Detecting Mild Cognitive Impairment via Digital Biomarkers of Cognitive Performance Found in Klondike Solitaire: A Machine-Learning Study

Karsten Gielis^a Marie-Elena Vanden Abeele^b Katrien Verbert^c Jos Tournoy^{d, e}
Maarten De Vos^f Vero Vanden Abeele^a

^ae-Media Research Lab, KU Leuven, Leuven, Belgium; ^bMemory Clinic, Jessa Hospital, Hasselt, Belgium; ^cDepartment of Computer Science, KU Leuven, Leuven, Belgium; ^dDepartment of Geriatric Medicine, University Hospital Leuven, Leuven, Belgium; ^eDepartment of Public Healthcare and Primary Care, Gerontology and Geriatrics, KU Leuven, Leuven, Belgium; ^fDepartment of Electrical Engineering, KU Leuven, Leuven, Belgium

Keywords

Mild cognitive impairment · Machine learning · Digital biomarkers · Klondike Solitaire · Cognitive monitoring · Dementia

Abstract

Background: Mild cognitive impairment (MCI) is a condition that entails a slight yet noticeable decline in cognition that exceeds normal age-related changes. Older adults living with MCI have a higher chance of progressing to dementia, which warrants regular cognitive follow-up at memory clinics. However, due to time and resource constraints, this follow-up is conducted at separate moments in time with large intervals in between. Casual games, embedded into the daily life of older adults, may prove to be a less resource-intensive medium that yields continuous and rich data on a patient's cognition. **Objective:** To explore whether digital biomarkers of cognitive performance, found in the casual card game Klondike Solitaire, can be used to train machine-learning models to discern games played by older adults living with MCI from their healthy counterparts. **Methods:** Digital biomarkers of cognitive performance were captured from 23 healthy older adults and 23 older adults living with MCI, each playing 3 games of Solitaire with 3 different deck shuffles.

These 3 deck shuffles were identical for each participant. Using a supervised stratified, 5-fold, cross-validated, machine-learning procedure, 19 different models were trained and optimized for F1 score. **Results:** The 3 best performing models, an Extra Trees model, a Gradient Boosting model, and a Nu-Support Vector Model, had a cross-validated F1 training score on the validation set of ≥ 0.792 . The F1 score and AUC of the test set were, respectively, >0.811 and >0.877 for each of these models. These results indicate psychometric properties comparative to common cognitive screening tests. **Conclusion:** The results suggest that commercial card games, not developed to address specific mental processes, may be used for measuring cognition. The digital biomarkers derived from Klondike Solitaire show promise and may prove useful to fill the current blind spot between consultations.

© 2021 The Author(s)
Published by S. Karger AG, Basel

Introduction

Mild cognitive impairment (MCI) is a condition where ≥ 1 cognitive domains are slightly impaired, but the instrumental activities of daily living are still intact [1, 2]. People with MCI have a higher chance of progressing to a form of

Table 1. Study inclusion and exclusion criteria

Inclusion criteria

A minimum age of 65 years
Living independently or semi-independently in own home, a service flat, or care home
Previous Solitaire experience
Fluent in written and spoken Dutch
No visual or motoric deficits
A stable medical condition

Exclusion criteria healthy group

MMSE <27, MoCA <26, or CDR >0

Exclusion criteria MCI group

Nonamnestic or single-domain MCI
MMSE <23

dementia, and MCI can also signal other neurologic or psychiatric diseases such as vascular disease or depression [1–3]. Therefore, the timely detection of patients with MCI is necessary to provide support and devise a (non)pharmaceutical management approach [1, 4]. While clinically valid, modern cognitive assessment is limited by the mode of administration, which is often a pen, paper, and stopwatch [5]. Such modes of administration require the continuous attention of a trained administrator, limiting the type and amount of data points captured, and making measurements vulnerable to administrator bias and white-coat effect [6, 7]. The consequent lack of accurate high-resolution data can make it difficult to make informed inferences of neuropsychological processes [5]. Cognitive assessment via digital biomarkers of cognitive performance could be an addition to the current cognitive toolset, by contributing to a more complete cognitive profile [5]. Digital biomarkers [8, 9] are user-generated physiological and behavioral measures, captured through connected digital devices, which can provide high-resolution, objective, and quantifiable cognitive data [10].

For MCI, the systems measuring digital biomarkers of cognitive performance can be categorized into 4 groups [10]: systems using dedicated or passive sensors, systems with wearable sensors, nondedicated technological solutions (e.g., software that captures text input), and dedicated or purposive technologies such as games. Games are in a unique position to yield digital biomarkers, as they are autelic in nature, i.e., played for the enjoyment they offer without the need of or request from a third person. Hence, they are intrinsically motivating and do not necessitate an administrator, thereby avoiding the white-coat effect and related biases. Moreover, they can provide different challenges with every playthrough while leaving the fundamental game rules intact [5]. This possibility of supplying novel challenges contrasts with the static prop-

erty of classical cognitive testing, i.e., administering them over a short period of time makes them more prone to learning effects [5].

Whereas previous research into games and cognition focused primarily on games specifically designed for the purpose of measuring 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 [11]. While both serious and COTS games may provide more interactive, immersive, and engaging experiences than traditional cognitive screening [12–14], 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 [14–16]. As such, this study explores whether Klondike Solitaire, an existing popular Solitaire card variant [17], can be used to detect differences in cognitive performance in 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 via machine learning supports the efficacy of COTS games to detect differences in cognitive performance on an individual level.

Materials and Methods

Participants

Participants with MCI were recruited from 2 leading memory clinics in Belgium and all had a clinical diagnosis of multiple-domain amnestic MCI according to Petersen's diagnostic criteria [18]. Healthy participants were recruited using a snowball sample

Table 2. Potential digital biomarkers of cognitive performance in Klondike Solitaire, divided into 5 categories.

Digital biomarker	Description	Aggregation	Data type (range)
<i>Result-based</i>			
Score	Final score of a game	value*	Integer (−∞, +∞)
Solved	Whether the game was completed or not	value*	Boolean
Game time	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, +∞)
<i>Performance-based</i>			
Successful move	Number of successful moves	percentage*	Double (0.00–100.00%)
Erroneous move	Number of erroneous moves	percentage*	Double (0.00–100.00%)
Rank error	Number of rank errors	percentage*	Double (0.00–100.00%)
Suit error	Number of suit errors	percentage*	Double (0.00–100.00%)
King error	Number of kings misplaced	percentage	Double (0.00–100.00%)
Ace error	Number of aces misplaced	percentage	Double (0.00–100.00%)
Pile move	Number of pile moves	percentage*	Double (0.00–100.00%)
Cards moved	Number of cards selected for each move	average*, median, SD	Double (0.00, + ∞)
Beta error	Number of pile moves with moves remaining on the board	percentage*	Double (0.00–100.00%)
King beta error	Number of missed opportunities to place a king on an empty spot	percentage	Double (0.00–100.00%)
Ace beta error	Number 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
<i>Time-based</i>			
Think time	Time spent thinking of a move, expressed in ms	average*, SD*, min*, max, median	Integer (0, +∞)
Think time successful	Time spent thinking of a successful move, expressed in ms	average, median, SD, min, max	Integer (0, +∞)
Think time erroneous	Time spent thinking of an erroneous move, expressed in ms	average, median, SD, min, max	Integer (0, +∞)
Move time	Time spent moving card(s), expressed in ms	average*, SD*, min*, max, median	Integer (0, +∞)
Move time successful	Time spent moving card(s) for a successful move, expressed in ms	average, median, SD, min, max	Integer (0, +∞)
Move time erroneous	Time spent moving card(s) for an erroneous move, expressed in ms	average, median, SD, min, max	Integer (0, +∞)
Total time	Total time to make a move, expressed in ms	average*, SD*, min*, max, median	Integer (0, +∞)
<i>Execution-based</i>			
Accuracy	Accuracy when selecting a card, defined by how close a card was touched to the center	average*, SD*, min*, max*, median	Double (0.00–100.00%)
Taps	Actuations on non-game or UI elements	sum*	Integer (0, +∞)
<i>Auxiliary-based</i>			
Undo move	Amount of undos requested	percentage	Double (0.00–100.00%)
Hint move	Amount of hints requested	percentage	Double (0.00–100.00%)

Remaining features used to train the models appear in bold type. * Remaining features after multicollinearity and zero value checks. SD, standard deviation; min, minimum; max, maximum.

starting from multiple senior citizen organizations. They were screened by using 2 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 [19–21]. The inclusion and exclusion criteria for both groups can be found in Table 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 3 games, resulting in a total of 138 games captured.

Study Overview

This study is part of an overarching study that assesses cognitive performance through meaningful play (ClinicalTrials.gov ID: NCT02971124). Every observation was conducted in the home of the participant between 9 a.m. and 5 p.m 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 [22] under the LGPL 3 license. In this application, cards requested from the pile came in 3s, 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 5-min introduction of the tablet and game was given. In addition, a practice game was played where questions to the researcher were allowed. Afterwards, 3 rounds of Klondike Solitaire, each with a different shuffle, were played in succession. To prevent unfair shuffle (dis)advantages, deck shuffles were identical for all participants for each round. These 3 shuffles were chosen beforehand by the researchers so that they were solvable and varied in difficulty. While playing these 3 rounds, no questions were allowed, and game play continued until the rounds were finished or until the participant indicated that they deemed no further moves were possible.

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 time stamp, touch coordinates, origin card information, destination card information, and the possibility of other moves on the board were logged. These game data were used to calculate the digital biomarkers of cognitive performance (Table 2). These digital biomarkers can be considered 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

Fig. 1. Machine Learning process based on the work of Raschka [23].

informative candidate digital biomarker can be obtained. This contextualization resulted in 61 candidate digital biomarkers of Klondike Solitaire (Table 2) to be classified in 1 of 5 categories: result-based, i.e., biomarkers related to performance at the end of a game; performance-based, i.e., biomarkers related to performance during the game; time-based, i.e., biomarkers related to time; execution-based, i.e., biomarkers related to the physical execution of moves; and auxiliary-based, i.e., biomarkers related to help features.

For model training, a machine-learning procedure was adapted from Raschka [23] (Fig. 1), using scikit-learn [24] 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 and 36 games from 12 participants in the test set). To prevent data leakage due to identity confounding [25], 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 [26]. In total, 26 features remained after selection (in bold type in Table 2). Afterwards, features were scaled using a Standard Scaler. As each algorithm has its inherent biases with none being superior to the rest, 19 classification models were trained, ranging from linear models like logistic regression up to nonlinear models like Gaussian Naïve Bayes [23]. The selection of our models was

Table 3. Demographic and neuropsychological data for both groups

	Demographic information	
	healthy (<i>n</i> = 23)	MCI (<i>n</i> = 23)
Age, years	70 (5.4)	80 (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 (0.65)	26.17 (1.75)
MoCA score	28.09 (1.28)	n.a.
CDR score	0 (0)	n.a.

Values are expressed as mean (SD), unless otherwise indicated. n.a., not available.

¹ Participants were categorized into 3 education groups according to the 1997 International Standard Classification of Education [27]: a. ISCED 1/2, b. ISCED 3/4, c. ISCED 5/6.

² Participants were categorized into 5 proficiency groups based on frequency of use: a. daily, b. weekly, c. monthly, d. yearly or less, e. never.

based on their maturity, popularity, and support available in the scikit-learn machine-learning library. To evaluate them during the training phase, the 5-fold, cross-validated F1 scores were compared. The hyperparameters of the 3 most performant models were further optimized. Ultimately, these 3 best performing models were evaluated on the test dataset.

Results

Study Population

In total, 46 participants (23 MCI and 23 healthy) were enrolled, resulting in 138 rounds of Klondike Solitaire being captured. Demographic and basic neuropsychological data of both groups can be found in Table 3.

Model Performance

The average results of all selected digital biomarkers of cognitive performance across the 3 rounds for both groups can be found in Table 4. The 5-fold, cross-validated F1 validation score of the 19 initial base models was, on average, 0.738. The validation performance metrics of the 3 best fine-tuned 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 with an AUC of 0.877 for the Extra Trees classifier. Confusion matrixes and ROC curves for these 3 models can be found in Figure 2.

Discussion

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

In total, 136 rounds were collected from 46 participants (23 healthy and 23 diagnosed with MCI). Derived digital biomarkers were used to train 19 diverse machine-learning models optimized for the F1 score (the harmonic mean of precision and recall). The choice of optimizing for the F1 score was 2-fold. First, the possible damage of false negatives, as well as false positives, is significant. False negatives, i.e., older adults with MCI being classified as healthy, could postpone diagnosis, thereby leading to longer undetected disease progression. False positives, i.e., healthy older adults being classified as having MCI, could have an equally detrimental impact; a misdiagnosis of cognitive impairment could further spiral depression in a healthy older adult. Second, F1 score is a robust parameter for unbalanced datasets. Should these studies be expanded to real-life settings where MCI and healthy populations are not equal in size, this scoring parameter will likely still be of relevance to other researchers.

After hyperparameter fine-tuning, the 5-fold, cross-validated F1 training score on the validation set was >0.792 for each of the 3 selected models. When evaluated on the test set, each of these models had an F1 score >0.811 and an AUC >0.877. The ROC curves of each model also revealed promising decision thresholds to maximize sensitivity (true positive rate) and specificity (1-false positive rate). It can also be noted that the 3 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) [24]. 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 does not have to

Table 4. Average (SD) performance scores of both groups across all rounds

Candidate biomarker	Healthy	MCI
Result-based		
Score	565.22 (896.92)	−56.3 (1,032.16)
Solved	28 of 69 games solved	10 of 69 games solved
Game time	266,107.33 (100,546.06)	422,283.35 (243,018.32)
Total moves	68.49 (17.45)	72.59 (28.54)
Performance-based		
Successful move percentage	95.37 (4.28)	87.45 (15.86)
Erroneous move percentage	3.65 (3.62)	6.62 (6.7)
Rank error percentage	1.85 (2.34)	4.51 (6.18)
Suit error percentage	2.33 (2.74)	3.59 (4.83)
Pile move percentage	47.36 (16.93)	56.66 (16.34)
Average cards moved	1.29 (0.21)	1.19 (0.2)
Beta error percentage	45.25 (27.83)	57.37 (29.98)
Final beta error	0.13 (0.34)	0.33 (0.47)
Time-based		
Average think time	2,765.71 (734.83)	4,514.78 (1,749.75)
Standard deviation think time	1,999.72 (812.16)	3,544.32 (2,181.62)
Minimum think time	957.04 (223.42)	1,289.55 (573.65)
Average move time	722.16 (169.82)	1,050.45 (426.31)
Standard deviation move time	440.04 (383.42)	943.64 (872.37)
Minimum move time	376.35 (97.09)	458.03 (140.38)
Average total time	3,767.28 (992.82)	5,666.61 (2,221.33)
Standard deviation total time	2,560.54 (1,123.73)	4,191.06 (2,576.13)
Minimum total time	741.12 (234.66)	842.41 (414.25)
Execution-based		
Average accuracy	79.43 (4.73)	74.51 (4.68)
Standard deviation accuracy	9.74 (2.67)	10.68 (2.63)
Minimum accuracy	51.88 (18.47)	49.06 (13.72)
Maximum accuracy	96.07 (2.33)	92.58 (4.48)
Taps	0.77 (1.41)	6.61 (12.84)

hinge 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 MCI from their healthy counterparts, lending support for their use in detecting cognitive decline.

The performance metrics of our models appear to be in line with those of current neuropsychological screening tests. Two of the most common screening tests for discriminating MCI from healthy individuals are the MoCA [19] and the MMSE [20]. In a systematic review by Pinto et al. [28], a mean AUC of 0.883 was found for the MoCA and 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 previous studies using serious games. Valladares-Rodríguez et al. [29] investigated the use of machine-learning models to distinguish between healthy older adults, older adults with MCI, and older adults with Alzheimer's disease. Their serious game set Panoramix consists of 7 games based on 7 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 of 1.00 for MCI, and a specificity of 0.7 for MCI. Direct comparison with this study is, however, problematic, due to the different inclusion criteria, the absence of a hold-out test set, and the ternary classification.

Although this study focused on discerning healthy older adults from older adults with MCI using measurements from a single point in time, our findings may well have a bearing on their use in frequent cognitive monitoring. As pointed out by Piau et al. [10], perhaps the biggest

Fig. 2. Test performance metrics on a per game basis.

shortcoming of today's neuropsychological examination is that it is taken at separate points in time at large intervals. This makes the results vulnerable to temporary alterations of motivation or cognition (e.g., stress or tiredness). As argued by Pavel et al. [5], the general principles of measurement may be extended to psychological processes. By increasing the number of measurements, uncertainty due to imperfections in the tool can be reduced, and natural variations in cognition caused by the characteristics of the phenomenon can be detected. The spatial and temporal richness of data derived from longitudinal game play may allow for a more detailed cognitive profile, and could signal events where cognition has been altered (e.g., the impact of a changed medication regimen or trauma) [5]. In addition, personal cognitive baselines can be created which allow the individual to be compared with themselves as opposed to normative data [9]. These cognitive baselines could be used to detect subtle cognitive fluctuations, an early indicator of cognitive change [5, 10, 30].

There are limitations to this study that should be addressed in future work. In particular, the small sample size prevents us from drawing any absolute conclusions. This might have led to potential bias in the test set, explaining the performance discrepancy between the test and validation set. In addition, discrepancies in age, tablet experience, and Klondike Solitaire experience in the 2 groups may have confounded the 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.

Conclusion

This study set out to investigate the suitability of the card game Klondike Solitaire to detect MCI via machine learning. The major finding 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 the altered cognition 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.

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.

Statement of Ethics

This study was 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 obtaining 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.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

This study received funding from the KU Leuven Impulse Fund IMP/16/025 and the Flemish Government (AI Research Program).

Author Contributions

K.G. cocreated the digital biomarkers, recruited healthy participants, coded the Android app, processed the study data, and codrafted the first manuscript version. J.T. is PI of the clinical study and designed the protocol used. M.-E.V.A. and J.T. recruited participants with MCI and supervised the clinical validity of the study. K.V. and M.D. supervised the technical validity and contributed to the machine-learning pipeline. V.V.A. cocreated the digital biomarkers, codrafted the first version of the manuscript, and supervised the whole study. All authors critically revised and added comments to the manuscript.

References

- 1 Petersen RC, Lopez O, Armstrong MJ, Getchius TS, Ganguli M, Gloss D, et al. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology*. 2018 Jan;90(3):126–35.
- 2 Anderson ND. State of the science on mild cognitive impairment (MCI). *CNS Spectr*. 2019 Feb;24(1):78–87.
- 3 Summers MJ, Saunders NL. Neuropsychological measures predict decline to Alzheimer's dementia from mild cognitive impairment. *Neuropsychology*. 2012 Jul;26(4):498–508.

- 4 Tangelos EG, Petersen RC. Mild Cognitive Impairment in Geriatrics. *Clin Geriatr Med*. 2018 Nov;34(4):563–89.
- 5 Pavel M, Jimison H, Hagler S, McKanna J. Using Behavior Measurement to Estimate Cognitive Function Based on Computational Models. In: Patel VL, Arocha JF, Ancker JS, editors. *Cognitive Informatics in Health and Biomedicine: Understanding and Modeling Health Behaviors*. Cham: Springer International Publishing; 2017. pp. 137–63.
- 6 Overton M, Pihlgård M, Elmståhl S, Walla P. Test administrator effects on cognitive performance in a longitudinal study of ageing. *Cogent Psychol*. 2016 Dec;3(1):1260237.
- 7 Schlemmer M, Desrichard O. Is Medical Environment Detrimental to Memory? A Test of White Coat Effect on Older People's Memory Performance. *Clin Gerontol*. 2018 Jan-Feb;41(1):77–81.
- 8 Torous J, Rodriguez J, Powell A. The New Digital Divide for Digital BioMarkers. *Digit Biomark*. 2017 Sep;1(1):87–91.
- 9 Dorsey ER, Papapetropoulos S, Xiong M, Kiebert K. The First Frontier: Digital Biomarkers for Neurodegenerative Disorders. *Digit Biomark*. 2017 Jul;1(1):6–13.
- 10 Piau A, Wild K, Mattek N, Kaye J. 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. *J Med Internet Res*. 2019 Aug;21(8):e12785.
- 11 Mandryk RL, Birk MV. The Potential of Game-Based Digital Biomarkers for Modeling Mental Health. *JMIR Ment Health*. 2019 Apr;6(4):e13485.
- 12 Valladares-Rodríguez S, Pérez-Rodríguez R, Anido-Rifón L, Fernández-Iglesias M. Trends on the application of serious games to neuropsychological evaluation: A scoping review. *J Biomed Inform*. 2016 Dec;64:296–319.
- 13 Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. *JMIR Serious Games*. 2016 Jul;4(2):e11.
- 14 Toril P, Reales JM, Ballesteros S. Video game training enhances cognition of older adults: a meta-analytic study. *Psychol Aging*. 2014 Sep;29(3):706–16.
- 15 Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. *J Educ Psychol*. 2013;105(2):249–65.
- 16 Baniqued PL, Lee H, Voss MW, Basak C, Cosman JD, Desouza S, et al. Selling points: what cognitive abilities are tapped by casual video games? *Acta Psychol (Amst)*. 2013 Jan;142(1):74–86.
- 17 Boot WR, Moxley JH, Roque NA, Andringa R, Charness N, Czaja SJ, et al. Exploring Older Adults' Video Game Use in the PRISM Computer System. *Innov Aging*. 2018 Apr;2(1):igy009.
- 18 Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangelos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*. 1999 Mar;56(3):303–8.
- 19 Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005 Apr;53(4):695–9.
- 20 Folstein MF, Folstein SE, McHugh PR. "Minimal state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975 Nov;12(3):189–98.
- 21 Morris JC. Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. *Int Psychogeriatr*. 1997;9 Suppl 1:173–6; discussion 177–8.
- 22 Bielefeld T. TobiasBielefeld/Simple-Solitaire: [Android] Simple Solitaire game collection. [cited 2020 Apr 2]. Available from: <https://github.com/TobiasBielefeld/Simple-Solitaire> <https://doi.org/10.1017/S1041610297004870>.
- 23 Raschka S. *Python Machine Learning*. Birmingham Mumbai: Packt Publishing open source; 2016.
- 24 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
- 25 Chaibub Neto E, Pratap A, Perumal TM, Tummalacherla M, Snyder P, Bot BM, et al. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *NPJ Digit Med*. 2019 Oct;2(1):99.
- 26 Rojo D, Htun NN, Verbert K. GaCoVi: A Correlation Visualization to Support Interpretability-Aware Feature Selection for Regression Models. *EuroVis 2020 - Short Papers*. 2020;5 pages.
- 27 UNESCO United Nations Educational, Scientific and Cultural Organization. International Standard Classification of Education, ISCED 1997. In: Hoffmeyer-Zlotnik JH, Wolf C, editors. *Advances in Cross-National Comparison*. Boston (MA): Springer US; 2003. pp. 195–220.
- 28 Pinto TC, Machado L, Bulgacov TM, Rodrigues-Júnior AL, Costa ML, Ximenes RC, et al. 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? *Int Psychogeriatr*. 2019 Apr;31(4):491–504.
- 29 Valladares-Rodríguez S, Anido-Rifón L, Fernández-Iglesias MJ, Facal-Mayo D. A Machine Learning Approach to the Early Diagnosis of Alzheimer's Disease Based on an Ensemble of Classifiers. In: Misra S, Gervasi O, Murgante B, Stankova E, Korkhov V, Torre C, et al., editors. *Computational Science and Its Applications – ICCSA 2019*. Cham: Springer International Publishing; 2019. pp. 383–96.
- 30 Zonderman AB, Dore GA. Risk of dementia after fluctuating mild cognitive impairment: when the yo-yoing stops. *Neurology*. 2014 Jan;82(4):290–1.