library", *Dr. Dobb's Journal of Software Tools*, 2000.

[172] A. Elyasaf, A. Hauptman, and M. Sipper, "Evolutionary Design of FreeCell Solvers", *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 4, pp. 270–281, Dec. 2012, ISSN: 1943-068X, 1943-0698. DOI: 10.1109/TCIAIG.2012.2210423. [Online]. Available: http://ieeexplore.ieee.org/document/6249736/ (visited on 02/04/2019).

[173] P. L. Baker, "Using FreeCell to teach mathematics", English, *The Mathematics Teacher; Reston*, vol. 96, no. 6, pp. 406–410, Sep. 2003, ISSN: 00255769. [Online]. Available: https://search.proquest.com/docview/204636284/abstract/51B7A28418804E39PQ/1 (visited on 02/11/2019).

[174] M. W. H. Chen, "The Hardness Analysis of FreeCell Games", en, Bacherlor's Thesis, The Hardness Analysis of FreeCell Games, Amsterdam, 2015.

[175] A. Elyasaf, A. Hauptman, and M. Sipper, "GA-FreeCell: Evolving solvers for the game of FreeCell", en, in *Proceedings of the 13th annual conference on Genetic and evolutionary computation - GECCO '11*, Dublin, Ireland: ACM Press, 2011, p. 1931, ISBN: 978-1-4503-0557-0. DOI: 10.1145/2001576.2001836. [Online]. Available: http://portal.acm.org/citation.cfm?doid=2001576.2001836 (visited on 02/11/2019).

[176] R. Jilani, A. Crampton, D. Kitchin, and M. Vallati, "Have a little patience: Let planners play cards", English, in *Proceedings of the 34th Workshop of the UK Planning and Scheduling Special Interest Group*, L. Chrpa, S. Parkinson, and M. Vallati, Eds., ser. UK Planning and Scheduling Special Interest Group 2016. Vol. 1782, 2017.

[177] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system", En, *Nature*, vol. 381, no. 6582, p. 520, Jun. 1996, ISSN: 1476-4687. DOI: 10.1038/381520a0. [Online]. Available: https://www.nature.com/articles/381520a0 (visited on 02/22/2019).

[178] N. A. Bishop, T. Lu, and B. A. Yankner, "Neural mechanisms of ageing and cognitive decline", en, *Nature*, vol. 464, pp. 529–535, Mar. 2010, ISSN: 1476-4687. DOI: 10.1038/nature08983. [Online]. Available: https://www.nature.com/articles/nature08983 (visited on 02/21/2019).

[179] T. Hedden and J. D. E. Gabrieli, "Insights into the ageing mind: A view from cognitive neuroscience", en, *Nature Reviews Neuroscience*, vol. 5, no. 2, pp. 87–96, Feb. 2004, ISSN: 1471-0048. DOI: 10.1038/nrn1323. [Online]. Available: https://www.nature.com/articles/nrn1323 (visited on 02/21/2019).

[180] J. H. Morrison and M. G. Baxter, "The ageing cortical synapse: Hallmarks and implications for cognitive decline", en, *Nature Reviews Neuroscience*, vol. 13, no. 4, pp. 240–250, Apr. 2012, ISSN: 1471-0048. DOI: 10.1038/nrn3200. [Online]. Available: https://www.nature.com/articles/nrn3200 (visited on 02/21/2019).

[181] R. D. Seidler, J. A. Bernard, T. B. Burutolu, B. W. Fling, M. T. Gordon, J. T. Gwin, Y. Kwak, and D. B. Lipps, "Motor Control and Aging: Links to Age-Related Brain Structural, Functional, and Biochemical Effects", *Neuroscience and biobehavioral reviews*, vol. 34, no. 5, pp. 721–733, Apr. 2010, ISSN: 0149-7634. DOI: 10.1016/j.neubiorev.2009.10.005. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2838968/ (visited on 02/21/2019).

[182] M. S. Collection, *Freecell*. [Online]. Available: https://zone.msn.com/en/mssolitairecollection/default.htm (visited on 04/09/2019).

[183] X. Xu, S. Xu, L. Jin, and E. Song, "Characteristic analysis of Otsu threshold and its applications", *Pattern Recognition Letters*, vol. 32, no. 7, pp. 956–961, May 2011, ISSN: 0167-8655. DOI: 10.1016/j.patrec.2011.01.021. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865511000365 (visited on 02/20/2019).

[184] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A Simple Deep Learning Baseline for Image Classification?", *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015, ISSN: 1057-7149, 1941-0042. DOI: 10.1109/TIP.2015.2475625. [Online]. Available: http://ieeexplore.ieee.org/document/7234886/ (visited on 05/02/2019).

[185] Y. Tang, J. Wang, B. Gao, E. Dellandrea, R. Gaizauskas, and L. Chen, "Large Scale Semi-Supervised Object Detection Using Visual and Semantic Knowledge Transfer", in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 2119–2128, ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.233. [Online]. Available: http://ieeexplore.ieee.org/document/7780602/ (visited on 05/02/2019).

[186] S. Zhou, Q. Chen, and X. Wang, "Discriminative Deep Belief Networks for image classification", in *2010 IEEE International Conference on Image Processing*, Hong Kong, Hong Kong: IEEE, Sep. 2010, pp. 1561–1564. DOI:

10.1109/ICIP.2010.5649922. [Online]. Available: http://ieeexplore.ieee.org/document/5649922/ (visited on 05/02/2019).

[187] C. N. Harada, M. C. Natelson Love, and K. L. Triebel, "Normal Cognitive Aging", en, *Clinics in Geriatric Medicine*, vol. 29, no. 4, pp. 737–752, Nov. 2013, ISSN: 07490690. DOI: 10.1016/j.cger.2013.07.002. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0749069013000591 (visited on 03/20/2019).

[188] W. H. Organisation, *Ageing and health*, en. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (visited on 03/14/2019).

[189] R. Vandenberghe, "Cognitive aging and Alzheimer's disease", en, *Postgraduate Medical Journal*, vol. 81, no. 956, pp. 343–352, Jun. 2005, ISSN: 0032-5473. DOI: 10.1136/pgmj.2004.028290. [Online]. Available: http://pmj.bmj.com/cgi/doi/10.1136/pgmj.2004.028290 (visited on 03/26/2019).

[190] T. Niikura, H. Tajima, and Y. Kita, "Neuronal cell death in Alzheimer's disease and a neuroprotective factor, humanin", eng, *Current Neuropharmacology*, vol. 4, no. 2, pp. 139–147, Apr. 2006, ISSN: 1570-159X.

[191] W. R. Boot, A. F. Kramer, D. J. Simons, M. Fabiani, and G. Gratton, "The effects of video game playing on attention, memory, and executive control", *Acta psychologica*, vol. 129, no. 3, pp. 387–398, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001691808001200 (visited on 09/27/2015).

[192] E. S. Association, "Essential facts about the computer and video game industry", Entertainment Software Association, Tech. Rep., 2016. [Online]. Available: http://essentialfacts.theesa.com/Essential-Facts-2016.pdf;.

[193] B. De Schutter, "Never Too Old to Play: The Appeal of Digital Games to an Older Audience", *Games and Culture: A Journal of Interactive Media,*, vol. in progress, 2008. (visited on 03/01/2009).

[194] V. O. Puntmann, "How-to guide on biomarkers: Biomarker definitions, validation and applications with examples from cardiovascular disease", en, *Postgraduate Medical Journal*, vol. 85, no. 1008, pp. 538–545, Oct. 2009, ISSN: 0032-5473. DOI: 10.1136/pgmj.2008.073759. [Online]. Available: http://pmj.bmj.com/cgi/doi/10.1136/pgmj.2008.073759 (visited on 07/09/2019).

[195] D. J. Selkoe and J. Hardy, "The amyloid hypothesis of Alzheimer's disease at 25 years", *EMBO Molecular Medicine*, vol. 8, no. 6, pp. 595–608, Jun. 2016, ISSN: 1757-4676. DOI: 10.15252/emmm.201606210. [Online]. Available: https://www.embopress.org/doi/full/10.15252/emmm.201606210 (visited on 07/09/2019).

[196] Microsoft, *Microsoft Casual Games*. [Online]. Available: http://zone.msn.com/microsoftcasualgames/#solitaire (visited on 06/22/2017).

[197] Y. Hindes, M. R. Schoenberg, and D. H. Saklofske, "Intelligence", en, in *Encyclopedia of Clinical Neuropsychology*, J. S. Kreutzer, J. DeLuca, and B. Caplan, Eds., New York, NY: Springer New York, 2011, pp. 1329–1335. DOI: 10.1007/978-0-387-79948-3_1061. [Online]. Available: http://link.springer.com/10.1007/978-0-387-79948-3_1061 (visited on 03/25/2019).

[198] T. Salthouse, "Consequences of Age-Related Cognitive Declines", en, *Annual Review of Psychology*, vol. 63, no. 1, pp. 201–226, Jan. 2012, ISSN: 0066-4308, 1545-2085. DOI: 10.1146/annurev-psych-120710-100328. [Online]. Available: http://www.annualreviews.org/doi/10.1146/annurev-psych-120710-100328 (visited on 03/27/2019).

[199] A. Adan, "Cognitive Performance and Dehydration", en, *Journal of the American College of Nutrition*, vol. 31, no. 2, pp. 71–78, Apr. 2012, ISSN: 0731-5724, 1541-1087. DOI: 10.1080/07315724.2012.10720011. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07315724.2012.10720011 (visited on 06/06/2020).

[200] T. A. Salthouse, "What and when of cognitive aging", *Current directions in psychological science*, vol. 13, no. 4, pp. 140–144, 2004.

[201] W. R. Boot and A. F. Kramer, *The Brain-Games Conundrum: Does Cognitive Training Really Sharpen the Mind?* -, 2014. [Online]. Available: http://dana.org/Cerebrum/2014/The_Brain-Games_Conundrum__Does_Cognitive_Training_Really_Sharpen_the_Mind_/ (visited on 09/27/2015).

[202] Nintendo, *Brain Age*, Kyoto, Japan, May 2005.

[203] L. Labs, *Lumosity*, San Francisco, CA, 2011. [Online]. Available: https://www.lumosity.com/en/ (visited on 04/09/2019).

[204] M. Merzenich, *Brain Fitness Program*, 2016. [Online]. Available: https://www.brainhq.com.

[205] M. Rusk, *Lumosity to Pay $2 Million to Settle FTC Deceptive Advertising Charges for Its "Brain Training" Program*, en, Jan. 2016. [Online]. Available: https://www.ftc.gov/news-events/press-releases/2016/01/lumosity-pay-2-million-settle-ftc-deceptive-advertising-charges (visited on 04/09/2019).

[206] G. Sala, K. S. Tatlidil, and F. Gobet, "Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation.", en, *Psychological Bulletin*, vol. 144, no. 2, pp. 111–139, Feb. 2018, ISSN: 1939-1455, 0033-2909. DOI: 10.1037/bul0000139. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/bul0000139 (visited on 07/09/2019).

[207] T. M. Shah, M. Weinborn, G. Verdile, H. R. Sohrabi, and R. N. Martins, "Enhancing Cognitive Functioning in Healthly Older Adults: A Systematic Review of the Clinical Significance of Commercially Available Computerized Cognitive Training in Preventing Cognitive Decline", en, *Neuropsychology Review*, vol. 27, no. 1, pp. 62–80, Mar. 2017, ISSN: 1573-6660. DOI: 10.1007/s11065-016-9338-9. [Online]. Available: https://doi.org/10.1007/s11065-016-9338-9 (visited on 04/09/2019).

[208] J. A. Anguera *et al.*, "Video game training enhances cognitive control in older adults", *Nature*, vol. 501, no. 7465, pp. 97–101, Sep. 2013, ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature12486. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature12486 (visited on 08/14/2016).

[209] C. S. Green and D. Bavelier, "Action video game modifies visual selective attention", *Nature*, vol. 423, no. 6939, pp. 534–537, May 2003, ISSN: 00280836. DOI: 10.1038/nature01647. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature01647 (visited on 08/14/2016).

[210] S. L. Tennstedt and F. W. Unverzagt, "The ACTIVE Study: Study Overview and Major Findings", en, *Journal of Aging and Health*, vol. 25, no. 8_suppl, 3S–20S, Dec. 2013, ISSN: 0898-2643, 1552-6887. DOI: 10.1177/0898264313518133. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0898264313518133 (visited on 07/26/2019).

[211] E. Donchin, M. Fabiani, and A. Sanders, "The learning strategies program: An examination of the strategies in skill acquisition", *Acta Psychologica, 71*, 1989. [Online]. Available: http://scholarcommons.usf.edu/psy_facpub/169/ (visited on 09/27/2015).

[212] H. M. Blumen, D. Gopher, J. Steinerman, and Y. Stern, "Training Cognitive Control in Older Adults with the Space Fortress Game: The Role of Training Instructions and Basic Motor Ability", English, *Frontiers in Aging Neuroscience*, vol. 2, 2010, ISSN: 1663-4365. DOI: 10.3389/fnagi.2010.00145. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnagi.2010.00145/full (visited on 04/09/2019).

[213] A. Coutrot *et al.*, "Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance", en, *PLOS ONE*, vol. 14, no. 3, L. Zamarian, Ed., e0213272, Mar. 2019, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0213272. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0213272 (visited on 07/04/2019).

[214] T. Tong, M. Chignell, P. Lam, M. C. Tierney, and J. Lee, "Designing Serious Games for Cognitive Assessment of the Elderly", en, *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, vol. 3, no. 1, pp. 28–35, Jun. 2014, ISSN: 2327-8595, 2327-8595. DOI: 10.1177/2327857914031004. [Online]. Available: http://journals.sagepub.com/doi/10.1177/2327857914031004 (visited on 06/07/2020).

[215] V. Vanden Abeele, P. Tierens, B. De Schutter, L. De-Wit, and L. Geurts, "Game-based Experiments on Human Visual Attention.", in *Foundations of Digital Games*, Chania, Greece, 2015.

[216] S. Hagler, H. B. Jimison, and M. Pavel, "Assessing Executive Function Using a Computer Game: Computational Modeling of Cognitive Processes", *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1442–1452, Jul. 2014, ISSN: 2168-2194, 2168-2208. DOI: 10.1109/JBHI.2014.2299793. [Online]. Available: http://ieeexplore.ieee.org/document/6732879/ (visited on 06/07/2020).

[217] H. Silva Neto, J. Cerejeira, and L. Roque, "Cognitive screening of older adults using serious games: An empirical study", en, *Entertainment Computing*, vol. 28, pp. 11–20, Dec. 2018, ISSN: 18759521. DOI: 10.1016/j.entcom.2018.08.002. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1875952118300326 (visited on 07/11/2019).

[218] T. Tong, M. Chignell, M. C. Tierney, and J. Lee, "A Serious Game for Clinical Assessment of Cognitive Status: Validation Study", *JMIR Serious Games*, vol. 4, no. 1, May 2016, ISSN: 2291-9279. DOI: 10.2196/games.5006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4902858/.

[219] L. Galarneau, "The elearning edge: Leveraging interactive technologies in the design of engaging, effective learning experiences", *Proceedings of e-Fest*, pp. 1–10, 2004.

[220] S. Ballesteros, A. Prieto, J. Mayas, P. Toril, C. Pita, L. Ponce De León, J. M. Reales, and J. Waterworth, "Brain training with non-action video games enhances aspects of cognition in older adults: A randomized controlled trial", *Frontiers in Aging Neuroscience*, vol. 6, Oct. 2014, ISSN: 1663-4365. DOI: 10.3389/fnagi.2014.00277. [Online]. Available:

http://journal.frontiersin.org/article/10.3389/fnagi.2014.00277/abstract (visited on 08/14/2016).

[221] W. R. Boot, M. Champion, D. P. Blakely, T. Wright, D. J. Souders, and N. Charness, "Video Games as a Means to Reduce Age-Related Cognitive Decline: Attitudes, Compliance, and Effectiveness", *Frontiers in Psychology*, vol. 4, 2013, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00031. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00031/abstract (visited on 08/14/2016).

[222] M. Duggan, *Which Americans play video games and who identifies as a "gamer"*, en-US, Dec. 2015. [Online]. Available: https://www.pewinternet.org/2015/12/15/who-plays-video-games-and-identifies-as-a-gamer/ (visited on 04/08/2019).

[223] A. Brown, S. 800Washington, and D. 2.-4.-4. |. M.-8.-8. |. F.-4.-4. |. M. Inquiries, *Who plays video games? Younger men, but many others too*, en-US, Sep. 2017. [Online]. Available: https://www.pewresearch.org/fact-tank/2017/09/11/younger-men-play-video-games-but-so-do-a-diverse-group-of-other-americans/ (visited on 04/08/2019).

[224] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy Array: A Structure for Efficient Numerical Computation", *Computing in Science & Engineering*, vol. 13, no. 2, pp. 22–30, Mar. 2011, ISSN: 1521-9615. DOI: 10.1109/MCSE.2011.37. [Online]. Available: http://ieeexplore.ieee.org/document/5725236/ (visited on 06/07/2020).

[225] T. Kluyver *et al.*, "Jupyter Notebooks - a publishing format for reproducible computational workflows", in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, F. Loizides and B. Schmidt, Eds., IOS Press, 2016, pp. 87–90.

[226] W. McKinney, "Pandas: A foundational Python library for data analysis and statistics", *Python for High Performance and Scientific Computing*, vol. 14, 2011.

[227] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[228] M. Waskom, *Seaborn: Statistical data visualization*. [Online]. Available: https://seaborn.pydata.org/ (visited on 04/08/2019).

[229] Scikit-learn, *Cross-validation: Evaluating estimator performance - scikit-learn 0.20.3 documentation*. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html (visited on 04/09/2019).

[230] ——, *Validation curves: Plotting scores to evaluate models - scikit-learn 0.20.3 documentation*. [Online]. Available: https://scikit-learn.org/stable/modules/learning_curve.html (visited on 04/09/2019).

[231] P. Saxena, *MACHINE LEARNING work->flow*, Jan. 2019. [Online]. Available: https://medium.com/@harshkr21august/machine-learning-work-flow-2e8f38da0511 (visited on 04/09/2019).

[232] J. R. Thompson, *John Tukey (1915-2000): Deconstructing Statistics*, en. [Online]. Available: http://www.casid.info/ACAS00-02/ACAS00/ThompsonJames/ThompsonJames.pdf.

[233] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique", en, *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, ISSN: 1076-9757. DOI: 10.1613/jair.953. [Online]. Available: https://www.jair.org/index.php/jair/article/view/10302 (visited on 04/09/2019).

[234] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed, ser. Springer series in statistics. New York, NY: Springer, 2009.

[235] M. E. Rice and G. T. Harris, "Comparing Effect Sizes in Follow-Up Studies: ROC Area, Cohen's d, and r", en, *Law and Human Behavior*, vol. 29, no. 5, pp. 615–620, Oct. 2005, ISSN: 1573-661X. DOI: 10.1007/s10979-005-6832-7. [Online]. Available: https://doi.org/10.1007/s10979-005-6832-7 (visited on 04/09/2019).

[236] L. E. Bourne Jr and R. A. Yaroush, "Stress and cognition: A cognitive psychological perspective", NASA/CR, Tech. Rep., 2003. [Online]. Available: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20040034070.pdf (visited on 08/03/2017).

[237] B. Wallace, R. Goubran, F. Knoefel, M. Petriu, and A. McAvoy, "Design of games for measurement of cognitive impairment", in *2014 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Jun. 2014, pp. 117–120. DOI: 10.1109/BHI.2014.6864318.

[238] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The motivational pull of video games: A self-determination theory approach.", *Motivation and Emotion*, vol. 30, no. 4, pp. 347–363, 2006, ISSN: 1573-6644(Electronic),0146-7239(Print). DOI: 10.1007/s11031-006-9051-8.

[239] S. Swarbrick, N. Wragg, S. Ghosh, and A. Stolzing, "Systematic Review of miRNA as Biomarkers in Alzheimer's Disease", eng, *Molecular Neurobiology*, vol. 56, no. 9, pp. 6156–6167, Sep. 2019, ISSN: 1559-1182. DOI: 10.1007/s12035-019-1500-y.

[240] J. Cannon-Bowers, C. Bowers, and J. Rodrigues, Eds., *Serious Game Design and Development: Technologies for Training and Learning*, ser. Advances in Multimedia and Interactive Technologies. IGI Global, 2010. DOI: 10.4018/978-1-61520-739-8. [Online]. Available: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-61520-739-8 (visited on 12/03/2019).

[241] D. Munoz and N. Weishaupt, "Vascular Dementia", en, in *The Cerebral Cortex in Neurodegenerative and Neuropsychiatric Disorders*, Elsevier, 2017, pp. 119–139, ISBN: 978-0-12-801942-9. DOI: 10.1016/B978-0-12-801942-9.00005-7. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780128019429000057 (visited on 03/30/2020).

[242] C. Reitz, C. Brayne, and R. Mayeux, "Epidemiology of Alzheimer disease", en, *Nature Reviews Neurology*, vol. 7, no. 3, pp. 137–152, Mar. 2011, ISSN: 1759-4758, 1759-4766. DOI: 10.1038/nrneurol.2011.2. [Online]. Available: http://www.nature.com/articles/nrneurol.2011.2 (visited on 03/30/2020).

[243] R. Tampi, D. Tampi, A. Ghori, and M. Durning, "Mild cognitive impairment: A comprehensive review", *Healthy Aging Research*, 2015, ISSN: 22617434. DOI: 10.12715/har.2015.4.39. [Online]. Available: http://www.har-journal.com/archives/1529 (visited on 03/30/2020).

[244] T. C. C. Pinto, L. Machado, T. M. Bulgacov, A. L. Rodrigues-Júnior, M. L. G. Costa, R. C. C. Ximenes, and E. B. Sougey, "Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly?", en, *International Psychogeriatrics*, vol. 31, no. 04, pp. 491–504, Apr. 2019, ISSN: 1041-6102, 1741-203X. DOI: 10.1017/S1041610218001370. [Online]. Available: https://www.cambridge.org/core/product/identifier/S1041610218001370/type/journal_article (visited on 11/04/2019).

[245] R. C. Petersen, "Mild Cognitive Impairment", *Continuum : Lifelong Learning in Neurology*, vol. 22, no. 2 Dementia, pp. 404–418, Apr. 2016, ISSN: 1080-2371. DOI: 10.1212/CON.0000000000000313. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5390929/ (visited on 07/03/2019).

[246] J. C. Morris, "Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type", en, *International Psychogeriatrics*, vol. 9, no. S1, pp. 173–176, Dec. 1997, ISSN: 1041-6102, 1741-203X. DOI: 10.1017/S1041610297004870. [Online].

Available: https://www.cambridge.org/core/product/identifier/S1041610297004870/type/journal_article (visited on 03/30/2020).

[247] P. H. Robert *et al.*, "Recommendations for the use of Serious Games in people with Alzheimer's Disease, related disorders and frailty", *Frontiers in Aging Neuroscience*, vol. 6, Mar. 2014, ISSN: 1663-4365. DOI: 10.3389/fnagi.2014.00054. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnagi.2014.00054/abstract (visited on 01/09/2020).

[248] A. Escandon, N. Al-Hammadi, and J. E. Galvin, "Effect of cognitive fluctuation on neuropsychological performance in aging and dementia", *Neurology*, vol. 74, no. 3, pp. 210–217, Jan. 2010, ISSN: 0028-3878. DOI: 10.1212/WNL.0b013e3181ca017d. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2809035/.

[249] V. Joshi, B. Wallace, A. Shaddy, F. Knoefel, R. Goubran, and C. Lord, "Metrics to monitor performance of patients with mild cognitive impairment using computer based games", in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Las Vegas, NV: IEEE, Feb. 2016, pp. 521–524, ISBN: 978-1-5090-2455-1. DOI: 10.1109/BHI.2016.7455949. [Online]. Available: http://ieeexplore.ieee.org/document/7455949/ (visited on 09/03/2020).

[250] C. Blake and I. P. Gent, "The Winnability of Klondike and Many Other Single-Player Card Games", *ArXiv*, vol. abs/1906.12314, 2019.

[251] M. H. Goadrich and J. Droscha, "Improving Solvability for Procedurally Generated Challenges in Physical Solitaire Games Through Entangled Components", *IEEE Transactions on Games*, pp. 1–1, 2019, ISSN: 2475-1502, 2475-1510. DOI: 10.1109/TG.2019.2918223. [Online]. Available: https://ieeexplore.ieee.org/document/8721547/ (visited on 10/03/2019).

[252] L. Longpré and P. McKenzie, "The complexity of Solitaire", en, *Theoretical Computer Science*, vol. 410, no. 50, pp. 5252–5260, Nov. 2009, ISSN: 03043975. DOI: 10.1016/j.tcs.2009.08.027. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0304397509006100 (visited on 08/17/2016).

[253] A. W. Roscoe, "Card games as pointer structures: Case studies in mobile CSP modelling", *ArXiv*, vol. abs/1611.08418, 2016.

[254] X. Yan, P. Diaconis, P. Rusmevichientong, and B. V. Roy, "Solitaire: Man Versus Machine", in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds., MIT Press, 2005, pp. 1553–1560. [Online]. Available: http://papers.nips.cc/paper/2568-solitaire-man-versus-machine.pdf.

[255] *Basic Klondike (Solitaire) Tips*, en-US, Mar. 2013. [Online]. Available: https://cardgameheaven.com/basic-klondike-solitaire-tips/ (visited on 10/03/2019).

[256] *Klondike Solitaire Strategy*. [Online]. Available: https://www.bvssolitaire.com/rules/klondike-solitaire-strategy.htm (visited on 10/03/2019).

[257] *Winning at Klondike*. [Online]. Available: https://semicolon.com/Solitaire/Articles/Klondike.html (visited on 10/03/2019).

[258] J. E. Tan, E. Strauss, and E. M. S. Sherman, "Clinical Dementia Rating", in *Encyclopedia of Clinical Neuropsychology*, J. S. Kreutzer, J. DeLuca, and B. Caplan, Eds., New York, NY: Springer New York, 2011, pp. 587–590, ISBN: 978-0-387-79948-3. DOI: 10.1007/978-0-387-79948-3_533. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-79948-3_533.

[259] A. Diamond, "Executive Functions", en, *Annual Review of Psychology*, vol. 64, no. 1, pp. 135–168, Jan. 2013, ISSN: 0066-4308, 1545-2085. DOI: 10.1146/annurev-psych-113011-143750. [Online]. Available: http://www.annualreviews.org/doi/10.1146/annurev-psych-113011-143750 (visited on 05/03/2018).

[260] E. L. Glisky, "Changes in Cognitive Function in Human Aging", eng, in *Brain Aging: Models, Methods, and Mechanisms*, ser. Frontiers in Neuroscience, D. R. Riddle, Ed., Boca Raton (FL): CRC Press/Taylor & Francis, 2007, ISBN: 978-0-8493-3818-2. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK3885/ (visited on 07/10/2018).

[261] A. M. Owen, "Cognitive planning in humans: Neuropsychological, neuroanatomical and neuropharmacological perspectives", *Progress in Neurobiology*, vol. 53, no. 4, pp. 431–450, Nov. 1997, ISSN: 0301-0082. DOI: 10.1016/S0301-0082(97)00042-7. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301008297000427 (visited on 12/06/2018).

[262] P. Verhaeghen and J. Cerella, "Aging, executive control, and attention: A review of meta-analyses", eng, *Neuroscience and Biobehavioral Reviews*, vol. 26, no. 7, pp. 849–857, Nov. 2002, ISSN: 0149-7634.

[263] K. Gielis, *ExplanationSolitaire_short*, Nov. 2019. DOI: 10.6084/m9.figshare.10304780.v1. [Online]. Available: https://figshare.com/articles/ExplanationSolitaire_Short/10304780 (visited on 11/14/2019).

[264] ——, *PlayerActionsSolitaire_short*, Nov. 2019. DOI: 10.6084/m9.figshare.10304801.v1. [Online]. Available: https://figshare.com/articles/PlayerActionsSolitaire_Short/10304801 (visited on 11/14/2019).

[265] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability.", en, *Psychological Bulletin*, vol. 86, no. 2, pp. 420–428, 1979, ISSN: 1939-1455, 0033-2909. DOI: 10.1037/0033-2909.86.2.420. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.86.2.420 (visited on 10/08/2019).

[266] *IBM SPSS Statistics 23 Documentation*, en, CT701, CT709, CT711, CT712, CT713, CT807, CT708, Mar. 2015. [Online]. Available: https://www.ibm.com/support/pages/ibm-spss-statistics-23-documentation (visited on 11/14/2019).

[267] A. von Eye, *Analyzing Rater Agreement: Manifest Variable Methods*, en, 1st ed. Psychology Press, Apr. 2014, ISBN: 978-1-4106-1102-4. DOI: 10.4324/9781410611024. [Online]. Available: https://www.taylorfrancis.com/books/9781410611024 (visited on 11/14/2019).

[268] K. A. Hallgren, "Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial", eng, *Tutorials in Quantitative Methods for Psychology*, vol. 8, no. 1, pp. 23–34, 2012, ISSN: 1913-4126.

[269] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research", en, *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, Jun. 2016, ISSN: 15563707. DOI: 10.1016/j.jcm.2016.02.012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1556370716000158 (visited on 10/08/2019).

[270] T. Bielefeld, *[Android] Simple Solitaire game collection. Contribute to TobiasBielefeld/Simple-Solitaire development by creating an account on GitHub*, original-date: 2016-10-27T16:51:18Z, Aug. 2019. [Online]. Available: https://github.com/TobiasBielefeld/Simple-Solitaire (visited on 08/16/2019).

[271] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020. [Online]. Available: https://www.R-project.org/.

[272] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4", *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015. DOI: 10.18637/jss.v067.i01.

[273] X. A. Harrison *et al.*, "A brief introduction to mixed effects modelling and multi-model inference in ecology", en, *PeerJ*, vol. 6, e4794, May 2018, ISSN: 2167-8359. DOI: 10.7717/peerj.4794. [Online]. Available: https://peerj.com/articles/4794 (visited on 06/10/2020).

[274] B. Winter, "Linear models and linear mixed effects models in R with linguistic applications", *arXiv:1308.5499 [cs]*, Aug. 2013, arXiv: 1308.5499. [Online]. Available: http://arxiv.org/abs/1308.5499 (visited on 06/09/2020).

[275] S. Nakagawa and H. Schielzeth, "A general and simple method for obtaining R2 from generalized linear mixed-effects models", en, *Methods in Ecology and Evolution*, vol. 4, no. 2, R. B. O'Hara, Ed., pp. 133–142, Feb. 2013, ISSN: 2041210X. DOI: 10.1111/j.2041-210x.2012.00261.x. [Online]. Available: http://doi.wiley.com/10.1111/j.2041-210x.2012.00261.x (visited on 08/14/2020).

[276] *Modern statistical methods for HCI*. New York, NY: Springer Berlin Heidelberg, 2016, ISBN: 978-3-319-26631-2.

[277] M. Ángeles Quiroga *et al.*, "Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can!", en, *Intelligence*, vol. 53, pp. 1–7, Nov. 2015, ISSN: 01602896. DOI: 10.1016/j.intell.2015.08.004. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S016028961500104X (visited on 07/14/2020).

[278] G. Csukly *et al.*, "The Differentiation of Amnestic Type MCI from the Non-Amnestic Types by Structural MRI", *Frontiers in Aging Neuroscience*, vol. 8, Mar. 2016, ISSN: 1663-4365. DOI: 10.3389/fnagi.2016.00052. [Online]. Available: http://journal.frontiersin.org/Article/10.3389/fnagi.2016.00052/abstract (visited on 01/22/2020).

[279] S. MacLean and S. Hatcher, "Constructing the (Healthy) Neoliberal Citizen: Using the Walkthrough Method "Do" Critical Health Communication Research", *Frontiers in Communication*, vol. 4, p. 52, Oct. 2019, ISSN: 2297-900X. DOI: 10.3389/fcomm.2019.00052. [Online]. Available: https://www.frontiersin.org/article/10.3389/fcomm.2019.00052/full (visited on 08/15/2020).

[280] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, "Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses", en, *Behavior Research Methods*, vol. 41, no. 4, pp. 1149–1160, Nov. 2009, ISSN: 1554-351X, 1554-3528. DOI: 10.3758/BRM.41.4.1149. [Online]. Available: http://link.springer.com/10.3758/BRM.41.4.1149 (visited on 09/09/2020).

[281] A. J. Larner, "Effect Size (Cohen's d) of Cognitive Screening Instruments Examined in Pragmatic Diagnostic Accuracy Studies", en, *Dementia and Geriatric Cognitive Disorders Extra*, vol. 4, no. 2, pp. 236–241, Jul. 2014, ISSN: 1664-5464. DOI: 10.1159/000363735. [Online]. Available: https://www.karger.com/Article/FullText/363735 (visited on 09/09/2020).

[282] M. Overton, M. Pihlsgård, and S. Elmståhl, "Test administrator effects on cognitive performance in a longitudinal study of ageing", *Cogent Psychology*, vol. 3, no. 1, P. Walla, Ed., p. 1 260 237, Dec. 2016, Publisher: Cogent OA, ISSN: null. DOI: 10.1080/23311908.2016.1260237. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/23311908.2016.1260237.

[283] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, "Mild cognitive impairment: Clinical characterization and outcome", eng, *Archives of Neurology*, vol. 56, no. 3, pp. 303–308, Mar. 1999, ISSN: 0003-9942. DOI: 10.1001/archneur.56.3.303.

[284] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician", *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, Nov. 1975, ISSN: 0022-3956. DOI: 10.1016/0022-3956(75)90026-6. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0022395675900266 (visited on 09/21/2015).

[285] S. Raschka, *Python Machine Learning*, eng, ser. Community experience distilled. Birmingham Mumbai: Packt Publishing open source, 2016, OCLC: 927507196, ISBN: 978-1-78355-513-0.

[286] E. Chaibub Neto *et al.*, "Detecting the impact of subject characteristics on machine learning-based diagnostic applications", en, *npj Digital Medicine*, vol. 2, no. 1, p. 99, Dec. 2019, ISSN: 2398-6352. DOI: 10.1038/s41746-019-0178-x. [Online]. Available: http://www.nature.com/articles/s41746-019-0178-x (visited on 06/09/2020).

[287] D. Rojo, N. N. Htun, and K. Verbert, "GaCoVi: A Correlation Visualization to Support Interpretability-Aware Feature Selection for Regression Models", *EuroVis 2020 - Short Papers*, 5 pages, 2020, Artwork Size: 5 pages ISBN: 9783038681069 Publisher: The Eurographics Association Version Number: 127-131. DOI: 10.2312/EVS.20201060. [Online]. Available: https://diglib.eg.org/handle/10.2312/evs20201060 (visited on 07/27/2020).

[288] UNESCO United Nations Educational, Scientific and Cultural Organization, "International Standard Classification of Education, ISCED 1997", en, in *Advances in Cross-National Comparison*, J. H. P. Hoffmeyer-Zlotnik and C. Wolf, Eds., Boston, MA: Springer US, 2003, pp. 195–220.

DOI: 10.1007/978-1-4419-9186-7_10. [Online]. Available: http://link.springer.com/10.1007/978-1-4419-9186-7_10 (visited on 04/01/2020).

[289] S. Valladares-Rodríguez, L. Anido-Rifón, M. J. Fernández-Iglesias, and D. Facal-Mayo, "A Machine Learning Approach to the Early Diagnosis of Alzheimer's Disease Based on an Ensemble of Classifiers", en, in *Computational Science and Its Applications - ICCSA 2019*, S. Misra *et al.*, Eds., vol. 11619, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, pp. 383–396. DOI: 10.1007/978-3-030-24289-3_28. [Online]. Available: http://link.springer.com/10.1007/978-3-030-24289-3_28 (visited on 07/29/2020).

[290] A. B. Zonderman and G. A. Dore, "Risk of dementia after fluctuating mild cognitive impairment: When the yo-yoing stops", en, *Neurology*, vol. 82, no. 4, pp. 290–291, Jan. 2014, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.0000000000000065. [Online]. Available: http://www.neurology.org/cgi/doi/10.1212/WNL.0000000000000065 (visited on 08/30/2017).

[291] M. S. Collection, *Get Microsoft Solitaire Collection*, en-us. [Online]. Available: https://www.microsoft.com/en-us/p/microsoft-solitaire-collection/9wzdncrfhwd2 (visited on 02/21/2019).

[292] J. Kennes and C. De Dobbeleer, "Classification and Regression of Age using Gameplay Data", English, M.S. thesis, KU Leuven, Leuven, Jun. 2019.

[293] M. Roosen and B. Hermans, "Data analytics for Klondike Solitaire through Image Recognition", English, M.S. thesis, KU Leuven, Leuven, Jun. 2018.

[294] J. Stessens and J. Schurmans, "Developing a Computer Vision Toolkit for Measuring Digital Biomarkers of Cognitive Performance in Card Gameplay: A Deep Neural Network Approach", English, M.S. thesis, KU Leuven, Leuven, Jun. 2020.

[295] R. E. Brown, "Hebb and Cattell: The Genesis of the Theory of Fluid and Crystallized Intelligence", *Frontiers in Human Neuroscience*, vol. 10, Dec. 2016, ISSN: 1662-5161. DOI: 10.3389/fnhum.2016.00606. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnhum.2016.00606/full (visited on 03/25/2019).

[296] J. Wobbrock and J. Kientz, "Research contribution in human-computer interaction", *interactions*, vol. 23, pp. 38–44, Apr. 2016. DOI: 10.1145/2907069.

[297] K. M. Langa and D. A. Levine, "The Diagnosis and Management of Mild Cognitive Impairment: A Clinical Review", en, *JAMA*, vol. 312, no. 23, p. 2551, Dec. 2014, ISSN: 0098-7484. DOI: 10.1001/jama.2014.13806. [Online]. Available: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2014.13806 (visited on 12/09/2020).

[298] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions", in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[299] A. Messalas, Y. Kanellopoulos, and C. Makris, "Model-agnostic interpretability with shapley values", in *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 2019, pp. 1–7. DOI: 10.1109/IISA.2019.8900669.

[300] S. Few, *Information dashboard design: the effective visual communication of data*, 1st ed. Beijing ; Cambride [MA]: O'Reilly, 2006, OCLC: ocm63676286, ISBN: 978-0-596-10016-2.

[301] H. Kaduszkiewicz *et al.*, "Prognosis of Mild Cognitive Impairment in General Practice: Results of the German AgeCoDe Study", en, *The Annals of Family Medicine*, vol. 12, no. 2, pp. 158–165, Mar. 2014, ISSN: 1544-1709, 1544-1717. DOI: 10.1370/afm.1596. [Online]. Available: http://www.annfammed.org/cgi/doi/10.1370/afm.1596 (visited on 01/12/2021).

[302] B. Greenberg, J. Sherry, K. Lachlan, K. Lucas, and A. Holmstrom, "Orientations to video games among gender and age groups", *Simulation & Gaming - Simulat Gaming*, vol. 41, pp. 238–259, Apr. 2010. DOI: 10.1177/1046878108319930.

[303] J. Sherry, B. Greenberg, K. Lucas, and K. Lachlan, "Video game uses and gratifications as predictors of use and game preference", in. Jan. 2006, vol. 8, pp. 213–224.

[304] R. M. Bauer, G. L. Iverson, A. N. Cernich, L. M. Binder, R. M. Ruff, and R. I. Naugle, "Computerized Neuropsychological Assessment Devices: Joint Position Paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology", en, *Archives of Clinical Neuropsychology*, vol. 27, no. 3, pp. 362–373, May 2012, ISSN: 0887-6177, 1873-5843. DOI: 10.1093/arclin/acs027. [Online]. Available: https://academic.oup.com/acn/article-lookup/doi/10.1093/arclin/acs027 (visited on 08/23/2018).

[305] A. Piau, Z. Steinmeyer, M. Cesari, J. Kornfeld, Z. Beattie, J. Kaye, B. Vellas, and F. Nourhashemi, "Intrinsic Capacitiy monitoring by digital biomarkers in Integrated Care for Older People (ICOPE)", *The Journal of Frailty & Aging*, pp. 1–7, 2020, ISSN: 22734309. DOI: 10.14283/jfa.2020.51. [Online]. Available: https://link.springer.com/article/10.14283/jfa.2020.51 (visited on 05/03/2021).

[306] L. Beglinger, B. Gaydos, O. Tangphaodaniels, K. Duff, D. Kareken, J. Crawford, P. Fastenau, and E. Siemers, "Practice effects and the use of alternate forms in serial neuropsychological testing", en, *Archives of Clinical Neuropsychology*, vol. 20, no. 4, pp. 517–529, Jun. 2005, ISSN: 08876177. DOI: 10.1016/j.acn.2004.12.003. [Online]. Available: https://academic.oup.com/acn/article-lookup/doi/10.1016/j.acn.2004.12.003 (visited on 05/03/2021).

[307] *Digital health: scaling healthcare to the world*. New York, NY: Springer Berlin Heidelberg, 2017, ISBN: 978-3-319-61445-8.

[308] A. S. Lu, "Serious Games for Healthcare: Applications and Implications", en, *Games for Health Journal*, vol. 2, no. 5, pp. 313–314, Oct. 2013, ISSN: 2161-783X, 2161-7856. DOI: 10.1089/g4h.2013.0062. [Online]. Available: http://www.liebertpub.com/doi/10.1089/g4h.2013.0062 (visited on 01/09/2021).

[309] B. Davies, "The right not to know and the obligation to know", en, *Journal of Medical Ethics*, vol. 46, no. 5, pp. 300–303, May 2020, ISSN: 0306-6800, 1473-4257. DOI: 10.1136/medethics-2019-106009. [Online]. Available: https://jme.bmj.com/lookup/doi/10.1136/medethics-2019-106009 (visited on 01/09/2021).

[310] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate", en, *Big Data & Society*, vol. 3, no. 2, p. 205 395 171 667 967, Dec. 2016, ISSN: 2053-9517, 2053-9517. DOI: 10.1177/2053951716679679. [Online]. Available: http://journals.sagepub.com/doi/10.1177/2053951716679679 (visited on 01/09/2021).

[311] R. Geirhos, J.-H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, and F. A. Wichmann, "Shortcut Learning in Deep Neural Networks", *arXiv:2004.07780 [cs, q-bio]*, May 2020, arXiv: 2004.07780. [Online]. Available: http://arxiv.org/abs/2004.07780 (visited on 06/09/2020).

[312] D. Lupton, "The digitally engaged patient: Self-monitoring and self-care in the digital health era", en, *Social Theory & Health*, vol. 11, no. 3, pp. 256–270, Aug. 2013, ISSN: 1477-8211, 1477-822X. DOI: 10.1057/sth.2013.10. [Online]. Available: http://link.springer.com/10.1057/sth.2013.10 (visited on 01/09/2021).

[313] O. C. Okonkwo *et al.*, "Medical decision-making capacity in mild cognitive impairment: A 3-year longitudinal study", en, *Neurology*, vol. 71, no. 19, pp. 1474–1480, Nov. 2008, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/01.wnl.0000334301.32358.48. [Online]. Available: http://www.neurology.org/cgi/doi/10.1212/01.wnl.0000334301.32358.48 (visited on 01/09/2021).

[314] A. Elragal and R. Klischewski, "Theory-driven or process-driven prediction? Epistemological challenges of big data analytics", en, *Journal of Big Data*, vol. 4, no. 1, p. 19, Dec. 2017, ISSN: 2196-1115. DOI: 10.1186/s40537-017-0079-2. [Online]. Available: http://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0079-2 (visited on 08/15/2020).

[315] K. Gielis, "Screening for Mild Cognitive Impairment through Digital Biomarkers of Cognitive Performance in Games", en, in *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, Barcelona Spain: ACM, Oct. 2019, pp. 7–13, ISBN: 978-1-4503-6871-1. DOI: 10.1145/3341215.3356332. [Online]. Available: https://dl.acm.org/doi/10.1145/3341215.3356332 (visited on 01/26/2021).

[316] K. Gielis, "Assessment of cognitive performance in elderly life via meaningful play", in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, 2019, pp. 1–2. DOI: 10.1109/ICHI.2019.8904861.

[317] K. Gielis, F. Brito, J. Tournoy, and V. Vanden Abeele, "Screening Mild Cognitive Impairment in Older Adults via Meaningful Play", en, in *CHI PLAY 2017 Position Paper - Position Paper of the Annual Symposium on Computer-Human Interaction in Play*, Amsterdam, Oct. 2017.

[318] K. Gielis and V. Vanden Abeele, *Analyzing Cognitive Functioning via Card Game Play: Presenting a Toolkit and a Case Study*, Forum on Video Games for Mental Health, CHI Play, Melbourne, Oct. 2018.

[319] K. Gielis, J. Tournoy, K. Verbert, and V. Vanden Abeele, *Screening for mild cognitive impairment in older adults through card games*, en, Games4MentalHealth, Leuven, Belgium, Feb. 2018. [Online]. Available: https://www.kuleuven.be/gamesformentalhealth.

[320] K. Gielis, J. Tournoy, K. Verbert, and V. Vanden Abeele, *Screening for Mild Cognitive Impairment in Older Adults through Card Games*, Symposium Veerkrachtig Ouder Worden, Leuven, Belgium, Sep. 2017.

[321] K. Gielis, J. Kennes, C. De Dobbeleer, S. Puttemans, and V. Vanden Abeele, *We built an Image Processing Toolkit for Card Games to Capture Digital Biomarkers of Cognitive Performance*, ICHI, Xi'an, 2019.

[322] K. Gielis, *Dr. Solitaire MCI Dataset*, Feb. 2021. [Online]. Available: DOI:10.6084/m9.figshare.14057951.

[323] ——, *Dr. Solitaire Aging Dataset*, Feb. 2021. [Online]. Available: DOI:10.6084/m9.figshare.14057960.

[324] K. Gielis, *Dr. Solitaire Computer Vision App*, Feb. 2021. [Online]. Available: DOI:10.6084/m9.figshare.13705291.

[325] ——, *Dr. Solitaire Android App*, Feb. 2021. [Online]. Available: DOI:10.6084/m9.figshare.13705360.

List of publications

Papers in International Journals (blind peer review)

Published K. Gielis, M.-E. Vanden Abeele, K. Verbert, J. Tournoy, M. De Vos, and V. Vanden Abeele, "Detecting Mild Cognitive Impairment through Digital Biomarkers of Cognitive Performance found in Klondike Solitaire: A Machine Learning Study", *Digital Biomarkers*, Jan. 2021, ISSN: 2504-110X. DOI: 10.1159/000514105

In Review K. Gielis *et al.*, "Dissecting Digital Card Games to Yield Digital Biomarkers for the Assessment of Mild Cognitive Impairment: A Methodological Approach and Exploratory Study", *Journal of Medical Internet Research*, Jan. 2021

Papers in Proceedings of International Conferences (blind peer review)

Published K. Gielis, J. Kennes, C. D. Dobbeleer, S. Puttemans, and V. V. Abeele, "Collecting Digital Biomarkers on Cognitive Health Through Computer Vision and Gameplay: An Image Processing Toolkit for Card Games", in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, Xi'an, China: IEEE, Jun. 2019, pp. 1–12, ISBN: 978-1-5386-9138-0. DOI: 10.1109/ICHI.2019.8904511

Published K. Gielis, K. Verbert, J. Tournoy, and V. Vanden Abeele, "Age? It's in the Game: An Exploratory Study on Detection of Cognitive Aging through Card Games", en, in *Proceedings of the Annual Symposium on Computer-*

Human Interaction in Play, Barcelona Spain: ACM, Oct. 2019, pp. 651–664, ISBN: 978-1-4503-6688-5. DOI: 10.1145/3311350.3347193

Abstracts in Proceedings of International Conferences

K. Gielis, F. Brito, J. Tournoy, and V. Vanden Abeele, "Can Card Games Be Used to Assess Mild Cognitive Impairment?: A Study of Klondike Solitaire and Cognitive Functions", en, in *CHI PLAY '17 Extended Abstracts*, Amsterdam, The Netherlands: ACM Press, 2017, pp. 269–276, ISBN: 978-1-4503-5111-9. DOI: 10.1145/3130859.3131328

K. Gielis, "Screening for Mild Cognitive Impairment through Digital Biomarkers of Cognitive Performance in Games", en, in *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, Barcelona Spain: ACM, Oct. 2019, pp. 7–13, ISBN: 978-1-4503-6871-1. DOI: 10.1145/3341215.3356332. [Online]. Available: https://dl.acm.org/doi/10.1145/3341215.3356332 (visited on 01/26/2021)

K. Gielis, "Assessment of cognitive performance in elderly life via meaningful play", in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, 2019, pp. 1–2. DOI: 10.1109/ICHI.2019.8904861

Position Papers

K. Gielis, F. Brito, J. Tournoy, and V. Vanden Abeele, "Screening Mild Cognitive Impairment in Older Adults via Meaningful Play", en, in *CHI PLAY 2017 Position Paper - Position Paper of the Annual Symposium on Computer-Human Interaction in Play*, Amsterdam, Oct. 2017

K. Gielis and V. Vanden Abeele, *Analyzing Cognitive Functioning via Card Game Play: Presenting a Toolkit and a Case Study*, Forum on Video Games for Mental Health, CHI Play, Melbourne, Oct. 2018

Posters and Presentations

K. Gielis, J. Tournoy, K. Verbert, and V. Vanden Abeele, *Screening for mild cognitive impairment in older adults through card games*, en, Games4MentalHealth, Leuven, Belgium, Feb. 2018. [Online]. Available: `https://www.kuleuven.be/gamesformentalhealth`

K. Gielis, J. Tournoy, K. Verbert, and V. Vanden Abeele, *Screening for Mild Cognitive Impairment in Older Adults through Card Games*, Symposium Veerkrachtig Ouder Worden, Leuven, Belgium, Sep. 2017

K. Gielis, J. Kennes, C. De Dobbeleer, S. Puttemans, and V. Vanden Abeele, *We built an Image Processing Toolkit for Card Games to Capture Digital Biomarkers of Cognitive Performance*, ICHI, Xi'an, 2019

Datasets

K. Gielis, *Dr. Solitaire MCI Dataset*, Feb. 2021. [Online]. Available: `DOI:10.6084/m9.figshare.14057951`

K. Gielis, *Dr. Solitaire Aging Dataset*, Feb. 2021. [Online]. Available: `DOI:10.6084/m9.figshare.14057960`

Source Code

K. Gielis, *Dr. Solitaire Computer Vision App*, Feb. 2021. [Online]. Available: `DOI:10.6084/m9.figshare.13705291`

K. Gielis, *Dr. Solitaire Android App*, Feb. 2021. [Online]. Available: `DOI:10.6084/m9.figshare.13705360`

FACULTY OF ENGINEERING TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE
E-MEDIA RESEARCH LAB
Andreas Vesaliusstraat 13
B-3000 Leuven
https://iiw.kuleuven.be/onderzoek/emedia/