Research article

Journal of Alzheimer's Disease
2025, Vol. 108(S1) S141–S158
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13872877251376939
journals.sagepub.com/home/alz
Sage | IOS Press

Machine learning model for predicting the conversion to dementia using the Cube Copying Test

Mio Shinozaki[1,2,3], Hiroyuki Hishida[4], Yasuyuki Gondo[2], Michio Yamamoto[2,5,6], Takashi Suzuki[7], Rina Miura[8], Takashi Sakurai[9], Akinori Takeda[10] and Yutaka Arahata[1]

Abstract

Background: Early detection of dementia requires highly accurate and efficient screening tests that minimize patient burden.

Objective: To develop a machine learning model predicting dementia conversion within 3–5 years using Cube Copying Test (CCT) drawings at baseline.

Methods: This retrospective study analyzed CCT drawing data from 767 patients at the Center for Comprehensive Care and Research on Memory Disorders (2011–2020). Of the 2303 patients who met the inclusion criteria, 534 were excluded due to mild cognitive impairment (MCI) persistence, pending diagnoses, or new neurovascular diseases, while 1002 were lost to follow-up. Eligibility criteria included a baseline Mini-Mental State Examination (MMSE) score ≥24, absence of dementia diagnosis or anti-dementia medication intake, and completion of a 3–5-year follow-up without meeting exclusion criteria.

Results: Of 767 patients, 457 converted to dementia (318 with Alzheimer's disease, 116 with dementia with Lewy bodies, and 23 with frontotemporal dementia) within 3–5 years, while 310 did not. The model achieved an area under the curve of 0.85 for predicting dementia conversion. Shapley Additive exPlanations analysis identified PatchCore-derived features as the strongest predictors, distinguishing drawing patterns of converters and non-converters.

Conclusions: In patients who convert to Alzheimer's disease, dementia with Lewy bodies, or frontotemporal dementia, the very early stages of constructional apraxia-like symptoms already exist at the preclinical stage or MCI stage. Applying deep learning-based anomaly-detection models can detect these early drawing distortions that differ from normal aging and contribute to improving the performance of dementia-conversion prediction.

Keywords

agraphia, Alzheimer's disease, artificial intelligence, cognitive dysfunction/diagnosis, early diagnosis, lewy body disease, machine learning, neuropsychological tests

Received: 3 September 2024; accepted: 26 July 2025

[1]Department of Neurology, National Center for Geriatrics and Gerontology, Aichi, Japan
[2]Graduate School of Human Sciences, The University of Osaka, Osaka, Japan
[3]Japan Society for the Promotion of Science, Tokyo, Japan
[4]MathWorks Japan, Tokyo, Japan
[5]RIKEN AIP, Tokyo, Japan
[6]Data Science and AI Innovation Research Promotion Center, Shiga University, Shiga, Japan
[7]Center for Mathematical Modeling and Data Science, The University of Osaka, Osaka, Japan
[8]Department of Psychiatry, National Center for Geriatrics and Gerontology, Aichi, Japan
[9]Research Institute, National Center for Geriatrics and Gerontology, Aichi, Japan
[10]Center for Comprehensive Care and Research on Memory Disorders, National Center for Geriatrics and Gerontology, Aichi, Japan

Corresponding author:
Mio Shinozaki, Department of Neurology, National Center for Geriatrics and Gerontology, 7-430 Morioka-cho, Obu-City, Aichi 474-8511, Japan.
Email: shinozaki@ncgg.go.jp

Introduction

Early detection of dementia is essential to delaying its onset and progression. Advances in dementia treatments targeting the early stages of mild cognitive impairment (MCI)[1] have amplified the importance of early detection before conversion to dementia. Dementia progresses along a continuum from normal cognitive function (NC) to subjective cognitive decline (SCD), MCI, and dementia.[2–4] Cognitive decline, marked by the accumulation of amyloid-β and other neuropathological changes, begins approximately 10–20 years before an MCI diagnosis,[5–7] with accelerated decline starting 3–7 years prior.[8] Therefore, current efforts focus on identifying patients at high risk of dementia even before the MCI stage.

For instance, in Alzheimer's disease (AD), early detection with over 90% accuracy can be achieved through a combination of amyloid positron emission tomography and cerebrospinal fluid biomarkers.[9] However, these diagnostic tools are limited by high costs, invasiveness, and accessibility issues. By contrast, neuropsychological tests are cost-effective, non-invasive, and safe; yet, the accuracy of a single test is often inadequate, necessitating the combination of multiple assessments to reliably identify individuals at high risk.[10] Although comprehensive neuropsychological test batteries can achieve high diagnostic accuracy, they are time-consuming and place a considerable burden on examinees, making them impractical for large-scale community-based screening programs that play a crucial role in early detection. Therefore, a critical need exists for the development of tools that can accurately detect early signs of future dementia conversion using only a minimal number of test items. Particularly for community-based population-screening programs, such tools must be rapid, minimally invasive, and cost-effective while providing highly accurate and efficient screening with minimal patient burden.

Early signs of cognitive decline manifest in domains such as episodic memory, working memory, language, visuospatial abilities, and executive functions.[5,7,11] Among these, visuospatial function often declines earlier than other cognitive domains, such as memory in major dementia subtypes, including AD and dementia with Lewy bodies (DLB).[5,7,12] As a result, recent research has increasingly focused on visuospatial cognitive function as a critical indicator for early dementia detection.[12,13]

Early visuospatial decline is commonly assessed through tasks such as the Cube Copying Test (CCT), Clock Drawing Test, and Double Pentagon Test. Among these, the CCT is widely adopted in clinical practice due to its sensitivity to subtle changes, making it a reliable tool for detecting cognitive impairment.[14–16] However, age-related changes can significantly influence CCT performance, even in the absence of dementia. For example, Ericsson et al. demonstrated that CCT accuracy declines with age, reporting that only 42% of individuals aged 75–79 years and 24% of those aged ≥90 years could copy correctly.[16] Moreover, demographic variables such as sex and years of education also impact performance; women typically score lower than men,[17] and individuals with more years of education tend to perform better on drawing tasks.[18]

In other words, even older adults with NC may experience drawing distortions with aging, and non-converters may sometimes produce drawings that deviate from the model cube drawing. Therefore, to utilize drawing tests related to visuospatial function for early detection of dementia, distinguishing between drawing distortions caused by normal aging and subtle pathological drawing distortions that precede dementia conversion is necessary.

However, traditionally, tests such as the CCT have relied on qualitative evaluation, where scoring is fundamentally based on visual judgments of whether specific criteria related to deviations from the model cube drawing are met. This approach has the limitation that the reliable detection of minor abnormalities is hindered because it is influenced by the scorer's experience and biases. Additionally, since normal aging can also cause deviations from the model cube drawing, sufficient discrimination between drawing distortions associated with normal aging and those specific to patients who will eventually convert to dementia was not always possible.

Conversely, artificial intelligence (AI) technology can overcome the limitations of manual scoring by accurately, quantitatively, and objectively extracting and analyzing features from images. Machine learning models utilizing image data have the advantage of effectively leveraging AI capabilities compared to other modalities, such as numerical data.[19] Furthermore, by combining deep learning-based anomaly-detection models, it may be possible to extract not only features related to deviations from the model cube drawing, but also features that distinguish pathological distortions from drawing distortions caused by normal aging, the latter of which have been difficult to extract and quantify with traditional qualitative assessment methods.

Additionally, medical institutions often face difficulties in obtaining sufficient normal data. While machine learning with small datasets often presents challenges, these limitations can be overcome by leveraging high-precision industrial anomaly-detection models designed to identify minor distortions and significant structural defects. Specifically, PatchCore,[20] which achieved state-of-the-art performance in industrial applications in 2021, demonstrates exceptional results even with limited data. To our knowledge, no studies in neuropsychology have applied this model to detect distortions in patient drawings. However, such advanced anomaly-detection models developed in industrial domains could be highly effective for classification tasks in clinical research, particularly when access to normal data is restricted, as in our study.

Therefore, this study aimed to develop a machine learning model to identify patients at high risk of dementia conversion within 3–5 years using only CCT drawings, by dramatically enhancing the predictive performance of the CCT through the application of advanced AI technology.

Methods

Participants for the analysis

This retrospective study employed an opt-out procedure, as many patients had passed away, relocated, or were transferred to other institutions, making the acquisition of individual consent impractical. The study adhered to the principles of the Declaration of Helsinki and was approved by the Research Ethics Committee of the National Center for Geriatrics and Gerontology (approval number 1449).

Between January 2011 and December 2020, patients aged ≥60 years who visited the Center for Comprehensive Care and Research on Memory Disorders at the National Center for Geriatrics and Gerontology were included in the study. Inclusion criteria were a baseline Mini-Mental State Examination (MMSE) score of ≥24, the availability of CCT image data, no dementia diagnosis, and no pharmacological treatment for dementia. Individuals with visual impairments (such as cataracts or glaucoma), essential tremors, schizophrenia or delusional disorders, mood disorders, delirium, alcohol or substance dependence, intellectual disabilities, higher brain dysfunction, developmental disorders, epilepsy, subarachnoid hemorrhage, stroke, subdural hematoma, epidural hematoma, multiple cerebral infarctions, brain tumors, normal pressure hydrocephalus, Parkinson's disease, or other conditions affecting drawing or cognitive function were excluded at baseline.

Patients at baseline ranged from those who were cognitively normal to those with MCI. Those diagnosed with AD, DLB, or frontotemporal dementia (FTD) within 3–5 years from baseline were classified as "converters." Patients who did not meet the criteria for MCI or dementia at the 3–5-year follow-up and whose cognitive function remained within the normal range, as determined through comprehensive physician assessments, brain imaging, and neuropsychological tests, were classified as "non-converters." Diagnoses were based on established clinical criteria for AD,[21] DLB,[22] and FTD,[23] as well as the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria for major neurocognitive disorders. For the exclusion of cases diagnosed with MCI at the 3–5-year follow-up, judgment was made based on Petersen's criteria[24] or DSM-5 criteria for minor neurocognitive disorders. Specifically, amnestic MCI was diagnosed based on criteria that included a Clinical Dementia Rating score of ≥0.5, a Wechsler Memory Scale-Revised Logical Memory I score of ≤13, or a Logical Memory II score of ≤8. For non-amnestic MCI, physicians made comprehensive judgments based on Clinical Dementia Rating items related to executive function and judgment, as well as information obtained from family members about changes in daily functioning during follow-up.

The observation period was set at 3–5 years to accommodate variations in follow-up visits, as not all patients were followed up at the same time intervals. The CCT was not used in the diagnostic classification of converters and non-converters at the 3–5-year follow-up.

Participants who were lost to follow-up at 3–5 years ($n =$ 1002) were excluded from the study. This group likely included individuals who no longer sought medical consultations due to symptom resolution, those who were transferred to nearby medical institutions, those who relocated, those admitted to care facilities, and those who passed away. However, as these participants did not attend follow-up visits, detailed background information about them is unavailable. Participants with unclear clinical progression or unconfirmed diagnoses of dementia (such as those with MCI or suspected dementia) during the 3–5-year follow-up were excluded. Therefore, while all cases included in the analysis had confirmed conversion to dementia, in some cases, the specific subtype (possible or probable diagnosis of AD, DLB, or FTD) was not definitively determined. MCI represents a gray zone[3,4] in which some patients progress to dementia, while 4–15% may revert even in clinical populations.[4,25,26] Therefore, participants diagnosed with MCI at the 3–5-year follow-up were excluded. Additionally, participants who developed new neurovascular conditions during the 3–5-year observation period, such as subarachnoid hemorrhage, stroke, subdural or epidural hematoma, brain tumors, meningiomas, or head injuries, were excluded because distinguishing whether cognitive decline resulted from these events was not feasible (Figure 1).

Image features

We analyzed the image data from CCT drawings created at baseline. Patients were shown the model cube drawing (4 cm on the long side, 1.9 cm on the diagonal, positioned 5.5 cm from the top of an A4 paper in portrait orientation) and instructed to copy it below the model using a pencil (Figure 2). The test was conducted individually by a clinical psychologist in a quiet room with no time limit. If a patient redrew the drawing, the clinical psychologist asked them to select one, and the chosen drawing was used for analysis. After the tests, the papers were saved as PDF files. During the study, the PDF data were extracted, and the clinical psychologists' notes were removed. Only the patients' drawings were cropped and retained for analysis. Since the study primarily aimed to identify characteristic distortions in drawings by older adults at high risk of converting to dementia, drawings that exhibited the closing-in phenomenon, where participants traced over the model drawing, were excluded from the analysis.

Related features

Since CCT performance is influenced by age, sex, and years of education,[17,27] we included age at baseline, sex (male =

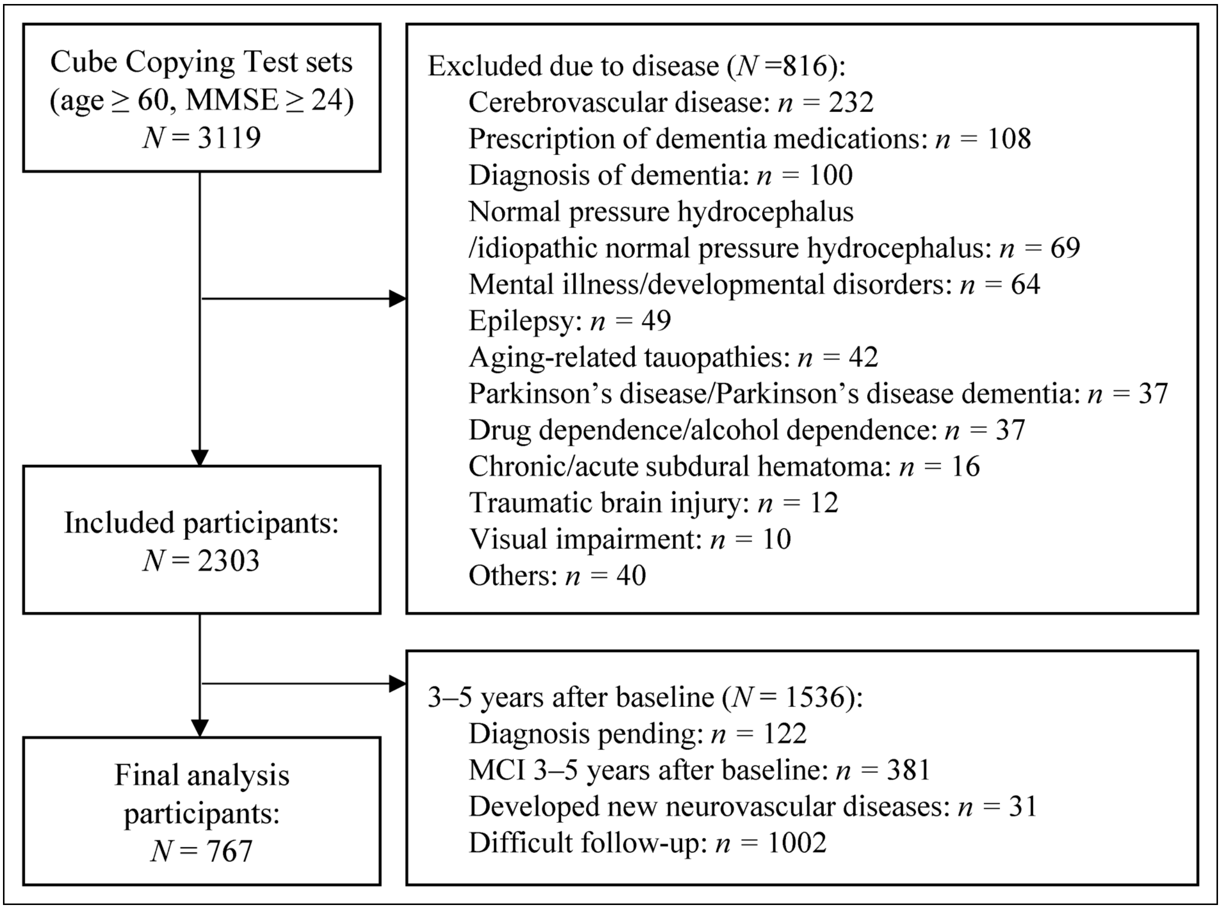
Cube Copying Test sets
(age ≥ 60, MMSE ≥ 24)
N = 3119
Excluded due to disease (N =816):
Cerebrovascular disease: n = 232
Prescription of dementia medications: n = 108
Diagnosis of dementia: n = 100
Normal pressure hydrocephalus /idiopathic normal pressure hydrocephalus: n = 69
Mental illness/developmental disorders: n = 64
Epilepsy: n = 49
Aging-related tauopathies: n = 42
Parkinson's disease/Parkinson's disease dementia: n = 37
Drug dependence/alcohol dependence: n = 37
Chronic/acute subdural hematoma: n = 16
Traumatic brain injury: n = 12
Visual impairment: n = 10
Others: n = 40
Included participants:
N = 2303
3–5 years after baseline (N = 1536):
Diagnosis pending: n = 122
MCI 3–5 years after baseline: n = 381
Developed new neurovascular diseases: n = 31
Difficult follow-up: n = 1002
Final analysis participants:
N = 767

Figure 1. The process of screening participants for analysis. MMSE: Mini-Mental State Examination; MCI: mild cognitive impairment.

0, female = 1), and years of education since elementary school as related features in the model.

Design of the machine learning model

Dataset division. We randomly selected 33% of the entire non-converter dataset as the core image dataset for generating features using anomaly-detection models (PatchCore and convolutional autoencoders [CAE]). Subsequently, 80% of the remaining non-converter images (67% of the total) and all images of converters (100%) were randomly designated as training data, while the remaining 20% were set aside as testing data (Figures 3 and 4).

Image data pre-processing. All image data were pre-processed by first removing stains on the paper and any notes made by the clinical psychologist. The images were then cropped to the bounding box of the drawing. The cropped images were resized so that the longer side measured 200 pixels and were centered on a 220 × 220-pixels white background. Finally, the image size was adjusted to 224 × 224 pixels. The same pre-processing steps were applied to the model cube image to ensure consistency in the matching-score calculation (Figure 3).

Data augmentation. Data augmentation was applied to all images except the test dataset. Images drawn by non-converters were augmented 7 times, while images drawn by converters were augmented 3 times.

Since no prior studies on data augmentation for dementia prediction using CCT images were found, the parameters in this study were independently determined through trial-and-error experimentation. The parameters were empirically selected by testing various parameter combinations and verifying the generated images as values were adjusted. The selection process aimed to balance the suppression of overfitting and the improvement of generalization performance while ensuring that transformations did not alter the clinical characteristics of the drawings, such as converting drawings associated with normal aging into pathological patterns or vice versa. Specifically, one of three rotation types (180° rotation, 90° rotation with a horizontal flip, or −90° rotation with a horizontal flip) was randomly selected and applied. Additionally, Gaussian noise was randomly added to the images. While maintaining the longer side of the images, the shorter side was randomly scaled by −5% to 5%. The images were rotated within the range of −3° to 3° and then binarized. The images

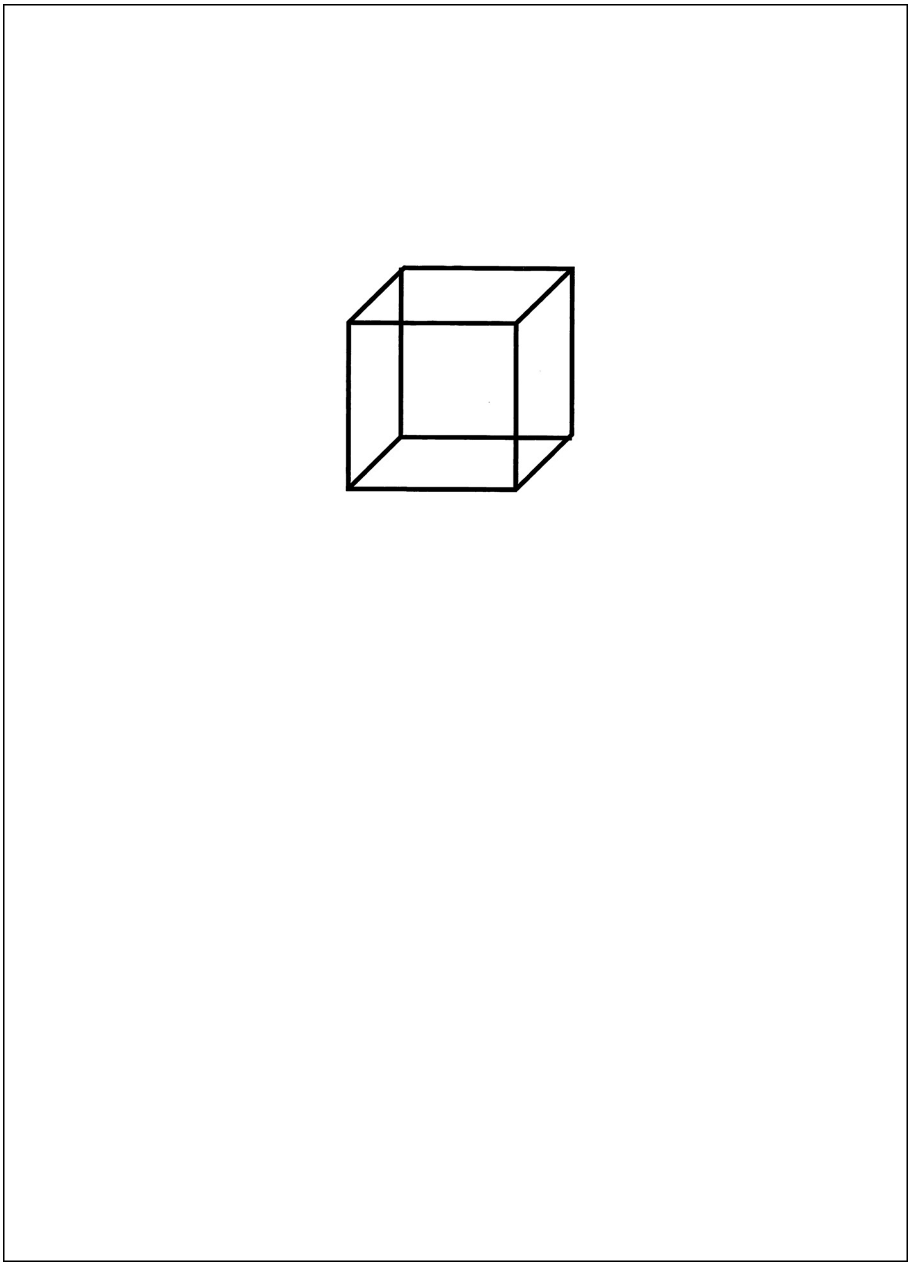

Figure 2. Cube Copying Test paper used for testing.

were randomly shifted by −7 to 7 pixels along the x and y axes. Random affine transformations were applied, and the display range was adjusted to center the transformed images. Finally, the augmented images were saved in PNG format with a size of 224 × 224 pixels (Figure 3).

Local features

The scoring criteria for the CCT, as described in the instruction manual for the Japanese version of the Montreal Cognitive Assessment,[28] specify the presence of all necessary lines, absence of unnecessary lines, preservation of parallel relationships between lines, and similarity in their lengths. Since these characteristics are clearly defined, they can be quantitatively analyzed using conventional image-processing techniques based on a rule-based approach. Accordingly, the following procedure was employed to extract these features (Figure 4).

The images were first converted to grayscale, and the Hough transform was applied to detect line segments. Lines were classified as horizontal (angles within −10° to 10° or 170° to 190°), vertical (angles within 80° to 100° or −100° to −80°), or diagonal (all other angles) based on their angles relative to the horizontal direction. After extracting line endpoints, the following features were

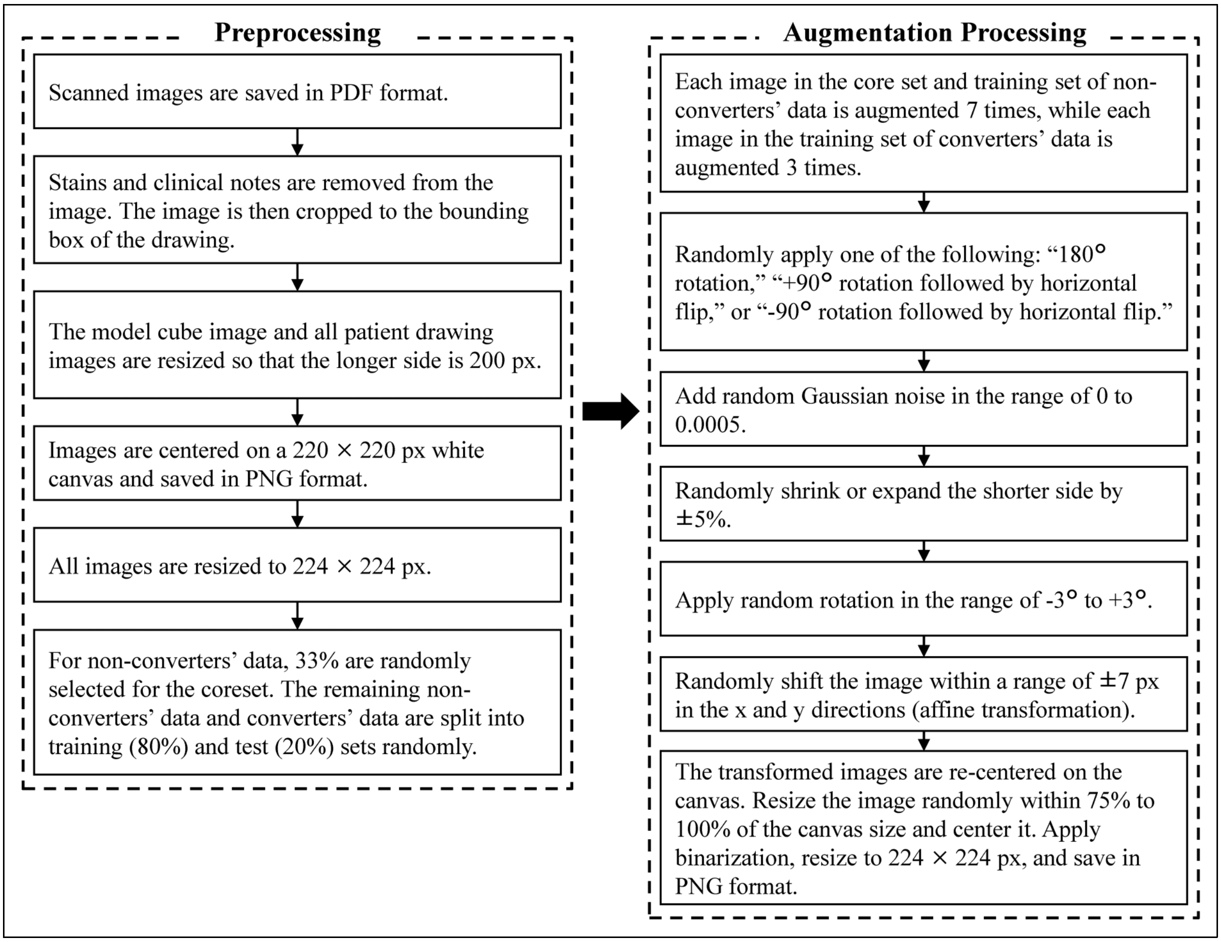
Preprocessing
Scanned images are saved in PDF format.
Stains and clinical notes are removed from the image. The image is then cropped to the bounding box of the drawing.
The model cube image and all patient drawing images are resized so that the longer side is 200 px.
Images are centered on a 220 × 220 px white canvas and saved in PNG format.
All images are resized to 224 × 224 px.
For non-converters' data, 33% are randomly selected for the coreset. The remaining non-converters' data and converters' data are split into training (80%) and test (20%) sets randomly.
Augmentation Processing
Each image in the core set and training set of non-converters' data is augmented 7 times, while each image in the training set of converters' data is augmented 3 times.
Randomly apply one of the following: "180° rotation," "+90° rotation followed by horizontal flip," or "-90° rotation followed by horizontal flip."
Add random Gaussian noise in the range of 0 to 0.0005.
Randomly shrink or expand the shorter side by ±5%.
Apply random rotation in the range of -3° to +3°.
Randomly shift the image within a range of ±7 px in the x and y directions (affine transformation).
The transformed images are re-centered on the canvas. Resize the image randomly within 75% to 100% of the canvas size and center it. Apply binarization, resize to 224 × 224 px, and save in PNG format.

Figure 3. Overview of data pre-processing and augmentation processing.

calculated for horizontal, vertical, and diagonal lines: the number of lines, average and variance of line lengths, and average and variance of the angles between lines of the same type (as a measure of line parallelism). Additionally, the number of vertices was calculated. For the number of lines in each direction (horizontal, vertical, and diagonal) and the number of vertices, the differences from the reference values (four lines in each direction and eight vertices) were computed. The absolute values of these differences were used as feature values.

Global features

The instruction manual for the Japanese version of the Montreal Cognitive Assessment[28] specifies that the CCT must be drawn in three dimensions. However, its definition is not sufficiently clear. Moreover, normal aging can introduce distortions in CCT drawings,[16–18] suggesting that even non-converters may produce images that deviate from the ideal cube representation. To address this issue, we employed both rule-based and data-driven approaches to quantify not only features related to deviations from the model cube drawing (matching score), but also differences from images drawn by non-converters included in the core set through feature extraction using deep learning-based anomaly-detection models (PatchCore scores and reconstruction errors; Figure 4).

Matching score. Similarity to the model cube image was assessed using the Speeded-Up Robust Features algorithm.[29] This algorithm detects feature points in both images, extracts their descriptors, and matches them. The matching score, generated based on the number of matched feature points, represents a rule-based approach that directly compares drawings to a predefined reference image.

PatchCore score. A core set comprising 33% of non-converter images was used to establish a reference distribution. These features quantified the dissimilarity between test samples (67% of non-converters and 100% of converters) and the reference distribution derived from non-converter images. After extracting image features using a pre-trained ResNet50 model, the PatchCore score,[20] defined as the distance to the nearest point within the core set, was computed and saved as a feature. Additionally, anomaly-detection results were visualized to confirm the appropriateness of the model. A higher PatchCore score signifies a greater deviation from the non-converter core set. This represents

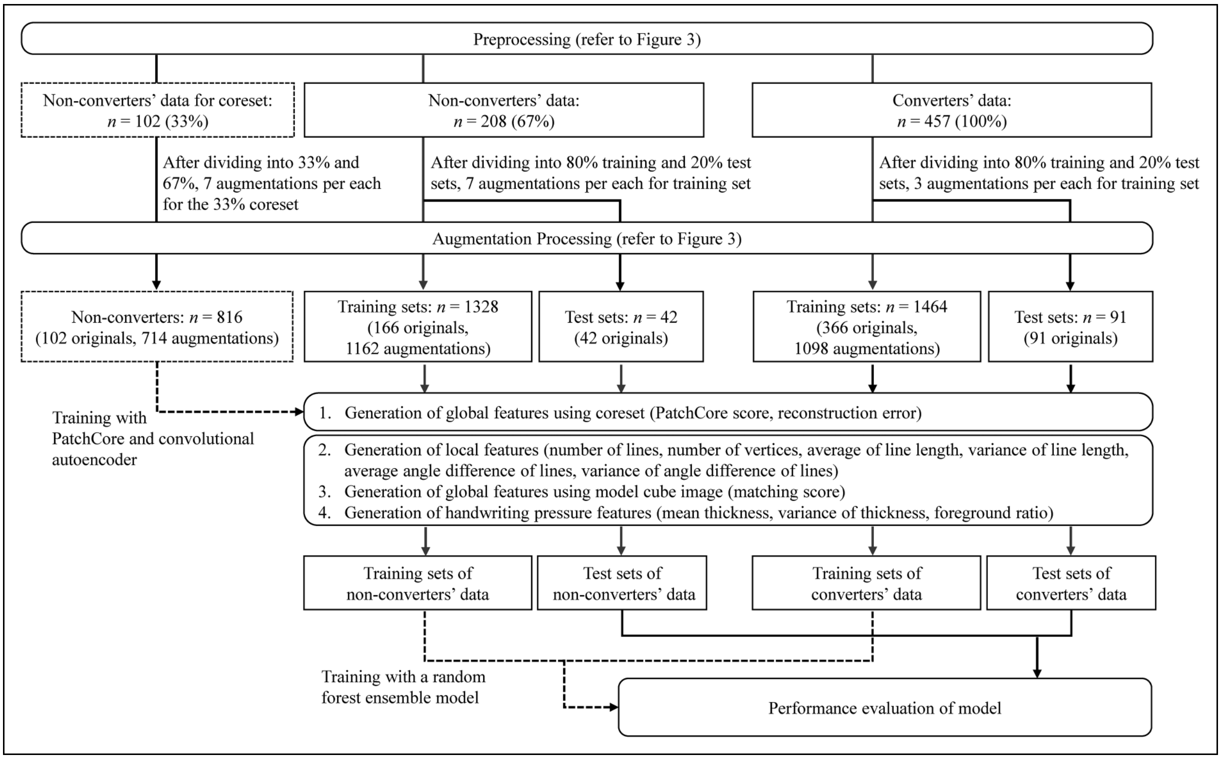
Preprocessing (refer to Figure 3)
Non-converters' data for coreset: n = 102 (33%)
Non-converters' data: n = 208 (67%)
Converters' data: n = 457 (100%)
After dividing into 33% and 67%, 7 augmentations per each for the 33% coreset
After dividing into 80% training and 20% test sets, 7 augmentations per each for training set
After dividing into 80% training and 20% test sets, 3 augmentations per each for training set
Augmentation Processing (refer to Figure 3)
Non-converters: n = 816 (102 originals, 714 augmentations)
Training sets: n = 1328 (166 originals, 1162 augmentations)
Test sets: n = 42 (42 originals)
Training sets: n = 1464 (366 originals, 1098 augmentations)
Test sets: n = 91 (91 originals)
Training with PatchCore and convolutional autoencoder
1. Generation of global features using coreset (PatchCore score, reconstruction error)
2. Generation of local features (number of lines, number of vertices, average of line length, variance of line length, average angle difference of lines, variance of angle difference of lines)
3. Generation of global features using model cube image (matching score)
4. Generation of handwriting pressure features (mean thickness, variance of thickness, foreground ratio)
Training sets of non-converters' data
Test sets of non-converters' data
Training sets of converters' data
Test sets of converters' data
Training with a random forest ensemble model
Performance evaluation of model

Figure 4. Feature extraction and machine learning flow.

a data-driven approach that leverages deep learning-based feature extraction.

Reconstruction error. Similar to the PatchCore score, 33% of the non-converter images were used as a core set to capture structural differences from non-converters in the remaining samples. After extracting image features using a pre-trained ResNet18 model, CAE was trained on the core set. Hyperparameters were optimized using Bayesian optimization. These included the hidden layer size, number of epochs, L2 weight regularization, and sparsity proportion. The reconstruction error, calculated based on the difference between the input and reconstructed images, was saved as a feature. A higher reconstruction error indicates greater structural deviation from the core set of non-converters, representing another data-driven approach using deep learning.

Handwriting pressure features

Since dementia is associated with reduced handwriting pressure,[30] the average and variance of line thickness, as well as the ratios of foreground pixels, were calculated to quantify handwriting pressure as a feature (Figure 4).

Image feature selection, model training, and evaluation

Missing values were estimated and imputed using the k-nearest neighbor method. Related features (age, sex, and years of education) were analyzed using the Mann–Whitney U test for continuous variables and chi-squared test for categorical variables to assess differences between the converter and non-converter groups. These variables were subsequently included in the model as covariates to adjust for their effects. The differences in image features between non-converters and converters were analyzed using two-sided analysis of covariance. Correlation coefficients between selected image features were calculated to identify and address multicollinearity. Pairs of features with an absolute correlation coefficient of 0.8 or higher were identified, and one feature from each pair was excluded. The mean and standard deviation (SD) were calculated for each feature using the labels. If more than half of the features for an image data point exceeded ±3 SD, that data point was considered an outlier and excluded.

The selected image and related features were integrated using a Random Forest ensemble model. This method was chosen based on several advantages: it is a well-established and representative approach, effectively handles nonlinear feature combinations, performs well with limited training data, provides implicit feature selection through random feature sampling that emphasizes optimal feature combinations, offers interpretability through feature importance analysis, and balances versatility with ease of implementation. During model selection, the performances of XGBoost and Support Vector Machine (SVM) were also compared. Random Forest was ultimately selected due to its superior performance.

Optimal hyperparameters, including the number of learning cycles, maximum number of splits, minimum leaf size, and number of variables to sample, were identified through Bayesian optimization using the training dataset. These hyperparameters were then applied to train the Random Forest ensemble model. To reduce overfitting, Shapley Additive exPlanations (SHAP) values were calculated to identify features with minimal contributions. Image features with SHAP values below a set threshold were excluded, and the model was retrained using the remaining features. To ensure consistency and avoid data leakage, the test data were processed using the same methods as the training data. The final model was trained, and predictions were made independently of the test data using the same parameters determined from the training dataset.

A 10-times five-fold cross-validation was performed to evaluate the accuracy and area under the curve (AUC) of the final model. Accuracy, sensitivity, specificity, F1 score, and AUC were calculated for the simple model, which used only related features (age, sex, and years of education), and the final model, which included image features in addition to the related features. Additionally, 95% confidence intervals (CIs) for the model metrics were computed using bootstrap resampling. The AUCs of the two models were compared using a two-sided DeLong test.

SHAP values, representing the contribution of each feature to the final model, were calculated. Absolute SHAP values were computed to evaluate the average impact of each feature on the predictions, and the mean and SD for each feature were calculated. Furthermore, a beeswarm plot was created for each label using SHAP values to visually interpret the distribution and impact of features.

Finally, the Mann–Whitney *U* test was used to examine differences in image features between misclassified and correctly classified images in the test dataset. Statistical tests were conducted with a significance level of $p < 0.05$. Effect sizes were calculated using Cohen's *d* or Cliff's δ for continuous variables and Cramér's *V* for categorical variables.

Software

The analysis was performed using MATLAB R2024a® from The MathWorks®. The Deep Learning Toolbox (version 24.1), Image Processing Toolbox (version 24.1), Statistics and Machine Learning Toolbox (version 24.1), Parallel Computing Toolbox (version 24.1), and Computer Vision Toolbox add-ons for MATLAB® were utilized. SHAP values were calculated using the Python SHAP library (version 0.46.0).

Results

In total, 767 participants met the criteria, comprising 457 converters (318 with AD, 116 with DLB, and 23 with FTD) and 310 non-converters (Table 1). While some participants with DLB and FTD had comorbid AD, the distribution of dementia subtypes among study participants (AD: 69.6%, DLB: 25.4%, FTD: 5.0%) aligned well with previously reported prevalence rates in the literature (AD: 60–77%, DLB: 15–20%, and FTD: 5–15%).[31–35] Cardiovascular disease was observed in 178 (38.9%) of the converters and 115 (37.1%) of the non-converters.

The analysis of covariance revealed significant differences between converters and non-converters for all but one image feature. For local features, converters demonstrated greater absolute differences from the reference values (four lines in each direction) in the number of lines detected across all directions (horizontal, vertical, and diagonal) compared to non-converters. Converters also exhibited larger absolute differences from the reference value (eight vertices) in the number of vertices. Additionally, converters had shorter horizontal and vertical lines and longer diagonal lines than non-converters. For all line types, converters showed greater variance in line lengths, larger average angle differences between lines of the same type, and greater variance in angle differences for horizontal and diagonal lines.

Regarding global features, converters had lower matching scores, higher PatchCore scores, and greater reconstruction errors. For handwriting pressure features, converters displayed lower mean line thicknesses, greater variance in thickness, and lower foreground ratios than non-converters (Table 2).

To prevent multicollinearity, the correlation coefficients for all image feature values were examined, and no feature pairs with absolute values of 0.8 or higher were found. Outlier detection was performed, but no outliers were identified. Consequently, 22 image features were selected for inclusion in the initial model.

Using the 22 selected image features and three related features, hyperparameters were optimized through Bayesian optimization, and the initial Random Forest ensemble model was trained. SHAP values were calculated to refine the model. One image feature (number of vertical lines) with SHAP values below the threshold of 0.0005 was excluded. The final model was retrained using the remaining 21 image features and three related features as inputs.

During model selection, XGBoost showed a training AUC of 1.00 but exhibited lower generalization with a test AUC of 0.73, indicating overfitting. The SVM model yielded moderate results with a test AUC of 0.83, accuracy 0.75, sensitivity 0.76, specificity 0.74, and F1 score 0.81. While SVM achieved slightly higher specificity, Random Forest demonstrated superior performance across all other metrics, with smaller discrepancies between training and test data.

Following 10 iterations of five-fold cross-validation, the final Random Forest model achieved an AUC of 0.85 (95% CI: 0.78–0.91) for the test data (Table 3). The ROC curves

Table 1. Characteristics of the participants ($n = 767$).

	Label				
Mean ± SD or *N* (%)	Converters ($n = 457$)	Non-converters ($n = 310$)	*U*-statistic or chi-squared value	*p*	Effect size
Age (y)	77.6 ± 5.9	73.9 ± 5.8	−46477.500	<0.001	0.344
Sex (female)	299 (65.4%)	162 (52.3%)	13.358	<0.001	0.132
Years of education (y)	11.3 ± 2.6	12.5 ± 2.5	−91210.000	<0.001	−0.288
MMSE	26.1 ± 1.7	28.2 ± 1.8	−114210.500	<0.001	−0.612
Cardiovascular disease	178 (38.9%)	115 (37.1%)			
Diagnosis 3–5 y after baseline					
Alzheimer's disease	318 (69.6%)				
Dementia with Lewy bodies	116 (25.4%)				
Frontotemporal dementia	23 (5.0%)				
No cognitive impairment		310 (100%)			

Sex was analyzed using χ^2 test with Cramér's *V* as the effect size. Age, years of education, and MMSE scores were analyzed using the Mann–Whitney *U* test with Cliff's δ as the effect size. SD: standard deviation; MMSE: Mini-Mental State Examination. Hypertension is included in cardiovascular disease.

Table 2. Differences between converters and non-converters after adjusting for related features (ANCOVA).

Feature	Converters ($n = 1464$) Mean (SD)	Non-converters ($n = 1328$) Mean (SD)	*p*	Cohen's *d*
Local features				
Number of lines (Abs)				
Horizontal	0.477 (0.772)	0.175 (0.439)	<0.001	0.475
Vertical	0.449 (0.744)	0.193 (0.487)	<0.001	0.404
Diagonal	1.480 (1.274)	1.222 (1.124)	<0.001	0.214
Number of vertices (Abs)	2.059 (2.730)	1.102 (2.090)	<0.001	0.391
Average line length				
Horizontal	111.104 (26.588)	117.687 (20.527)	<0.001	−0.275
Vertical	113.533 (25.982)	122.158 (20.146)	<0.001	−0.369
Diagonal	88.141 (88.071)	72.265 (62.758)	<0.001	0.206
Variance of line length				
Horizontal	1063.765 (1487.302)	502.276 (900.328)	<0.001	0.452
Vertical	1014.968 (1488.599)	547.037 (963.425)	<0.001	0.370
Diagonal	5550.603 (17183.455)	2618.515 (12008.967)	<0.001	0.196
Average angle difference of lines				
Horizontal	3.236 (2.388)	2.582 (1.723)	<0.001	0.312
Vertical	2.612 (1.696)	2.228 (1.620)	<0.001	0.231
Diagonal	25.322 (12.373)	21.646 (12.052)	<0.001	0.301
Variance of angle difference of lines				
Horizontal	10.031 (16.989)	6.572 (7.966)	<0.001	0.257
Vertical	7.841 (35.705)	6.826 (48.928)	0.601	0.024
Diagonal	587.061 (458.210)	449.658 (381.093)	<0.001	0.325
Global features				
Matching score	3.903 (2.203)	5.149 (2.438)	<0.001	−0.538
PatchCore score	17.290 (1.927)	15.978 (1.470)	<0.001	0.761
Reconstruction error	0.040 (0.008)	0.036 (0.006)	<0.001	0.632
Handwriting pressure features				
Mean thickness	3.478 (0.079)	3.497 (0.059)	<0.001	−0.267
Variance of thickness	0.459 (0.036)	0.452 (0.035)	<0.001	0.186
Foreground ratio	0.071 (0.006)	0.072 (0.004)	<0.001	−0.186

The sample sizes (*n*) reflect the number of images after data augmentation.
ANCOVA was conducted after standardization, but the means and standard deviations in the table are presented as raw, pre-standardized values. Abs: absolute value; ANCOVA: analysis of covariance; SD: standard deviation.

are shown in Figure 5. The difference in AUC between the final model and a simple model using only the related variables was evaluated using the DeLong test. The final model demonstrated a significantly higher AUC ($p < 0.001$, Cohen's $d = 1.307$).

SHAP values indicated that among the image features, the PatchCore score, reconstruction error, and variance in horizontal and vertical line lengths substantially contributed to the final model's predictions (Table 4, Figure 6).

Finally, for the test dataset, differences in image features between misclassified and correctly classified images were analyzed by labels (Table 5). The results showed that for converters, misclassified images exhibited smaller absolute differences from the reference values in the number of lines detected in horizontal and diagonal directions, as well as in the number of vertices, compared to correctly classified images. Misclassified images also exhibited longer horizontal and vertical lines, shorter diagonal lines, and smaller variances in line lengths within each line type (horizontal, vertical, and diagonal). Additionally, misclassified images had smaller averages and variances in angle differences among vertical and diagonal lines. Regarding global features, misclassified images had higher matching scores, lower PatchCore scores, and fewer reconstruction errors. For handwriting pressure features, misclassified images exhibited higher mean thickness, lower variance in thickness, and higher foreground ratios.

For non-converters, misclassified images exhibited shorter horizontal and vertical lines and longer diagonal lines compared to correctly classified images. Additionally, misclassified images showed greater variance in line lengths among diagonal lines. Global features of misclassified images included higher PatchCore scores and greater reconstruction errors. Handwriting pressure features of misclassified images showed lower mean thickness, greater variance in thickness, and lower foreground ratios compared to correctly classified images.

Table 3. Performance evaluation of test data.

	Developed model		Simple model	
Accuracy	0.77	[0.70–0.83]	0.43	[0.35–0.51]
Sensitivity	0.80	[0.72–0.88]	0.48	[0.38–0.58]
Specificity	0.71	[0.56–0.84]	0.31	[0.18–0.46]
F1 score	0.83	[0.76–0.88]	0.54	[0.44–0.62]
AUC	0.85	[0.78–0.91]	0.31	[0.22–0.41]

Values in brackets are 95% confidence intervals. AUC: area under the curve.

Discussion

In this study, a machine learning model was developed to predict dementia conversion within 3–5 years using only CCT drawings from patients with NC or MCI at baseline. The model achieved an AUC of 0.85, sensitivity 0.80,

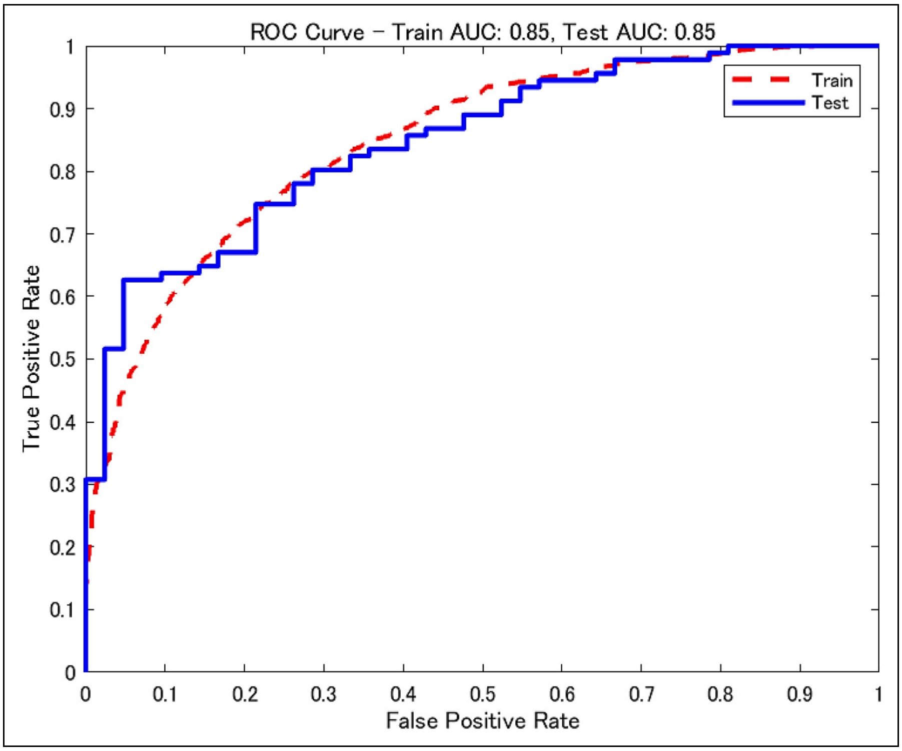

Figure 5. ROC curve for the final model. AUC: area under the curve; ROC: receiver operating characteristic.

Table 4. Mean and standard deviation of absolute SHAP values for features selected in the final model.

	Mean Abs (SD Abs)
Local features	
Number of lines (Abs)	
Horizontal	0.00169 (0.00140)
Diagonal	0.00173 (0.00182)
Number of vertices (Abs)	0.00573 (0.00431)
Average line length	
Horizontal	0.00877 (0.00945)
Vertical	0.00992 (0.00723)
Diagonal	0.00399 (0.00206)
Variance of line length	
Horizontal	0.04426 (0.02022)
Vertical	0.03207 (0.01494)
Diagonal	0.00403 (0.00265)
Average angle difference of lines	
Horizontal	0.00388 (0.00279)
Vertical	0.00105 (0.00084)
Diagonal	0.00534 (0.00227)
Variance of angle difference of lines	
Horizontal	0.00764 (0.00450)
Vertical	0.00374 (0.00227)
Diagonal	0.00568 (0.00368)
Global features	
Matching score	0.02352 (0.01237)
PatchCore score	0.12919 (0.04482)
Reconstruction error	0.08193 (0.04205)
Handwriting pressure features	
Mean thickness	0.00753 (0.00530)
Variance of thickness	0.00220 (0.00159)
Foreground ratio	0.01228 (0.01216)
Related features	
Age	0.00736 (0.00625)
Sex	0.00245 (0.00152)
Years of education	0.00622 (0.00646)

Abs: absolute value; SHAP: Shapley Additive exPlanations; SD: standard deviation.

and specificity 0.71 using CCT image features and related features (age, sex, and years of education).

Among the tested algorithms, Random Forest demonstrated the most stable generalization performance and was less prone to overfitting compared to XGBoost and SVM. Given the relatively small dataset and the wide variety of feature types, including categorical variables and deep learning-derived image features, Random Forest was considered the most suitable approach for this study.

To our knowledge, this is the first study to develop a machine learning approach for identifying individuals at high risk of dementia conversion using CCT alone, without requiring time-consuming neuropsychological test batteries or invasive biomarker assessments.

Previous studies using CCT data have primarily focused on distinguishing dementia status at the time of testing. One study reported an AUC of 0.73, with a sensitivity of 81.9% and specificity of 53.9%, by manually scoring features such as the number of vertices, edge lengths, and three-dimensionality to classify dementia versus non-dementia.[36] Another study achieved an AUC of 0.78, accuracy 0.77–0.78, and F1 score 0.50–0.51 in classifying MCI versus normal cognition using deep learning with CCT.[37] In contrast, our study tackled the more challenging task of predicting future dementia conversion by detecting subtle drawing differences between future converters and non-converters.

To enhance predictive performance, our approach not only considered traditional features related to geometric deviations from the model cube drawing, but also utilized deep learning-based anomaly-detection models to extract features that distinguish pathological drawing distortions from drawing distortions arising from normal aging. In particular, PatchCore, an industrial-grade anomaly-detection model, demonstrated highly effective performance in detecting structural anomalies in drawings, distinguishing pathological changes from normal aging, thereby significantly improving model accuracy.

Our findings suggest that very early signs of constructional apraxia-like symptoms are already present during the preclinical or MCI stages in individuals who will eventually convert to AD, DLB, or FTD and that these subtle changes can be detected using high-precision AI technology. Visuospatial dysfunction in patients with dementia encompasses difficulties in object recognition, spatial orientation, figure-ground discrimination, visual integration, and visual attention.[38–41] These impairments are often reflected in CCT drawings as inappropriate sizes, inaccurate line lengths and shapes, failures in connecting lines or positioning, lack of parallelism, missing elements, unnecessary lines, and distortions or simplifications of three-dimensional structures.[38,39,41–43]

In this study, converters' drawings exhibited greater discrepancies from the reference values for the number of lines in all directions (horizontal, vertical, and diagonal) and the number of vertices compared to those of non-converters. These deviations likely reflect characteristics such as drawing with multiple discontinuous lines rather than a single straight line, producing wavy lines, endpoints failing to overlap at a single point, and lines that were inappropriately segmented or connected at incorrect points.

Converters also exhibited shorter horizontal and vertical lines and longer diagonal lines than non-converters, along with greater variance in the lengths of all line types. This pattern suggests a distorted drawing style characterized by compromised perspective and reduced accuracy. Additionally, the average angle differences between lines of the same type, as well as the variance of angle differences between horizontal and diagonal lines, were larger. These findings indicate that converters' drawings lacked parallelism, reflecting inconsistency in their drawing styles (Figure 7).

The lower matching scores observed in converters further support the distinct drawing characteristics associated with progression to dementia. Conversely, non-

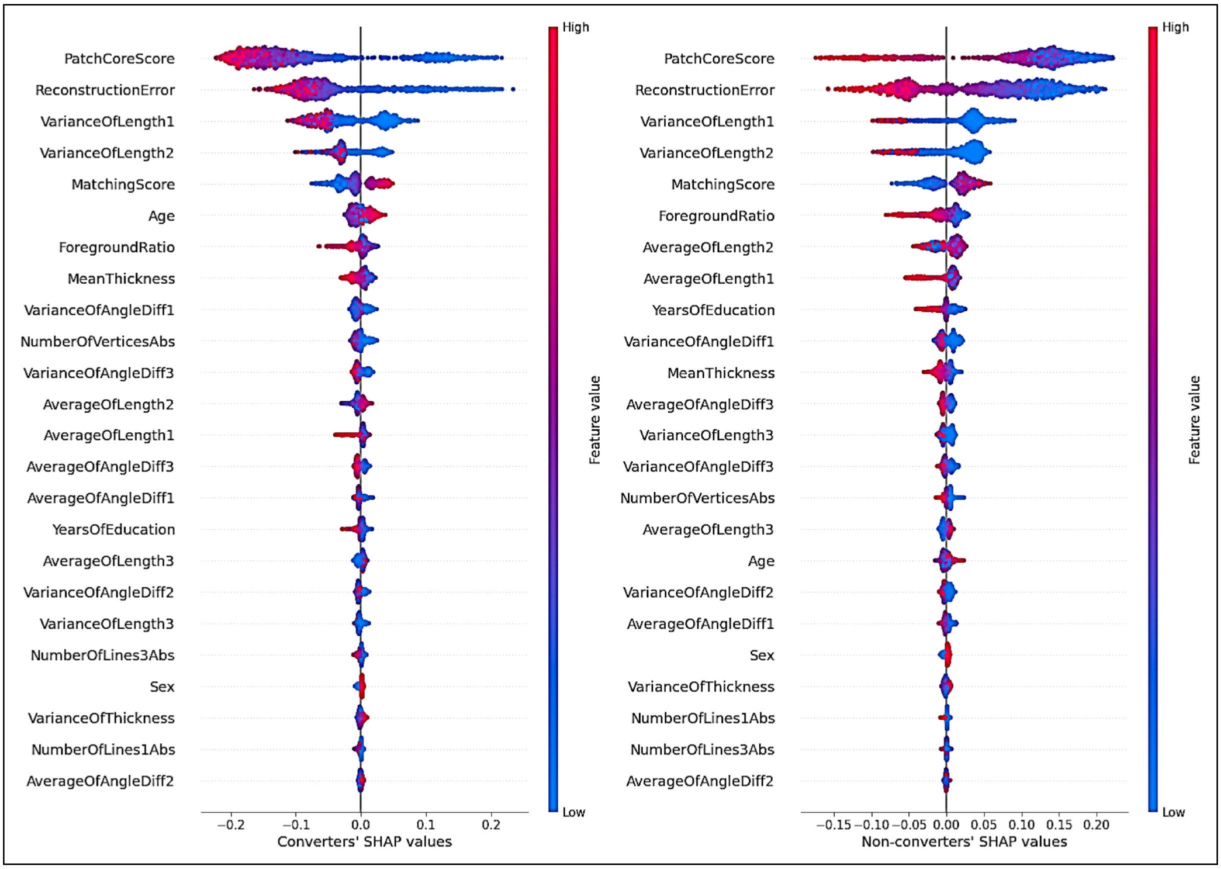

Figure 6. Shapley additive exPlanations (SHAP) beeswarm plot by label for features selected in the final model.

converters demonstrated higher matching scores, suggesting that while aging may cause some drawing distortions, as noted in previous studies, their drawings remained closer to the model cube drawing than those of converters. The higher PatchCore scores and reconstruction errors of the CAE using the core set indicate that converters exhibit structurally different drawing styles, not only compared to the model cube drawing but also to non-converters. While normal aging can lead to distortions in drawings and a reduction in the accuracy of the CCT,[16] this study demonstrates that these distortions have structurally distinct characteristics compared to those observed in individuals at high risk of progressing to dementia within 3–5 years.

Additionally, consistent with previous studies,[30] converters exhibited weaker handwriting pressure, drew thinner lines, and showed greater variability in line darkness compared to non-converters. The association between lower grip strength and increased dementia risk has been previously highlighted,[44–46] and the results of this study align with these findings.

Compared to DLB, AD typically presents with milder visuospatial impairments or constructional deficits, while FTD shows even milder visuospatial impairments than AD and DLB, as reported in previous studies.[15,47–50] Despite most converters in this study being patients with AD, the model achieved an AUC of 0.85 based on a single test result related to visuospatial impairment. This result supports those of previous studies[12,13] emphasizing the importance of visuospatial function as an indicator for the early detection of dementia.

The SHAP values indicated that among the image features, those generated using the PatchCore algorithm contributed the most to the model's predictions. Visualization of the anomaly-detection points identified by the PatchCore algorithm confirmed the model's appropriateness, as these points closely corresponded to areas that clinical psychologists typically focus on during assessments (Figure 8). This study demonstrated that PatchCore's anomaly-detection approach, originally developed for industrial applications, can be effectively repurposed to analyze visuospatial cognitive impairments in drawing tests. Additionally, the reconstruction error of the CAE and the variances in horizontal and vertical line lengths contributed significantly to the model's predictions. These results indicate that both global structural features, particularly those related to differences from drawing distortions associated with normal aging, and local features, particularly those related to parallelism and consistency, are critical for accurate prediction.

Table 5. Differences in image features between misclassified and correctly classified images by label.

	Converters ($n = 91$)					Non-converters ($n = 42$)				
	Misclassified ($n = 18$) Mean (SD)	Correctly classified ($n = 73$) Mean (SD)	Mann–Whitney *U*-statistic	*p*	Cliff's δ	Misclassified ($n = 12$) Mean (SD)	Correctly classified ($n = 30$) Mean (SD)	Mann–Whitney *U*-statistic	*p*	Cliff's δ
Local features										
Number of lines (Abs)										
Horizontal	0.167 (0.514)	0.630 (0.874)	−881.500	0.010	−0.342	0.250 (0.452)	0.067 (0.254)	−147.000	0.107	0.183
Diagonal	0.500 (0.618)	1.644 (1.295)	−1008.500	<0.001	−0.535	1.583 (0.996)	1.033 (0.999)	−125.500	0.112	0.303
Number of vertices (Abs)	0.556 (0.616)	2.356 (2.359)	−1010.000	<0.001	−0.537	0.833 (1.030)	0.500 (0.777)	−148.500	0.325	0.175
Average line length										
Horizontal	134.837 (17.243)	110.026 (25.509)	−260.500	<0.001	0.604	120.761 (16.077)	136.307 (10.920)	−289.000	0.003	−0.606
Vertical	121.719 (14.970)	101.970 (22.329)	−297.500	<0.001	0.547	98.692 (17.879)	126.170 (15.753)	−314.000	<0.001	−0.744
Diagonal	58.305 (8.550)	89.822 (52.795)	−969.500	0.002	−0.476	92.714 (62.321)	58.699 (16.524)	−86.000	0.009	0.522
Variance of line length										
Horizontal	325.591 (691.455)	1715.410 (1872.001)	−1077.500	<0.001	−0.640	876.972 (1287.105)	383.281 (809.974)	−127.000	0.144	0.294
Vertical	545.858 (1125.767)	1281.306 (1556.254)	−935.000	0.006	−0.423	241.156 (294.579)	174.104 (486.608)	−116.000	0.077	0.356
Diagonal	339.927 (406.161)	7454.334 (16260.060)	−1146.000	<0.001	−0.744	2245.602 (3160.454)	962.961 (3046.138)	−59.000	0.001	0.672
Average angle difference of lines										
Horizontal	2.868 (2.032)	4.209 (4.267)	−815.000	0.117	−0.240	2.585 (1.681)	2.358 (1.305)	−177.000	0.945	0.017
Vertical	1.641 (0.948)	3.098 (2.333)	−1021.000	<0.001	−0.554	2.706 (1.635)	2.007 (1.073)	−126.000	0.136	0.300
Diagonal	13.263 (9.284)	26.563 (9.868)	−1121.000	<0.001	−0.706	19.923 (9.886)	16.846 (8.435)	−156.000	0.513	0.133
Variance of angle difference of lines										
Horizontal	6.409 (6.678)	17.964 (53.857)	−853.000	0.051	−0.298	7.160 (5.730)	4.641 (5.258)	−128.000	0.152	0.289
Vertical	2.297 (2.015)	25.276 (144.780)	−1066.000	<0.001	−0.623	6.651 (5.695)	4.544 (5.063)	−133.000	0.195	0.261
Diagonal	271.869 (342.326)	641.750 (419.728)	−1069.000	<0.001	−0.627	322.801 (194.253)	357.422 (235.139)	−197.000	0.646	−0.094
Global features										
Matching score	6.222 (1.865)	3.890 (2.183)	−268.500	<0.001	0.591	4.000 (1.537)	5.667 (2.721)	−246.500	0.063	−0.369
PatchCore score	15.789 (0.795)	18.257 (1.719)	−1245.000	<0.001	−0.895	17.089 (0.916)	15.713 (1.211)	−52.000	<0.001	0.711
Reconstruction error	0.034 (0.004)	0.042 (0.008)	−1047.000	<0.001	−0.594	0.040 (0.006)	0.031 (0.004)	−36.000	<0.001	0.800
Handwriting pressure features										
Mean thickness	3.566 (0.049)	3.479 (0.074)	−215.000	<0.001	0.673	3.467 (0.085)	3.552 (0.056)	−295.000	0.001	−0.639
Variance of thickness	0.414 (0.026)	0.455 (0.035)	−1084.000	<0.001	−0.650	0.463 (0.040)	0.420 (0.032)	−69.000	0.002	0.617
Foreground ratio	0.076 (0.004)	0.072 (0.005)	−361.000	0.003	0.451	0.069 (0.004)	0.076 (0.003)	−326.000	<0.001	−0.811

Abs: absolute value; SD: standard deviation.

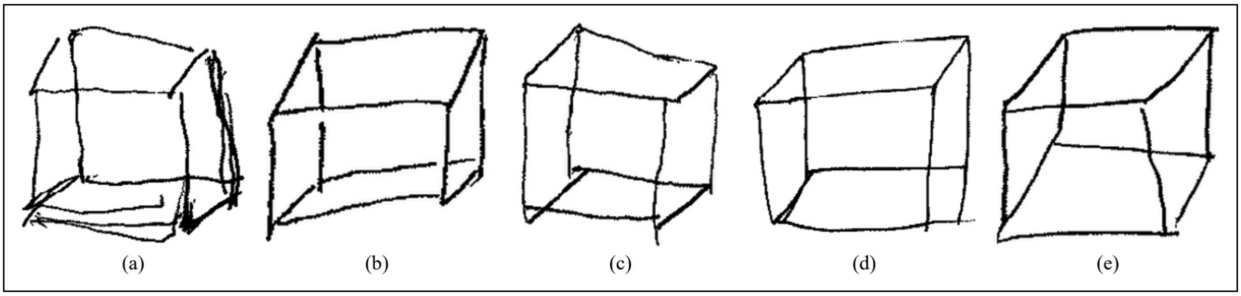

Figure 7. Characteristics of converters' drawings. (a) Multiple discontinuous, wavy lines instead of a single straight line. (b) Lines segmented at intersections with other lines. (c) Endpoints do not converge at a single point. (d) Incorrect connection positions of endpoints. (e) Lengths and angles between lines of the same type differ, causing a lack of parallelism.

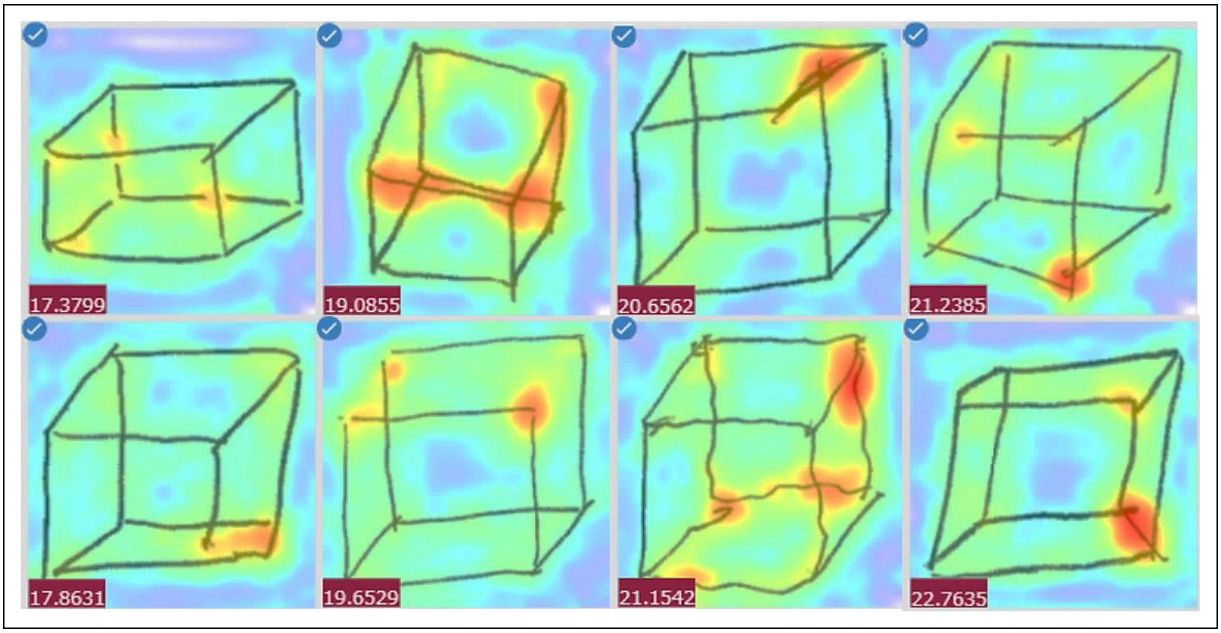

Figure 8. Visualization of anomaly-detection locations using PatchCore.

Conversely, analysis of misclassified cases revealed that converter and non-converter images shared features with their misclassified categories. In other words, some converters display drawing styles similar to those of non-converters and vice versa. This suggests that even with advanced AI technology, misclassification remains possible. Some high-risk individuals may present with minimal visuospatial impairments and produce relatively undistorted drawings, whereas cognitively normal individuals may create significantly distorted drawings due to factors such as their educational background. Therefore, to improve the reliability of dementia-risk assessment, CCT should be administered in conjunction with other cognitive domain evaluations, such as memory tests.

Our findings suggest that AI-enhanced CCT technology has the potential to serve as a practical solution for community-based, large-scale screening, providing a rapid, minimally invasive, and cost-effective option. Beyond these advantages, the CCT's non-verbal nature makes it uniquely suitable for cognitive assessment across diverse populations, including individuals with hearing impairments, non-native speakers, and immigrant populations, contributing to more equitable and accessible cognitive screening. This aspect will become increasingly important as healthcare systems continue to serve diverse communities with varying levels of language proficiency and cultural backgrounds.

Future integration into digital platforms could facilitate remote cognitive assessments, reduce geographical barriers, and support the broader adoption of dementia screening in both clinical and community settings.

Limitations

While the model demonstrated good sensitivity (0.80), its specificity was relatively low (0.71). The data were collected retrospectively from a single memory clinic in Japan, which may have introduced several sources of bias and limited the generalizability and external validity of our findings. This is recognized as a major limitation of our study.

Patients with subjective memory complaints have a 2.07 times higher relative risk of progressing to dementia than those without such complaints.[51] Additionally, depression is a known risk factor for dementia.[52] The study participants were patients who visited the clinic multiple times due to perceived cognitive abnormalities reported by themselves or their families. Consequently, the sample may have included a higher proportion of individuals with subjective memory complaints or depressive tendencies compared to the general population, indicating a potentially high-risk sample.

Analysis of misclassified cases revealed that some non-converters exhibited drawing styles similar to those of converters. In this study, we attempted to exclude MCI cases based on follow-up diagnostic results at 3–5 years according to Petersen's criteria[24] or DSM-5; however, complete exclusion was difficult, particularly for non-amnesic MCI, and some may have been included in the non-converter group.

From a data availability perspective, this database began collection in 2011, limiting the maximum available follow-up period to approximately 10 years. Additionally, the consultation rate during the pre-dementia conversion stage was low, and long-term longitudinal data were very limited. While extending the follow-up period would enable analysis of longer-term progression patterns, prolonged follow-up increases the risk of new medical events such as stroke, trauma, and depression, which could obscure the causal relationship with baseline image features. Considering these factors, the observation period for this study was set at 3–5 years from baseline.

Although some variations exist across studies, reports indicate that the annual conversion rate from MCI to dementia and the recovery rate from MCI to normal cognition are each approximately 15%[26,53,54]; approximately half of the patients with MCI are expected to show clear outcomes within 3–5 years. This period setting was judged to capture relatively definitive cognitive outcomes rather than short-term temporary changes. Nevertheless, progression to dementia can take longer than this timeframe,[5–7] and some patients classified as non-converters may have developed the condition later. Additionally, patients with MCI or an undetermined diagnosis after 3–5 years, as well as those who were challenging to follow up with, were excluded. Among those difficult to follow up, some may have been institutionalized or deceased, while others might have experienced improved or stable cognitive function, and therefore, did not return to the clinic. This selection bias may have influenced the study's results.

The data used in this study were collected from a single facility in Japan. The CCT is influenced by the educational content and standards of a country or region,[18] and the findings may reflect Japan-specific factors, such as mathematics education. The SHAP value for years of education, as shown in Table 4, was 0.00622, which was relatively small compared to the image features. These results may differ in countries or regions where drawing cubes is not a common practice in school education.

Despite these limitations, our model achieved a high AUC of 0.85 using only CCT images, demonstrating strong predictive performance with a rapid, non-invasive, and low-burden assessment. While the feature-extraction methodology is adaptable through calibration and may be applicable across cultural contexts using region-specific data, extensive external validation remains essential for clinical translation. To address these limitations and ensure broader applicability, we are preparing multi-center collaborative validation studies. These upcoming efforts will involve diverse populations across various healthcare settings and cultural backgrounds, leveraging larger sample sizes to enhance the robustness, generalizability, and real-world applicability of our model.

Conclusion

A machine learning model was developed to predict dementia conversion within 3–5 years using only CCT drawings from patients with NC or MCI at baseline. The model achieved strong predictive performance with an AUC of 0.85.

To our knowledge, this is the first study to develop a machine learning approach for identifying individuals at high risk of dementia conversion using CCT alone, without requiring time-consuming neuropsychological test batteries or invasive biomarker assessments. Notably, PatchCore, an industrial-grade anomaly-detection model, exhibited highly effective performance at detecting structural anomalies in drawings that distinguish pathological changes from normal aging.

Our findings indicate that very early signs of constructional apraxia-like symptoms are already present during the preclinical or MCI stages in individuals who will eventually convert to AD, DLB, or FTD and that these subtle changes can be detected using high-precision AI technology.

For community-based population-screening programs, rapid, minimally invasive, and cost-effective assessment tools that can provide highly accurate and efficient screening while minimizing patient burden are essential. When combined with minimal additional assessments such as brief memory tests, this AI-enhanced CCT technology has potential applications for developing practical screening tools suitable for real-world implementation.

Acknowledgements

We sincerely thank the patients and their families for providing the data. We also extend our gratitude to Ms. Shigemi Yamamoto

(Department of Neurology, National Center for Geriatrics and Gerontology) for organizing the data, Assoc. Prof. Takashi Nakazawa (Emerging Media Initiative, Kanazawa University) for his programming guidance, and Prof. Nobuhiko Asakura (Center for Mathematical Modeling and Data Science, The University of Osaka) and Assoc. Prof. Atsushi Koike (Graduate School of Information Sciences, Tohoku University) for their advice on performance improvement.

ORCID iDs

Mio Shinozaki https://orcid.org/0000-0002-3103-7004
Hiroyuki Hishida https://orcid.org/0009-0000-4025-0088
Yasuyuki Gondo https://orcid.org/0000-0002-9805-2807
Michio Yamamoto https://orcid.org/0000-0003-1272-7090
Takashi Suzuki https://orcid.org/0000-0002-0203-5587
Rina Miura https://orcid.org/0009-0009-0934-9333
Takashi Sakurai https://orcid.org/0000-0002-2369-3095
Akinori Takeda https://orcid.org/0000-0002-2761-3240
Yutaka Arahata https://orcid.org/0000-0001-7810-8984

Ethical considerations

The study adhered to the principles of the Declaration of Helsinki and was approved by the Research Ethics Committee of the National Center for Geriatrics and Gerontology (approval number 1449).

Consent to participate

This retrospective study employed an opt-out procedure, as many patients had passed away, relocated, or were transferred to other institutions, making the acquisition of individual consent impractical.

Consent for publication

This article contains data derived from individual participants. As the study was retrospective and many patients were deceased, had relocated, or had been transferred, obtaining written informed consent for publication from each individual was impractical. In accordance with the Declaration of Helsinki, the Research Ethics Committee of the National Center for Geriatrics and Gerontology approved the study and granted a waiver of individual consent for publication (approval number 1449). An opt-out procedure was implemented, allowing patients or their legally authorized representatives to decline the use of their data. All non-essential identifying information has been omitted to protect participant confidentiality. We confirm that documentation of the approved waiver and opt-out procedures is maintained in our institutional research records.

Author contribution(s)

Mio Shinozaki: Conceptualization; Formal analysis; Funding acquisition; Methodology; Project administration; Software; Validation; Visualization; Writing – original draft; Writing – review & editing.
Hiroyuki Hishida: Formal analysis; Methodology; Software; Supervision; Validation; Writing – review & editing.
Yasuyuki Gondo: Supervision; Writing – review & editing.
Michio Yamamoto: Methodology; Supervision; Writing – review & editing.
Takashi Suzuki: Methodology; Supervision.
Rina Miura: Data curation; Supervision.
Takashi Sakurai: Conceptualization; Data curation; Resources; Supervision; Writing – review & editing.
Akinori Takeda: Data curation; Resources; Supervision; Writing – review & editing.
Yutaka Arahata: Conceptualization; Data curation; Project administration; Resources; Supervision; Writing – review & editing.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Grant-in-Aid for JSPS Fellows (Grant Number 23KJ1482) and Research Funding for Longevity Sciences (20–43) from the National Center for Geriatrics and Gerontology, Japan.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

The dataset will be made publicly available on GitHub upon publication. Due to ongoing application development, the source code is not publicly available at this time but can be provided by the corresponding author upon reasonable request for academic research purposes.

References

1. Sevigny J, Chiao P, Bussière T, et al. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature* 2016; 537: 50–56.
2. Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA Research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement* 2018; 14: 535–562.
3. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med* 2004; 256: 183–194.
4. Petersen RC. Mild cognitive impairment. *N Engl J Med* 2011; 364: 2227–2234.
5. Caselli RJ, Langlais BT, Dueck AC, et al. Neuropsychological decline up to 20 years before incident mild cognitive impairment. *Alzheimers Dement* 2019; 16: 512–523.
6. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; 7: 280–292.

7. Weintraub S, Wicklund AH and Salmon DP. The neuropsychological profile of Alzheimer disease. *Cold Spring Harb Perspect Med* 2012; 2: a006171.
8. Karr JE, Graham RB, Hofer SM, et al. When does cognitive decline begin? A systematic review of change point studies on accelerated decline in cognitive and neurological outcomes preceding mild cognitive impairment, dementia, and death. *Psychol Aging* 2018; 33: 195–218.
9. Counts SE, Ikonomovic MD, Mercado N, et al. Biomarkers for the early detection and progression of Alzheimer's disease. *Neurotherapeutics* 2017; 14: 35–53.
10. Belleville S, Fouquet C, Duchesne S, et al. Detecting early preclinical Alzheimer's disease via cognition, neuropsychiatry, and neuroimaging: qualitative review and recommendations for testing. *J Alzheimers Dis* 2014; 42: S375–S382.
11. Jessen F. Subjective and objective cognitive decline at the pre-dementia stage of Alzheimer's disease. *Eur Arch Psychiatry Clin Neurosci* 2014; 264: 3–7.
12. Johnson DK, Storandt M, Morris JC, et al. Longitudinal study of the transition from healthy aging to Alzheimer disease. *Arch Neurol* 2009; 66: 1254–1259.
13. Bäckman L, Jones S, Berger AK, et al. Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology* 2005; 19: 520–531.
14. Ericsson K, Forssell L, Amberla K, et al. Graphic skills used as an instrument for detecting higher cortical dysfunctions in old age. *Hum Mov Sci* 1991; 10: 335–349.
15. Salimi S, Irish M, Foxe D, et al. Visuospatial dysfunction in Alzheimer's disease and behavioural variant frontotemporal dementia. *J Neurol Sci* 2019; 402: 74–80.
16. Ericsson K, Forssell LG, Holmén K, et al. Copying and handwriting ability in the screening of cognitive dysfunction in old age. *Arch Gerontol Geriatr* 1996; 22: 103–121.
17. Paganini-Hill A and Clark LJ. Preliminary assessment of cognitive function in older adults by clock drawing, box copying, and narrative writing. *Dement Geriatr Cogn Disord* 2007; 23: 74–81.
18. Gaestel Y, Amieva H, Letenneur L, et al. Cube drawing performances in normal ageing and Alzheimer's disease: data from the PAQUID elderly population-based cohort. *Dement Geriatr Cogn Disord* 2005; 21: 22–32.
19. Javeed A, Dallora AL, Berglund JS, et al. Machine learning for dementia prediction: a systematic review and future research directions. *J Med Syst* 2023; 47: 17.
20. Roth K, Pemula L, Zepeda J, et al. Towards total recall in industrial anomaly detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* New Orleans, LA, 2022, pp.14318–14328.
21. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* 1984; 34: 939.
22. McKeith IG, Dickson DW, Lowe J, et al. Diagnosis and management of dementia with Lewy bodies. *Neurology* 2005; 65: 1863–1872.
23. McKhann GM, Albert MS, Grossman M, et al. Clinical and pathological diagnosis of frontotemporal dementia: report of the work group on frontotemporal dementia and pick's disease. *Arch Neurol* 2001; 58: 1803–1809.
24. Petersen RC, Doody R, Kurz A, et al. Current concepts in mild cognitive impairment. *Arch Neurol* 2001; 58: 1985–1992.
25. Gainotti G, Quaranta D, Vita MG, et al. Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease. *J Alzheimers Dis* 2014; 38: 481–495.
26. Koepsell TD and Monsell SE. Reversion from mild cognitive impairment to normal or near-normal cognition: risk factors and prognosis. *Neurology* 2012; 79: 1591–1598.
27. Staios M, Nielsen TR, Kosmidis MH, et al. Validity of visuoconstructional assessment methods within healthy elderly Greek Australians: quantitative and error analysis. *Arch Clin Neuropsychol* 2022; 38: 598–607.
28. Suzuki H and Fujiwara Y. Instruction manual of Japanese version of Montreal Cognitive Assessment (MoCA-J), 2010. https://s50b45448262f1812.jimcontent.com/download/version/1558490455/module/11363501891/name/MoCA-Instructions-Japanese_2010.pdf (accessed 4 June 2024).
29. Bay H, Tuytelaars T and Van Gool L. *SURF: speeded up robust features.* Berlin, Heidelberg: Springer, 2006, pp.404–417.
30. Yamada Y, Kobayashi M, Shinkawa K, et al. Characteristics of drawing process differentiate Alzheimer's disease and dementia with Lewy bodies. *J Alzheimers Dis* 2022; 90: 693–704.
31. DeTure MA and Dickson DW. The neuropathological diagnosis of Alzheimer's disease. *Mol Neurodegener* 2019; 14: 32.
32. Barker WW, Luis CA, Kashuba A, et al. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the state of Florida brain bank. *Alzheimer Dis Assoc Disord* 2002; 16: 203–212.
33. Knopman DS, Mastri AR, Frey WH, et al. Dementia lacking distinctive histologic features. *Neurology* 1990; 40: 251.
34. Hansen L, Salmon D, Galasko D, et al. The Lewy body variant of Alzheimer's disease: a clinical and pathologic entity. *Neurology* 1990; 40: 1.
35. Campbell S, Stephens S and Ballard C. Dementia with Lewy bodies: clinical features and treatment. *Drugs Aging* 2001; 18: 397–407.
36. Mathew R, Renjith N and Mathuranath PS. A new scoring system and norms for, and the performance of cognitively-unimpaired older adults on the cube copying test. *Neurol India* 2018; 66: 1644–1648.
37. Ruengchaijatuporn N, Chatnuntawech I, Teerapittayanon S, et al. An explainable self-attention deep neural network for detecting mild cognitive impairment using multi-input digital drawing tasks. *Alzheimers Res Ther* 2022; 14: 111.
38. Mendez MF, Mendez MA, Martin R, et al. Complex visual disturbances in Alzheimer's disease. *Neurology* 1990; 40: 439–439.

39. Parsey CM and Schmitter-Edgecombe M. Quantitative and qualitative analyses of the clock drawing test in mild cognitive impairment and Alzheimer disease: evaluation of a modified scoring system. *J Geriatr Psychiatry Neurol* 2011; 24: 108–118.
40. Quental NBM, Brucki SMD and Bueno OFA. Visuospatial function in early Alzheimer's disease: preliminary study. *Dement Neuropsychol* 2009; 3: 234–240.
41. Trojano L and Gainotti G. Drawing disorders in Alzheimer's disease and other forms of dementia. *J Alzheimers Dis* 2016; 53: 31–52.
42. Palmqvist S, Hansson O, Minthon L, et al. Practical suggestions on how to differentiate dementia with Lewy bodies from Alzheimer's disease with common cognitive tests. *Int J Geriatr Psychiatry* 2009; 24: 1405–1412.
43. Palmqvist S, Hansson O, Minthon L, et al. The usefulness of cube copying for evaluating treatment of Alzheimer's disease. *Am J Alzheimers Dis Other Demen* 2008; 23: 439–446.
44. Buchman AS, Wilson RS, Boyle PA, et al. Grip strength and the risk of incident Alzheimer's disease. *Neuroepidemiology* 2007; 29: 66–73.
45. Buchman AS, Schneider JA, Leurgans S, et al. Physical frailty in older persons is associated with Alzheimer disease pathology. *Neurology* 2008; 71: 499–504.
46. Wang L, Larson EB, Bowen JD, et al. Performance-based physical function and future dementia in older people. *Arch Intern Med* 2006; 166: 1115–1120.
47. Karantzoulis S and Galvin JE. Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Rev Neurother* 2011; 11: 1579–1591.
48. Bondi MW, Edmonds EC and Salmon DP. Alzheimer's disease: past, present, and future. *J Int Neuropsychol Soc* 2017; 23: 818–831.
49. Salmon DP, Galasko D, Hansen LA, et al. Neuropsychological deficits associated with diffuse Lewy body disease. *Brain Cogn* 1996; 31: 148–165.
50. Cronin-Golomb A. Visuospatial function in Alzheimer's disease and related disorders. In: Budson AE and Kowall NW (eds) *The handbook of Alzheimer's disease and other dementias*. Hoboken, NJ: Wiley-Blackwell, 2011, pp.457–482.
51. Mitchell AJ, Beaumont H, Ferguson D, et al. Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: meta-analysis. *Acta Psychiatr Scand* 2014; 130: 439–451.
52. Gauthier S, Reisberg B, Zaudig M, et al. International psychogeriatric association expert conference on mild cognitive impairment. Mild cognitive impairment. *Lancet* 2006; 367: 1262–1270.
53. Roberts R and Knopman DS. Classification and epidemiology of MCI. *Clin Geriatr Med* 2013; 29: 753–772.
54. Petersen RC, Roberts RO, Knopman DS, et al. Mild cognitive impairment: ten years later. *Arch Neurol* 2009; 66: 1447–1455.