Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning Approach to Predict Cognitive Domains with Process Parameters in Digital Drawing Tasks

Christian Mychajliw

christian.mychajliw@med.uni-tuebingen.de

Universitätsklinikum Tübingen

Franz Wortha

Loughborough University

Ulrike Sünkel

German Center for Mental Health (DZPG)

Gerhard W. Eschweiler

Universitätsklinikum Tübingen

Kristina Dawidowsky

University of Tübingen

Anna-Katharina von Thaler

Kiel University

Andreas J. Fallgatter

German Center for Mental Health (DZPG)

Walter Maetzler

Kiel University

Florian Metzger

Universitätsklinikum Tübingen

Heiko Holz

Ludwigsburg University of Education

Article

Keywords: CERAD, CERAD-plus NAB, neuropsychology, dementia, MCI, MiND, Clock Drawing Test, Trail Making Test, Figure Copying Task, Digital, Tablet, cognitive domains, executive functioning, affinity propagation, machine learning, regression models, digital drawing tasks, artificial neural networks

Posted Date: May 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9453201/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Digital drawing data offers a promising, time-efficient addition to the neuropsychological assessment of neurodegenerative diseases, providing a more detailed understanding of early deficits and helping to differentiate neurodegenerative disease subtypes. This study investigates the predictive power of digitally drawing data for domain-level cognitive deficits. We analyzed a single time point assessments from 937 participants in a longitudinal observational study. We extracted 5 data-driven cognitive domains from a neuropsychological assessment battery using affinity propagation: attention, executive/language, intrusions, memory, and visuospatial, which corresponded with conventional theory-driven domains and differentiate between healthy individuals and those with mild cognitive impairment. Machine learning used to predict these data-driven domains from process parameters of digital drawing data showed substantial predictive value for the executive/language domain and moderate value for the memory domain. These findings underscore the potential of digital drawing data for time-efficient diagnostic purposes, particularly in assessing time-sensitive and graphomotoric facets of cognitive impairment.

Introduction

The emergence of potential disease-modifying causal treatments, such as amyloid-targeting antibodies [1–4], has heightened the societal and clinical urgency for the early diagnosis of neurodegenerative diseases and particularly different subtypes of dementias[5–7]. Neurodegeneration results from diverse etiologies with distinct pathological trajectories. In Alzheimer's disease (AD), for instance, the accumulation of amyloid plaques and tau fibrils leads to synaptic loss and the progressive deterioration of cognitive functions[8,9].

Approaches for more differentiated diagnostics are necessary to better understand and differentiate symptoms, pathologies, and progression of these different types of dementia[4,9,10]. Even with rapid developments in laboratory and imaging diagnostics, neuropsychological diagnostic procedures remain an important source of information due to low cost, high accessibility, and the proximity of the examinations to the functions of everyday life. Digital assessments are particularly promising in this context as they can leverage process data not accessible in conventional paper-based assessments to improve diagnostic properties[11]. Yet investigations into the predictive value of these digital data for specific cognitive domain-level performance remain scarce.

This study addresses this research gap by employing a data-driven investigation of the domain structure within digitally recorded drawing data and the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuropsychological assessment battery (NAB)[12–14]. CERAD NAB is a test battery internationally used in dementia diagnostics[12–14], such as the widely used extended German version CERAD-plus NAB[15,16]. Digital recordings of drawings were generated by performing drawing tasks on a tablet using a digitizer-pen.

Utilizing a large cross-sectional sample from a longitudinal cohort study, we first identify a robust domain structure in the CERAD-plus NAB using affinity propagation. We then examine how digital drawing data from three common tasks, the Clock Drawing Test (CDT) respectively the digital CDT (dCDT), the Figure Copy Test (FCT) respectively the digital FCT (dFCT), and the Trail Making Test (TMT) respectively the digital TMT (dTMT), can predict these domains when used alongside key control variables. By highlighting the predictive power of digital process data, this study contributes to the refined differentiation of cognitive deficits and underscores the promise of novel digital tools for the early diagnosis of dementia.

Cognitive Domains

Cognition is an overarching construct that encompasses multiple domains. In AD, stage-specific cognitive changes arise from the progressive neurodegeneration of different brain regions[17,18]. Subtypes of dementia differ in patterns of cognitive decline. Thus, a global cognition score is insufficient to describe specific types of neurodegeneration. To distinguish between different areas of cognition, domains and subdomains are used[19]. Since neuropsychological assessments aim to identify deficits in specific domains, the classification of domains and the mapping of impaired domains to neurodegenerative diseases and different types of dementia are important not only in conceptualization, but also in the construction and interpretation of the test procedures. Both the methodology and the models for distinguishing cognitive domains are vastly different[14,19]. In addition, tasks do not map one-to-one onto cognitive domains, as their completion relies on multiple overlapping cognitive functions and several domains are involved in executing a task (e.g., drawing figures in a recall task requires both memory and visuoconstruction).

Cognitive Domains in MCI/MiND

Dementia disorders in general are categorized based on the cause and the severity of neurodegeneration. Early and mild stages of cognitive impairment are referred to as mild cognitive impairment (MCI) [10,20]or, more recently in DSM-5, as mild neurocognitive disorder (MiND)[21–23]. Severe and later stages are classified as major neurocognitive disorder (MaND). Both MiND and MaND are specified by the underlying neuropathological process. In the case of Alzheimer's disease, “MaND due to Alzheimer’s disease” corresponds to Alzheimer's dementia (AD)[7,24,25]. With MCI and MiND, different forms and stages are distinguished based on cognitive domains. Thus, the relevance of the concept of different domains is further demonstrated by the development of the conceptualization of MCI/MiND and Dementia/MaND: With new findings in research, these are evolving concepts[7]. The concept changes from MCI to MiND: Accordingly, MiND is classified in the DSM-5 via 331.83[26,27] and in the ICD-11 as 6D71[28]. The new term is also associated with a change in the concept[21,29]: While Petersen[30] made a distinction between single/multi-domain MCI and amnestic/non-amnestic MCI[31,32], the DSM-5 speaks of MiND at the domain level as an impairment in *one or more domains* and specifies their neurodegenerative etiology with, e.g., “due to Alzheimer’s Disease” or “due to Lewy body disease”.

Overall, six domains are described in the DSM-5: complex attention, executive function, learning and memory, language, perceptual-motor, and social cognition. Further, assessments as well as specific deficits are characterized and described for each domain[7,25].

Cognitive Domains in the CERAD-plus NAB

For the CERAD-plus NAB, different factor structures, domains, and total scores have been proposed to arrange the individual tasks into a rational framework and to gain insights into cognitive areas across tasks, and the domains have been studied differently over time, depending on the research question and the sample studied[14]. The assignment of tasks to domains is ambiguous and inconsistent and the data for some domains described in DSM-5 are unavailable based on the CERAD-plus NAB, e.g., social cognition is not examined in CERAD-plus NAB. Thus, each task can load on different domains, and various domains are involved in completing each task. Fillenbaum & Mohs[14] also describe different approaches to the factorial structure of the CERAD-plus NAB. There is therefore no standardized structuring of the CERAD-plus NAB into domains or factors. The number of domains as well as their definition and construction from the CERAD-plus NAB is based on numerous assumptions. On the one hand, these domains can be differentiated based on theoretical and observed clinical characteristics. On the other hand, statistical associations between different tasks are also of interest.

Different *theory-driven* approaches have been used to categorize subtests of the CERAD-plus NAB into cognitive domains, as distinct domains and patterns of impairment are relevant for different disorders. E.g., various types of classification are used for PD[33–37] and MCI[37]. Most of these are disorder-specific, e.g., for MCI due to Alzheimer's Disease or PD-related cognitive dysfunction. For instance, whereas early stages of Alzheimer's focus on memory-related facets, PD focuses on processing-related aspects[37]. This reiterates that the domain structure is ambiguous and the domains measured are partially overlapping due to the complexity of the items (e.g., word list task measuring both memory and language functions).

To date, only a limited number of *data-driven* domain classification approaches have been reported to categorizing subtests into domains based on CERAD-plus NAB data[14,38–41]. There are also relevant differences here due to different approaches, samples, and questions. Some researchers used cluster analyses or principal component analyses, while others used exploratory or confirmatory factor analyses, see Fillenbaum & Mohs for an overview[14]. Data-driven approaches found between one and five factors using different factor and component analyses with diverse samples, summarized in Fillenbaum 2023[14,39–47]. The selected number of domains depended on various predetermined criteria and assumptions about the construct and number of domains[14]. However, we use a novel approach for the data-driven classification of domains: In the *affinity propagation*[48], the algorithm determines the ideal number of clusters and does not depend on an a priori specified number of expected domains. Further, it can be applied to similarity matrices that can capture non-linear relationships, e.g., distance correlations.

Thus, it can reveal complex domain structures that are not obtainable through conventional methods (like factor analytical approaches).

Digital Data

Conventional paper-based examinations in neuropsychiatric diagnostics can be time-consuming, strenuous for the participants, and typically only capture the outcome of a task, omitting any process-related information (e.g., graphomotoric and processing speed). Simple digital devices such as tablets can record extensive additional data, including drawing processes (e.g., pressure and movements), response and processing times and patterns, speech data, as well as audio recordings[49–53]. These digital data provide levels of information not available in conventional paper-based neuropsychological examinations and are increasingly enriching dementia diagnostics. The additional collection of digital data could ultimately lead to the identification of those test components that are highly relevant for diagnosis and reduce the duration of testing (and thus also the burden on patients), provided that the collection of digital data then becomes an integral part of testing. Ultimately, this can result in shortened examinations and the enrichment of the diagnostic process with additional information[54].

Drawing Tasks

Numerous *drawing tasks* are frequently used in the neuropsychological diagnosis of neurodegenerative diseases, including different forms of dementia. For example, figures are to be drawn, objects marked or connected[55–58]. In the following, we will focus on three drawing tasks, all of which have been extensively researched at various levels. In this study, we use digital recordings of the following drawing tasks (recorded on a tablet, see Methods/Digital data section for details):

Firstly, in **FCTs**, a figure is to be copied from a template with a pen. These figures can vary in complexity: from a simple circle via a spiral[59–61] to the Rey Osterrieth Complex Figure[62–64]. With these tests, visuospatial and visuoconstructive functions can be assessed[65,66]. New scorings and analytical approaches were introduced[67–70], which enable more fine-grained differentiation between the symptoms of various neurodegenerative diseases, e.g., the QSPT for specific drawing conspicuities in Lewy body dementia[70–72] and the CUPRO for retrograde procedural memory in Parkinson's disease[67].

Secondly, various versions of **TMTs** are used in neuropsychological diagnostics to help identify executive deficits in individuals with neurodegenerative diseases[73]. They differ in length, difficulty, and complexity, as well as in the stimuli used. The stimuli which are to be interconnected with a pen can consist e.g., of numbers only, numbers and letters alternating, or more complex patterns with, e.g., additional number words[74]. These tests are primarily designed to examine executive functions, but also attention and working speed. There are also digital versions of different TMT paradigms that can not only check for correctness, but also collect information about processing speed and search behavior[56,75–79], which can lead to higher diagnostic accuracy[76,77,79]. The conventional paper-based version of the TMT with TMT-A

and TMT-B is often used: In part A, numbers are to be connected in ascending order, and in part B, numbers and letters are to be connected in an alternating pattern[16,73,80,81]. For cognitively impaired people, this can take more than 10 minutes. The widely used Montreal Cognitive Assessment (MoCA) contains a short version of the TMT with alternating numbers and letters[82].

Thirdly, one of the digital drawing tasks that is frequently researched is the **CDT**. Drawing a clock involves many different cognitive skills. The paper-and-pencil test version has been in application for decades[83] and many studies about the digital clock drawing test have been published in recent years with different approaches[84,85]. For the clock drawing test, Müller et al[86] identified correlations between specific parameters such as conspicuous time-in-air (vs. time on surface) and early cognitive impairment. In other works, these are referred to as think time and ink time[87]. These parameters provide a better prediction than the paper-based CDT itself and are comparable to the CERAD-plus NAB[86]. Using a similar approach, Kobayashi et al.[88] combined drawing features of the sentence task of the Mini Mental State Examination (MMSE), pentagon test (MMSE), CDT, TMT-A and TMT-B in one Machine Learning (ML) model[56,88] to predict diagnostic categories, i.e. Healthy Control (HC), MCI, and AD, to find associations with cognitive parameters. They used seven different and distinct tasks representing cognitive domains without combining these tasks. Dion[89] used parameters of the dCDT to predict the domains processing speed, working memory, language, and declarative memory. Each domain consists of the results of several established tests.

Machine Learning in Different Drawing Tasks

ML algorithms used with drawing tasks data have been increasingly used in recent years[76,90,91]. There are a multitude of approaches to analyzing drawing data digitally. Observing healthy and cognitively impaired individuals drawing a clock in the CDT or copying a figure or completing the TMT, it will come as no surprise that parameters in the process are also diagnostically relevant. The strategies of the drawing process can be included as qualitative parameters for this purpose. Quantifiable parameters can also be used in digital drawing, such as pen pressure or duration of the movement of the pen above the surface (time-in-air). Digital recording of the drawing process using a tablet and pen thus offers a major benefit compared to paper-based drawing. Some ML approaches used scans of finished paper-based drawings (i.e. the results) of CDTs, for example using pattern recognition[92]. Other approaches include the course of the drawing process[93]. Others also incorporate additional digital time and drawing process data based on digital process data[94]. This data can be recorded with a digital pen writing on paper[95] or on a tablet with a digitizer pen[88]. Digital assessments therefore provide a wealth of new data that can be incorporated into analyses for predicting diagnoses and symptoms more accurately when parameters are identified that can distinguish between healthy individuals and those affected[50]. Further, dCDT data can be utilized to distinguish between MCI subtypes[96].

The captured data depends on the digital test used[94]. As different studies use heterogenous tasks, sampling techniques and data sets, the results to date have also differed considerably in most aspects

examined, but with good diagnostic performance[85]. At the same time, many possible digital parameters are available, and it is important to select the relevant parameters.

Different data is thus collected in various tasks of the assessments, which can represent added value when making a diagnosis. There are some approaches to combine drawing data in different tasks[88] to pool strengths and information parameters of different drawing tasks and to obtain information about different cognitive domains. Since different cognitive functions are required to solve the various tasks, we assume that there is added information value: Language-based tasks require different cognitive functions than memory tasks or drawing tasks, for example. Since different types of dementia have different patterns of neuropsychological deficits regarding cognitive domains[14,97], the aim is to combine these data to differentiate between symptom patterns and to identify those digital parameters that enable comparable accuracy and richness of diagnostic information[7]. Extensive test procedures can be strenuous for older participants. Therefore, a helpful use of the methods described is to collect additional and more meaningful data in a shorter time.

Prediction Models

In recent years, the number of studies using ML and other artificial intelligence (AI) approaches in neuropsychological diagnostics has increased[51,98]. Many previous publications focused on the evaluation of individual tests, particularly the clock drawing test. Analyses at the level of cognitive domains, however, provide more precise insights into the severity and progression of neurodegeneration. Because different forms of dementia affect distinct cognitive domains, this approach supports clearer differentiation and more differential prediction of disease progression. A more comprehensive approach should therefore allow to identify the parameters that can be used to make the best possible prediction.

Research Questions

To close the research gap, this paper addresses the following research questions:

- Which data-driven domains in the CERAD-plus NAB assessments can be identified through affinity propagation?
- How well do the data-driven domains correspond to the theory-driven domain structure and how well do they differentiate between (mild) cognitively impaired and healthy individuals?
- Can digital drawing data of the CDT, FCT, and TMT predict the data-driven domains when controlling for age, sex, and years of education?

Methods

Present study

In this article, we present a two-stage analytical approach using two different sources of data collected in the TREND study, a longitudinal cohort study for early recognition of neurodegeneration. Firstly, this is conventional paper-based CERAD-plus NAB data, and secondly, digital drawing data from drawing on a tablet. In the first stage, we apply affinity propagation on the conventional paper-based data to derive a data-driven classification of cognitive domains. In the second stage, we investigate how well specific digital drawing features predict the previously identified domains using artificial neural networks. For an illustration of the study procedure, see Figure 1.

TREND Study

The TREND study is a longitudinal observational study assessing risk and prodromal markers for AD and PD[99]. Beginning in 2009, 1,201 participants aged 50+ years were included in the study. The TREND cohort consists of an at-risk cohort (participants with prodromal markers such as REM sleep behavior disorder, depression, and/or hyposmia) and a cohort without specific prodromal markers and risk factors for neurodegeneration. All participants had to be to be free of diagnosed severe neurodegenerative diseases at the start of the study. Inclusion and exclusion criteria as well as study implementation and conduct have been described previously[99]. Participants are examined every two years with an extensive test battery which includes, among other things, MoCA, CERAD-plus NAB and other neurological, motor and neuropsychiatric assessments. The digital assessment of neuropsychological tests presented here was added later during the study and was only used from 2019 (TREND wave “FU5”) onwards. The data is notable for its use of a large, regional, at-risk sample followed in a longitudinal design since 2009. This long-term tracking allows for highly accurate individual characterization using available longitudinal data (since 2009). However, digital neuropsychological test data is only available from the fifth follow-up (FU5 in 2019) onward.

Sample

The initial sample for this study was derived from the TREND study. This study focused on the three most recent measurement waves (FU3 to FU5), which were collected between 2015 (FU3) and 2023 (FU5). These three measurement waves included 959 individuals. To ensure the validity of neuropsychological assessments, we excluded 22 non-native speakers, leaving a sample of N = 937 individuals. This study focused on the three most recent measurement waves (FU3 to FU5). The most recent wave, FU5, was specifically chosen for its inclusion of digital drawing data, while the two preceding waves (FU3 and FU4) were used to validate the clustering analysis. The number of participants varied across these assessments due to the longitudinal nature of the study, with a sample of 937 at FU3, 853 at FU4, and 770 at FU5. For the data-driven domain analysis, we computed the cluster solutions across all participants with partially complete data at each measurement wave to compute pairwise complete distance correlations. However, the domain score calculation for FU5 was limited to participants with complete CERAD-plus NAB assessments, resulting in a sample of N = 761 at FU5.

Finally, the machine learning analyses were restricted to the subset of participants with complete digital drawing data for all investigated tasks, yielding a final sample of *n* = 729 for these predictions. This final sample represents a subset of the larger cohort and is the primary focus of the predictive modeling results.

Participants and Sample Demographics

Descriptive statistics for the sample are displayed in Table 1 detailing the full available data set at FU5 as well as the subsample with complete drawing data using for ML analysis.

Table 1. Sample demographics at Follow-up 5 (FU5)

	Full Sample (FU5)	Machine Learning Predictions		
	n	Mean (SD)	n	Mean (SD)
Age (years)	761	72.43 (6.18)	729	72.33 (6.09)
Sex	761	53% female	729	53% female
Education Years	761	14.59 (2.64)	729	14.61 (2.62)
MoCA (corrected for education)	759	24.47 (3.15)	727	24.51 (3.11)

Assessments

Cognitive examinations performed were the German versions of the MoCA[100–103] and the Consortium to Establish a Registry for Alzheimer's Dementia plus test battery CERAD-plus NAB[12,15,16,104,105]. CERAD-plus NAB is regular CERAD, which has been expanded to include additional tasks: TMT and S-words (words starting with the letter S), a word fluency task[16]. For the global analyses, a z-score based CERAD-plus NAB total score according to TS2 in Lillig et al.[36] based on normative data from Aebi[15] was used.

Different parallel versions of the tests were used in the respective study rounds.

Digital data

From the beginning of the sixth wave (FU5) of the TREND-Study in 2019, we started the digital acquisition of digital drawing data as part of the TuCAN project. Drawing Data was recorded by an application developed especially for this purpose on a “Samsung Galaxy Tab A (2016) with S Pen” tablet (model SM-P580) using a “Staedtler Noris digital classic 180 22” pen, which comes very close to a real pencil in feel and use. The original assessments remained unaltered with the modification that the drawing tasks were recorded digitally on the tablet. The screen on the tablet where the participants drew can be seen for the three tasks in Figure 2. With this setup, we captured data of the drawing process,

including parameters like pen pressure and acceleration. Of various digitized tests, the three digitally recorded drawing tasks from the MoCA were used for this study i.e., TMT, FCT, CDT.

The data was captured in tabular format and contains parameters used to record the drawing process. The variables include the 2D coordinates on the screen and above, speeds, acceleration, and pen pressure levels for each time stamp. The sequence of movements is segmented into events and for each event for which the previously described data is available.

Study data were collected and managed using REDCap electronic data capture tools [106,107] hosted at the university hospital Tübingen.

Data Preprocessing and Feature Extraction

Before the main analysis, the data was preprocessed using the following steps: (1) The raw surface and hover movements (sampled at 120Hz) were categorized into individual movements (i.e., a stroke). (2) Intermediate measures including the starting and stopping coordinates, distance covered, average speed etc. were calculated for each movement. (3) The features (see Table 9, Appendix) were aggregated across all surface and hover movements respectively. This procedure was applied to all digital drawing tasks (dTMT, dFCT, dCDT).

Analytical Approach

This study utilized a two-step analytical approach to investigate the predictive value of digital drawing data for cognitive assessment. First, a data-driven clustering approach was used to identify the underlying domain structure of the CERAD-plus NAB. Second, machine learning models were employed to evaluate the predictive value of digital drawing data from three MoCA tasks — the Clock Drawing Task (dCDT), Figure Copying Task (dFCT), and Trail Making Task (dTMT) — for each of the identified domains.

Affinity propagation [48] was used to reveal a data-driven domain structure of the measures within the CERAD-plus NAB. This algorithm was selected as it does not require a pre-specified number of clusters and is robustly applicable to various similarity matrices. To ensure the stability of the cluster solution, the analysis was performed on data from the three most recent waves of the longitudinal study. One variable, “Constructional Praxis figure copying overlapping rectangles“, was excluded from the analysis due to a lack of variance across all three time points.

We applied affinity propagation to the distance correlation matrix of the 30 remaining CERAD-plus NAB raw-scores, cf. Table 10. Distance correlations were chosen to capture both linear and non-linear relationships and were completed on pairwise complete data at each measurement wave. The main diagonal of the distance matrix, representing each measure's self-dissimilarity, was set to 0.03, as a value of 0 yielded a single-cluster solution. The stability of the clustering solution was assessed by

varying this parameter from 0.01 to 0.05 in 0.01 increments. The number of clusters remained stable across all time points in this range.

To establish the most reliable domain structure, affinity propagation was applied to the distance correlation matrices from each measurement wave. Only items consistently assigned to the same cluster across all time points were included in the final scales. Domain scores were calculated by first standardizing all items into a unit range (0 to 1) to ensure equal contribution from each item to the domain. Items where lower scores indicated better performance were inverted by subtracting the standardized score from 1. The domain scores were computed as the mean of standardized raw scores. The content validity of the derived scales was estimated by computing their correlations with established theory-driven CERAD-plus NAB domain structures from the literature[108] and with the MoCA total score adjusted for years of education. Furthermore, we compared the scores of individuals with and without cognitive impairment on both the data-driven and established theory-driven scales. Cognitive status was defined based on a MoCA total score cutoff of 23 or higher for healthy individuals[103,109]. Permutation tests were used to compare the mean and median scores, as they are robustly applicable without distributional assumptions.

Additionally, conventional paper-based CERAD-plus NAB domain scores were composed adapted from Roberts [108] of the following CERAD-plus NAB subtests: Global cognition: MMSE; Executive domain: TMT-B; Memory domain: Word list recall, Figure recall; Language domain: Semantic fluency, Phonematic fluency, Boston naming test; Visuospatial domain: Figure copying.

For the second phase of the analysis, machine learning models were used to assess the predictive value of digital drawing data for the five cognitive domains identified in the previous step. First, to establish a performance baseline, a separate set of models was created to predict domain scores using only control variables (age, sex, and education years). Subsequently, separate models were trained for each drawing task (dCDT, dFCT, and dTMT) and each cognitive domain using drawing data in conjunction with the control variables to assess the added value of the drawing parameters for predicting the domain scores, cf. Table 9.

Preliminary analysis involved the evaluation of several machine learning algorithms, including decision trees, random forests, elastic net approaches, and Huber regressions, as well as ensemble methods such as extreme gradient boosting. These models were also tested in conjunction with various feature engineering and sampling techniques. As the pattern of results remained consistent across the tested models and techniques, and as Artificial Neural Networks (ANNs) consistently demonstrated superior predictive performance, we solely focus on ANNs in the following sections. The performance of the tested models converged to a consistent range, suggesting that the superior ANN results reflect the inherent predictive limit for this dataset.

Two distinct ANN architectures were employed. For the baseline models, a multi-layer perceptron with a single hidden layer of eight neurons was used to predict the five data-driven domains from the control

variables. For the models predicting cognitive domains using digital drawing data, a more complex architecture was chosen. This network incorporated a bottleneck structure (akin to an autoencoder) to transform the high-dimensional drawing process data into a lower-dimensional representation before expanding it for optimal prediction. All features and target variables were z-standardized as part of the model pipeline prior to being passed into the network to ensure a consistent scale and to optimize the training process.

To ensure a robust and unbiased performance estimate, a 5x5-fold nested cross-validation procedure was used for model evaluation and hyperparameter tuning. This approach prevents data leakage by ensuring that the data used for final model assessment is not used during the training or hyperparameter tuning phases. The dataset was first partitioned into five non-overlapping folds. In each of the five repetitions, one fold was designated as the test set, while the remaining four constituted the training set. Within each outer training set, another five-fold cross-validation was performed to select the optimal hyperparameters. The hyperparameters tuned were the L1/L2 mix (alpha, with values of 0, 0.25, 0.50, 0.75, and 1.0) and the weight regularization penalty (lambda, with values of 0.01, 0.1, 0, 1, and 10). The best hyperparameter combination was chosen based on its performance on the inner validation folds. The model with the selected optimal hyperparameters was then trained on the full outer training set and evaluated on the held-out test set. The final performance metrics were calculated by averaging the results across all 25 test folds (5 folds x 5 repetitions), providing a reliable estimate of the model's ability to generalize to unseen data.

To provide a comprehensive evaluation of model performance, we reported three distinct metrics: the coefficient of determination (R2), the mean absolute percentage error (MAPE), and the root mean squared error (RMSE). The **R2** score was used to quantify the proportion of variance in the cognitive domain scores that our model was able to explain, providing a measure of overall goodness-of-fit. The **RMSE** and **MAPE** were included to assess the magnitude of the prediction error in absolute terms. The RMSE, being sensitive to large errors, provided an indication of the model's performance on outliers. In contrast, the MAPE offered a relative error metric, expressing the average prediction error as a percentage of the true value, which made it easier to interpret the error magnitude across different cognitive domains. Together, these metrics provided a holistic view of our model's predictive accuracy and its ability to generalize to new data.

Results

Data-driven domain identification

Affinity propagation applied to the distance correlation matrices of the CERAD-plus NAB measures consistently yielded a five-cluster solution across all three measurement time points (FU3, FU4, and FU5), except for the FU5 analysis at a self-similarity of 0.05, which favored a six-cluster solution. Based on this stability, we selected the five-cluster solution for subsequent analysis.

To enhance the robustness of the cluster membership, we retained only those measures that were consistently assigned to the same cluster across all three waves. This procedure led to the exclusion of six measures ("MMSE writing a sentence", "Figure Copying „Diamond“", "Word List delayed recall intrusions", "Figure Recall „circle“", "Figure Recall „overlapping rectangles“", "Figure Recall „pentagons“") due to inconsistent cluster assignments. The final domain structure consisted of five distinct clusters, which were labeled based on their constituent measures: attention (2 measures), memory (9 measures), intrusions (3 measures), visuospatial ability (4 measures), and executive functioning and language (henceforth executive/language; 6 measures, see Table 11 for a detailed overview of domain assignments).

The internal consistency of each domain scale was then assessed using Cronbach's Alpha. The memory (α=0.736), intrusions (α=0.641), and executive/language (α=0.728) domains showed acceptable internal consistency. The visuospatial ability domain had limited internal consistency (α=0.445), while the two-item attention domain resulted in a negative Cronbach's Alpha (α=−0.423), suggesting the items were not measuring a unitary construct. Given the acceptable internal consistency of most domains, the average value for each domain was used for subsequent analyses.

Content Validity of the Data-Driven Domains

Correlation analyses were conducted to assess the content validity of the data-driven domain structure identified in the previous step. The complete correlation matrix is displayed in Table 2.

All five data-driven domain scores were significantly and positively correlated with both the z-score based CERAD-plus NAB total score (TS2) and the sum score for the MoCA (all *p*s < .001). Further, the domains that could be directly mapped onto an established theory-driven CERAD-plus NAB domain structure showed strong positive relationships. Specifically, the memory (r=0.70, p<.001), visuospatial (r=0.59, p<.001), and executive/language domains (r=0.65, p<.001) all exhibited high correlations with their established theory-driven CERAD-plus NAB domain counterparts. However, the intrusions domain showed no significant relation to the conventional theory-driven language domain (r=0.06, p=.100), which was expected given the distinct nature of this data-driven construct.

Further, permutation tests were used to compare the mean scores of the five data-driven cognitive domains between cognitively healthy individuals and those with cognitive impairment based on a MoCA cutoff of 23. This non-parametric approach was selected to provide a robust assessment of group differences without making assumptions about the underlying data distribution.

Results for all five domains indicated a highly significant difference between the two groups (all ps < .001). For each domain, the mean score for the cognitively impaired group was significantly lower than the mean score for the healthy control group. Specifically, the mean scores for the impaired and healthy groups were as follows: for the attention domain, impaired (M = 0.64) and healthy (M = 0.66); for memory, impaired (M = 0.69) and healthy (M = 0.80); for intrusions, impaired (M = 0.87) and healthy (M =

0.92); for visuospatial, impaired (M = 0.80) and healthy (M = 0.87); and for executive/language, impaired (M = 0.63) and healthy (M = 0.75), see Table 3.

Machine Learning Analysis

To establish a baseline for predictive performance, multilayer perceptrons were trained using only control variables (i.e., age, gender, and years of education) to predict scores in each of the five cognitive domains. As detailed in Table 4, these baseline models showed strong predictive power for the executive/language domain (e.g., mean R2=0.23), and good performance for the memory domain (mean R2=0.21). In contrast, control variables offered very limited predictive power for the visuospatial domain (mean R2=0.05) and were virtually ineffective for the attention and intrusions domains (mean R2<0.02 for both).

The predictive performance of the ANNs utilizing digital drawing data and control variables is presented in Figure 3 (see Table 5, Table 6, and Table 7 for a detailed breakdown). The results showed that, irrespective of the drawing task, the attention, intrusions, and visuospatial domains could not be predicted with meaningful accuracy using digital drawing data. For these domains, the models failed to surpass the control variable baseline.

However, the digital drawing data proved highly predictive for the executive/language and memory domains. The models trained on the dCDT and dTMT data outperformed the baseline models for the executive/language domain (mean R2 of 0.34 and 0.37, respectively), and the FCT data was also highly predictive (mean R2 of 0.24). For the memory domain, only the ANN using TMT data was able to match the corresponding model using only control variables (mean R2 of 0.21 vs. the baseline's 0.21). The CDT and FCT models did not improve on the memory baseline.

To assess the contribution of individual features to the predictive models for the executive/language domain, Shapley values were computed for the ANNs trained on both the dCDT and dFCT digital drawing data. These values were calculated to quantify the marginal contribution of each feature to the model's output for each prediction, with the mean absolute Shapley value providing an overall measure of feature importance.

As detailed in Table 8, results consistently indicated that age was the strongest predictor in both models, with mean absolute Shapley values (|s|) greater than 0.023. Other control variables also contributed significantly. For the dCDT predictions, the most important drawing features were the mean straight-line distance of hover movements (|s|=0.008) and the adjusted duration (|s|=0.007). In contrast, the dFCT predictions were primarily driven by the total drawing duration (|s|=0.020) and the total distance drawn (|s|=0.008). The low magnitude of the Shapley values across all features likely reflects the high complexity of the ANN architecture and the influence of regularization on the model's weights.

Discussion

Principal Findings

Domain structure

The first phase of our analysis ought to establish a data-driven cognitive domain structure from the CERAD-plus NAB. Using affinity propagation, we consistently identified a robust five-cluster solution, which was confirmed across three consecutive measurement waves with varying sample sizes. This finding is particularly notable as it suggests a stable domain structure even as the cohort's composition changes due to attrition and conversion to cognitive impairment. The five domains identified were: attention, executive/language function, memory, visuospatial ability, and intrusions.

Our data-driven classification aligns with established, theory-driven models of cognitive function[14,36,37], reinforcing the plausibility of a multi-domain concept of cognition as described in the DSM-5. While the clustering of CERAD-plus NAB data can be sensitive to sample characteristics and parameter selection, the consistency of our findings across measurement waves adds to their generalizability.

A notable deviation from most established data-driven classifications was the emergence of an “intrusions“ domain. The domain consists of the words that were mentioned during the word list recall but did not appear in the word list and is typically not treated as a separate construct. An exception can be found in Röhr[45]. They also examined a German-language sample. The intrusions also form a separate domain in exploratory factor analyses. But these are later excluded again because they did not load significantly on the factors[45]. The fact that intrusions formed a distinct cluster suggests a unique underlying cognitive phenomenon that is not fully captured by existing memory or language domains. This highlights the potential of data-driven approaches to uncover novel aspects of cognitive functioning that may be clinically relevant. We did not use German-language standard values[15] as we used age, sex, and education as covariates/baseline features to assess their contribution to the predictions. It must be noted though that the executive/language domain we identified has a substantial language component (correlation with conventional theory-driven CERAD-plus NAB language domain $r = .71$, $p < .001$ and with conventional theory-driven executive domain $r = .65$, $p < .001$). This represents a deviation from the typical DSM-5 domain structure and highlights a strong overlap between language and executive tasks. For instance, phonematic and semantic language fluency tasks also require significant executive control.

The results of our correlational analysis and permutation tests further support the validity of these data-driven domains by demonstrating significant differences between cognitively healthy and impaired individuals across all five domains. This confirms that the identified domains are not merely statistical artifacts but reflect meaningful, clinically relevant differences in cognitive performance.

The findings from this first phase of our analysis provide a solid foundation for the second phase, as they confirm that a data-driven approach yields a reliable and clinically meaningful cognitive domain structure. This structure is a valid basis for investigating the predictive relationships between fine-grained drawing data and specific aspects of cognitive performance.

Predictions with digital data

Our results provide strong evidence for the predictive value of digital drawing data, particularly for certain cognitive domains, and offer several key insights that challenge previous findings in the literature. We hypothesized that digital drawing data would primarily relate to domains of **executive function** and **visuospatial ability**, with weaker or no associations with **memory** and other domains. Our findings partially confirm this hypothesis, demonstrating strong predictive power for executive function but not for visuospatial ability. We argue that despite a strong language component in our executive/language domain, the predictive value of the digital drawing data is primarily attributable to executive functioning. Associations with memory were lower as expected as the best performing model for this domain did not provide substantial performance increases over the baseline model, solely using age, sex and education years.

A key distinction of this study is our focus on predicting continuous domain scores rather than binary outcomes (e.g., classifying healthy vs. impaired individuals). This **regression approach** provides a more nuanced understanding of the relationship between drawing process data and cognitive performance. This is particularly important given that most prior research has utilized a classification paradigm on small, often artificially balanced samples. By using a **longitudinal cohort study with a risk sample**, our results are more representative of a typical population and can be more readily generalized to clinical practice.

The predictive success of the digital drawing TMT data for the executive/language function and memory domains is particularly noteworthy. Our results highlighted that a digital short form of the dTMT (from the MoCA) strongly predicts extensive assessments of executive functioning and language. This finding suggests that fine-grained process data from brief, digital tasks could serve as a powerful and time-efficient alternative to longer, more burdensome established paper-based assessments. The TMT's predictive power for the memory domain, while limited, is also a novel finding that warrants further investigation. Additionally, while the present study solely focused on the predictive value of digital drawing data and selected demographic variables, future work should investigate how the predictive performance can be further enhanced by including further paper-based and digital markers (e.g., speech data from the assessment[110,111] or word list recall[54]. Contrary to our initial expectations, the digital drawing data from the dCDT, dFCT, and dTMT showed very limited predictive power for the **visuospatial ability** domain. This finding appears to contradict some previous studies. We posit that this may be due to differences in data and methodology. Our study relied solely on process-based digital data rather than categorical, a-priori categorized image data of the clock drawings. Additionally, the lack of variance in the CERAD-plus NAB visuospatial scores within our sample may have limited the models' ability to find a meaningful signal. The low variance suggests that these specific scores might not be sufficiently sensitive to subtle differences in visuospatial ability within a largely cognitively healthy cohort. The figures were scored by experienced study staff according to the criteria specified in the CERAD-plus NAB. However, some of the criteria are very vague and left to the subjective interpretation of the evaluating staff (e.g., “sides of approximately equal length”, “overlap resembles the original”). Based on

our experience evaluating thousands of tests, we believe that the operationalization of the figure evaluation in the CERAD-plus NAB is insufficient. The given criteria do not distinguish well between perfect and deficient drawings. The use of digital drawings is therefore in itself an enrichment for differentiating between these drawings, not least because it is more objective than the evaluation described above.

In conclusion, our study confirms the potential of digital drawing data as a biomarker for cognitive performance, particularly for executive functioning and language. It highlights the importance of moving beyond simple classification models and demonstrates that a regression approach in a representative, at-risk sample can yield valuable insights. Specifically, accurately predicting cognitive decline in specific domains is a core component of diagnostic procedures as, for instance, different forms of dementia show varying levels of deficits in specific cognitive domains (e.g., in most FTD cases executive dysfunctions are more present in early stages whereas in early AD stages memory dysfunctions tend to manifest more commonly). Thus, differential regression approaches, such as the one presented in this study, have great potential to inform and enhance the diagnostic process. The results also suggest that the choice of digital drawing tasks and the type of data collected are critical for accurately predicting specific cognitive domains.

Relevance of this study

On the one hand, the methodological approach of mathematically determining the structure of the domains is important, as the previous classifications are very heterogeneous. In the future, it will become increasingly important to make differentiated statements about cognitive deficits at the domain level. On the other hand, statements about which cognitive functions are represented by the drawing process are important.

Predictive research in clinical settings is primarily focused on classification problems, particularly differentiating healthy and unhealthy individuals. While this approach has merit, it does not reflect the complex nature of some diseases. For instance, (mild) cognitive impairment and dementia incorporate a wealth of differential symptoms that crucially inform optimal treatment. Thus, we opted for a regression approach predicting individuals' scores in several cognitive domains. At the same time, this work is distinctive in that the drawing data used does not predict global measures but rather makes predictions at the domain level. On this basis, it can be assumed that the associations between drawing data and cognition in previous studies predicting global measures can be attributed primarily to executive function and, to some extent, to memory.

Strengths and Limitations

To the best of our knowledge, this is the first study to use a data-driven ML approach to divide the CERAD-plus NAB into domains and examine domain-specific associations with the three combined drawing tasks of the MoCA. The domain classification proved to be stable over three follow-ups, underlining its robustness in face of systematic dropouts of impaired individuals. Further, the identified

data-driven domains corresponded highly to established theory-driven domains reported in the literature and differentiated MCI and healthy participants well. The rigorous predictive modelling approach showed substantial predictive value of drawing data for selected domains of cognition.

Compared to most other samples, it is particularly noteworthy that a large sample was examined in this study. Moreover, the sample investigated in the present study is more naturalistic than most comparable studies as we did not artificially balance impaired and healthy individuals. At the same time, the results of the study are specific to the sample examined and are not entirely representative. It may therefore also be necessary to transfer the method of domain classification to other samples instead of the result.

Future Research

Future studies can show to what degree our results can be applied to other and larger samples. It will also be of interest to analyze trajectories when individuals are examined several times with the same tests. A more differentiated view of the domain structure may help in the future to tailor drug and antibody treatments more precisely to cognitive profiles and to better allocate the treatment and monitor its progress.

Thus, it can be assumed that correlations found with global measures can be primarily attributed to executive functions, depending on the data analyzed (i.e., finished drawings or process parameters).

On the one hand, our results can be used to improve and differentiate the analysis of existing data and tasks. On the other hand, our results can be used to consider how tasks should be arranged in future assessments to gain as much insight as possible into neuropsychological domains with as little testing effort as possible.

Conclusion

Our work builds a promising basis for supporting early differential diagnosis of cognitive decline, which becomes increasingly important as potential treatments become available. Specifically, our data-driven approach revealed a domain structure related to cognitive decline that is largely congruent with diagnostic standards but also highlights facets that have been scarcely researched so far (e.g., intrusions). Moreover, our machine learning predictions demonstrated the added value of digital drawing data to predict these domain scores, particularly for executive/language abilities. This presents a promising foundation for the development of digital assessment that can significantly enhance diagnostic and enable timely interventions.

Declarations

Contributions

C.M., F.W., U.S., G.W.E., A.Kv.T., A.J.F., W.M., F.M., and H.H. conceptualized the study. Methodology was developed by C.M., F.W., and F.M. Validation was performed by C.M., F.W., and U.S. Formal analysis was

conducted by F.W. Software was provided by K.D. and H.H. Investigation was carried out by C.M., U.S., A.Kv.T., A.J.F., W.M., F.M., and C.M. Data curation was done by F.W. and U.S. Resources were contributed by A.J.F., W.M., F.M., and G.W.E.. C.M. and F.W. wrote the original draft. All authors (C.M., F.W., U.S., K.D., G.W.E., A.Kv.T., A.J.F., W.M., F.M., H.H.) reviewed and edited the manuscript. Visualization was done by C.M. and F.W. Project administration was led by C.M., U.S., and A.Kv.T. Supervision was provided by G.W.E., A.J.F., W.M., F.M., and H.H. Funding was acquired by C.M., F.W., K.D., G.W.E., and F.M.

Acknowledgements

We would like to thank the study participants for the valuable time they devoted to the study over the years. We would also like to express our gratitude to Yannick Gerstorfer for the initial programming of the data recording app. We are grateful to student assistants, doctoral students and federal volunteers for supporting the numerous assessments and help with data entry.

Funding Declaration

[TuCAN: BMBF FKZ 16SV8120, TREND: ES 151/4-1, HE 8926/1-1; AOBJ: 675915 (and other fundings)] FW was supported by UKRI Economic and Social Research Council (grant number ES/W002914/1).

Ethics Declarations

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Medical Faculty of the University of Tübingen (reference number 90/2009BO2). All participants provided written informed consent prior to participation in the study. Additional study procedures introduced later (e.g., digital assessment of neuropsychological test data) were approved through an ethics amendment, and participants provided additional written informed consent.

References

1. Sims, J. R. *et al.* Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *JAMA* **330**, 512–527 (2023).
2. van Dyck, C. H. *et al.* Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine* **388**, 9–21 (2023).
3. European Medicines Agency. Leqembi (lecanemab). An overview of Leqembi and why it is authorised in the EU. https://www.ema.europa.eu/en/documents/overview/leqembi-epar-medicine-overview_en.pdf (2025).
4. Self, W. K. & Holtzman, D. M. Emerging diagnostics and therapeutics for Alzheimer disease. *Nat. Med.* **29**, 2187–2199 (2023).
5. Livingston, G. *et al.* Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *The Lancet* **404**, 572–628 (2024).

6. Rasmussen, J. & Langerman, H. Why We Need Early Diagnosis. *Degener. Neurol. Neuromuscul. Dis.* **9**, 123–130 (2019).
7. Alzola, P. *et al.* Neuropsychological Assessment for Early Detection and Diagnosis of Dementia: Current Knowledge and New Insights. *J. Clin. Med.* **13**, 3442 (2024).
8. Alzheimer, A. Über eine eigenartige Erkrankung der Hirnrinde. *Allgemeine Zeitschrift für Psychiatrie und psychisch-gerichtliche Medizin* **64**, 146–148 (1907).
9. Zhang, J. *et al.* Recent advances in Alzheimer's disease: Mechanisms, clinical trials and new drug development strategies. *Signal Transduct. Target. Ther.* **9**, (2024).
10. Alzola, P. *et al.* Neuropsychological Assessment for Early Detection and Diagnosis of Dementia: Current Knowledge and New Insights. *J. Clin. Med.* **13**, 3442 (2024).
11. Bonvino, A. *et al.* Digital Tools for Mild Cognitive Impairment: A Systematic Review and Meta-analysis of Diagnostic Accuracy and Methodological Challenges. *Neuropsychol. Rev.* https://doi.org/10.1007/s11065-025-09680-8 (2025) doi:10.1007/s11065-025-09680-8.
12. Morris, J. C., Mohs, R. C., Rogers, H., Fillenbaum, G. G. & Heyman, A. Consortium To Establish a Registry for Alzheimer's Disease (CERAD) clinical and neuropsychological assessment of Alzheimer's disease. *Psychopharmacol. Bull.* **24**, 641–652 (1988).
13. Morris, J. C. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**, 1159–1165 (1989).
14. Fillenbaum, G. G. & Mohs, R. C. CERAD (Consortium to Establish a Registry for Alzheimer's Disease) Neuropsychology Assessment Battery: 35 Years and Counting. *Journal of Alzheimer's Disease* **93**, 1–27 (2023).
15. Aebi, C. Validierung der neuropsychologischen Testbatterie CERAD-NP: eine Multi-Center Studie. *Dissertation* https://doi.org/10.5451/unibas-002728525 (2002) doi:10.5451/unibas-002728525.
16. Schmid, N. S., Ehrensperger, M. M., Berres, M., Beck, I. R. & Monsch, A. U. The Extension of the German CERAD Neuropsychological Assessment Battery with Tests Assessing Subcortical, Executive and Frontal Functions Improves Accuracy in Dementia Diagnosis. *Dement. Geriatr. Cogn. Dis. Extra* **4**, 322–334 (2014).
17. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
18. Braak, H. & Del Tredici, K. The preclinical phase of the pathological process underlying sporadic Alzheimer's disease. *Brain* **138**, 2814–2833 (2015).
19. Harvey, P. D. Domains of cognition and their assessment. *Dialogues Clin. Neurosci.* **21**, 227–237 (2019).
20. Petersen, R. C. & Negash, S. Mild cognitive impairment: an overview. *CNS Spectr.* **13**, 45–53 (2008).
21. Stokin, G. B., Krell-Roesch, J., Petersen, R. C. & Geda, Y. E. Mild Neurocognitive disorder: An old wine in a new bottle. *Harv. Rev. Psychiatry* **23**, 368–376 (2015).

22. Sachs-Ericsson, N. & Blazer, D. G. The new DSM-5 diagnosis of mild neurocognitive disorder and its relation to research in mild cognitive impairment. *Aging Ment. Health* **19**, 2–12 (2015).
23. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (5th Ed.)*. (Washington, DC:, 2013).
24. Pirlo, G. *et al.* Early Diagnosis of Neurodegenerative Diseases by Handwritten Signature Analysis. in *LNCS* vol. 9281 290–297 (Springer International Publishing Switzerland, 2015).
25. World Health Organization (WHO). *International Statistical Classification of Diseases and Related Health Problems, Eleventh Revision (ICD-11)*. (World Health Organization, 2019).
26. Sachs-Ericsson, N. & Blazer, D. G. The new DSM-5 diagnosis of mild neurocognitive disorder and its relation to research in mild cognitive impairment. *Aging Ment. Health* **19**, 2–12 (2015).
27. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (5th Ed.)*. (Washington, DC:, 2013).
28. World Health Organization (WHO). *International Statistical Classification of Diseases and Related Health Problems, Eleventh Revision (ICD-11)*. (World Health Organization, 2019).
29. Bermejo-Pareja, F. *et al.* Predementia constructs: Mild cognitive impairment or mild neurocognitive disorder? A narrative review. *Int. J. Geriatr. Psychiatry* **36**, 743–755 (2021).
30. Petersen, R. C. *et al.* Mild cognitive impairment: clinical characterization and outcome. *Arch. Neurol.* **56**, 303–309 (1999).
31. Petersen, R. C. *et al.* Mild cognitive impairment: A concept in evolution. *J. Intern. Med.* **275**, 214–228 (2014).
32. Petersen, R. C. & Morris, J. C. Mild Cognitive Impairment as a Clinical Entity and Treatment Target. *Arch. Neurol.* **62**, 1160–1163; discussion 1167 (2005).
33. Becker, S. *et al.* Cognitive-Driven Activities of Daily Living Impairment as a Predictor for Dementia in Parkinson Disease: A Longitudinal Cohort Study. *Neurology* **99**, E2548–E2560 (2022).
34. Becker, S. *et al.* Assessment of cognitive-driven activity of daily living impairment in non-demented Parkinson's patients. *J. Neuropsychol.* **14**, 69–84 (2020).
35. Liepelt-Scarfone, I. *et al.* Influence of different cut-off values on the diagnosis of mild cognitive impairment in Parkinson's disease. *Parkinsons Dis.* **2011**, (2011).
36. Lillig, R. *et al.* A new CERAD total score with equally weighted z-scores and additional executive and non-amnestic „CERAD-Plus" tests enhances cognitive diagnosis in patients with Parkinson's disease: Evidence from the LANDSCAPE study. *Parkinsonism Relat. Disord.* **90**, 90–97 (2021).
37. Ophey, A. *et al.* Cognitive profiles of patients with mild cognitive impairment due to Alzheimer's versus Parkinson's disease defined using a base rate approach: Implications for neuropsychological assessments. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **13**, 1–12 (2021).
38. Morris, J. C. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**, 1159–1165 (1989).

39. Collie, A., Shafiq-Antonacci, R., Maruff, P., Tyler, P. & Currie, J. Norms and the effects of demographic variables on a neuropsychological battery for use in healthy ageing Australian populations. *Australian and New Zealand Journal of Psychiatry* **33**, 568–575 (1999).
40. Strauss, M. E. & Fritsch, T. Factor structure of the CERAD neuropsychological battery. *Journal of the International Neuropsychological Society* **10**, 559–565 (2004).
41. Jones, S. N. & Ayers, C. R. Psychometric properties and factor structure of an expanded CERAD neuropsychological battery in an elderly VA sample. *Archives of Clinical Neuropsychology* **21**, 359–365 (2006).
42. Aguirre-Acevedo, D. C. *et al.* Diagnostic accuracy of CERAD total score in a Colombian cohort with mild cognitive impairment and Alzheimer's disease affected by E280A mutation on presenilin-1 gene. *Int. Psychogeriatr.* **28**, 503–510 (2016).
43. Guruje, O. *et al.* The CERAD Neuropsychological Test Battery: norms from a Yoruba-speaking Nigerian sample. *West Afr. J. Med.* **14**, 29—33 (1995).
44. Lee, J. H. *et al.* Development of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K): Clinical and neuropsychological assessment batteries. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* **57**, P47–P53 (2002).
45. Röhr, F. *et al.* Relationship between Lipoprotein (a) and cognitive function – Results from the Berlin Aging Study II. *Sci. Rep.* **10**, 1–10 (2020).
46. Unverzagt, F. W. *et al.* Effects of age, education, and gender on CERAD neuropsychological test performance in an african American sample. *Clin. Neuropsychol.* **10**, 180–190 (1996).
47. Beisteiner, R. *et al.* Transcranial Pulse Stimulation with Ultrasound in Alzheimer's Disease—A New Navigated Focal Brain Therapy. *Advanced Science* **7**, (2020).
48. Brusco, M. J., Steinley, D., Stevens, J. & Cradit, J. D. Affinity propagation: An exemplar-based tool for clustering in psychological research. *British Journal of Mathematical and Statistical Psychology* **72**, 155–182 (2019).
49. Henkel, C., Seibert, S. & Nichols Widmann, C. Current Advances in Computerized Cognitive Assessment for Mild Cognitive Impairment and Dementia in Older Adults: A Systematic Review. *Dement. Geriatr. Cogn. Disord.* 1–11 (2024) doi:10.1159/000541627.
50. Koo, B. M. & Vizer, L. M. Mobile Technology for Cognitive Assessment of Older Adults: A Scoping Review. *Innov. Aging* **3**, 1–14 (2019).
51. Veneziani, I. *et al.* Applications of Artificial Intelligence in the Neuropsychological Assessment of Dementia: A Systematic Review. *J. Pers. Med.* **14**, 113 (2024).
52. Polgar, S. *et al.* Towards the Development of a Tablet-Based Screening for Cognitive Symptoms of Neuropsychiatric Disorders. in *MobileHCI 2018 - Beyond Mobile: The Next 20 Years - 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, Conference Proceedings Adjunct* 253–260 (Association for Computing Machinery, Inc, 2018). doi:10.1145/3236112.3236147.

53. Mychajliw, C. *et al.* Performance Differences of a Touch-Based Serial Reaction Time Task in Healthy Older Participants and Older Participants With Cognitive Impairment on a Tablet: Experimental Study. *JMIR Aging* **7**, e48265 (2024).
54. Jannati, A. *et al.* Digital Clock and Recall is superior to the Mini-Mental State Examination for the detection of mild cognitive impairment and mild dementia. *Alzheimers Res. Ther.* **16**, 1–14 (2024).
55. Unger, S., Bayram, Z., Anderle, L. & Ostermann, T. Random Forest Classification of Cognitive Impairment Using Digital Tree Drawing Test (dTDT) Data. in *Proceedings of the 13th International Conference on Data Science, Technology and Applications* 585–592 (SCITEPRESS - Science and Technology Publications, 2024). doi:10.5220/0012859100003756.
56. Yamada, Y. *et al.* Characteristics of Drawing Process Differentiate Alzheimer's Disease and Dementia with Lewy Bodies. *Journal of Alzheimer's Disease* **90**, 693–704 (2022).
57. Hosseini-Kivanani, N. *et al.* Predicting Alzheimer's Disease and Mild Cognitive Impairment with Off-line and On-line House Drawing Tests. in *2024 IEEE 20th International Conference on e-Science (e-Science)* 1–10 (IEEE, 2024). doi:10.1109/e-Science62913.2024.10678661.
58. Greco, C. *et al.* Discriminative Power of Handwriting and Drawing Features in Depression. *Int. J. Neural Syst.* **34**, 2350069 (2024).
59. Yoon, H. & Ahn, M. Quantification of Movement Error from Spiral Drawing Test. *Sensors* **23**, 3043 (2023).
60. Lopez-de-Ipiña, K. *et al.* Automatic Analysis of Archimedes' Spiral for Characterization of Genetic Essential Tremor Based on Shannon's Entropy and Fractal Dimension. *Entropy* **20**, 531 (2018).
61. Alty, J. E., Cosgrove, J., Jamieson, S., Smith, S. L. & Possin, K. L. Which figure copy test is more sensitive for cognitive impairment in Parkinson's disease: Wire cube or interlocking pentagons? *Clin. Neurol. Neurosurg.* **139**, 244–246 (2015).
62. Zhang, X. *et al.* A tablet-based multi-dimensional drawing system can effectively distinguish patients with amnestic MCI from healthy individuals. *Sci. Rep.* **14**, 982 (2024).
63. Zhang, X. *et al.* Overview of the Complex Figure Test and Its Clinical Application in Neuropsychiatric Disorders, Including Copying and Recall. *Front. Neurol.* **12**, 1–10 (2021).
64. Frigeni, M. *et al.* Tablet-based Rey–Osterrieth Complex Figure copy task: a novel application to assess spatial, procedural, and kinematic aspects of drawing in children. *Sci. Rep.* **14**, 1–12 (2024).
65. Jennings, E. E. & Rogers, S. A. 18 Illustrating Alzheimer's: The Role of Visuospatial Abilities in Figure Copying. *Journal of the International Neuropsychological Society* **29**, 228–229 (2023).
66. Salimi, S. *et al.* Can visuospatial measures improve the diagnosis of Alzheimer's disease? *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **10**, 66–74 (2018).
67. Pauly, L. *et al.* Retrograde Procedural Memory in Parkinson's Disease: A Cross-Sectional, Case-Control Study. *J. Parkinsons Dis.* **12**, 1013–1022 (2022).
68. Park, I., Kim, Y. J., Kim, Y. J. & Lee, U. Automatic, qualitative scoring of the interlocking pentagon drawing test (PDT) based on U-net and mobile sensor data. *Sensors (Switzerland)* **20**, 1–25 (2020).

69. Beretta, L. *et al.* Two distinct pathological substrates associated with MMSE-pentagons item deficit in DLB and AD. *Neuropsychologia* **133**, 107174 (2019).
70. Caffarra, P. *et al.* The qualitative scoring MMSE pentagon test (QSPT): A new method for differentiating dementia with Lewy Body from Alzheimer's disease. *Behavioural Neurology* **27**, 213–220 (2013).
71. Galbiati, A. *et al.* Qualitative scoring of the pentagon test: A tool for the identification of subtle cognitive deficits in isolated rem sleep behavior disorder patients. *Archives of Clinical Neuropsychology* **34**, 1113–1120 (2019).
72. Mitolo, M. *et al.* The New Qualitative Scoring MMSE Pentagon Test (QSPT) as a Valid Screening Tool between Autopsy-Confirmed Dementia with Lewy Bodies and Alzheimer's Disease. *Journal of Alzheimer's Disease* **39**, 823–832 (2014).
73. Bowie, C. R. & Harvey, P. D. Administration and interpretation of the Trail Making Test. *Nat. Protoc.* **1**, 2277–2281 (2006).
74. Guo, Y. A selective review of the ability for variants of the Trail Making Test to assess cognitive impairment. *Applied Neuropsychology:Adult* **29**, 1634–1645 (2022).
75. Simfukwe, C., Youn, Y. C., Kim, S. Y. & An, S. S. Digital trail making test-black and white: Normal vs MCI. *Applied Neuropsychology:Adult* **29**, 1296–1303 (2022).
76. Karande, S. & Kulkarni, V. Prognosis of Dementia Using Early Fusion Approach with Digital Clock Drawing and Trail-Making Tests. *Journal of Computer Science* **20**, 898–908 (2024).
77. Park, S. Y. & Schott, N. The trail-making-test: Comparison between paper-and-pencil and computerized versions in young and healthy older adults. *Applied Neuropsychology:Adult* **29**, 1208–1220 (2022).
78. Jang, J.-W., Kim, K., Baek, M. J. & Kim, S. A Comparison of Five Types of Trail Making Test in Korean Elderly. *Dement. Neurocogn. Disord.* **15**, 135 (2016).
79. Libon, D. J. *et al.* Dysexecutive difficulty and subtle everyday functional disabilities: the digital Trail Making Test. *Front. Neurol.* **15**, 1–9 (2024).
80. Reitan, R. M. *Trail Making Test: Manual for Administration and Scoring*. (Reitan Neuropsychology Laboratory, Tuscon, AZ, 1992).
81. Reitan, R. M. The relation of the Trail Making Test to organic brain damage. *J. Consult. Psychol.* **19**, 393–394 (1955).
82. Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).
83. Pinto, E. & Peters, R. Literature review of the Clock Drawing Test as a tool for cognitive screening. *Dement. Geriatr. Cogn. Disord.* **27**, 201–213 (2009).
84. Thompson, L. I. *et al.* Digital Clock Drawing as an Alzheimer's Disease Susceptibility Biomarker: Associations with Genetic Risk Score and APOE in Older Adults. *Journal of Prevention of Alzheimer's Disease* **11**, 79–87 (2024).

85. Chan, J. Y. C. *et al.* Evaluation of Digital Drawing Tests and Paper-and-Pencil Drawing Tests for the Screening of Mild Cognitive Impairment and Dementia: A Systematic Review and Meta-analysis of Diagnostic Studies. *Neuropsychol. Rev.* **32**, 566–576 (2022).
86. Müller, S. *et al.* Diagnostic value of digital clock drawing test in comparison with CERAD neuropsychological battery total score for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. *Sci. Rep.* **9**, 3543 (2019).
87. Cohen, J. *et al.* Digital clock drawing: Differentiating 'thinking' versus 'doing' in younger and older adults with depression. *Journal of the International Neuropsychological Society* **20**, 920–928 (2014).
88. Kobayashi, M. *et al.* Automated Early Detection of Alzheimer's Disease by Capturing Impairments in Multiple Cognitive Domains with Multiple Drawing Tasks. *Journal of Alzheimer's Disease* **88**, 1075–1089 (2022).
89. Dion, C. *et al.* Cognitive Correlates of Digital Clock Drawing Metrics in Older Adults with and without Mild Cognitive Impairment. *Journal of Alzheimer's Disease* **75**, 73–83 (2020).
90. Jiménez-Mesa, C. *et al.* Automatic Classification System for Diagnosis of Cognitive Impairment Based on the Clock-Drawing Test. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 13258 LNCS 34–42 (Springer Science and Business Media Deutschland GmbH, 2022).
91. Jiménez-Mesa, C. *et al.* Using Explainable Artificial Intelligence in the Clock Drawing Test to Reveal the Cognitive Impairment Pattern. *Int. J. Neural Syst.* **33**, 1–16 (2023).
92. Chen, S. *et al.* Automatic dementia screening and scoring by applying deep learning on clock-drawing tests. *Sci. Rep.* **10**, 1–11 (2020).
93. Spenciere, B. & Charchat-Fichman, H. Strategies Classification of the Clock-Drawing Test Construction. *American Journal of Gerontology and Geriatrics* **1**, 1010 (2018).
94. Davoudi, A. *et al.* Classifying Non-Dementia and Alzheimer's Disease/Vascular Dementia Patients Using Kinematic, Time-Based, and Visuospatial Parameters: The Digital Clock Drawing Test. *Journal of Alzheimer's Disease* **82**, 47–57 (2021).
95. Davis, R., Libon, D. J., Au, R., Pitman, D. & Penney, D. L. *THink: Inferring Cognitive Status from Subtle Behaviors*. (2014).
96. Matusz, E. F. *et al.* Dissociating Statistically Determined Normal Cognitive Abilities and Mild Cognitive Impairment Subtypes with DCTclock. *Journal of the International Neuropsychological Society* **29**, 148–158 (2023).
97. Harvey, P. D. Domains of cognition and their assessment. *Dialogues Clin. Neurosci.* **21**, 227–237 (2019).
98. Aditya Shastry, K. & Sanjay, H. A. Artificial Intelligence Techniques for the effective diagnosis of Alzheimer's Disease: A Review. *Multimed. Tools Appl.* **83**, 40057–40092 (2023).
99. Berg, D. Is pre-motor diagnosis possible? – The European experience. *Parkinsonism Relat. Disord.* **18 Suppl 1**, S195–S198 (2012).

100. Thomann, A. E. *et al.* The Montreal Cognitive Assessment: Normative Data from a German-Speaking Cohort and Comparison with International Normative Samples. *Journal of Alzheimer's Disease* **64**, 643–655 (2018).
101. Nasreddine, Z. S., Bartusch, S. M. & Zipper, S. G. *Montreal Cognitive Assessment (MOCA)*. (2004).
102. Julayanont, P., Phillips, N., Chertkow, H. & Nasreddine, Z. S. Montreal Cognitive Assessment (MoCA): concept and clinical review. *Cognitive screening Instruments: A practical approach* 111–151 (2013).
103. Thomann, A. E., Berres, M., Goettel, N., Steiner, L. A. & Monsch, A. U. Enhanced diagnostic accuracy for neurocognitive disorders: A revised cut-off approach for the Montreal Cognitive Assessment. *Alzheimers Res. Ther.* **12**, 1–10 (2020).
104. Ehrensperger, M. M., Berres, M., Taylor, K. I. & Monsch, A. U. Early detection of Alzheimer's disease with a total score of the German CERAD. *J. Int. Neuropsychol. Soc.* **16**, 910–20 (2010).
105. Chandler, M. J. *et al.* A total score for the CERAD neuropsychological battery. *Neurology* **65**, 102–106 (2005).
106. Harris, P. A. *et al.* Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).
107. Harris, P. A. *et al.* The REDCap consortium: Building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).
108. Roberts, R. O. *et al.* The Mayo Clinic Study of Aging: Design and Sampling, Participation, Baseline Measures and Sample Characteristics. *Neuroepidemiology* **30**, 58–69 (2008).
109. Carson, N., Leach, L. & Murphy, K. J. A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *Int. J. Geriatr. Psychiatry* **33**, 379–388 (2018).
110. Tang, Q. *et al.* Optimized sequential classification models for mild cognitive impairment screening based on handwriting and speech data. *Journal of Alzheimer's Disease* **107**, 112–127 (2025).
111. Yang, J. *et al.* Combining Speech and Drawing Data for Alzheimer's Disease Detecting. *ACM International Conference Proceeding Series* 51–60 (2023) doi:10.1145/3638884.3638893.

Tables

Tables 1 to 8 are available in the supplementary files section

Figures

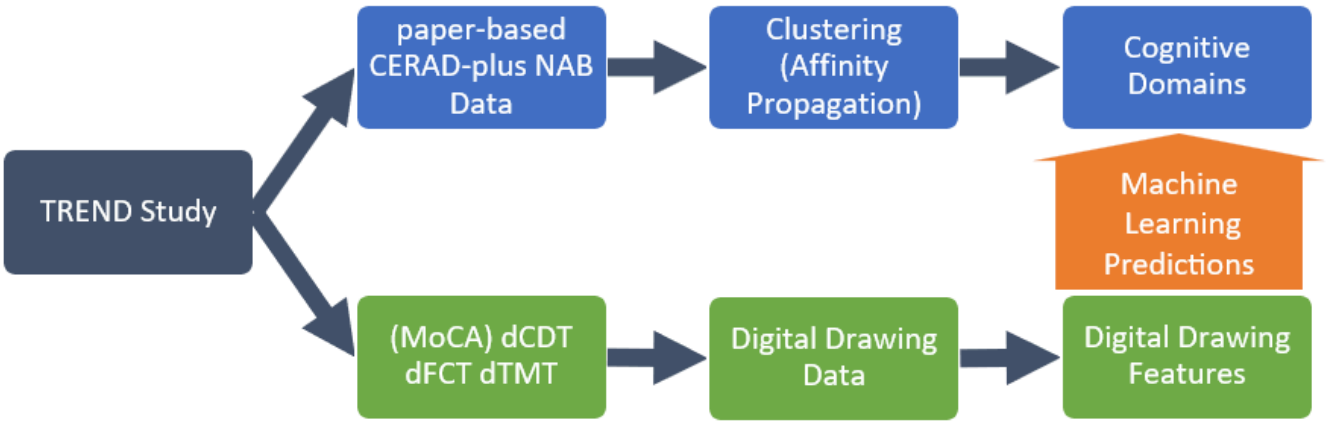
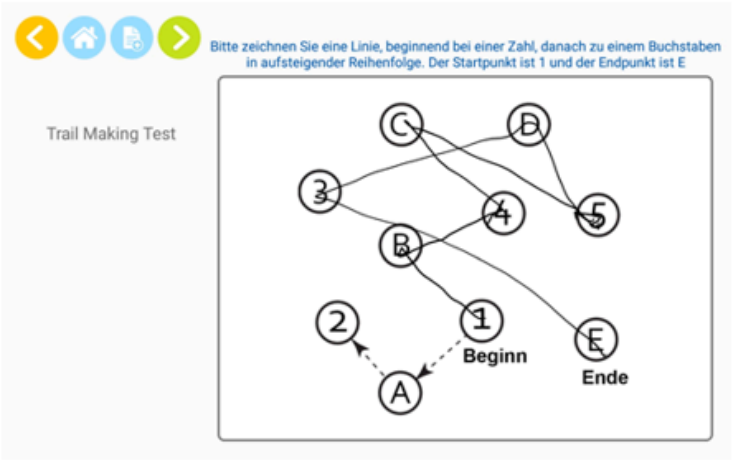


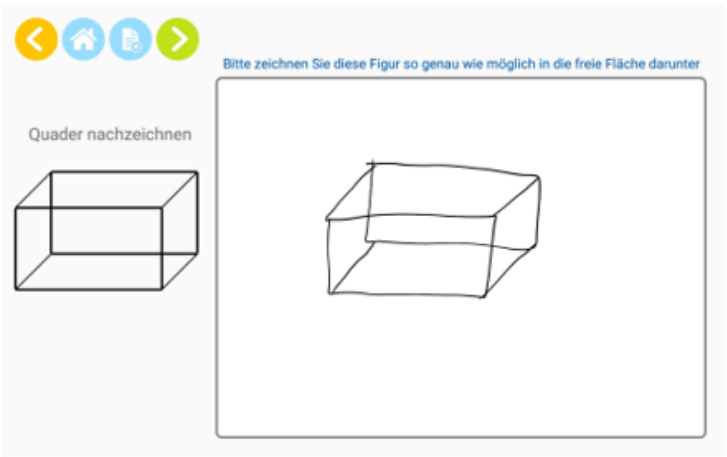
Figure 1

Illustration of the study procedure

A)



B)



C)

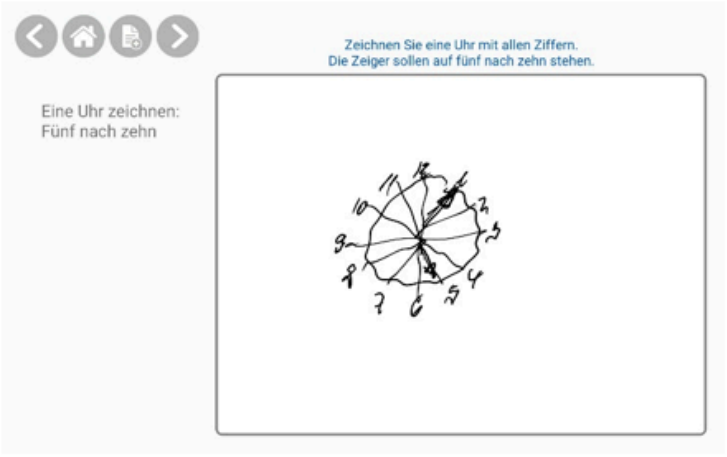


Figure 2

Examples from the drawing tasks: A) Trail Making Task (dTMT) B) Figure Copying Task (dFCT) C) Clock Drawing Task (dCDT) how it looked on the screen for the participants when drawing.

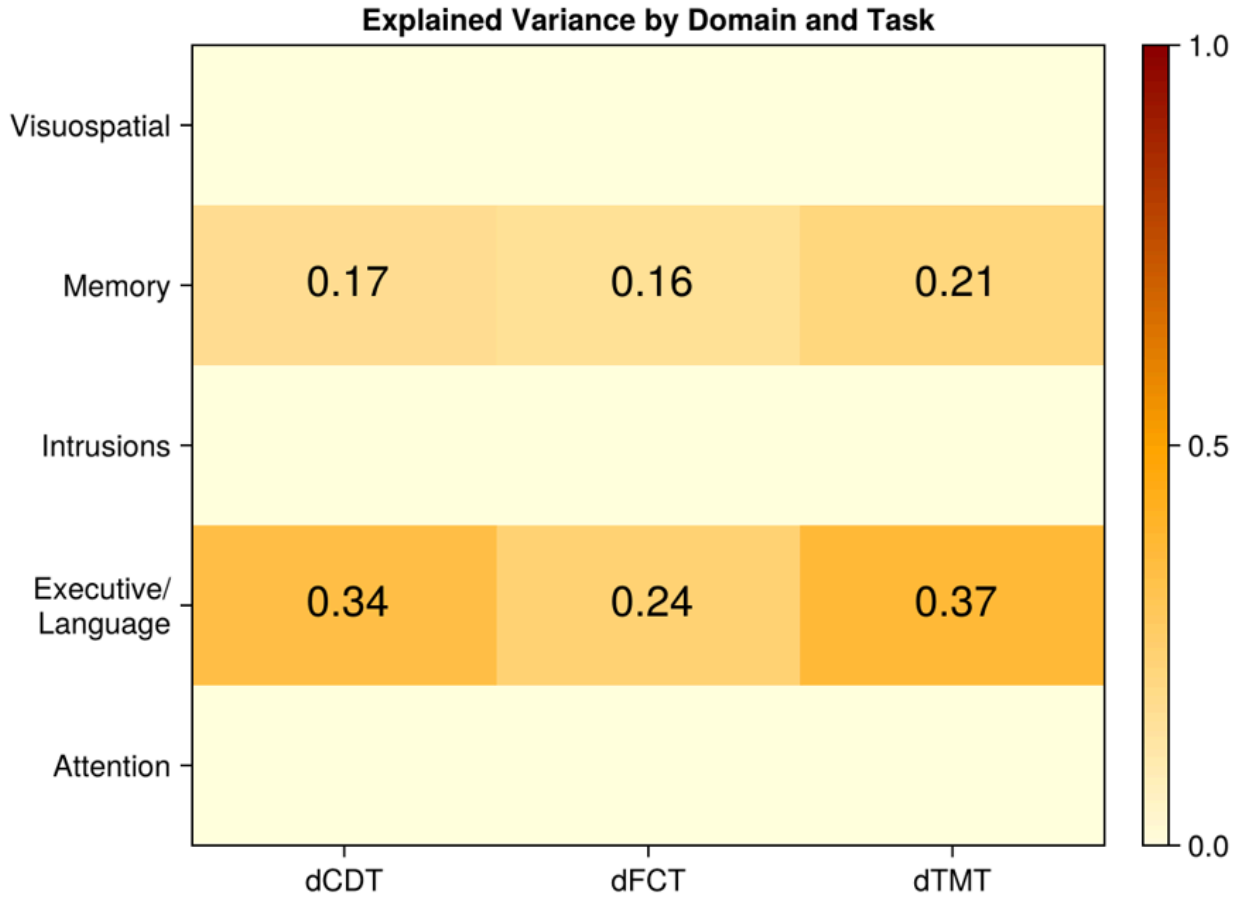


Figure 3

Illustration of the explained variance for each data-driven domain by task. Note. Negative R^2 values, indicating performance worse than predicting the sample mean, were omitted in this figure.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendi1.docx
- Tables.docx