Classifying Non-Dementia and Alzheimer's disease/ Vascular Dementia Patients Using Kinematic, Time-Based, and Visuospatial Parameters: The Digital Clock Drawing Test

Anis Davoudi[1], **Catherine Dion**[2], **Shawna Amini**[2], **Patrick J. Tighe**[3], **Catherine C. Price**[2,3], **David J. Libon**[#4], **Parisa Rashidi**[#1]

[1]Department of Biomedical Engineering, University of Florida, Gainesville, Florida, USA

[2]Department of Clinical and Health Psychology, University of Florida, Gainesville, FL, USA

[3]Department of Anesthesiology, University of Florida, Gainesville, Florida, USA

[4]Departments of Geriatrics and Gerontology and Psychology, New Jersey Institute for Successful Aging, School of Osteopathic Medicine, Rowan University, New Jersey, USA

[#] These authors contributed equally to this work.

Abstract

Background: Advantages of digital clock drawing metrics for dementia subtype classification needs examination.

Objective: To assess how well kinematic, time-based, and visuospatial features extracted from the digital Clock Drawing Test (dCDT) can classify a combined group of Alzheimer's disease/ Vascular Dementia patients versus healthy controls (HC), and classify dementia patients with Alzheimer's disease (AD) versus vascular dementia (VaD).

Methods: Healthy, community-dwelling control participants (n= 175), patients diagnosed clinically with Alzheimer's disease (n= 29), and vascular dementia (n= 27) completed the dCDT to command and copy with hands to *10 after 11"*. Thirty-seven dCDT command and 37 copy dCDT features were extracted and subjected to Random Forest machine learning analysis.

Results: When HC participants were compared to participants with dementia, optimal area under the curve was achieved using models that combined both command and copy dCDT features (AUC= 91.52%). Similarly, when AD versus VaD participants were compared, optimal area under the curve was, again, achieved with models that combined both command and copy features (AUC= 76.94%). Subsequent follow-up analyzes of a corpus of 10 variables of interest complied using a Gini Index found that groups could be dissociated based on kinematic, time-based, and visuospatial features.

Conclusions: The dCDT is able to operationally define graphomotor output that cannot be measured using traditional paper and pencil test administration in older health controls and

Conflict of interest: The authors report no conflict of interest.

participants with dementia. These data suggest that kinematic, time-based, and visuospatial behavior obtained using the dCDT may provide much needed neurocognitive biomarkers that may be able to identify and tract dementia syndromes.

Introduction

Alzheimer's disease (AD) and vascular dementia (VaD) are two common types of dementia. Current diagnostic studies for the diagnosis and classification of AD and VaD include cerebrospinal fluid assays, neuroimaging studies of degenerative alterations, and comprehensive neuropsychological testing. Yet, these methods can be expensive, invasive, and time-consuming, which limits their potential utility of widespread screening for neurodegenerative disease [1, 2]. A short, noninvasive test that can identify and screen patients at risk for neuropsychological impairment associated with these dementia disorders could help with diagnostic decision-making; and may be useful for treatment planning as disease-modifying medication becomes available.

Previous research has investigated how the Clock Drawing Test (CDT) can detect underlying neurocognitive impairment in dementia [3–6]. The traditional CDT is comprised of two conditions. In the command condition, the patient is asked, *"to draw the clock face, put in all of the numbers and set the hands to read 10 after 11."* This is followed by the copy condition where patients are asked to copy a model clock [7]. Regarding dementia identification, the standard paper and pencil administration of the clock drawing test has high specificity, but low sensitivity [6, 8]. Recently, Spenciere and colleagues (2017) examined a multitude of paper and pencil clock scoring systems and concluded that many scoring systems are able to distinguish between clinical groups. However, many paper and pencil scoring systems are difficult to operationalize, and there is no consensus regarding optimal clock drawing criteria. This latter observation is illustrated by comparing the scoring systems used by Rouleau et al [9], Royal et al [10], and Cosentino et al. [11]. These clock drawing systems differ with respect to administration procedures, scoring criteria, and the clinical groups that were assessed. Nonetheless, all three studies show meaningful between-group differences that are able to define neurocognitive constructs that underlie dementia syndromes[12].

Over the past decade, a digital version of the clock drawing test (dCDT) has been introduced by the Clock Sketch Consortium [13]. Digital clock drawing protocols use inexpensive digital pen technology. This obviates many of the problems associated with traditional pencil and paper clock drawing scoring systems. Moreover, this technology is able to capture a multitude of behaviors that cannot otherwise be extracted using traditional pen and paper tests. For example, all pen strokes are time stamped and latencies between pen strokes are recorded. This results in precise measurement of all constructional elements such as size, length, pen pressure, and drawing velocity throughout the entire test. The richness of this

information allows for a very detailed analysis of process and errors [14, 15] that can inform researchers and clinicians regarding the underlying neurocognitive impairment.

While previous studies have evaluated the use of the traditional clock drawing test for the detection and classification of dementia subtypes [9–11], the advantages of the dCDT have yet to be fully explored for this purpose. A previous research study conducted by Muller and colleagues used several digitally-obtained features measuring kinematic behavior such as pen pressure and drawing velocity extracted from digital pen data. Logistic regression analyses were able to differentiate patients with amnestic mild cognitive impairment (aMCI) versus mild Alzheimer's dementia (mAD) [16]. Subsequent research [17] using machine-learning models were able to differentiate patients with memory impairment, vascular cognitive disorders, and healthy individuals. In a third study, researchers extracted 350 features from digital clock drawings from 163 patients. Neural networks and information theory-based feature selection methods were able to classify 91.42% of AD versus non-MCI patients [18].

In the present study, we evaluated how well machine learning analysis using Random Forest [19] models can differentiate between dementia patients diagnosed with Alzheimer's Disease (AD), Vascular Dementia (VaD) associated with MRI evidence of subcortical white matter alterations, and non-demented elder controls.

Methods

Participants.

Data were acquired from two prospective research investigations approved by the University of Florida's Institutional Review Board; and, from an investigation conducted at Rowan University. Written informed consent was obtained from all participants with all investigations conducted in accordance with the Declaration of Helsinki.

Healthy Control (HC) Participants.

Inclusion criteria were: age 55 or older, English as primary language, intact instrumental activities of daily living (IADLs), and baseline neuropsychological testing negative for cognitive impairment per Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition [20]. Exclusion criteria were: the presence of neurodegenerative disorders; major medical illness including head trauma or heart disease that could induce encephalopathy; major psychiatric disorders; documented learning disabilities; a seizure disorder or other significant neurological illness; less than a sixth-grade education, and history of substance abuse. The Telephone Interview for Cognitive (TICS) Status [21] was used to screen for dementia. An in-person appointment was completed where neuropsychological and clock drawing protocols were administered along with an assessment of medical comorbidities (Charlson Comorbidity Index [22]). The presence of anxiety, depression, and ADL/IADL abilities were also assessed. Data were reviewed by a licensed clinical neuropsychologist, double scored, and double data entered for accuracy.

Dementia Participants.

Individuals with dementia were evaluated at the New Jersey Institute for Successful Aging (NJISA), Memory Assessment Program, School of Osteopathic Medicine, Rowan University. Individuals were seen by a neuropsychologist, psychiatrist, and social worker. Inclusion criteria included age 55 and up, with exclusion criteria the same as that for HC participants except neither the TICS nor the Charlson Comorbidity Index were administered. In addition to neuropsychological assessment, all dementia participants were evaluated with the Mini Mental State Exam [23], serum studies, and an MRI scan of the brain. Additional exclusion criteria included vitamin B12, folate, or thyroid deficiency. Individuals with AD and VaD associated with this study have been described in prior research studies [24]. As described in prior reports [25, 26], these individuals were diagnosed with either AD (n= 29) or VaD (n= 27) using standard diagnostic criteria, respectively.

Digital Clock Drawing Test (dCDT) Parameters.

The dCDT yields a corpus of over 2000 variables. In the current research 37 dCDT features were used for classification. This corpus of dCDT variables was chosen based on prior research [18] demonstrating the capacity of these variables to dissociate between clinical groups (Table 1).

Clock face area was [27] calculated with the formula (pi*(average of two radii)^2). The presence of anchor digits [28] was determined whether participants initially drew the digits 12, 3, 6, and 9 inside the clock face before drawing the remaining digits. Kinematic parameters including mean pen pressure, and the ratio of mean pen pressure divided by mean drawing velocity. All dCDT features described in Table 1 were extracted from the digitally acquired drawings from both the command and the copy test conditions. When a feature could not be calculated (e.g., the ratio of the hour hand length and minute hand length when a hand is missing) the feature was classified as missing. We handled missing data using Multivariate Imputation by Chained Equations (MICE) [29].

Machine Learning Analysis.

As stated above, our primary goal was to use Random Forest as a classification model to determine how well dCDT features from the three data sets (1) command, (2) copy, and (3) combined command/copy test conditions) could classify individuals into their respective diagnostic categories based upon two binary classification problems, e.g., (1) all dementia participants versus HC classification, and (b) AD versus VaD (Figure 1). Random Forest is a machine learning algorithm that has shown strong performance in public health studies [30, 31]. Random Forest models generally do not make any assumption about the attribute of distributions and also consider the interactions between the attributes. Unlike logistic regression models, Random Forest models are not limited to learning linear decision boundaries. Additionally, their ensemble and boosted approach can accommodate overfitting.

We used nested cross-validation for the development of Random Forest classifiers and five-fold cross-validation for both inner and outer folds in classifying dementia versus HC. Due to the modest sample size, we used three-fold cross-validation for both inner and outer folds

to classify AD versus VaD. In addition to accuracy, we reported several other performance metrics including area under the ROC curve (AUC), sensitivity, specificity, and the F1 score, since accuracy alone is not sufficient to reflect the performance of the model for imbalanced datasets. We repeated the experiments 50 times and reported confidence intervals for all performance metrics. We also reported the ranking of variables' importance in the final models in terms of mean decrease in Gini index. The Gini index was used to express the probability that a random element in a data set is randomly and incorrectly labeled according to the distribution of the labels. Thus, the importance of variables are ranked based on how much they each decrease the Gini index in splitting trees in a random forest model.

Descriptive Statistics.

Chi-Square analyses or *t-tests* were used to compare groups on demographic and the top-ten clock variables selected using variable importance from Random Forest models. Significance was set at $p < 0.05$. Bonferroni correction was applied to account for multiple comparisons. All analyses were performed using R version 3.5.2.

Results

Demographic and Clinical Data.

Our final dataset comprised 231 individuals (56 dementia, 175 HC; Table 2). Dementia individuals were older (dementia age = 80.04, HC age = 68.37; $p < 0.001$); had fewer years of education (dementia = 12.75, HC = 16.39, $p < 0.001$); and scored lower on the MMSE (dementia total = 22.35, HC total = 28.81, $p < 0.001$). The AD and VaD dementia groups did not differ for age. However, the VaD group was comprised of more female individuals (female/male in VaD: 85.18%, in AD: 51.72%; $p < 0.016$) with fewer years of education (VaD = 11.85 vs. AD = 13.61; $p < 0.033$; Table 3).

Machine Learning Analysis: HC versus dementia.

None of the participants produced perseverations when drawing digits; or drew digits past the number 12, or a single line connecting the numbers 11 and 2 (i.e., the ten-eleven stroke). These errors, along with missing hands and more-than-two hands variables were excluded because they were absent or occurred very infrequently (command = 2.16% and 0.43%; copy = 0% and 1.30%, respectively). Random Forest classifiers were developed using clock features extracted from command clocks, copy clocks, and combining command and copy clocks. All three models performed similarly in detecting dementia versus HC participants. However, models developed using data from both the command and copy test condition performed best; e.g., area under the curve medians differentiating between HC and dementia were command (89.76%), copy (87.54%), and command/copy (91.52%; Table 4).

Machine Learning Analysis: AD versus VaD.

The models performed similarly in classifying AD and VaD patients. Models developed using both the command and copy test conditions performed slightly better than models developed using data from command or copy clocks (AUC medians - command = 74.82%, copy = 73.70%, and command & copy = 76.94%; Table 5). Figures 2 and 3 show the median and 95% confidence interval of the top five variables for models using a single clock

condition; and the top 10 variables for models using both clock conditions according to their Gini index (ranked based on the median value).

Clock Features of Interest: HC versus Dementia.

Figure 2 shows the median and 95% confidence interval of the 10 top variables of interest for models using both command and copy features according to their Gini index (ranked based on the median value). Table 6 displays between-group analyses for these 10 variables. Features that statistically differentiated between HC versus all dementia participants were approximately equally represented between the two clock drawing test conditions (command= 4; copy= 3). As displayed in Table 6, groups were differentiated based on a combination of features measuring kinematic operations, decision-making latency, and the location of features within the clock face. Pen pressure (command & copy) was greater for HC compared to dementia. Post-clock face latency was slower for dementia as compared to HC. Finally, precision regarding the location of the hour and minute hands in relation to the clock face, center dot was more accurate for HC compared to dementia.

Clock Features of Interest: AD versus VaD.

Figure 3 shows the median and 95% confidence interval of the 10 top variables of interest for models using both clock conditions together according to their Gini index (ranked based on the median value). In this analysis, there were perhaps, more features from the copy as compared to the command test condition that differentiated between groups (copy= 4; command= 2). As displayed in Table 7 groups were differentiated based on a combination of features measuring kinematic operations and drawing speed. Pen pressure for both test conditions was greater for VaD compared to AD patients. The ratio of pen pressure and drawing velocity was greater for VaD compared to AD consistent with reduced pen pressure along with slower drawing speed. Finally, total drawing time was slower for VaD compared to AD in the copy test condition.

Discussion

The purpose of the current research was to examine how well features from the dCDT analyzed with machine learning using Random Forrest could correctly classify HC versus AD/ VaD dementia patients; and, patients diagnosed clinically with AD versus VaD. It is widely believed as symptoms and behavior suggesting either MCI or dementia emerge decades before sufficient functional and neuropsychological impairment occurs that would permit a clinical diagnosis. At such time that disease-modifying medication is available, the putative target population will most certainly be individuals early in the course of their illness. An important question revolves around the neuropsychological methods or techniques that are able to identify emergent AD/ VaD syndromes.

The MMSE and MoCA are commonly administered neuropsychological tests used to screen for neurocognitive impairment. The problems associated with the MMSE and MoCA include the lack of granularity of the behavior that can be assessed and scored. The dCDT obviates this problem. The time necessary to administer the test is very reasonable; and, as shown above, a large corpus of highly nuanced behavior can be measured. As compared to

prior digital clock drawing research, the data described above achieved comparable or better rates of classifying HC versus dementia individuals [15, 25]. Our classification rates are also comparable to prior research comparing HC versus individuals diagnosed with MCI recruited from a memory clinic [18].

Differences between Healthy Control and Dementia Participants -

From the entire corpus of approximately 2000 dCDT features, a circumscribed corpus of 37 command/ copy dCDT features were selected and assessed with machine learning analyses. Ten variables of interest emerged. The decision to use the corpus of the 37 features described above was guided by the fact that these variables measure very common and necessary behavior for successful test performance.

In the analysis classifying HC versus all dementia individuals, features that were able to dissociate between groups included kinematic parameters that measured command and copy pen pressure. In both clock drawing test conditions, pen pressure was lower when dementia patients were compared to HC participants. These data are consistent with prior clock drawing research [15]. These data are also similar to prior research associating reduced pen pressured among patients with Parkinson's disease (PD) versus HC [32] when PD patients were asked to draw an Archimedean spiral. Accurate output using a pen requires the production of comparatively gross strokes such as the clock face, and the production of discrete strokes such as digits and the clock hands [33]. The current research was able to only analyze mean pen pressure for the entire drawing. However, it is possible that greater between-group differences would have emerged to the extent pen pressure data could have been obtained from each portion of the drawing (i.e., clock face, number placement, time setting).

Time-Based dCDT Parameters -

In the copy condition only, participants with dementia produced lower Post-Clock Face Latency (PCFL). PCFL is one of a number of decision-making latencies obtained from the dCDT [34] and measures the time necessary to transition from one portion of the test to the next portion of the test. Most often PCFL measures the time between the completion of the clock face and drawing either the center dot or the number 12. Slower PCFL for dementia versus HC participants in the copy test condition might suggest the need for greater scanning back and forth from the clock model to the patient's drawing. If this is true, then slower copy PCFL suggests that copying a model of a clock might not be as automatized as expected. Concomitant measures of visual scanning behavior while patients are drawing would help disambiguate this issue. Also, in the copy condition only dementia, as compared to HC individuals were less accurate in drawing both hands in relation to the center dot.

AD versus VaD Patients -

Many of the variables of interest described above also dissociated AD versus VaD patients. Patients with AD obtained lower scores for command/ copy pen pressure compared to VaD. However, participants with VaD, relative to AD, appeared to exert greater pen pressure in the context of slower velocity. VaD patients also produced slow total time to completion in the copy test condition compared to AD patients. Both of these findings are consistent with

previous research showing that VaD patients produce less output as a function of time [35]. Prior research has argued that the neurocognitive constructs underlying clock drawing to command and clock are complimentary but not identical. These observations are consistent with the data described above showing that optimal classification was achieved using a combination of command and copy features. Finally, a question that needs to be addressed revolves around the practicality of extracting the behavior described above. Does the analysis of nuanced graphomotor output truly offer a clinical advantage? The data described above appears to answer this question in the affirmative. However, we acknowledge that additional research examining other clinical groups needs to be undertaken.

The current research is not without weakness. First, we recognize the need for replication in larger samples and correction for age/education differences between dementia and the HC participants. There is a small but statistically significant group difference in education. It is possible that lower education may negatively impact clock drawing performance; education is considered neuroprotective against the emergence of dementia [36, 37]. There is also an age difference between dementia and HC, with the HC group younger. A post-hoc analysis showed us, however, a similar model (lower AUC, 87.89 for command and copy; 87.24 command; 86.16 copy) for participants age 70 and up (67 HC, and 52 dementia). Although dementia groups did not differ for age, this is another factor that can potentially influence clock drawing behavior between and within dementia participants. Second, while we opted to use Random Forest classification models because of their overall strong performance in other health domain studies, the performance of other machine learning models needs to be evaluated. While Random Forest model performance is not affected significantly by collinearity among variables, variable importance will, nonetheless, be impacted; the reported variable importance values need to be interpreted with caution. Because of the modest and uneven sample size, weighting the outcome classes and under-sampling/ oversampling techniques may also improve the performance of the models. Third, marginally different recruitment procedures were used to recruit healthy control and dementia participants (e.g., the HC group did not complete laboratory measurements examining vitamin B12, folate, or thyroid deficiency, that may have resulted in exclusion). Finally, most of the participants in the current research were white and right-handed, suggesting the need to replicate the findings reported above with participants from other ethnic/demographic communities.

Despite these limitations, the current research has several strengths including data from two common dementia subtypes and the analysis of features that measure kinematic behavior. These data suggest that machine learning analyses of digitally acquired output using the dCDT could provide much needed neuropsychological biomarkers that might be used to identify and monitor changes in cognition over time for dementia syndromes. Future studies need to examine dCDT variables for validity and cross comparison across different test modalities (pen capture versus tablet) and dCDT variable value for differentiating HC from neurodegenerative diseases including those with and without dementia (e.g., PD; [36])

Acknowledgements:

Bibliography

[1]. Bernstein A, Rogers KM, Possin KL, Steele NZ, Ritchie CS, Kramer JH, Geschwind M, Higgins JJ, Wohlgemuth J, Pesano R (2019) Dementia assessment and management in primary care settings: a survey of current provider practices in the United States. BMC health services research 19, 919. [PubMed: 31783848]

[2]. Zverova M (2018) Alzheimer's disease and blood-based biomarkers - potential contexts of use. Neuropsychiatr Dis Treat 14, 1877–1882. [PubMed: 30050302]

[3]. Kirby M, Denihan A, Bruce I, Coakley D, Lawlor BA (2001) The clock drawing test in primary care: sensitivity in dementia detection and specificity against normal and depressed elderly. International Journal of Geriatric Psychiatry 16, 935–940. [PubMed: 11607936]

[4]. Reiner K, Eichler T, Hertel J, Hoffmann W, Thyrian JR (2018) The Clock Drawing Test: A Reasonable Instrument to Assess Probable Dementia in Primary Care? Current Alzheimer Research 15, 38–43. [PubMed: 28891446]

[5]. Ahn T-B, Na B, Lee D (2018) Visuospatial dysfunction in cube copying test and clock drawing test in non-demented Parkinson's disease: a volumetric analysis. Parkinsonism & Related Disorders 46, e50.

[6]. Palsetia D, Rao GP, Tiwari SC, Lodha P, De Sousa A (2018) The clock drawing test versus mini-mental status examination as a screening tool for dementia: A clinical comparison. Indian journal of psychological medicine 40, 1. [PubMed: 29403122]

[7]. Mainland BJ, Shulman KI (2017) Clock drawing test In Cognitive Screening Instruments Springer, pp. 67–108.

[8]. Roalf DR, Moberg PJ, Xie SX, Wolk DA, Moelter ST, Arnold SE (2013) Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. Alzheimer's & dementia : the journal of the Alzheimer's Association 9, 529–537.

[9]. Royall DR, Cordes JA, Polk M (1998) CLOX: an executive clock drawing task. Journal of Neurology, Neurosurgery & Psychiatry 64, 588.

[10]. Rouleau I, Salmon DP, Butters N, Kennedy C, McGuire K (1992) Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. Brain and Cognition 18, 70–87. [PubMed: 1543577]

[11]. Cosentino S, Jefferson A, Chute DL, Kaplan E, Libon DJ (2004) Clock Drawing Errors in Dementia: Neuropsychological and Neuroanatomical Considerations. Cognitive and Behavioral Neurology 17.

[12]. Spenciere B, Alves H, Charchat-Fichman H (2017) Scoring systems for the Clock Drawing Test: A historical review. Dementia & neuropsychologia 11, 6–14. [PubMed: 29213488]

[13]. Libon DJ, Penney DL, Davis R, Tabby DS, Eppig J, Nieves C, Bloch A, Donohue J, Brennan L, Rife K (2014) Deficits in processing speed and decision making in relapsing-remitting multiple-sclerosis: the digital clock drawing test (dCDT). J Mult Scler 1, 2.

[14]. Kaplan E (1988) The process approach to neuropsychological assessment. Aphasiology 2, 309–311.

[15]. Kaplan E (1990) The process approach to neuropsychological assessment of psychiatric patients. The Journal of Neuropsychiatry and Clinical Neurosciences.

[16]. Muller S, Herde L, Preische O, Zeller A, Heymann P, Robens S, Elbing U, Laske C (2019) 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. [PubMed: 30837580]

[17]. Souillard-Mandar W, Davis R, Rudin C, Au R, Libon DJ, Swenson R, Price CC, Lamar M, Penney DL (2016) Learning Classification Models of Cognitive Conditions from Subtle Behaviors in the Digital Clock Drawing Test. Mach Learn 102, 393–441. [PubMed: 27057085]

[18]. Binaco R, Calzaretto N, Epifano J, McGuire S, Umer M, Emrani S, Wasserman V, Libon DJ, Polikar R Machine Learning Analysis of Digital Clock Drawing Test Performance for Differential Classification of Mild Cognitive Impairment Subtypes Versus Alzheimer's Disease. Journal of the International Neuropsychological Society, 1–11.

[19]. Breiman L (2001) Random Forests. Machine Learning 45, 5–32.

[20]. American Psychiatric Association., American Psychiatric Association. DSM-5 Task Force. (2013) Diagnostic and statistical manual of mental disorders : DSM-5, American Psychiatric Association, Washington, D.C.

[21]. Cook SE, Marsiske M, McCoy KJ (2009) The use of the Modified Telephone Interview for Cognitive Status (TICS-M) in the detection of amnestic mild cognitive impairment. Journal of geriatric psychiatry and neurology 22, 103–109. [PubMed: 19417219]

[22]. Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40, 373–383. [PubMed: 3558716]

[23]. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research 12, 189–198. [PubMed: 1202204]

[24]. Emrani S, Lamar M, Price CC, Wasserman V, Matusz E, Au R, Swenson R, Nagele R, Heilman KM, Libon DJ (2020) Alzheimer's/Vascular Spectrum Dementia: Classification in Addition to Diagnosis. Journal of Alzheimer's Disease, 1–9.

[25]. Price CC, Jefferson AL, Merino JG, Heilman KM, Libon DJ (2005) Subcortical vascular dementia: integrating neuropsychological and neuroradiologic data. Neurology 65, 376–382. [PubMed: 16087901]

[26]. Price CC, Garrett KD, Jefferson AL, Cosentino S, Tanner JJ, Penney DL, Swenson R, Giovannetti T, Bettcher BM, Libon DJ (2009) Leukoaraiosis severity and list-learning in dementia. The Clinical neuropsychologist 23, 944–961. [PubMed: 19370451]

[27]. Hizel LP, Warner ED, Wiggins ME, Tanner JJ, Parvataneni H, Davis R, Penney DL, Libon DJ, Tighe P, Garvan CW, Price CC (2019) Clock Drawing Performance Slows for Older Adults After Total Knee Replacement Surgery. Anesth Analg 129, 212–219. [PubMed: 30273231]

[28]. Lamar M, Ajilore O, Leow A, Charlton R, Cohen J, GadElkarim J, Yang S, Zhang A, Davis R, Penney D, Libon DJ, Kumar A (2016) Cognitive and connectome properties detectable through individual differences in graphomotor organization. Neuropsychologia 85, 301–309. [PubMed: 27037044]

[29]. Buuren SV, Groothuis-Oudshoorn K (2010) mice: Multivariate imputation by chained equations in R. Journal of statistical software, 1–68.

[30]. Tu W, Chen PA, Koenig N, Gomez D, Fujiwara E, Gill MJ, Kong L, Power C (2020) Machine learning models reveal neurocognitive impairment type and prevalence are associated with distinct variables in HIV/AIDS. Journal of NeuroVirology 26, 41–51. [PubMed: 31520320]

[31]. Kanerva N, Kontto J, Erkkola M, Nevalainen J, Mannisto S (2018) Suitability of random forest analysis for epidemiological research: Exploring sociodemographic and lifestyle-related risk

factors of overweight in a cross-sectional design. Scand J Public Health 46, 557–564. [PubMed: 29082809]

[32]. Zham P, Kumar DK, Dabnichki P, Poosapadi Arjunan S, Raghav S (2017) Distinguishing Different Stages of Parkinson's Disease Using Composite Index of Speed and Pen-Pressure of Sketching a Spiral. Front Neurol 8, 435. [PubMed: 28932206]

[33]. HoRrE S, Shibata K (2018) Quantitative evaluation of handwriting: factors that affect pen operating skils. The Journal of Physical Therapy Science 30, 971–975. [PubMed: 30154583]

[34]. Piers RJ, Devlin KN, Ning B, Liu Y, Wasserman B, Massaro JM, Lamar M, Price CC, Swenson R, Davis R (2017) Age and graphomotor decision making assessed with the digital clock drawing test: the Framingham Heart Study. Journal of Alzheimer's Disease 60, 1611–1620.

[35]. Lamar M, Price CC, Davis KL, Kaplan E, Libon DJ (2002) Capacity to maintain mental set in dementia. Neuropsychologia 40, 435–445. [PubMed: 11684176]

[36]. Dion C, Frank BE, Crowley SJ, Hizel LP, Rodriguez K, Tanner J, Libon DJ, Price CC (2021) Parkinson's Disease Cognitive Phenotypes Show Unique Clock Drawing Features when Measured with Digital Technology. J Parkinsons Dis 11(2), 1–13.. [PubMed: 33492249]
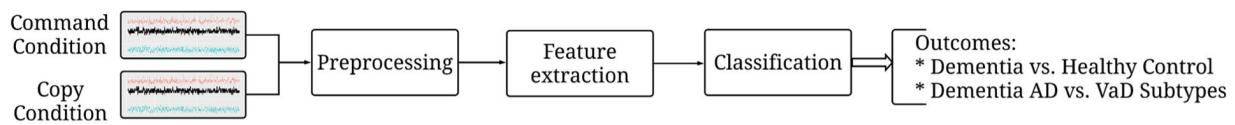
Figure 1.
Analysis flow. The analysis in the study included preprocessing the raw pen movement data, extracting features, and developing and testing classifiers.
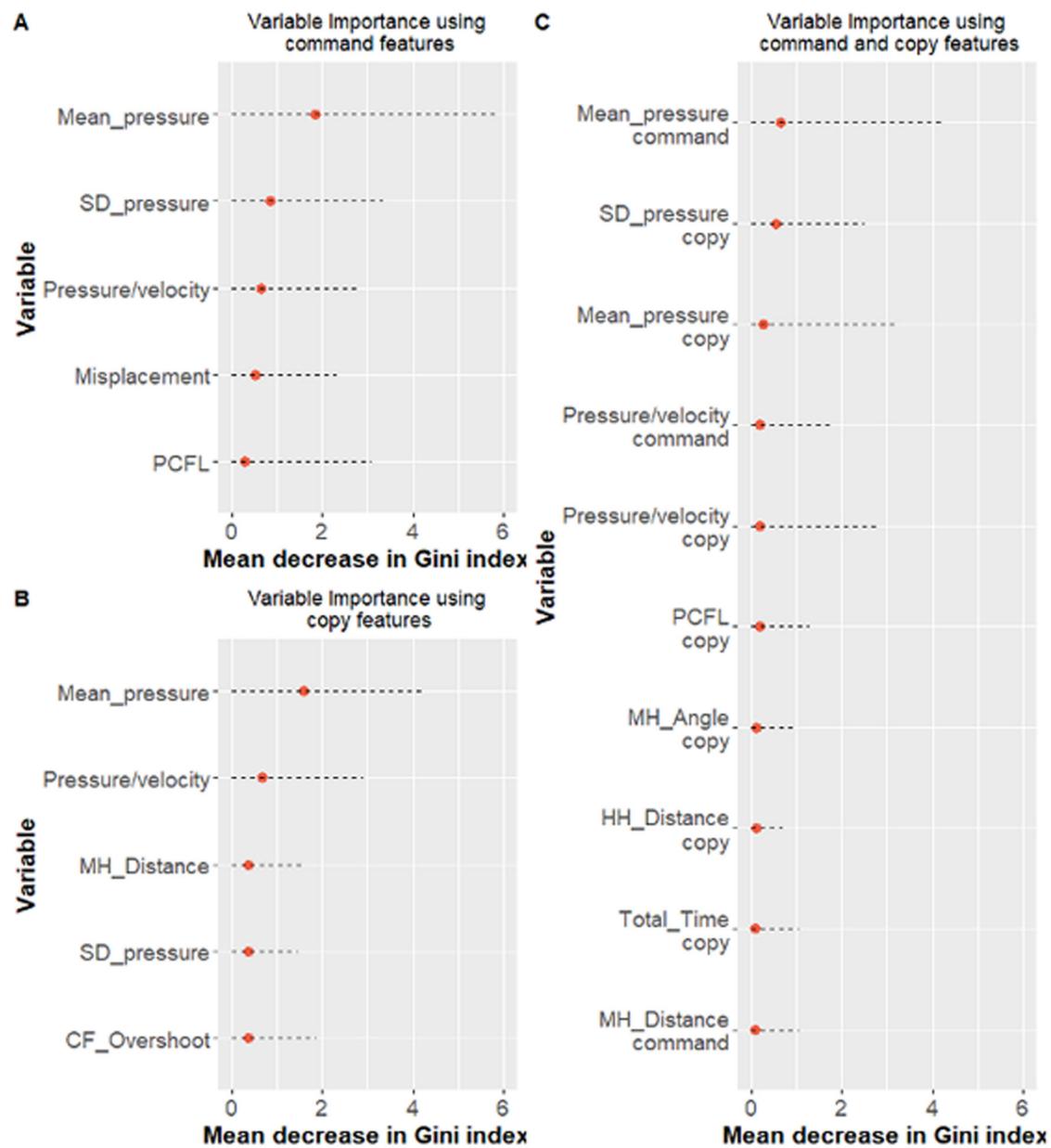
Figure 2.
Ranking of the top ten variables ranked based on their median in 50 runs for Random Forest models for classification of dementia vs HC using A) command data, B) copy data, and C) command and copy data together.
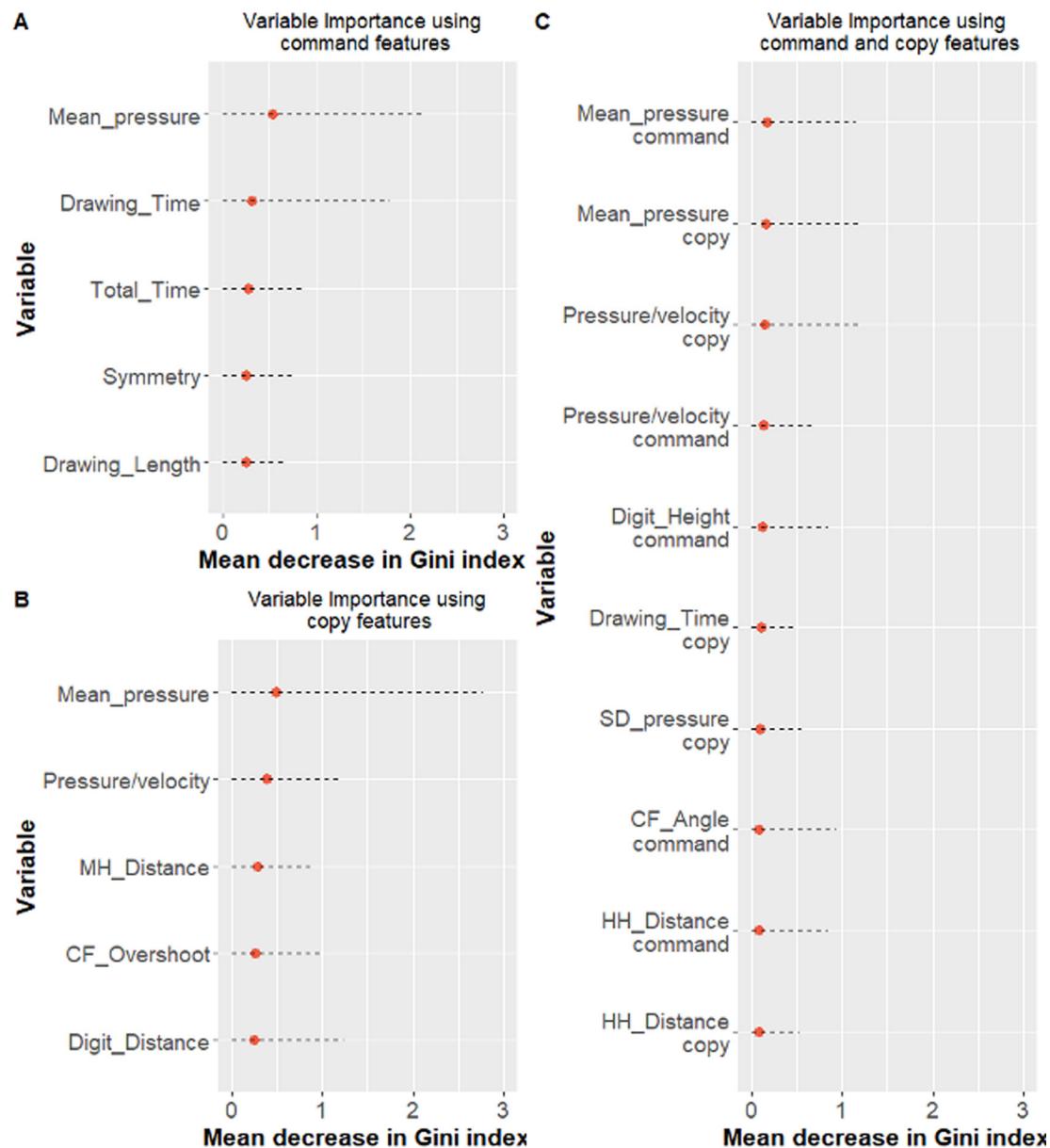
Figure 3.
Ranking of the top ten variables ranked based on their median in 50 runs for Random Forest models for classification of dementia subtypes using A) command data, B) copy data, and C) command and copy data together.

Table 1.

Digital Clock Drawing Test Feature description.

| Variable Name	Variable Description
Total number of strokes	Total number of strokes used in drawing all components of this clock
Total completion time (sec)	Total time to draw all clock components
Clockface number of strokes	Total number of strokes used to draw the clockface
Clockface total time (sec)	Time to draw the clockface
Clockface overshoot distance (mm)	The minimum distance between start of the longest clockface stroke and the end, divided by the total length of the stroke. Gives a normalized measure of how far apart or overlapped the closest points near the end of the stroke are
Clockface overshoot angle (degrees)	Number of degrees between the start and end of the longest stroke in the first clockface, measured from the center of the ellipse. Positive: end of stroke goes past the start, negative: end does not go past the start
Hour Hand length (mm)	Distance from innermost to outermost point on the hand
Hour Hand distance from center (mm)	Distance from inner end of hour hand to center of ellipse
Hour Hand difference from ideal angle (degrees)	Difference between hour hand angle and the perfect angle
Minute Hand length (mm)	Distance from innermost to outermost point on the hand
Minute Hand distance from center (mm)	Distance from inner end of minute hand to center of ellipse
Minute Hand difference from ideal angle (degrees)	Difference between minute hand angle and the perfect angle
Hour Hand length/Minute Hand length	Represents the ratio of hour hand length the minute hand length
Post-Clock Face Latency (PCFL) (sec)	Delay between finishing the last stroke in the clockface and whatever is drawn next
Pre-First Hand Latency (PFHL) (sec)	Delay between whatever was drawn before the first clock hand and starting to draw that clock hand
Pre-Second Hand Latency (PSHL) (sec)	Delay between whatever was drawn before the second clock hand and starting to draw that clock hand
Total Drawing time (sec)	Total time with the pen drawing on the paper
Drawing length (mm)	Summation of total ink length of the drawn components
Average digit height (mm)	Average height of the bounding boxes for all numbers on the clock
Average digit width (mm)	Average width of the bounding boxes for all numbers on the clock
Average digit distance from circumference (mm)	Average distance from each digit to the nearest point on the clock face
Average digit misplacement (degrees)	A summation of each digit's distance from ideal placement, divided by the number of digits
Clockface area (mm²)	Area of a circle fitted to the drawn clockface
Symmetry	Measure of how circular or oblong the clockface is, using horizontal and vertical dimensions
Velocity (mm/sec)	Drawing length/Drawing time
Average pen pressure	Average pressure of the pen on the paper during the entire test
Standard deviation of pressure	Standard deviation of pressure of the pen on the paper during the entire test
Average pressure/velocity	Average pressure/Velocity
Any digit missing	Any digits not appearing on the clock
Any digit outside	Any digits placed outside the clockface
Any digit perseveration	Any digits repeated
Anchoring	Anchoring

Variable Name	Variable Description
Any hand missing	Any of the hour hand or minute hand missing
More than one for each hand	More than one hour-hand or more than one minute-hand
Hand crossed Out	Any of the hands crossed out
Any digit over 12	Any digits over 12 drawn
Ten-eleven stroke	Single line drawn connecting the numbers 11 and 2

Table 2.

Participants Demographic/ Clinical Characteristics.

Variable	All (n=231)	Dementia (n=56)	HC (n=175)	p-value
Age, mean (SD)	71.20 (7.72)	80.04 (6.28)	68.37 (5.76)	<0.0001
Female, N (%)	118 (51.08%)	38 (67.86%)	80 (45.71%)	0.0063
Race: white N (%)	223 (96.54%)	55 (98.21%)	168 (96.00%)	0.7122
Handedness	218 (94.78%)	56 (100.00%)	162 (93.10%)	0.0943
Education, mean (SD)	15.52 (3.09)	12.75 (3.07)	16.39 (2.54)	<0.0001
MMSE, mean (SD)	27.33 (3.16)	22.35 (2.71)	28.81 (1.07)	<0.0001

Table 3.

Demographic and Clinical Characteristics of Dementia Patients.

Variable	All (n=56)	AD (n=29)	VaD (n=27)	p-value
Age, mean (SD)	80.04 (6.28)	80.34 (6.03)	79.72 (6.65)	0.7163
Female, N (%)	38 (67.86%)	15 (51.72%)	23 (85.18%)	0.0167
Race: white N (%)	55 (98.21%)	28 (96.55%)	27 (100.00%)	1
Handedness	56 (100.00%)	29 (100.00%)	27 (100.00%)	-
Education, mean (SD)	12.75 (3.07)	13.61 (2.81)	11.85 (3.12)	0.0330
MMSE, mean (SD)	22.35 (2.71)	22.24 (3.35)	22.44 (2.02)	0.7931

Table 4.

Model performance reported for the dementia vs. HC classification – AUC: Area under the ROC curve.

Model: Dementia vs. Healthy	Accuracy	AUC	Specificity	Sensitivity	F1 score
Command	90.48 (85.91, 93.74)	89.76 (84.22, 93.92)	97.71 (94.86, 98.86)	69.64 (52.19, 82.14)	77.95 (64.66, 86.32)
Copy	89.18 (84.18, 92.54)	87.54 (81.54, 92.83)	96.57 (93.84, 98.29)	66.07 (51.79, 78.17)	73.84 (61.35, 83.13)
Command & Copy	91.34 (87.64, 94.71)	91.52 (86.74, 95.17)	97.71 (96.0, 99.43)	71.43 (57.95, 83.53)	80.21 (69.61, 88.37)

Table 5.

Model performance reported for the VaD vs. AD subtype classification – AUC: Area under the ROC curve.

Model: Dementia subtypes	Accuracy	AUC	Specificity	Sensitivity	F1 score
Command	74.11 (53.57, 85.31)	74.82 (53.39, 85.70)	72.41 (55.17, 86.21)	74.07 (48.15, 95.46)	74.58 (51.07, 85.19)
Copy	73.21 (58.93, 85.31)	73.70 (59.08, 85.31)	79.31 (62.07, 89.66)	66.67 (48.15, 58.19)	72.00 (53.86, 84.70)
Command & Copy	76.78 (60.71, 87.10)	76.94 (60.75, 87.48)	79.31 (55.17, 88.87)	74.07 (56.39, 92.59)	76.00 (58.54, 87.38)

Table 6.

Ranking of the top ten command and copy clocks variables - All dementia patients versus HC participants

Variable	All (n=231)	Dementia (n=56)	HC (n=175)	t-value	df	p-value
1. Mean pressure (command)	95.16 (32.23)	52.14 (40.95)	108.93 (7.54)	10.322	56.197	<0.0001 * **D<HC**
2. Standard deviation of pressure (command)	23.27 (5.87)	19.38 (9.00)	24.51 (3.68)	4.1556	60.976	0.0001 * **D<HC**
3. Mean pressure (copy)	97.61 (29.99)	57.31 (38.02)	110.51 (6.46)	10.425	56.021	<0.0001 * **D<HC**
4. Average pressure/velocity (command)	3.03 (1.82)	2.90 (3.02)	3.07 (1.21)	0.42607	60.738	0.6716 **ns**
5. Average pressure/velocity (copy)	3.33 (2.03)	3.30 (3.51)	3.35 (1.24)	0.10497	59.47	0.9168 **ns**
6. Post clockface latency (PCFL) (copy)	1.50 (1.48)	2.23 (2.57)	1.27 (0.76)	−2.7643	58.095	0.0076 **D>HC**
7. Minute hand difference from ideal angle (copy)	−2.19 (12.30)	−0.34 (20.96)	−2.79 (7.73)	−.85617	59.861	0.3953 **ns**
8. Hour hand distance from center (copy)	3.26 (2.51)	3.98 (2.58)	3.04 (2.45)	−2.4081	88.833	0.0181 **D>HC**
9. Total completion time (copy)	30.86 (14.43)	42.50 (21.68)	27.14 (8.32)	−5.1806	60.265	<0.0001 * **D>HC**
10. Minute hand distance from center (command)	4.51 (4.56)	6.12 (4.97)	4.03 (4.33)	−2.7389	75.445	0.0077 **D>HC**

means and standard deviations;

*
signifies variables that remain significant after Bonferroni correction

Table 7.

Ranking of the top ten command and copy clocks variables comparing AD and VaD participants

Variable	AD (n=29)	VaD (n=27)	t-value	Df	p-value
1. Average pressure (command)	21.57 (31.56)	84.98 (17.66)	9.3618	44.575	<0.0001 * AD<VaD
2. Average pressure (copy)	30.23 (29.89)	86.39 (19.99)	8.3175	49.15	<0.0001 * AD<VaD
3. Average pressure/velocity (copy)	1.26 (1.31)	5.48 (3.83)	5.438	31.606	<0.0001 * AD<VaD
4. Average pressure/velocity (command)	0.92 (1.28)	5.01 (2.93)	6.6897	34.974	<0.0001 * AD<VaD
5. Average digit height (command)	5.67 (1.47)	5.15 (1.33)	−1.385	53.959	0.1718 ns
6. Total drawing time (copy)	14.36 (6.68)	20.73 (9.32)	2.922	46.872	0.0053 AD<VaD
7. Standard deviation of pressure (copy)	18.36 (7.18)	23.63 (4.27)	3.3576	46.153	0.0016 * AD<VaD
8. Clockface overshoot angle (command)	2.29 (1.50)	3.17 (3.05)	1.4703	40.065	0.1828 Ns
9. Hour hand distance from center (command)	5.85 (4.79)	7.25 (6.00)	0.9462	47.838	0.3488 Ns
10. Hour hand distance from center (copy)	3.37 (2.41)	4.63 (2.64)	1.861	52.598	0.0683 ns

mean and standard deviation;

*
signifies variables that remain significant after Bonferroni correction