Schizophrenia Research: Cognition 27 (2022) 100216

Contents lists available at ScienceDirect

Schizophrenia Research: Cognition

journal homepage: www.elsevier.com/locate/scog

Applying machine learning to smartphone based cognitive and sleep assessments in schizophrenia

Mark Kalinich [a,b,1], Senan Ebrahim [a,c,1], Ryan Hays [d], Jennifer Melcher [d], Aditya Vaidyam [d], John Torous [a,d,*]

[a] Harvard Medical School, Boston, MA, USA
[b] Watershed Informatics, Inc., Boston, MA, USA
[c] Delfina Inc., Boston, MA, USA
[d] Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA, USA

ARTICLE INFO

Keywords:
Machine learning
Schizophrenia
Gamification
Digital psychiatry

ABSTRACT

Background: Cognitive impairment in schizophrenia remains a chief source of functional disability and impairment, despite the potential for effective interventions. This is in part related to a lack of practical and easy to administer screening strategies that can identify and help triage cognitive impairment. This study explores how smartphone-based assessments may help address this need.

Methods: In this study, data was analyzed from 25 subjects with schizophrenia and 30 controls who engaged with a gamified mobile phone version of the Trails-B cognitive assessment in their everyday life over 90 days and complete a clinical neurocognitive testing battery at the beginning and end of the study. Machine learning was applied to the resulting dataset to predict disease status and neurocognitive function and understand which features were most important for accurate prediction.

Results: The generated models predicted disease status with high accuracy using static features alone (AUC = 0.94), with the total number of items collected and the total duration of interaction with the application most predictive. The addition of temporal data statistically significantly improved performance (AUC = 0.95), with the amount of idle time a significant new predictor. Correlates of sleep dysfunction were also predicted (AUC = 0.80), with similar feature importance.

Discussion: Machine learning enabled the highly accurate identification of subjects with schizophrenia versus healthy controls, and the accurate prediction of neurocognitive function. The addition of temporal data significantly improved the performance of these models, underscoring the value of smartphone-based assessments of cognition as a practical tool for assessing cognition.

1. Introduction

Cognitive impairment is a core symptom of schizophrenia that directly impacts a patient's clinical trajectory and quality of life. Patients may experience deficits in attention, working memory, verbal fluency, verbal learning/memory, and executive functioning (Shamsi et al., 2011). While these deficits are often assumed to be stable over the course of the illness, they are malleable and responsive to pharmacological and behavioral therapies. Clinician engagement with assessing cognition in schizophrenia is limited, as even the brief assessment batteries currently in clinical use require specific training and more time to complete than is available in an outpatient visit. This paper explores how smartphone-based assessments of cognition may offer a more practical tool for quickly assessing cognition with a focus on attention and memory.

Current research on cognition in schizophrenia is based largely on retrospective and sparse assessments. The impressive current research base highlights the opportunity for smartphone-based assessments. Retrospective research has identified early disruption of cognition as a predictor of long-term outcomes in schizophrenia that also may help predict clinical trajectories. Yet clinical experience and recent findings suggest that patient trajectories are not linear and that cognition will

* Corresponding author at: Beth Israel Deaconess Medical Center, 330 Brookline Ave., Boston, MA 02215, USA.
E-mail address: jtorous@bidmc.harvard.edu (J. Torous).
[1] These authors contributed equally to this work

https://doi.org/10.1016/j.scog.2021.100216
Received 12 June 2021; Received in revised form 9 September 2021; Accepted 19 September 2021
Available online 1 October 2021

vary across environments and situations. Patients with schizophrenia may not be able to accurately self-report their cognitive status (Potvin et al., 2017; Silberstein and Harvey, 2019; Treichler et al., 2019), and changes in cognition impact patients' ability to report symptoms (Takeuchi et al., 2016) – highlighting the need for objective metrics. Recent studies using technology like EMA/smartphones to capture real time symptom reports from patients have noted that a failure to control for momentary cognition is a weakness of current methods (Blum et al., 2015; Kimhy et al., 2020). Clinically, abrupt changes in both cognition and symptoms around relapse further highlight the need for accurate reporting and early indication of cognitive changes from baseline.

Computerized and tablet assessment of cognition offers a scalable means to assess cognition, and although not providing temporal data, it does provide a mechanism to facilitate temporal sampling. The Brief Assessment of Cognition (BACS) has been validated on a tablet to offer research assessments of verbal memory, working memory, processing speed, and reasoning/problem solving (Atkins et al., 2017). The computer-based CogState Research Battery (CSRB), which offers a test structure that follows MATRICS-recommended cognitive domains, has rendered performance in first episode schizophrenia comparable to traditional paper/pencil testing for working memory, executive functions, and social cognition domains (Benoit et al., 2015). In addition to providing a platform for electronic versions of previously developed neurocognitive assessment, the devices may be used to develop entirely new neurocognitive tests by leveraging captured metadata, such as the timing and location of screen touches. Still, while computerized and tablet-based assessments of cognition increase remote access to cognitive testing, they still lack the scalability of smartphone assessments (Hays et al., 2019) and the ability to provide more momentary assessments necessary for relapse prediction or just-in-time adaptive interventions.

Mobile technology, such as smartphones, offers a novel means to capture cognitive data on a true moment-by-moment basis, with the ability to offer repeated and longitudinal cognitive assessments in vivo. While mobile cognitive testing remains nascent, early studies have suggested that smartphone-based assessments are reliable between-person and within-person, in naturalistic settings (Sliwinski et al., 2018). While most research to date has focused on neurocognitive disorders (Chinner et al., 2018), efforts in psychiatric illness are emerging (Moore et al., 2017). Prior research by this team suggests that time and accuracy to complete a modified Trails-B task on the smartphone can separate controls from subjects with schizophrenia (Liu et al., 2019). Mobile phone keyboard activity patterns have also been correlated with mood state in bipolar disorder (Zulueta et al., 2018) and cognitive functioning in healthy controls (Dagum, 2018). These assessments rely on temporal metadata in that they record how a person types or completes a task. To date, there has not been exploration of how such temporal metadata can augment performance of smartphone cognitive assessments.

This study investigates an assessment of working memory and attention using a smartphone application and whether temporal metadata gathered during the task increases correlations with gold standard clinical tests, as compared to simple static metrics around completion time and accuracy.

2. Methods

2.1. Subject selection and measures

Subjects with clinically diagnosed schizophrenia were recruited along with healthy controls. Patient subjects were recruited from a community mental health center in the metro Boston area, and health control subjects were recruited from Craigslist and local colleges. These subjects were informed of the risks and data use practices and provided consent to the study under BIDMC IRB #2017P00359. Subjects were assessed at two distinct clinical visits (Visit 1 and 2) and instructed to utilize their smartphones with specific applications installed in the intervening period (the Study Period) which was 90 days in duration. Smartphone application data was collected during the Study Period and clinical data from these subjects was collected at Visits 1 and 2. Details of the dataset have been previously published and the sensor data explored, but not the cognitive results (Henson et al., 2020).

Clinical data included the following neuropsychiatric clinical measurements taken by trained staff from consenting subjects at each Visit: Social Functioning Scale (SFS), Patient Health Questionnaire-9 (PHQ-9) for Major Depressive Disorder, General Anxiety Disorder-7 (GAD-7), Behavior and Symptom Identification Scale (BASIS), Warning Signs Scale (WSS), Pittsburgh Sleep Quality Index (PSQI), interpersonal comparison (Comparison), the presence of false beliefs or delusions (Beliefs), Clinical Global Impression (CGI), Positive and Negative Syndrome Scale (PANSS), and Brief Assessment of Cognition in Schizophrenia (BACS). Subscores were also recorded separately for pertinent tests, notably including PSQI subscores of Subjective Quality, Sleep Latency, Sleep Duration, Sleep Efficiency, Sleep Disturbance, Sleep Medication Use, and Daytime Dysfunction. Of particular note is BACS, a neurocognitive test which assesses verbal memory, digit sequencing, motor skills, semantic fluency, letter fluency, symbol coding, and problem solving (through the Tower of London task). BACS was administered electronically, on an iPad. In total, the resulting dataset included application use data from the LAMP ("Learn, Assess, Manage, Prevent") smartphone application (Torous et al., 2019), as well as clinical measurements of neuropsychiatric function including depression, sleep quality, and neurocognitive ability.

2.2. Neurocognitive testing

The LAMP application includes the Jewels Pro game, modeled after the Trails B exam (Fig. 1), a validated instrument for assessing executive functioning. In this team's prior work with 17 patients with schizophrenia, a modified Cox proportional hazard model to demonstrate this cognitive task could separate healthy controls from patients with schizophrenia (Liu et al., 2019) and other teams have applied the task to characterize cognitive deficits in Parkinson's Disease (Weizenbaum et al., 2021).

Subjects were excluded if they did not successfully complete neurocognitive testing in both Visit 1 and Visit 2 or if they used the Jewels Pro game fewer than 5 times. Each use of the Jewels Pro game is subsequently referred to as an encounter. Subjects were able to

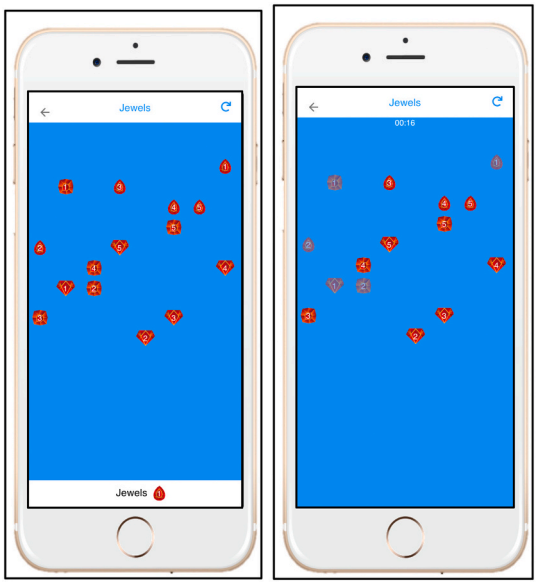

Fig. 1. Jewels Pro game (modified Trails B) within the mindLAMP mobile app.

spontaneously engage in encounters with the Jewels Pro game and with other games at any time of day, without incentives or disincentives from the study. Jewels Pro game encounters were isolated for further analysis.

2.3. Featurization

In order to extract maximum signal from the collected LAMP Jewels Pro encounter data to train highly performing models, the raw data was transformed into two sets of features: (1) static features; and (2) temporal features. Static features were extracted from each encounter that represented the total length of time spent on each encounter, various measurements of the performance of the subject on that encounter, and the number of attempts the subject made during the encounter.

To featurize a subject's performance in time, temporal features were introduced including the mean and variance of the amount of time a subject took to correctly click each jewel, and four measures of resilience. These resilience features were devised to featurize a subject's persistence after getting an answer incorrect in each episode. These four measures were as follows: (1) "corrected fraction": the fraction of incorrect items the subject gets right by the end of the episode; (2) "corrected time": the mean duration of time a subject takes to correct answers for incorrect items; (3) "resilience time": the mean duration of time a subject plays after a wrong item; (4) "resilience items": the mean number of items a subject plays after a wrong item.

2.4. Training models

The scikit-learn RandomForestClassifier was used with 100 estimators using Python v 3.3.8 on sklearn v 0.24.1 with the pre-set parameters. Then, leave-one-out cross-validation (LOOCV) was employed in order to estimate the performance of a model trained on the entirety of this dataset on predicting new data from a previously unseen subject. AUROC (area under the receiver operating characteristic) was used as the primary performance measurement. This process was repeated for 100 iterations, with a random subset of 5 encounters used for each subject for each iteration. Limiting the sample to 5 encounters per subject ensured equally weighted contributions of subjects when training the models. Furthermore, this sampling with replacement approach enabled cross-validated models to be compared across the entire dataset to reduce potential bias. Of the 85 patients with LAMP mobile application data, 55 had 5 or more LAMP encounters (e.g. separate application sessions).

2.5. Model interpretation

To interpret the impact of various static and temporal features on trained random forest models, the SHAP score (SHapley Additive exPlanation) was used. The SHAP score is a game theoretic score that estimates the marginal contribution of adding a feature to prediction performance. Advantages of this approach include greater interpretability by optimizing feature credit allocation and local interaction effects, as compared to model-agnostic methods. In this study, SHAP scores are used for the random forest models to identify which LAMP Jewels Pro user characteristics contribute most significantly to the model's successful prediction of clinical outcomes. For visualization, SHAP scores were plotted for 275 randomly selected episodes that were used to train a representative random forest classifier.

3. Results

3.1. Exploratory data analysis

88 subjects were recruited (45 subjects with schizophrenia, 43 controls) to participate in the study; 55 (25 subjects with schizophrenia, 30 controls) successfully both completed the battery of neurocognitive testing during the in-clinic pre- and post-study visits and had at least five encounters with the Jewels Pro game; these 55 subjects were included in subsequent analyses.

Performing principal component analysis (PCA) on the six static features demonstrated clear separation between encounters originating from subjects with schizophrenia and controls, although this separation is not linear (Fig. 2A). Additionally, 42 'temporal' features were engineered from the raw spatiotemporal data, including the mean and average time to successfully identify each Jewel, the idle time on the application, and the resilience a given subject has for recovering from an incorrect Jewel selection. PCA on the temporal data reveals a less clear separation of the subjects with schizophrenia from controls, potentially due to the increased noise introduced by the added new temporal dimensions (Fig. 1B). PCA was also performed on the gold-standard neurocognitive testing battery at both visits and showed clear separation of subjects with schizophrenia from controls at both visits (S. Fig. 1A–B). Interestingly, subjects with schizophrenia did not separate from controls on a PCA constructed from the change in neurocognitive scores between visit one and visit 2 (S. Fig. 1C).

3.2. Developing and characterizing machine learning classification models

Based on this promising exploratory data analysis, a random forest classifier was built using the approach outlined in Fig. 3 and described in 2.4 Methods to predict whether a given set of encounter data was generated by a subject with schizophrenia or a healthy control.

The models built on this summary data demonstrated excellent performance, with a median AUROC of 0.940 (IQR 0.921, 0.955; Fig. 3B). Next, this approach was applied to the temporal data. The subsequent median AUROC generated from models built from these temporal features was 0.953 (IQR 0.940, 0.969; Fig. 3C) and was found to be significantly different from the performance from the static data alone by Mann-Whitney U test with $p = 4.3E\text{-}16$. (S Fig. 2). Interestingly, models built from the neurocognitive functioning data from in-person visits produced similar AUCs (Visit 1: of 0.891; IQR 0.884-0.896; Visit 2: 0.932; IQR 0.929-0.934), respectively, highlighting the strong performance of mobile phone training data for classifying patients with schizophrenia.

In order to understand the relative contribution of each feature to the predictive power of the models, a single model was trained using the entirety of the data, and the SHAP score calculated for each feature. The SHAP score represents the impact of each feature on model output. The total number of Jewels collected was found to be the most important factor for predicting the identity of both subjects with schizophrenia and the control group, followed by the total number of attempts (Fig. 4A). Similar results were observed in the model trained from the temporal data; total Jewels collected was again the most important single feature; 'durr_diff,' or the difference in duration spent on all items and total episode time (an approximation of idle time with the application open), was found to be second-most important, followed by total attempts.

After the successful identification of subjects with schizophrenia from the control groups, the next step was to predict the results from subject neurocognitive testing for 63 clinical outcomes.. In order to mitigate the increase in false positives from multiple hypothesis testing, the highly conservative Bonferroni correction was used, which takes the unadjusted p value cutoff (0.05) and divides it by the number of tests performed (63), yielding a Bonferroni corrected alpha value of 7.94E-4. After performing the model training workflow from Fig. 3A, 1 of the 63 neurocognitive parameters, the Pittsburgh Sleep Quality Index (PSQI) Daytime Dysfunction met this predefined statistical significance (Fig. 5A). The PSQI is a validated measure of sleep quality, which is further sub-divided into 7 sub-categories, one of which is Daytime Dysfunction. The AUROC was found to be 0.795 (IQR 0.767,0.817). This distribution of AUROC for the trained model differed significantly by T-Test from the random model with $p < 1E\text{-}50$. The SHAP score revealed similar trends to the previous models; total jewels collected and 'durr_diff' were again the first and second most important factors,

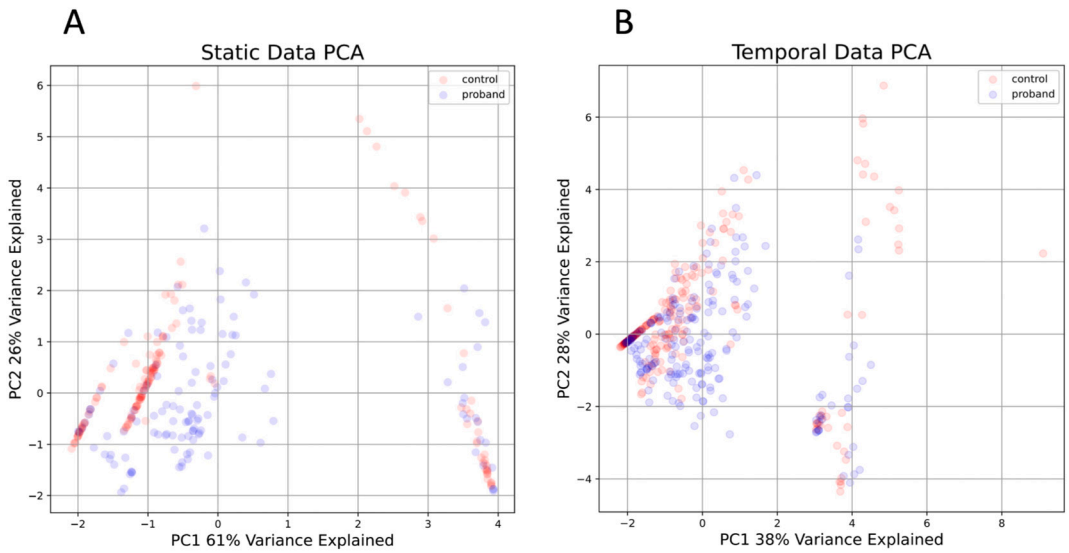

Fig. 2. PCAs for LAMP mobile data. (A) Principal component analysis (PCA) from the static data for each subject, downsampled to 5 encounters per subject, and labeled by disease status (proband = schizophrenia). The 6 variables included were duration, total attempts, total bonus collected, points, total jewels collected, and score. (B) PCA using temporal data for each subject, downsampled to 5 encounters per patient, and labeled by disease status.

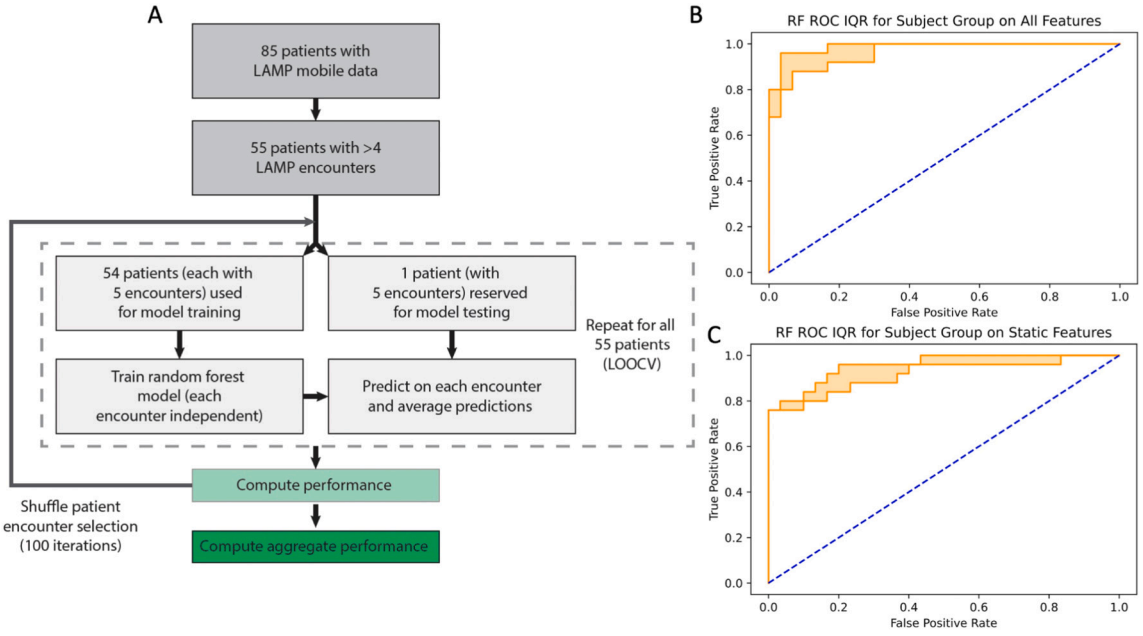

Fig. 3. Building and evaluating machine learning classification models (A). Flowchart for the machine learning approach employed for all prediction tasks. (B). Random forest ROC distribution for models trained on static features alone. The IQR of the ROC distribution is represented as the ROC at the 25th and the ROC at the 75th percentile. The blue line represents random performance (AUROC = 0.5). The corresponding IQR of the AUROCs is: (0.921, 0.955) (C). Random Forest ROC Distribution Training on Static and Temporal Features. The IQR of the ROC distribution is represented as the ROC at the 25th and the ROC at the 75th percentile. The blue line represents random performance (AUROC = 0.5). The corresponding IQR of the AUROCs is: (0.940, 0.969).

respectively, but now the maximum time taken to complete item 5 also served as a strong predictor for the model (Fig. 5B).

The IQR of each ROC distribution is represented as the ROC at the 25th and the ROC at the 75th percentile. The blue line represents random performance (AUROC = 0.5). The corresponding IQR of the Trained AUROCs is: (0.767,0.817). (B) SHAP Scores for All Features Predicting PSQI Daytime Dysfunction. The SHAP score represents the impact of each feature on model output. The actual value of the feature is displayed as an overlaid heatmap.

4. Discussion

This work demonstrates the power of improved data collection strategies coupled with machine learning approaches to generate a highly separable signal between populations with and without schizophrenia as well as to predict dysfunction related to the Pittsburgh Sleep Quality Index (PSQI). Notably, the addition of temporal data statistically significantly increased the performance of the model predictions. Although preliminary, this performance increase suggests that longitudinal monitoring of neurocognitive functioning of patients not only offers the potential for dynamic updates on a patient's condition, but also

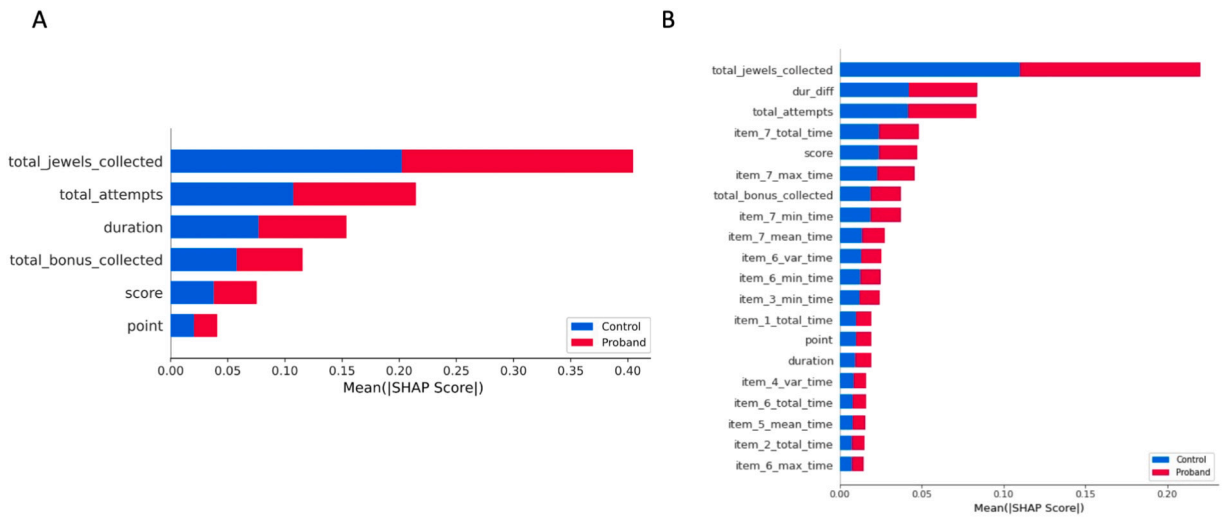

Fig. 4. SHAP scores for each model. (A) Static features. (B) All features.

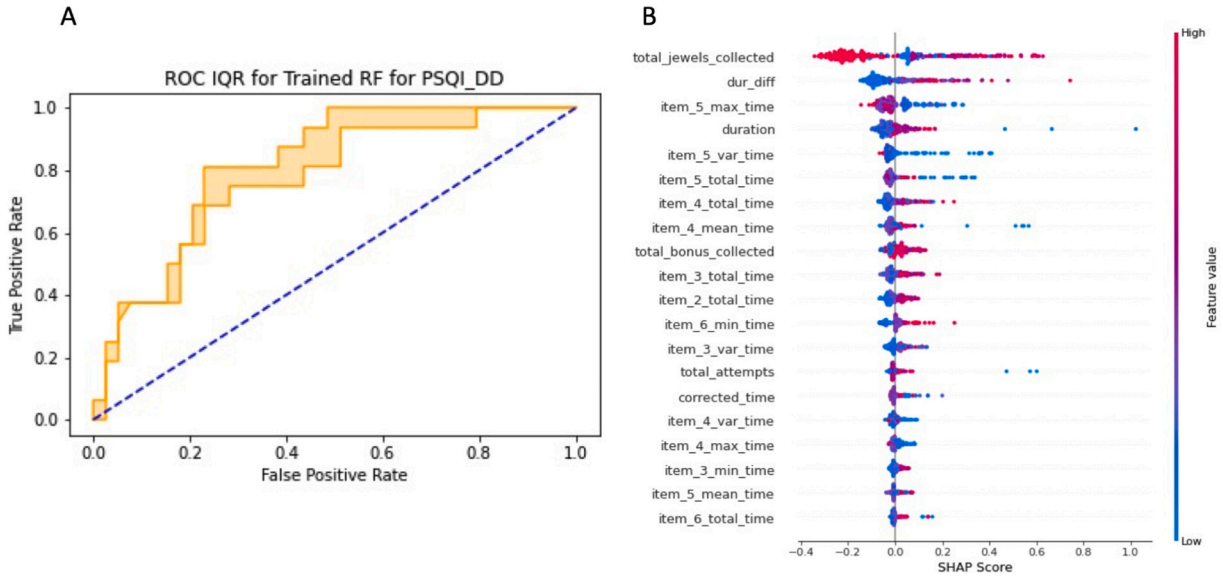

Fig. 5. PSQI daytime dysfunction prediction performance and interpretation. (A) Comparison of ROC Distribution in Random Forest Model for Trained vs. Control.

increases the predictive power at any given moment in time. Additionally, this machine learning workflow demonstrates the accurate prediction of more fine-grained cognitive functioning metrics, such as sleep dysfunction. These results are in line with prior work from this team demonstrating that survival analysis on temporal data on this task of time-series data led to statistically significant, but overlapping, distributions between patients with schizophrenia and healthy controls (Liu et al., 2019).

These new results suggest that remotely monitoring cognition across all serious mental illness is feasible through smartphone-based screening tools. While it is known that early intervention leads to improved outcomes in a variety of psychopathologies (McGorry and Mei, 2018), schizophrenia included (Kane et al., 2016), earlier interventions for cognition remain nascent. Implementing scalable screening tools, as demonstrated here, is the next step towards actualizing timely interventions for individuals experiencing neurocognitive impairment as a result of serious mental illness.

Further, the signal observed for predicting daytime sleep dysfunction, although weaker than that observed for predicting schizophrenia status, is tantalizing. It suggests that this approach, if appropriately refined and analyzed, potentially in parallel with other passive data streams such as geolocation and temporal data, could yield deeper insights into the more subtle neurocognition of subjects, and evaluate these states across time. Further research is required to elucidate the psychiatric dimensions most amenable to the described technology. Nevertheless, these findings outline the strength of leveraging machine learning approaches to mobile application data for the monitoring of psychiatric illness and prove that this approach could scale. The high AUCs for the neurocognitive functioning data with the smartphone scores at both visit 1 and visit 2 suggest that beyond being feasibility, this approach can be explored in the future regarding comparable predictive power.

While these results are encouraging, larger and more variegated patient use data is required to represent the heterogeneity of both baseline mental health and patient cell phone use. This study was limited by the sample size, sparse neurocognitive functioning data separated from the predictor data by up to several weeks, and a control group not age, gender, or SES-matched with the subjects with schizophrenia. This work lays the groundwork for establishing a low-cost, low-barrier longitudinal screening strategy for thought disorders.

To build on the assessment of the smartphone application as a detection tool for serious mental illness, future research should aim to grow the number of subjects in the study and to diversify age, gender, and SES-matched controls. Additionally, the length of monitoring time should be increased to create longitudinal time series, and an opportunity for deep learning approaches capable of leveraging those intra-patient time series. Finally, while this study focused primarily on attention and working memory, recent work has also shown that verbal learning assessment via smartphones for those with serious mental illness, including schizophrenia, is feasible and ecologically valid (Parrish et al., 2021). Future studies should incorporate verbal cognitive assessments adapted to smartphone use.

5. Conclusion

The machine learning models in this analysis were able to distinguish between those with and without schizophrenia at high accuracy, based on cognitive game performance within a smartphone app alone, with improved accuracy when temporal features were added. This work demonstrated how smartphones are uniquely suited to capture dynamic behavioral data which may serve as a biomarker for schizophrenia. The ability of this data to help predict functional symptoms highlights the clinical potential for early intervention in serious mental illness. Monitoring digital biomarkers such as this one present an opportunity to advance research in psychiatry as well as augment care today.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scog.2021.100216.

Role of funding source

MK and SE were supported by NIH grant T32GM007753 and the Harvard Medical Scientist Training Program. MK was additionally supported by F30CA224588. SE was additionally supported by NIH F31NS105161, K24NS088568, T32MH020017, T32GM007753, and the Paul & Daisy Soros Fellowship. JT is supported by NIMH K23MH116130.

Declaration of competing interest

MK and SE are currently affiliated with and have equity in Watershed Informatics, Inc. and Delfina Inc. respectively. This study was completed before the affiliation change, and neither organization has had any influence on the content of this article.

References

Atkins, A.S., Tseng, T., Vaughan, A., Twamley, E.W., Harvey, P., Patterson, T., Narasimhan, M., Keefe, R.S.E., 2017. Validation of the tablet-administered brief assessment of cognition (BAC App). Schizophr. Res. 181, 100–106.

Benoit, A., Malla, A.K., Iyer, S.N., Joober, R., Bherer, L., Lepage, M., 2015. Cognitive deficits characterization using the CogState Research Battery in first-episode psychosis patients. Schizophr. Res. https://doi.org/10.1016/j.scog.2015.03.006.

Blum, L.H., Vakhrusheva, J., Saperstein, A., Khan, S., Chang, R.W., Hansen, M.C., Zemon, V., Kimhy, D., 2015. Depressed mood in individuals with schizophrenia: a comparison of retrospective and real-time measures. Psychiatry Res. 227, 318–323.

Chinner, A., Blane, J., Lancaster, C., Hinds, C., Koychev, I., 2018. Digital technologies for the assessment of cognition: a clinical review. Evid. Based. Ment. Health 21, 67–71.

Dagum, P., 2018. Digital biomarkers of cognitive function. NPJ Digit. Med. https://doi.org/10.1038/s41746-018-0018-4.

Hays, R., Henson, P., Wisniewski, H., Hendel, V., Vaidyam, A., Torous, J., 2019. Assessing cognition outside of the clinic: smartphones and sensors for cognitive assessment across diverse psychiatric disorders. Psychiatr. Clin. N. Am. 42, 611–625.

Henson, P., Barnett, I., Keshavan, M., Torous, J., 2020. Towards clinically actionable digital phenotyping targets in schizophrenia. NPJ Schizophr. 6, 1–7.

Kane, J.M., Robinson, D.G., Schooler, N.R., Mueser, K.T., Penn, D.L., Rosenheck, R.A., Addington, J., Brunette, M.F., Correll, C.U., Estroff, S.E., Marcy, P., Robinson, J., Meyer-Kalos, P.S., Gottlieb, J.D., Glynn, S.M., Lynde, D.W., Pipes, R., Kurian, B.T., Miller, A.L., Azrin, S.T., Goldstein, A.B., Severe, J.B., Lin, H., Sint, K.J., John, M., Heinssen, R.K., 2016. Comprehensive versus usual Community Care for First-Episode Psychosis: 2-year outcomes from the NIMH RAISE early treatment program. Am. J. Psychiatry 173, 362–372.

Kimhy, D., Lister, A., Liu, Y., Vakhrusheva, J., Delespaul, P., Malaspina, D., Ospina, L.H., Mittal, V.A., Gross, J.J., Wang, Y., 2020. The impact of emotion awareness and regulation on psychotic symptoms during daily functioning. NPJ Schizophr. 6, 7.

Liu, G., Henson, P., Keshavan, M., Pekka-Onnela, J., Torous, J., 2019. Assessing the potential of longitudinal smartphone based cognitive assessment in schizophrenia: a naturalistic pilot study. Schizophr. Res. Cogn. 17, 100144.

McGorry, P.D., Mei, C., 2018. Early intervention in youth mental health: progress and future directions. Evid. Based. Ment. Health 21, 182–184.

Moore, R.C., Swendsen, J., Depp, C.A., 2017. Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. Int. J. Methods Psychiatr. Res. 26 https://doi.org/10.1002/mpr.1562.

Parrish, E.M., Kamarsu, S., Harvey, P.D., Pinkham, A., Depp, C.A., Moore, R.C., 2021. Remote ecological momentary testing of learning and memory in adults with serious mental illness. Schizophr. Bull. 47 (3), 740–750.

Potvin, S., Aubin, G., Stip, E., 2017. L'insight neurocognitif dans la schizophrénie. L'Encéphale. https://doi.org/10.1016/j.encep.2016.01.002.

Shamsi, S., Lau, A., Lencz, T., Burdick, K.E., DeRosse, P., Brenner, R., Lindenmayer, J.-P., Malhotra, A.K., 2011. Cognitive and symptomatic predictors of functional disability in schizophrenia. Schizophr. Res. 126, 257–264.

Silberstein, J., Harvey, P.D., 2019. Cognition, social cognition, and self-assessment in schizophrenia: prediction of different elements of everyday functional outcomes. CNS Spectr. 24, 88–93.

Sliwinski, M.J., Mogle, J.A., Hyun, J., Munoz, E., Smyth, J.M., Lipton, R.B., 2018. Reliability and validity of ambulatory cognitive assessments. Assessment 25, 14–30.

Takeuchi, H., Fervaha, G., Remington, G., 2016. Reliability of a patient-reported outcome measure in schizophrenia: results from back-to-back self-ratings. Psychiatry Res. https://doi.org/10.1016/j.psychres.2016.07.055.

Torous, J., Wisniewski, H., Bird, B., Carpenter, E., David, G., Elejalde, E., Keshavan, M., 2019. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. J. Technol. Behav. Sci. 4 (2), 73–85.

Treichler, E.B.H., Thomas, M.L., Bismark, A.W., Hochberger, W.C., Tarasenko, M., Nungaray, J., Cardoso, L., Joshi, Y.B., Zhang, W., Sprock, J., Swerdlow, N., Cohen, A. N., Light, G.A., 2019. Divergence of subjective and performance-based cognitive gains following cognitive training in schizophrenia. Schizophr. Res. 210, 215–220.

Weizenbaum, E.L., Fulford, D., Torous, J., Pinsky, E., Kolachalama, V.B., Cronin-Golomb, A., 2021. Smartphone-based neuropsychological assessment in Parkinson's disease: feasibility, validity, and contextually driven variability in cognition. J. Int. Neuropsychol. Soc. 1–13.

Zulueta, J., Piscitello, A., Rasic, M., Easter, R., Babu, P., Langenecker, S.A., McInnis, M., Ajilore, O., Nelson, P.C., Ryan, K., Leow, A., 2018. Predicting mood disturbance severity with Mobile phone keystroke metadata: a BiAffect digital phenotyping study. J. Med. Internet Res. 20, e241.