scientific reports

OPEN

Development of an eye-tracking system based on a deep learning model to assess executive function in patients with mental illnesses

Minah Kim[1,2,6], Jungha Lee[3,6], Soo Yong Lee[4], Minji Ha[3], Inkyung Park[3], Jiseon Jang[1], Moonyoung Jang[1,2], Sunghyun Park[1] & Jun Soo Kwon[1,2,3,5✉]

Patients with mental illnesses, particularly psychosis and obsessive−compulsive disorder (OCD), frequently exhibit deficits in executive function and visuospatial memory. Traditional assessments, such as the Rey−Osterrieth Complex Figure Test (RCFT), performed in clinical settings require time and effort. This study aimed to develop a deep learning model using the RCFT and based on eye tracking to detect impaired executive function during visuospatial memory encoding in patients with mental illnesses. In 96 patients with first-episode psychosis, 49 with clinical high risk for psychosis, 104 with OCD, and 159 healthy controls, eye movements were recorded during a 3-min RCFT figure memorization task, and organization and immediate recall scores were obtained. These scores, along with the fixation points indicating eye-focused locations in the figure, were used to train a Long Short-Term Memory + Attention model for detecting impaired executive function and visuospatial memory. The model distinguished between normal and impaired executive function, with an F_1 score of 83.5%, and identified visuospatial memory deficits, with an F_1 score of 80.7%, regardless of psychiatric diagnosis. These findings suggest that this eye tracking-based deep learning model can directly and rapidly identify impaired executive function during visuospatial memory encoding, with potential applications in various psychiatric and neurological disorders.

Keywords Eye tracking, Deep learning, Early psychosis, Obsessive−compulsive disorder, Executive function, Visuospatial memory

Psychiatric disorders commonly manifest from underlying brain dysfunctions, often resulting in cognitive deficits across various neuropsychological domains[1]. The Rey−Osterrieth Complex Figure Test (RCFT) has been used as a neuropsychological measure in clinical and research settings to evaluate visuospatial memory and executive function, such as organizational strategy and planning[2]. Notably, impairment of these functions is commonly reported in both patients with psychotic disorders and those with obsessive−compulsive disorder (OCD)[3–8]. Both patient groups often exhibit significantly lower RCFT immediate recall scores and organization scores than healthy controls (HCs), stemming from executive function deficits during visuospatial memory encoding[3,5,7], suggesting that these deficits could serve as transdiagnostic markers across these disorders. These cognitive deficits are also often observed in patients with various psychiatric and neurological disorders, such as bipolar disorder, Alzheimer's disease, Parkinson's disease, closed head injury, autism, epilepsy, encephalitis, multiple sclerosis, and ischemic stroke, who exhibit brain dysfunctions[9–16].

Although the well-established conventional RCFT is a valuable tool for assessing cognitive deficits across various psychiatric and neurological disorders, it has several limitations. The RCFT indirectly assesses executive function through a drawing task, which makes it challenging to link observable behavioral-level phenotypes, such as poor drawing, to underlying brain dysfunction. The RCFT can also be affected by a patient's visuomotor

[1]Department of Neuropsychiatry, Seoul National University Hospital, Seoul, Republic of Korea. [2]Department of Psychiatry, Seoul National University College of Medicine, 101 Daehak-no, Chongno-gu, Seoul 03080, Republic of Korea. [3]Department of Brain and Cognitive Sciences, Seoul National University College of Natural Sciences, Seoul, Republic of Korea. [4]Kim Jaechul Graduate School of Artificial Intelligence, KAIST, Daejeon, Republic of Korea. [5]Institute of Human Behavioral Medicine, SNU-MRC, Seoul, Republic of Korea. [6]These authors contributed equally: Minah Kim and Jungha Lee. ✉email: kwonjs@snu.ac.kr

function[17]; in addition, the administration and scoring systems of the RCFT are time-consuming, labor-intensive, complex, and subject to scoring variability among clinicians owing to human biases[18,19].

Considerable efforts have been made to overcome the limitations of the RCFT scoring system, including the development of an automated scoring system using photos of RCFT drawings and a deep learning algorithm[20-23], the implementation of a tablet-based digital drawing assessment[24], and the adoption of a simpler method for scoring organizational strategies (0 or 1 points)[25]. Although these approaches have made substantial advancements in addressing the complexity, labor intensity, and scoring variability of the scoring system, there are still limitations in its administration given that it is a time-consuming, visuomotor function-affected, indirect drawing test.

To address the remaining limitations, a previous study from our laboratory successfully identified an eye movement biomarker that can be used to rapidly and directly assess impaired organizational strategy during the RCFT in patients with OCD[26]. However, the application of the results of that study to other psychiatric and neurological disorders is limited because the biomarker relies solely on calculations of eye gaze distribution and is based on the assumption that patients with OCD exhibit weak central coherence, focusing on narrow details of the RCFT figure. Thus, there is a need to develop a data-driven deep learning-based assessment model that is not constrained by a single disease characteristic and exhibits increased speed, simplicity, and directness and to extend its applicability to a broader range of psychiatric and neurological disorders. The development of this model would also be consistent with recent interest in assessing cognitive functions such as visual memory and attention in patients with mental illnesses using eye tracking, as eye movements provide real-time insights into the cognitive activities involved in how gaze interacts with visuospatial stimuli during the visual encoding process[27-32].

Therefore, this study aimed to establish an assessment model using eye tracking and deep learning in patients with early psychosis, including those with first-episode psychosis (FEP), patients at clinical high risk (CHR) for psychosis, patients with OCD and HCs. The primary aim was to explore the effectiveness of the model as a rapid, simple, and direct assessment of impaired executive function in these patients. FEP and OCD patients were selected because of their notable impairments in executive function and visuospatial memory[3,5], whereas CHR patients were selected because they included both patients with early psychosis and those with nonspecific psychiatric symptoms, such as depression and anxiety[33]. We hypothesized that the RCFT assessment model developed in this study would be able to distinguish normal and impaired executive function as well as identify visuospatial memory impairment on the basis of eye movements during the memorization of the RCFT figure, regardless of the specific psychiatric diagnosis, with increased speed, simplicity, and directness.

Results

Participant characteristics

The demographic and clinical characteristics of the participants in each diagnostic group are summarized in Table 1. The demographic characteristics of the participants, stratified by normal or impaired executive function and normal or impaired visuospatial memory, are summarized in Table 2. The participants with normal executive function, as measured by the organization T score, and normal visuospatial memory, as measured by the immediate recall T score, had a greater intelligence quotient (IQ) than did the subjects with impaired executive function ($t = 2.801$, $p = 0.013$) and impaired visuospatial memory ($t = 6.832$, $p = <0.001$), respectively. There were more females than males with impaired executive function than with normal executive function ($\chi^2 = 4.620$, $p = 0.032$). Age, years of education, handedness and the proportion of participants wearing glasses were not different between the groups with normal and impaired executive function or visuospatial memory.

RCFT performance and long short-term memory (LSTM) + Attention model results

Analysis of covariance (ANCOVA) with IQ as a covariate revealed that participants with normal executive function had higher organization T scores ($F = 249.031$, $p < 0.001$) than did participants with impaired executive function. ANCOVA using sex and IQ as covariates revealed that the participants with normal executive function had higher total organization scores ($F = 50.382$, $p < 0.001$), fragmentation scores ($F = 4.301$, $p = 0.039$), and planning scores ($F = 102.373$, $p < 0.001$) than did the participants with impaired executive function. Similarly, ANCOVA with IQ as a covariate revealed that participants with normal visuospatial memory had higher immediate recall T scores ($F = 87.960$, $p < 0.001$), total immediate recall scores ($F = 200.415$, $p < 0.001$), configural presence scores ($F = 242.611$, $p < 0.001$), configural accuracy scores ($F = 70.500$, $p < 0.001$), cluster presence scores ($F = 89.203$, $p < 0.001$), cluster accuracy scores ($F = 52.323$, $p < 0.001$), and detail presence scores ($F = 39.270$, $p < 0.001$) than did participants with impaired visuospatial memory. There was no difference in immediate recall time ($F = 2.073$, $p = 0.151$) between participants with normal and impaired visuospatial memory. The ANCOVA results for RCFT performance between HCs and patients with normal and impaired executive function and visuospatial memory are described in Table S1 in the Supplementary material.

Table 3 shows the deep learning performance results of the LSTM + Attention model[34,35] for the two classifications. The model achieved F_1 scores of 83.5 and 80.7% and area under the receiver operating characteristic curve (AUROC) values of 60.7% (Fig. 1a) and 69.9% (Fig. 1b) for distinguishing between normal and impaired executive function and between normal and impaired visuospatial memory, respectively.

Model interpretation results

Eye gaze sequence maps are displayed in Fig. 1c and d. The order of gaze fixation was concentrated in a narrow area and horizontally distributed, with a low number of fixations in the patient with the lowest organization T score (i.e., < 20; Fig. 1c). Conversely, the order of gaze fixation was widely and evenly distributed across the figure in the patient with the highest organization T score (i.e., 70; Fig. 1d). The eye movement comparison results between the normal and impaired groups are provided in Tables 4 and 5.

	FEP (N = 96)	CHR (N = 49)	OCD (N = 104)	HC (N = 159)	Statistical analysis[a] χ² or F	P
Sex (male/female)	38/58	37/12	73/31	82/77	27.866	< 0.001**
Handedness (right/left)	78/3	41/1	97/7	149/8	1.565	0.667
Glasses (yes/no)	41/37	24/18	51/53	61/96	7.025	0.071
Age (years)	24.5 ± 5.0	20.9 ± 3.8	26.2 ± 6.6	23.8 ± 3.5	13.747	< 0.001**
Education (years)	13.4 ± 2.3	13.6 ± 2.0	13.9 ± 2.2	14.9 ± 1.8	8.3	< 0.001**
IQ	100.7 ± 17.4	100.35 ± 11.9	110.9 ± 14.1	117.3 ± 12.9	31.658	< 0.001**
PANSS						
Total	58.2 ± 21.8	–	–	–	–	–
Positive symptoms	13.2 ± 6.4	–	–	–	–	–
Negative symptoms	15.6 ± 6.8	–	–	–	–	–
General symptoms	29.4 ± 10.9	–	–	–	–	–
SIPS						
Positive	–	9.3 ± 3.7	–	–	–	–
Negative	–	12.8 ± 7.0	–	–	–	–
Disorganized	–	4.5 ± 3.4	–	–	–	–
General	–	7.3 ± 4.1	–	–	–	–
Y-BOCS						
Total	1.93 ± 5.8	6.1 ± 9.6	15.5 ± 6.3	–	–	–
Obsession	1.1 ± 3.2	3.6 ± 5.5	7.9 ± 3.4	–	–	–
Compulsion	0.8 ± 2.9	2.4 ± 4.5	7.6 ± 3.6	–	–	–
HAM-D score	5.39 ± 5.3	10.7 ± 7.1	6.4 ± 3.7	–	–	–
HAM-A score	7.0 ± 6.0	12. 8 ± 8.2	5.6 ± 2.9	–	–	–
RCFT						
Organization T score	56.5 ± 10.7	57.6 ± 6.9	59.0 ± 8.7	62.9 ± 7.5	5.8	0.001**
Organization total score	5.5 ± 1.8	6.6 ± 0.9	6.8 ± 1.2	7.3 ± 0.9	13.3	< 0.001**
Fragmentation score	3.5 ± 0.7	3.8 ± 0.4	3.6 ± 0.6	3.7 ± 0.5	1.730	0.161
Planning score	3.0 ± 1.1	2.8 ± 0.9	3.3 ± 1.0	3.6 ± 0.7	5.1	0.002**
Immediate recall time (s)	101.5 ± 46.5	100.4 ± 31.6	118.6 ± 46.3	113.7 ± 47.6	1.573	0.196
Immediate recall T score	53.2 ± 11.0	53.7 ± 9.8	55.7 ± 10.5	62.7 ± 9.8	5.9	0.001**
Immediate recall total score	13.0 ± 3.5	13.1 ± 3.1	13.7 ± 3.4	15.7 ± 2.7	1.413	0.239
Configural presence score	3.5 ± 0.8	3.6 ± 0.9	3.7 ± 0.7	3.9 ± 0.3	0.977	0.404
Configural accuracy score	2.8 ± 1.2	2.8 ± 1.2	2.8 ± 1.0	3.0 ± 0.8	0.212	0.888
Cluster presence score	2.8 ± 0.9	2.8 ± 0.9	3.2 ± 0.9	3.5 ± 0.7	2.429	0.066
Cluster accuracy score	2.1 ± 1.1	2.1 ± 1.0	1.9 ± 1.0	2.5 ± 1.1	0.901	0.441
Detail presence score	1.8 ± 1.2	1.9 ± 1.1	2.2 ± 1.7	2.7 ± 1.1	2.464	0.063

Table 1. Demographic and clinical characteristics of the participants grouped by disease diagnosis and their Rey–Osterrieth complex figure test (RCFT) results. [a]Analysis of variance and χ² analysis or Fisher's exact test for categorical data. The data are presented as the means ± standard deviations. *The mean difference is significant at the 0.05 level. **The mean difference is significant at the 0.005 level. *FEP* first-episode psychosis, *CHR* clinical high risk for psychosis, *OCD* obsessive−compulsive disorder, *HC* healthy control, *PANSS* Positive and Negative Syndrome Scale, *SIPS* Structured Interview for Prodromal Symptoms, *Y-BOCS* Yale−Brown Obsessive Compulsive Scale, *HAM-D* Hamilton Rating Scale for Depression, *HAM-A* Hamilton Rating Scale for Anxiety.

Discussion

This study aimed to develop an eye-tracking and deep learning-based RCFT assessment model for evaluating impaired executive function during visuospatial memory encoding in the RCFT that is faster, simpler, and more direct. The model achieved high performance in assessing impairment in early psychosis and OCD patients on the basis of their sequential eye movements while they were memorizing the RCFT figure. This assessment is performed regardless of specific psychiatric diagnoses, as this impairment is shared across these disorders. These results indicate that eye movements during the encoding of highly complex figures reflect executive function during visuospatial memory encoding, serving as a transdiagnostic biomarker of impairment in early psychosis and OCD. Additionally, since the assessment model in this study utilized a data-driven deep learning technique that does not require handcrafted feature selection according to specific disease hypotheses, there is the potential to extend the use of this model beyond early psychosis and OCD to other psychiatric and neurological disorders that also exhibit impaired executive function, poor performance in the RCFT in the form of disorganized and fragmented drawings, and difficulties in visuospatial integration[9–16].

Grouped by organization T score	Normal (N = 385)	Impaired (N = 23)	Statistical analysis[a]	
			χ² or T or F	P
Sex (male/female)	222/163	8/15	4.620	0.032*
Handedness (right/left)	346/17	19/2	0.989	0.279
Glasses (yes/no)	173/201	9/13	0.202	0.653
Age (years)	24.2 ± 5.1	24.1 ± 4.4	0.107	0.915
Education (years)	14.3 ± 2.1	13.4 ± 2.4	1.445	0.149
IQ	111.0 ± 14.8	93.3 ± 25.2	2.801	0.013*
Organization T score	60.5 ± 7.7	27.9 ± 6.9	249.031	< 0.001**
Organization total score	6.8 ± 1.3	4.3 ± 1.8	50.382	< 0.001**
Fragmentation score	3.6 ± 0.5	3.3 ± 0.5	4.301	0.039*
Planning score	3.4 ± 0.8	1.1 ± 1.5	102.373	< 0.001**
Diagnostic group[b]				
FEP	80 (20.8)	16 (69.6)	–	–
CHR	47 (12.2)	2 (8.7)	–	–
OCD	101 (26.2)	3 (13.0)	–	–
HC	157 (40.8)	2 (8.7)	–	–
Grouped by immediate recall T score	Normal (N = 381)	Impaired (N = 27)		
Sex (male/female)	217/164	13/14	0.795	0.373
Handedness (right/left)	340/17	25/2	0.374	0.634
Glasses (yes/no)	174/205	10/15	0.330	0.565
Age (years)	24.1 ± 5.1	25.5 ± 5.2	− 1.398	0.163
Education (years)	14.3 ± 2.1	13.8 ± 1.9	0.920	0.358
IQ	111.6 ± 14.6	89.7 ± 19.0	6.832	< 0.001**
Immediate recall time (s)	108.1 ± 45.2	95.3 ± 54.5	2.073	0.151
Immediate recall T score	58.9 ± 9.5	32.9 ± 10.4	87.960	< 0.001**
Immediate recall total score	14.8 ± 2.6	5.9 ± 2.0	200.415	< 0.001**
Configural presence score	3.9 ± 0.4	2.1 ± 0.8	242.611	< 0.001**
Configural accuracy score	3.0 ± 0.8	1.3 ± 1.2	70.500	< 0.001**
Cluster presence score	3.3 ± 0.8	1.5 ± 0.7	89.203	< 0.001**
Cluster accuracy score	2.3 ± 1.0	0.4 ± 0.8	52.323	< 0.001**
Detail presence score	2.4 ± 1.1	0.6 ± 0.5	39.270	< 0.001**
Diagnostic group[b]				
FEP	81 (21.3)	15 (55.6)	–	–
CHR	46 (12.1)	3 (11.1)	–	–
OCD	96 (25.2)	8 (29.6)	–	–
HC	158 (41.5)	1 (3.70)	–	–

Table 2. Demographic characteristics and Rey–Osterrieth complex figure test (RCFT) results compared between participants with normal and impaired executive function, as measured by the organization T score, and between participants with normal and impaired visuospatial memory, as measured by the immediate recall T score. [a]Independent t test or Welch's t test if the variances were not equal, analysis of covariance with IQ and/or sex as covariates, and χ² analysis or Fisher's exact test for categorical data. [b]Number (percentage) of patients who were diagnosed with each psychiatric disorder and healthy controls. The data are presented as the means ± standard deviations. *The mean difference is significant at the 0.05 level. **The mean difference is significant at the 0.005 level. *IQ* intelligence quotient, *FEP* first-episode psychosis, *CHR* clinical high risk for psychosis, *OCD* obsessive−compulsive disorder, *HC* healthy control.

Input data-model-trained with	Recall (sensitivity)	Precision	AUROC score	F_1 score
Fixation-LSTM-Organization T score	0.7860	0.9036	0.6073	0.8351
Fixation-LSTM-Immediate recall T score	0.7492	0.8987	0.6997	0.8069

Table 3. LongShort Term Memory (LSTM) + Attention model classification results of patients with normal and impaired executive function (organization T score) and patients with normal and impaired visuospatial memory (immediate recall T score). *AUROC* area under the receiver operating characteristic curve.
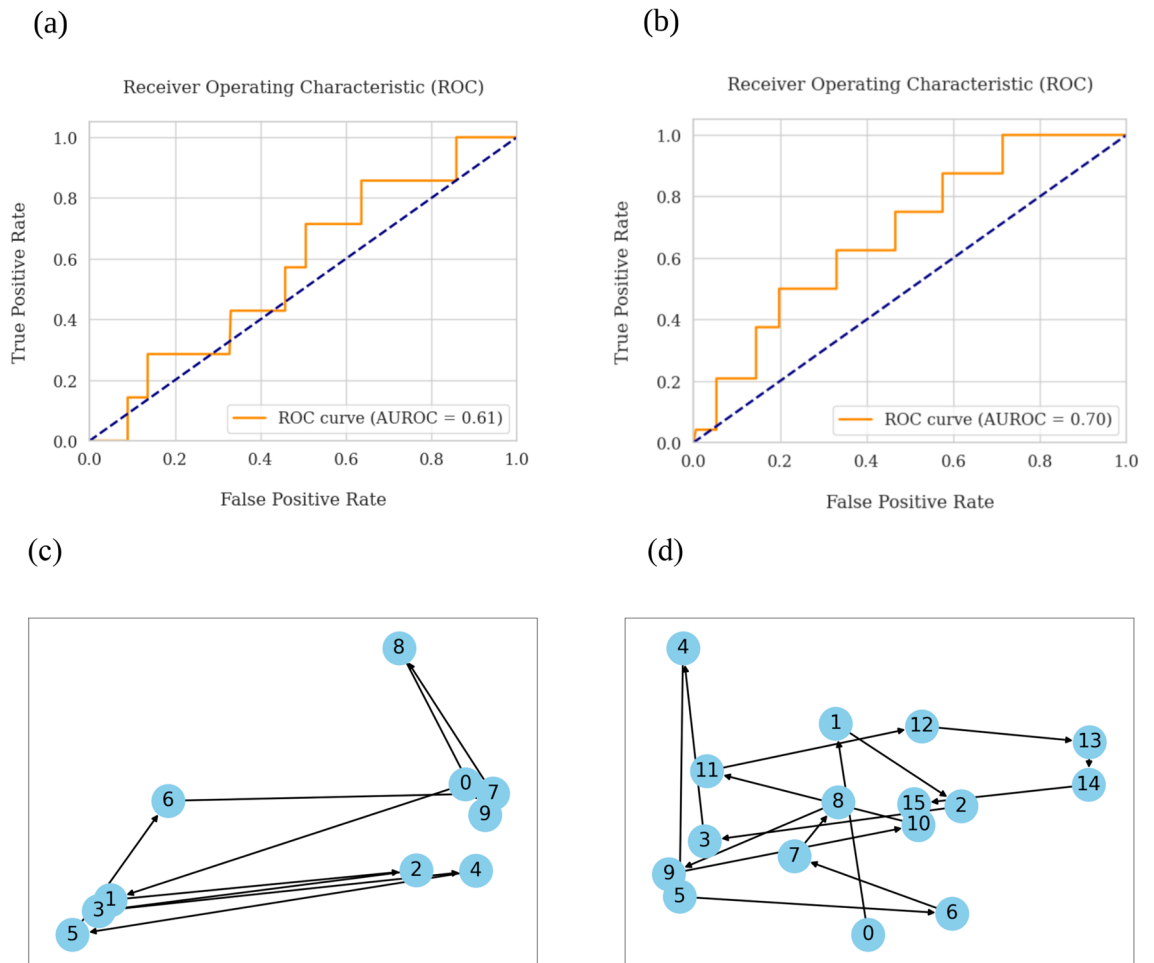
Figure 1. Gaze fixation sequence map showing the order of eye movements during the 3-min memorization of the Rey−Osterrieth Complex Figure Test (RCFT) figure. (**a**) Receiver operating characteristic (ROC) curve of the Long Short-Term Memory (LSTM) + Attention model classification for normal and impaired executive function. (**b**) ROC curve of the LSTM + Attention model classification for normal and impaired visuospatial memory. (**c**) The order of gaze fixation in the patient with the lowest organization T score (i.e., < 20). (**d**) The order of gaze fixation in the patient with the highest organization T score (i.e., 70). The numbers within the circles indicate the order of gaze fixation, starting from 0. Abbreviation: AUROC, area under the ROC curve.

	Normal[a]	Impaired[b]	Statistical analysis[c]	
	(N = 385)	(N = 23)	T	P
Number of fixations	486.3 ± 106.3	416.1 ± 107.1	3.078	0.002**
Average duration of fixations (ms)	311.9 ± 79.6	346.0 ± 115.7	− 1.395	0.176
Average saccade amplitude	3.4 ± 0.9	3.4 ± 0.9	0.203	0.839
Duration (ms)	191,026.0 ± 13,375.5	184,777.0 ± 5122.8	4.931	< 0.001**
Number of blinks	97.9 ± 62.7	95.87 ± 49.1	0.151	0.880
Number of saccades	485.7 ± 106.2	415.5 ± 107.1	3.077	0.002**

Table 4. Eye movement characteristics of participants with normal and impaired executive function measured by the organization T score. [a]Normal executive function, organization T score 40–70. [b]Impaired executive function, organization T score < 39. [c]Independent t test or Welch's t test if the variances were not equal. The data are presented as the means ± standard deviations. *The mean difference is significant at the 0.05 level. **The mean difference is significant at the 0.005 level.

	Normal[a]	Impaired[b]	Statistical analysis[c]	
	(N = 381)	(N = 27)	T	P
Number of fixations	486.0 ± 106.7	431.9 ± 106.6	2.546	0.011*
Average duration of fixations (ms)	314.1 ± 81.6	309.8 ± 92.0	0.258	0.797
Average saccade amplitude	3.4 ± 0.9	3.4 ± 0.9	0.049	0.961
Duration (ms)	191,024.8 ± 13,353.3	185,719.6 ± 7980.7	2.037	0.042*
Number of blinks	97.3 ± 62.5	104.6 ± 54.8	− 0.589	0.556
Number of saccades	485.3 ± 106.7	431.3 ± 106.6	2.541	0.011*

Table 5. Eye movement characteristics of the participants with normal and impaired visuospatial memory measured by immediate-recall T scores. [a]Normal visuospatial memory, immediate recall T score 40–70. [b]Impaired visuospatial memory, immediate recall T score < 39. [c]Independent t test or Welch's t test if the variances were not equal. The data are presented as the means ± standard deviations. *The mean difference is significant at the 0.05 level.

The model also offered a rapid and simple measure of impaired executive function, taking only 3 min from administration to assessment through computerized eye movement recording and deep learning analysis. This process is considerably faster and simpler than the traditional RCFT scoring system and the automated scoring systems reported in previous studies, which still require a prior drawing process[20–23]. This improved model is beneficial and easy to apply in real-world clinical and research settings, saving a significant amount of labor and time and reducing human scoring variability. Moreover, the model enabled a more direct assessment, as sequential eye movements reflected real-time visuospatial information processing[36] and indicated how subjects strategically encoded, planned, and organized the figure. Thus, eye-tracking measurements can bridge the gap between the behavioral-level phenotype and brain dysfunction by capturing the inner workings of executive function during visuospatial memory encoding.

In the interpretation of the results of the deep learning model, the gaze fixation sequence maps in Fig. 1c and d show distinct eye movement patterns between patients with the lowest and highest organization scores. In the patient with the lowest organization score, exploration of the figure was limited, and the patient lacked a strategy and focused simply on horizontal movements without attending to important areas. In contrast, in the patient with the highest organization score, exploration was more structured and extensive, capturing a larger picture and focusing sequentially on various critical areas. In addition, quantitative differences in eye movements revealed that participants with impaired executive function and visuospatial memory spent less time looking at the figure and exhibited fewer fixations and saccades than the participants in the normal group did, indicating less effective and comprehensive encoding of the visuospatial information within the figure (Tables 4 and 5). These ineffective, disorganized, and limited eye movement patterns in the impaired group seem to be consistent with previous research findings that patients with executive function deficits have difficulties processing the overall RCFT figure and utilize a fragmented and piecemeal approach[3,5,7]. Overall, these differences in eye movement patterns between the impaired and normal groups may have contributed to the ability of the deep learning model to distinguish between them.

There are several limitations in this study. First, this assessment model was initially developed using data from individuals with early psychosis and OCD and classified their functions into only normal and impaired, limiting its applicability to individuals with other psychiatric and neurological disorders. Additionally, the impairment criteria in this study were stringent, making it challenging to detect patients with mild impairment. Nonetheless, this model has the potential to expand beyond binary classification and include a wider range of psychiatric and neurological disorders. Second, the small number of participants in the impaired group resulted in highly imbalanced datasets, which limits the validity and reliability of the assessment model in this study. To address this, the split dataset was stratified to maintain class label proportions consistent with those of the original dataset, and synthetic minority oversampling technique (SMOTE)[37] data augmentation was implemented, as in previous studies with similarly imbalanced datasets[38–40]. However, our results should be interpreted with caution because of the relatively small and imbalanced sample size of the original dataset. Third, most patients were taking medication at the time of the eye-tracking RCFT. Therefore, it is necessary to consider the medication effect when interpreting the study results, as this study did not investigate the impact of medication on patients' RCFT performance or eye movement markers. However, given that the assessment model aims to encompass various psychiatric and neurological disorders in future research, these findings remain promising, as they indicate effectiveness of the model even in the presence of potential influences from medication. Fourth, there was a significant difference in IQ between the normal and impaired groups. Although the statistical group comparison was conducted with IQ as a covariate, the deep learning model, LSTM + Attention, does not account for or exclude the potential impact of cognitive function on eye movement markers in its classification.

Although the RCFT is a well-established tool for evaluating executive function during visuospatial memory encoding, its administration and scoring pose difficulties because of its time-consuming nature, indirect measurement, and scoring variability. While a previous study from our laboratory identified an eye movement biomarker to detect impaired executive function with enhanced speed and directness, its utility was limited to OCD because it was based on OCD-specific characteristics. Therefore, we developed an RCFT assessment model that integrated eye tracking and deep learning, which not only offered a more direct, rapid, and simplified evaluation

of impaired executive function but also demonstrated the potential for wider applicability to other disorders, as it was data driven and did not rely on singular disease hypotheses. Future studies could benefit from including various psychiatric and neurological disorders and utilizing explainable artificial intelligence to identify key features distinguishing between individuals with normal and impaired executive function during visuospatial memory encoding.

Methods

Participants

We analyzed data from 408 participants, including 96 patients with FEP, 49 patients at CHR for psychosis, 104 patients with OCD, and 159 HCs. FEP patients and CHR individuals were recruited from both the inpatient and outpatient clinics of the Department of Neuropsychiatry and the Seoul Youth Clinic (www.youthclinic.org) at Seoul National University Hospital (SNUH). In this study, the FEP patients included individuals who were diagnosed with schizophrenia, schizoaffective disorder, or schizophreniform disorder according to the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Axis I Disorders (SCID-I) with an illness duration of less than 2 years. Psychotic symptoms were assessed using the Positive and Negative Syndrome Scale (PANSS). To confirm the CHR status of the participants, the Structured Interview for Prodromal Symptoms (SIPS)[41] was used. Prodromal symptoms were assessed using the validated Korean version of the Scale of Prodromal Symptoms (SOPS)[41,42]. Patients with OCD were recruited from the OCD clinic (www.ocd.snu.ac.kr) at SNUH and fulfilled the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) criteria for OCD. The Yale−Brown Obsessive Compulsive Scale (Y-BOCS)[43] was used to evaluate the severity of OCD. The Hamilton Rating Scale for Depression (HAM-D)[44] and the Hamilton Rating Scale for Anxiety (HAM-A)[45] were used to evaluate the severity of depressive and anxious symptoms, respectively. The HCs were recruited using internet advertisements. To screen for the presence of psychiatric disorders or symptoms, the HCs underwent an assessment employing the Structured Clinical Interview for DSM-IV Non-Patient Version (SCID-NP). HCs with a past or current axis I diagnosis or first- to third-degree biological relatives with a history of psychotic disorders were not eligible to participate in this study. We assessed participants' IQ using the Korean version of the Wechsler Adult Intelligence Scale (K-WAIS)[46].

All participants were assessed according to the exclusion criteria, which included the presence of neurological conditions, significant head injuries, substance abuse or dependency (except for nicotine), and intellectual disability (IQ < 70). We provided thorough explanations of the research procedures and obtained written informed consent from all participants (IRB no. H-1110-009-380, H-1201-008-392). For participants younger than 18, consent was also obtained from their parents. This study adhered to the principles outlined in the Declaration of Helsinki (2013) and received approval from the Institutional Review Board of SNUH (IRB No. H-2306-210-1445).

Eye movement data acquisition and the RCFT

Eye movement data were obtained during an eye-tracking experiment in which participants were instructed to view and memorize the RCFT figure for a duration of 3 min (Fig. 2), as detailed by Kim et al.[26]. Briefly, the
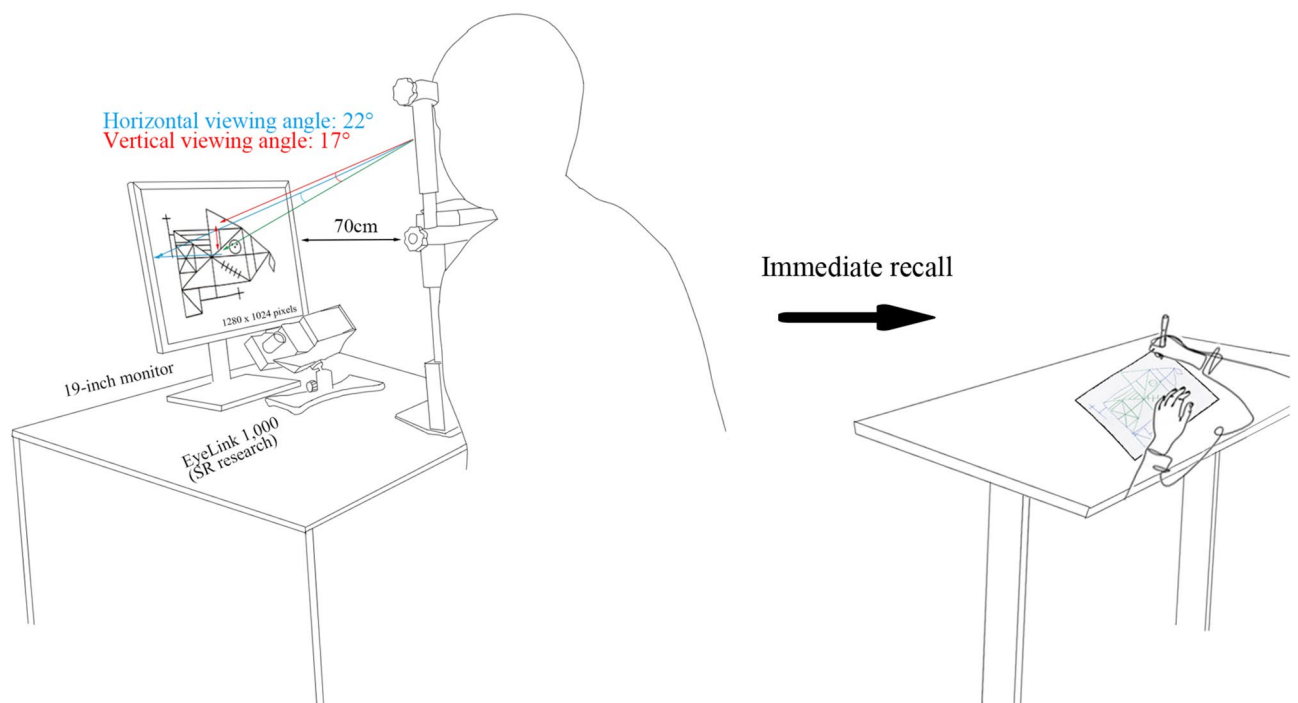

Figure 2. Eye tracking-based Rey−Osterrieth Complex Figure Test (RCFT) procedure used in this study. Memorization of an RCFT figure for 3 min was followed by immediate recall of the figure.

RCFT figure was presented on a 19-inch monitor with a screen resolution of 1280 × 1024 pixels using Experiment Builder v.2.1.45 software (SR Research, Ottawa, Ontario, Canada). During the experiment, the participant's head was positioned on a chin rest in a room with low lighting. The distance between the chin rest and the monitor was 70 cm, and the participant had a horizontal viewing angle of 22° and a vertical viewing angle of 17°. Before eye movements were measured, a nine-point calibration and verification process was conducted. The data were collected at a 1,000-Hz sampling rate and exported through the EyeLink 1000 (SR Research) eye tracking device.

The key parameters collected were gaze fixation point coordinates, indicating where the eyes briefly paused to focus and acquire new information[47], and time in milliseconds. Upon the completion of the eye-tracking session, the participants were instructed to reproduce the RCFT figure from memory, akin to the immediate recall condition of the RCFT. During this drawing task, response times were recorded, and an experimenter systematically tracked the participant's reproduction of the figure. This meticulous monitoring aimed to assess organizational strategies as a substitution for the RCFT copy condition. A skilled evaluator manually assessed the participants' drawings using the Boston Qualitative Scoring System (BQSS)[48]. This assessment aimed to evaluate the participants' organizational and immediate recall performance in the RCFT. The organization and immediate recall scores were subsequently categorized according to the BQSS clinical interpretation criteria as normal (score: 40–70) or impaired (score: < 39). The participants were grouped as follows: a normal executive function group with normal organization scores, a normal visuospatial memory group with normal immediate recall scores, an impaired executive function group with impaired organization scores and an impaired visuospatial memory group with impaired immediate recall scores.

LSTM + attention model

In this study, we utilized an LSTM model using Python and PyTorch[49] to analyze time series eye movement fixations during the memorization of the RCFT figure. The effectiveness of the LSTM model in handling sequential data was a key factor in its selection, especially since sequential temporal relationships might play a significant role in participants' effective memorization and organization of the RCFT figure. Additionally, the LSTM model was combined with an attention mechanism. Time series fixation sequences were input recursively into the model, facilitating the learning of patterns and relationships within sequential eye movement fixations. The acquired representations at each timestamp were summed by attention coefficients to obtain the final sequence representation. The resulting representations were fed into a single-layer classifier to determine the probabilities of sequences belonging to a specific class (normal or impaired). The model was trained with a sequence size of 32 fixation points, and the learning rate was set to $\alpha = 0.005$ over the training course. The fixation dataset was split 70/30 into training and testing sets, and the split was stratified to preserve class label proportions similar to those of the original dataset. The evaluation metrics used in the LSTM + Attention model included recall (sensitivity), precision, AUROC, and F_1 score. The AUROC and F_1 score were utilized to determine the accuracy of the model in highly imbalanced datasets, as in this study. The overall modeling workflow is described in Fig. 3.

Data augmentation

In this study, the dataset was highly imbalanced, with a significant disparity between the majority class (e.g., 385 participants in the normal group) and the minority class (e.g., 23 participants in the impaired group). Imbalanced datasets cause problems for learning algorithms that expect an even distribution across classes, leading to bias favoring the majority class[50]. To address this, data augmentation is commonly employed to achieve an ideal balance, e.g., a 50:50 ratio, by artificially expanding the training dataset for enhanced reliability. Thus, the minority class (impaired group) in our training datasets was oversampled using SMOTE. The imbalance was maintained in the test dataset to represent the real-world distribution.
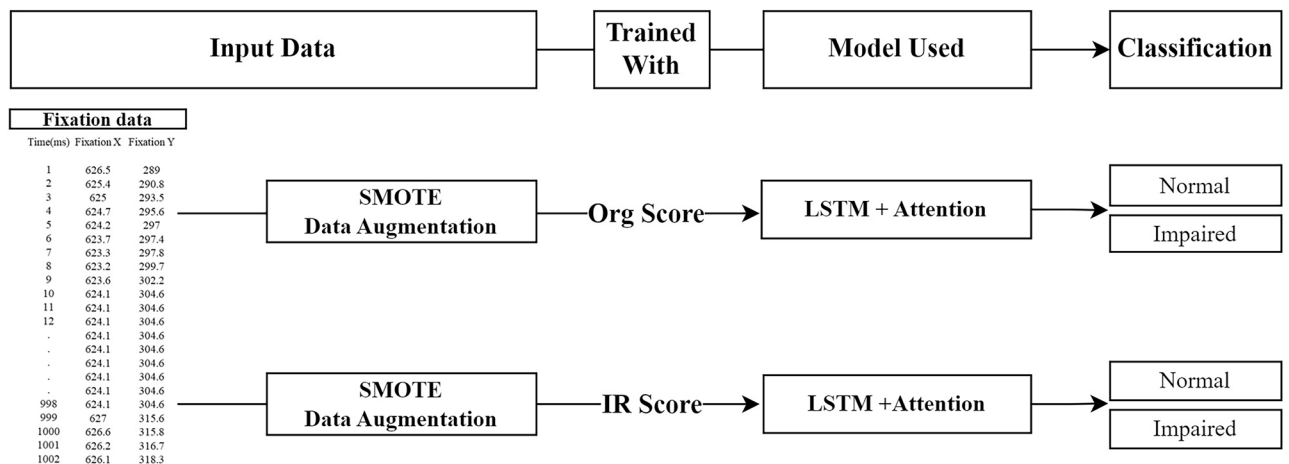

Figure 3. Overall workflow of modeling in this study. Abbreviations: Org, organization; IR, immediate recall; SMOTE, synthetic minority oversampling technique.

Statistical analysis

SPSS v.26.0 (IBM, Armonk, NY, USA) was used for the statistical analyses, and the significance level was set at $p < 0.05$. Comparisons of demographic and clinical characteristics across groups were performed using independent t tests or Welch's t tests if the variances were not equal for continuous variables and chi-square tests for categorical variables. Group comparisons of RCFT scores were performed using ANCOVA with IQ or IQ and sex as covariates.

Model interpretation

Owing to the limited interpretability of the decision-making process of the LSTM + Attention model, additional visual interpretations and statistical analyses were conducted to interpret the results of the model in this study. First, a gaze fixation sequence map was created to explore participants' visuospatial information processing and organization of the RCFT figure in sequence. Second, eye movement measures, including the number of fixations, average duration of fixation (ms), average saccade amplitude and duration (ms), and number of blinks and saccades, were compared between the normal and impaired groups to identify quantitative differences in eye movements. Saccades refer to rapid eye movements between fixations.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Keefe, R. S. The contribution of neuropsychology to psychiatry. *Am. J. Psychiatry* **152**, 6–15 (1995).
2. Rey, A. L'examen psychologique dans les cas d'encephalopathie traumatique. *Arch. Psychol.* **28**, 215–285 (1941).
3. Kim, M. S., Namgoong, Y. & Youn, T. Effect of organizational strategy on visual memory in patients with schizophrenia. *Psychiatry Clin. Neurosci.* **62**, 427–434 (2008).
4. Penadés, R., Catalán, R., Andrés, S., Salamero, M. & Gastó, C. Executive function and nonverbal memory in obsessive-compulsive disorder. *Psychiatry Res.* **133**, 81–90 (2005).
5. Savage, C. R. *et al.* Organizational strategies mediate nonverbal memory impairment in obsessive–compulsive disorder. *Biol. Psychiatry* **45**, 905–916 (1999).
6. Scala, S. *et al.* Verbal and visual–spatial memory impairment in youth at familial risk for schizophrenia or affective psychosis: A pilot study. *Schizophr. Res.* **144**, 122–128 (2013).
7. Seidman, L. J., Lanca, M., Kremen, W. S., Faraone, S. V. & Tsuang, M. T. Organizational and visual memory deficits in schizophrenia and bipolar psychoses using the Rey-Osterrieth complex figure: Effects of duration of illness. *J. Clin. Exp. Neuropsychol.* **25**, 949–964 (2003).
8. Shin, N. Y., Lee, T. Y., Kim, E. & Kwon, J. S. Cognitive functioning in obsessive-compulsive disorder: A meta-analysis. *Psychol. Med.* **44**, 1121–1130 (2014).
9. Bigler, E. D., Rosa, L., Schultz, F., Hall, S. & Harris, J. Rey-Auditory verbal learning and Rey-Osterrieth complex figure design performance in Alzheimer's disease and closed head injury. *J. Clin. Psychol.* **45**, 277–280 (1989).
10. Cardillo, R., Lievore, R. & Mammarella, I. C. Do children with and without autism spectrum disorder use different visuospatial processing skills to perform the Rey-Osterrieth complex figure test?. *Autism Res.* **15**, 1311–1323 (2022).
11. Correa-Ghisays, P. *et al.* Visual memory dysfunction as a neurocognitive endophenotype in bipolar disorder patients and their unaffected relatives. Evidence from a 5-year follow-up Valencia study. *J. Affect. Disord.* **257**, 31–37 (2019).
12. Dimitrov, I. *et al.* Application of the Rey-Osterrieth complex figure test for assessment of cognitive impairment in multiple sclerosis. *Scr. Sci. Med.* **47**, 59–64 (2015).
13. Finke, C. *et al.* Evaluation of cognitive deficits and structural hippocampal damage in encephalitis with leucine-rich, glioma-inactivated 1 antibodies. *JAMA Neurol.* **74**, 50–59 (2017).
14. Grossman, M. *et al.* Visual construction impairments in Parkinson's disease. *Neuropsychology* **7**, 536–547 (1993).
15. Hernandez, M. T. *et al.* Attention, memory, and behavioral adjustment in children with frontal lobe epilepsy. *Epilepsy Behav.* **4**, 522–536 (2003).
16. Schaapsmeerders, P. *et al.* Long-term cognitive impairment after first-ever ischemic stroke in young adults. *Stroke* **44**, 1621–1628 (2013).
17. VanGilder, J. L., Lohse, K. R., Duff, K., Wang, P. & Schaefer, S. Y. Evidence for associations between Rey-Osterrieth complex figure test and motor skill learning in older adults. *Acta Psychol.* **214**, 103261 (2021).
18. Fastenau, P. S., Bennett, J. M. & Denburg, N. L. Application of psychometric standards to scoring system evaluation: Is "new" necessarily "improved"?. *J. Clin. Exp. Neuropsychol.* **18**, 462–472 (1996).
19. Tupler, L. A., Welsh, K. A., Asare-Aboagye, Y. & Dawson, D. V. Reliability of the Rey-Osterrieth complex figure in use with memory-impaired patients. *J. Clin. Exp. Neuropsychol.* **17**, 566–579 (1995).
20. Langer, N. *et al.* The AI neuropsychologist: Automatic scoring of memory deficits with deep learning. *bioRxiv*. https://doi.org/10.1101/2022.06.15.496291 (2022).
21. Park, J. Y., Seo, E. H., Yoon, H. J., Won, S. & Lee, K. H. Automating Rey complex figure test scoring using a deep learning-based approach: A potential large-scale screening tool for cognitive decline. *Alzheimers Res. Ther.* **15**, 145 (2023).
22. Vogt, J. *et al.* Automated scoring of the Rey-Osterrieth complex figure test using a deep-learning algorithm. *Arch. Clin. Neuropsychol.* **34**, 836–836 (2019).
23. Youn, Y. C. *et al.* Use of the clock drawing test and the Rey-Osterrieth complex figure test-copy with convolutional neural networks to predict cognitive impairment. *Alzheimers Res. Ther.* **13**, 85 (2021).
24. Petilli, M. A., Daini, R., Saibene, F. L. & Rabuffetti, M. Automated scoring for a Tablet-based Rey Figure copy task differentiates constructional, organisational, and motor abilities. *Sci. Rep.* **11**, 14895 (2021).
25. Smith, S. R. *et al.* Psychometrics of a simple method for scoring organizational approach on the Rey-Osterrieth complex figure. *J. Neuropsychol.* **1**, 39–51 (2007).
26. Kim, M. *et al.* Eye movement as a biomarker of impaired organizational strategies during visual memory encoding in obsessive–compulsive disorder. *Sci. Rep.* **11**, 18402 (2021).
27. Yi, L. *et al.* Children with autism spectrum disorder scan own-race faces differently from other-race faces. *J. Exp. Child Psychol.* **141**, 177–186 (2016).

28. Bradley, M. C. *et al.* Obsessive–compulsive symptoms and attentional bias: An eye-tracking methodology. *J. Behav. Ther. Exp. Psychiatry* **50**, 303–308 (2016).
29. Haque, R. U. *et al.* VisMET: A passive, efficient, and sensitive assessment of visuospatial memory in healthy aging, mild cognitive impairment, and Alzheimer's disease. *Learn. Mem.* **26**, 93–100 (2019).
30. Vail, A. K. *et al.* Visual attention in schizophrenia: Eye contact and gaze aversion during clinical interactions. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* 490–497 (IEEE, 2017).
31. Navalón, P. *et al.* Attentional biases to emotional scenes in schizophrenia: An eye-tracking study. *Biol. Psychol.* **160**, 108045 (2021).
32. Pavisic, I. M. *et al.* Eye-tracking indices of impaired encoding of visual short-term memory in familial Alzheimer's disease. *Sci. Rep.* **11**, 8696 (2021).
33. Addington, J. *et al.* Comorbid diagnoses for youth at clinical high risk of psychosis. *Schizophr. Res.* **190**, 90–95 (2017).
34. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
35. Mnih, V., Heess, N., Graves, A. & Kavukcuoglu, K. Recurrent models of visual attention. In *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems* 2204–2212 (MIT Press, 2014).
36. Wolf, K. *et al.* Neurocognitive development of the resolution of selective visuo-spatial attention: Functional MRI evidence from object tracking. *Front. Psychol.* **9**, 1106 (2018).
37. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
38. Ahmed, Z. A. T. *et al.* Applying eye tracking with deep learning techniques for early-stage detection of autism spectrum disorders. *Data* **8**, 168 (2023).
39. Priyadarshini, I. Autism screening in toddlers and adults using deep learning and fair AI techniques. *Future Internet* **15**, 292 (2023).
40. Sims, S. & Conati, C. A neural architecture for detecting confusion in eye-tracking data. *arXiv preprint* arXiv:2003.06434 (2020).
41. Miller, T. J. *et al.* Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: Predictive validity, interrater reliability, and training to reliability. *Schizophr. Bull.* **29**, 703–715 (2003).
42. Jung, M. H. *et al.* The reliability and validity of the Korean version of the structured interview for prodromal syndrome. *Psychiatry Investig.* **7**, 257–263 (2010).
43. Goodman, W. K. *et al.* The Yale-Brown obsessive compulsive scale. I. Development, use, and reliability. *Arch. Gen. Psychiatry* **46**, 1006–1011 (1989).
44. Hamilton, M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* **23**, 56–62 (1960).
45. Hamilton, M. The assessment of anxiety states by rating. *Br. J. Med. Psychol.* **32**, 50–55 (1959).
46. Kim, Z. Two-and four-subtest short forms of the Korean-Weschler Adult Intelligence. *Scale Seoul J. Psychiatry* **19**, 121–126 (1994).
47. Blignaut, P. Fixation identification: The optimum threshold for a dispersion algorithm. *Atten. Percept. Psychophys.* **71**, 881–895 (2009).
48. Somerville, J., Tremont, G. & Stern, R. A. The Boston qualitative scoring system as a measure of executive functioning in Rey-Osterrieth complex figure performance. *J. Clin. Exp. Neuropsychol.* **22**, 613–621 (2000).
49. Paszke, A. *et al.* PyTorch: An imperative style, high-performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems* 8026–8037 (Curran Associates Inc., 2019).
50. Fotouhi, S., Asadi, S. & Kattan, M. W. A comprehensive data level analysis for cancer diagnosis on imbalanced data. *J. Biomed. Inform.* **90**, 103089 (2019).

Acknowledgements

Author contributions

Authors M.K., J.L., and J.S.K. contributed to the conception and design of the study. Authors M.K., J.L., M.H., I.P., J.J., M.J., and S.P. collected the data. Authors J.L. and S.Y.L. performed the data analysis. Authors M.K. and J.L. wrote the first draft of the manuscript. Authors S.Y.L., M.H., I.P., J.J., M.J., S.P., and J.S.K. interpreted the data and critically edited the manuscript. Authors M.K. and J.S.K. contributed to the conception of the study, interpreted the data, and provided critical comments regarding the manuscript. All the authors contributed to manuscript revision and read and approved the submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-68586-2.

Correspondence and requests for materials should be addressed to J.S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.