Combining Multimodal Behavioral Data of Gait, Speech, and Drawing for Classification of Alzheimer's Disease and Mild Cognitive Impairment

Yasunori Yamada[a,*], Kaoru Shinkawa[a], Masatomo Kobayashi[a], Vittorio Caggiano[b,1],
Miyuki Nemoto[c], Kiyotaka Nemoto[d] and Tetsuaki Arai[d]

[a]*Digital Health, IBM Research, Chuo, Tokyo, Japan*
[b]*Healthcare and Life Sciences, IBM Research, Yorktown Heights, NY, USA*
[c]*Department of Psychiatry, University of Tsukuba Hospital, Tsukuba, Ibaraki, Japan*
[d]*Department of Psychiatry, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan*

Handling Associate Editor: Alexandra König

Abstract.
Background: Gait, speech, and drawing behaviors have been shown to be sensitive to the diagnosis of Alzheimer's disease (AD) and mild cognitive impairment (MCI). However, previous studies focused on only analyzing individual behavioral modalities, although these studies suggested that each of these modalities may capture different profiles of cognitive impairments associated with AD.
Objective: We aimed to investigate if combining behavioral data of gait, speech, and drawing can improve classification performance compared with the use of individual modality and if each of these behavioral data can be associated with different cognitive and clinical measures for the diagnosis of AD and MCI.
Methods: Behavioral data of gait, speech, and drawing were acquired from 118 AD, MCI, and cognitively normal (CN) participants.
Results: Combining all three behavioral modalities achieved 93.0% accuracy for classifying AD, MCI, and CN, and only 81.9% when using the best individual behavioral modality. Each of these behavioral modalities was statistically significantly associated with different cognitive and clinical measures for diagnosing AD and MCI.
Conclusion: Our findings indicate that these behaviors provide different and complementary information about cognitive impairments such that classification of AD and MCI is superior to using either in isolation.

Keywords: Drawing, gait analysis, handwriting, machine learning, speech, voice, walking

INTRODUCTION

As the world's elderly population increases, the number of people living with dementia is increasing rapidly, becoming a serious health and social problem. Alzheimer's disease (AD) is the most common form of dementia and may account for an estimated

---

[1]Present address: Facebook AI Research, New York, NY, USA.

*Correspondence to: Yasunori Yamada, Digital Health, IBM Research, 19-21 Nihonbashi Hakozaki-cho, Chuo, Tokyo, 103-8510, Japan. Tel.: +81 80 6706 9381; E-mail: ysnr@jp.ibm.com.

60 to 80% of cases [1, 2]. Although no cure of AD is available, there is an urgent need for the early identification of AD, especially at its early stages, e.g., mild cognitive impairment (MCI), to enable the individual and their carers to prepare for the future and receive appropriate care to help manage symptoms [3]. A growing body of evidence also suggests modifying risk factors and interventions that could prevent or delay progression to AD [3–5]. However, even dementia diagnostic coverage worldwide remains so low that only 40–50% of people with dementia have been diagnosed even in high-income countries [6]. In this context, there is growing interest in developing accurate and easy-to-perform screening tools for the early identification of AD and MCI.

Gait, speech, and drawing behaviors individually have been shown to change in AD and MCI patients [7–27], and discrete characteristics of these behavioral impairments are associated with specific cognitive impairments [7–13, 18, 21, 22, 27]. For example, slower gait speed and greater gait variability (e.g., step-to-step fluctuations in time) occur even in the prodromal stages of AD and has been suggested to be particularly associated with impaired executive function [7–12]. Speech and language impairments, including changes in phonetic characteristics such as speech rate and pause as well as in linguistic characteristics such as lexical and semantic content, have been observed in the early stages of AD [15–18] and may enable to predict the onset of AD [14, 15, 20]. The associations between quantitative speech characteristics and specific cognitive impairments have also been shown. For example, increased pauses as well as a reduction in relevant information during a picture description task were correlated with impairments in episodic memory [18, 21]. Drawing tasks, such as trail-making and clock-drawing tests, have been widely used for screening cognitive impairment and dementia. In addition to the drawing outcome, the characteristics of the drawing process such as reduced drawing speed as well as increased pauses between drawings have been reported as statistically significant signatures for detecting AD and MCI [22–25] and impaired global cognition [22]. Therefore, these behavioral analyses are expected to help in the early detection of AD and MCI as useful screening tools. Furthermore, a gait test may be incorporated into the clinical practice by using sensors, such as an electronic walkway or wearable sensors [13, 27–30], and speech and drawing data can be collected during standard neuropsychological tests for the diagnosis of dementia [19, 20, 22–24, 31].

Although most existing research investigated each behavior in isolation, the heterogeneity of the brain regions involved in their execution [12, 13, 24, 27, 29, 32–36] as well as of their behavior-cognitive relationships [7–13, 18, 21, 22, 27] suggests that each of these behaviors may capture different and complementary profiles of cognitive impairments associated with AD. Therefore, we hypothesized that gait, speech, and drawing behaviors that provide complementary information about cognitive impairments can be combined to provide higher accuracy for AD and MCI detection.

Here, we aimed to investigate if combining behavioral data of gait, speech, and drawing can improve classification performance compared with the use of individual modality and if each of these behavioral data can be associated with different cognitive and clinical measures for the diagnosis of AD and MCI. To this end, we collected gait, speech, and drawing behaviors as well as cognitive and clinical measures from the same participants and investigated them by using statistical analysis and machine learning models.

METHODS

Participants

We recruited outpatients from the Department of Psychiatry, University of Tsukuba Hospital, the spouses of the patients, and other participants either through local recruiting agencies or advertisements in the community in Ibaraki, Japan. Inclusion criteria for the patients were a diagnosis of AD or MCI. The AD patients were in mild to moderate stages. MCI subtypes were not examined. Controls were age-matched to the patients. Patients were excluded if they had diagnoses of non-AD dementia (dementia with Lewy bodies, frontotemporal dementia, vascular dementia, etc.), non-AD MCI, or other serious diseases or disabilities that would interfere with the behavioral data collection. All examinations were conducted in Japanese. This study was conducted under the approval of the Ethics Committee, University of Tsukuba Hospital (H29-065 and R01–168) and followed the ethical code for research with humans as stated by the Declaration of Helsinki. All participants provided written informed consent to participate in this study.

Cognitive performance of all participants was assessed using the following neuropsychological

examinations conducted by neuropsychologists: the Mini-Mental State Examination (MMSE) for global cognition, immediate and delayed recall of the logical memory-story A of the Wechsler Memory Scale-Revised (LM-IA and LM-IIA) for episodic memory, the Frontal Assessment Battery (FAB) for executive function, Trail Making Test part-A (TMT-A) for information processing speed and Trail Making Test part-B (TMT-B) for executive function and attention, and the Clock Drawing Test (CDT) for visuospatial function. For clinical scales, all participants were administered the Clinical Dementia Rating (CDR), Geriatric Depression Scale (GDS), body mass index (BMI), Barthel Index of Activities of Daily Living (ADL), and Lawton Instrumental Activities of Daily Living (IADL). The severity of medial temporal lobe (MTL) atrophy was also evaluated using structural magnetic resonance imaging (MRI) (please see the Supplementary Material for more details).

Regarding the diagnosis of AD, MCI, and cognitively normal (CN), two psychiatrists (authors T.A. and K.N.), who were experts in dementia and blinded to the results of behavioral data analysis, discussed each case based on the clinical record as well as the neuropsychological scores and clinical scales, and confirmed the diagnoses of AD, MCI, and CN. AD patients fulfilled the National Institute on Aging and Alzheimer's Association (NIA-AA) criteria for probable AD dementia [37]. They also fulfilled the AD Neuroimaging Initiative (ADNI) criteria for AD in terms of MMSE, CDR, and LM-IIA scores [38], except that we included patients with moderate AD based on Benoit et al.'s criteria [39]. Specifically, AD patients fulfilled the following criteria: MMSE score 10–26; the CDR global score 0.5 or 1; and LM-IIA score ≤8 for 16 years of education (YoE), ≤4 for 8–15 YoE, and ≤2 for 0–7 YoE. MCI patients fulfilled the NIA-AA criteria for MCI [40]: 1) cognitive concern by the subject, informant, or clinician; 2) impairment in one or more cognitive domains; 3) essentially normal functional activities; and 4) absence of dementia. They also fulfilled at least two of the three ADNI criteria for MCI [38]: MMSE score 24–30; the CDR global score 0.5, with the memory box score being 0.5 or greater; and LM-IIA score ≤8 for 16 YoE, ≤4 for 8–15 YoE, and ≤2 for 0–7 YoE. CN participants did not fulfill the NIA-AA criteria for MCI or dementia [37, 40]. They also fulfilled at least two of the three ADNI criteria for CN [38]: MMSE score 24–30; the CDR global score 0; and LM-IIA score ≥9 for 16 YoE, ≥5 for 8–15 YoE, and ≥3 for 0–7 YoE.

Behavioral data collection and feature extraction

For the gait task, participants walked at their usual pace over nine meters with a marker-based motion capture system with an eight-camera Opti-Track Flex13, sampled at 120 Hz using OptiTrack Motive software 2.1.0 Beta 1 (NaturalPoint, Inc, Corvallis, OR, USA), and we analyzed three trials per participants. Speech and drawing tasks were selected from a set of neuropsychological tasks frequently used for detecting AD or MCI [19, 22–24, 31]. Participants performed five speech tasks with a tablet device (iPad Air 2): counting backwards, subtraction, phonemic and semantic verbal fluency, and picture description with the Cookie Theft picture adapted from the Boston Diagnostic Aphasia Examination [41]. Speech data were recorded by using the iPad's internal microphone (core audio format, 44,100 Hz, 16-bit). Participants also performed the following five drawing tasks during the neuropsychological assessment using a digitizing tablet and pen (Wacom Cintiq Pro 16; sampling rate: 180 Hz, pen pressure levels: 8192): writing a sentence about anything and the copy intersecting-pentagon item of MMSE (hereinafter called Sentence and Pentagon), TMT-A and TMT-B, and CDT.

We extracted behavioral features based on previous studies to facilitate interpretations (for gait [7, 8, 27, 33, 42], for speech [19, 31, 43–51], for drawing [22–25]). Here, we provide an overview of the behavioral features, and a full description is provided in the Supplementary Material. The 35 gait features consisted of those associated with pace (e.g., gait speed and step/stride length), rhythm (e.g., step/stride time), variability (e.g., step/stride time variability), left-right asymmetry (e.g., the difference between left-right step/stride time), and postural control (e.g., maximum toe clearance). The 84 speech features consisted of acoustic features (e.g., Mel-frequency cepstral coefficients (MFCCs)) and prosodic features (e.g., pitch variability and proportion of pause duration) extracted from audio data as well as linguistic features (e.g., proportion of mistakes in both calculation tasks) extracted from manually transcribed text data. Linguistic features during the picture description task included Honoré's statistics for measuring vocabulary richness [48] and the number of unique entities that a participant described in the picture, referred to as information units, for measuring informativeness [43, 44, 49–51]. The 60 drawing features consisted of those associated with kinematics (e.g., drawing speed), pressure-related (e.g., pressure vari-

ability), time-related (e.g., pause durations between drawings), and TMT-specific features (e.g., time duration between nodes). For drawing, we focused on features that represent the drawing process, and thus did not investigate features representing the final product of drawing such as linguistic aspects of the sentence and characteristics of the clock face in CDT. Variability features of all modalities were calculated using the standard deviation (SD), except that variabilities of step/stride length and writing pressure were calculated using the coefficient of variation (CV).

Classification model

The classification models were built on the basis of behavioral features using multiple types of machine learning models with automatic feature selection. The models included k-nearest neighbors [52], random forest [53], and support vector machine (SVM) [54] implemented using the Python package scikit-learn (version 0.23.2). For the feature selection, we used a sequential forward selection algorithm. Model performances were evaluated by using accuracy and the area under receiver operating characteristic curve (AuROC) obtained from 20 iterations of ten-fold cross-validation methods. A three-class AuROC was computed as defined by Hand and Till [55]. For more details including model parameters, please see the Supplementary Material.

Statistical analysis

Group differences between CN, MCI, and AD in demographics and cognitive/clinical measures were examined by using the chi-square test for sex and one-way analysis of variance (ANOVA) tests for other continuous data. Shapiro-Wilk tests were used to check normality assumption and behavioral features that did not fit the normal distribution were power transformed to normalize the distributions. Between-group comparisons of behavioral features were conducted with one-way ANOVAs adjusted for age and sex. Pairwise multiple comparisons (Bonferroni adjusted p values) were performed when comparing individual diagnostic groups. We did not perform multiple-comparison corrections across behavioral features, and the statistical significance was set to a two-sided $p < 0.05$.

Multiple linear regression controlling for age and sex as covariates was used to investigate associations between cognitive/clinical measures and behav-

ioral features selected in the classification model. Specifically, the dependent variable was one of the cognitive or clinical measures for MMSE, LM-IA, LM-IIA, FAB, TMT-A, TMT-B, CDT, GDS, ADL, IADL, and MTL atrophy. As for the independent variables, we first performed principal component analyses to limit multiple comparisons and extract uncorrelated behavioral factors. We then used the first two principal components (PCs) of each behavioral modality as the independent variables in addition to age and sex as covariates. Thus, the number of independent variables were eight.

RESULTS

Participant characteristics and behavioral data

A total of 135 participants were recruited for this study and met the inclusion criteria, and behavioral data collection was performed on them for all three behavioral modalities. 17 participants were excluded from this analysis because of problems in the data processing due to noise in the motion capture data. This yields a total of 118 participants: 47 CN participants, 45 MCI patients, and 26 AD patients. Table 1 lists the clinical and demographic information of all these participants. Regarding the demographics, age, and years of education did not differ between groups, while there was a significantly lower proportion of female participants in the MCI group compared with the control group (Table 1). For the cognitive and clinical measures, BMI, GDS, and ADL did not differ between groups, and other variables were different between the three diagnostic categories (Table 1).

For the behavioral data collection, all participants completed the three gait trials and five drawing tasks and 114 of the 118 participants completed all five speech tasks. For full information of the missing data, please see the Supplementary Material.

Differences in behavioral features between AD, MCI, and controls

In an adjusted model controlling for age and sex, 18 of the 35 gait features, 13 of the 84 speech features, and 27 of the 60 drawing features showed statistically significant differences between the diagnostic groups (all $p < 0.05$; Table 2 and Supplementary Table 1). Among these 58 features with significant differences between diagnostic groups, 51 features (i.e., 87.9%) showed larger changes from CN in AD compared with MCI, indicating gradual changes of behavioral features from CN to MCI and to AD.

Table 1
Demographic, cognitive, and clinical measures for CN, MCI, and AD

Characteristics	CN (N = 47)	MCI (N = 45)	AD (N = 26)	p
Age	72.3 ± 4.2	74.1 ± 4.8	75.2 ± 7.5	0.0692
Sex (% Female)	30/47 (63.8)[M]	17/45 (37.8)[C]	16/26 (61.5)	**0.0279**
Education	13.0 ± 2.0	13.7 ± 2.6	12.3 ± 2.5	0.0688
Body Mass Index	23.3 ± 2.8	23.1 ± 2.7	21.8 ± 3.5	0.1030
Mini-Mental State Examination[a]	27.6 ± 1.9[A]	26.8 ± 2.0[A]	20.1 ± 2.9[C,M]	**< 0.0001**
LM-IA[b]	11.4 ± 3.5[M,A]	7.4 ± 3.3[C,A]	2.1 ± 2.0[C,M]	**< 0.0001**
LM-IIA[c]	9.5 ± 3.1[M,A]	4.8 ± 3.2[C,A]	0.4 ± 0.9[C,M]	**< 0.0001**
Frontal Assessment Battery[d]	13.6 ± 2.5[A]	12.9 ± 3.2[A]	8.7 ± 2.8[C,M]	**< 0.0001**
Trail making test-part A [s]	34.8 ± 12.0[M,A]	43.1 ± 16.8[C,A]	68.7 ± 44.3[C,M]	**< 0.0001**
Trail making test-part B [s]	93.4 ± 51.0[M,A]	138.2 ± 76.0[C,A]	242.7 ± 74.3[C,M]	**< 0.0001**
Clock Drawing test[e]	6.7 ± 0.9[A]	6.6 ± 0.7[A]	5.5 ± 2.2[C,M]	**0.0003**
Geriatric Depression Scale[f]	2.9 ± 2.9	3.0 ± 2.8	4.0 ± 3.2	0.3355
Clinical Dementia Rating[†]	0.0 ± 0.1[M,A]	0.5 ± 0.0[C,A]	0.8 ± 0.2[C,M]	**< 0.0001**
Activity of Daily Living[g]	99.9 ± 0.7	99.3 ± 2.3	98.1 ± 5.6	0.0530
Instrumental Activity of Daily Living[h]	7.9 ± 0.4[M,A]	7.3 ± 1.0[C,A]	5.4 ± 2.1[C,M]	**< 0.0001**
Medial temporal lobe atrophy	0.8 ± 0.5[M,A]	1.2 ± 0.8[C,A]	2.4 ± 1.2[C,M]	**< 0.0001**

Data displayed as mean ± standard deviations were assessed using one-way ANOVAs, whereas number (percentage) were assessed using the chi-square test. Bold values highlight statistically significant differences. Pairwise multiple comparisons (Bonferroni adjusted *p* values) were performed when comparing individual diagnostic groups. Significant differences between groups are marked with C, M, or A (C: Different to CN, M: Different to MCI, A: Different to AD). [†]The CN group includes one participant with a CDR test score of 0.5. [a]The total possible score ranges from 0 to 30. [b]LM-IA: immediate recall of the logical memory-story A of the Wechsler memory scale-revised; the total possible score ranges from 0 to 25. [c]LM-IIA: delayed recall of the logical memory-story A of the Wechsler memory scale-revised; the total possible score ranges from 0 to 25. [d]The total possible score ranges from 0 to 18. [e]The total possible score ranges from 0 to 7. [f]The total possible score ranges from 0 to 15. [g]The total possible score ranges from 0 to 100. [h]The total possible score ranges from 0 to 8.

Pairwise comparisons with Bonferroni corrections identified statistically significant differences in behavioral features between CN participants and patients with AD or MCI (Table 2 and Supplementary Table 1). We found features with statistically differences between CN participants and AD patients from all categories of behavioral features in each modality and from all tasks investigated in this study. Specifically, compared with CN participants, AD patients walked with slower gait speed related to pace ($p < 0.001$); with slower stride time related to rhythm ($p = 0.027$); with greater stride time variability ($p = 0.044$); with greater step time left-right asymmetry ($p = 0.043$); and with lower maximum toe clearance related to postural control ($p < 0.001$). Speech of AD patients showed significant changes in acoustic features such as lower MFCC1 during the picture description task ($p = 0.023$); in prosodic features such as less pitch variability during the counting backwards task ($p = 0.009$) and greater proportion of pause durations both in spontaneous speech during the picture description task ($p = 0.001$) and in word production during the semantic verbal fluency task ($p < 0.001$); and in linguistic features such as increased proportion of mistakes during both calculation tasks ($p < 0.001$) and lower vocabulary richness measured by Honoré's statistics and less

informativeness measured by the total number of information units normalized by the length of the speech during the picture description task ($p = 0.019$, $p = 0.009$). Drawings of AD patients showed significant differences in kinematics such as slower drawing speed in TMT-B ($p < 0.001$); in pressure-related features such as greater pressure variability in both TMT-A ($p = 0.010$) and TMT-B ($p < 0.001$); in time-related features such as longer pause duration between drawings normalized by total stroke length in all five tasks (Sentence: $p = 0.005$, Pentagon: $p = 0.023$, TMT-A and TMT-B: $p < 0.001$; CDT: $p = 0.003$); and in TMT-specific features such as longer time duration between nodes in both TMT-A ($p = 0.008$) and TMT-B ($p = 0.026$).

MCI patients had also statistically significantly different features from CN participants in parts of behavioral feature categories and in parts of tasks. Gait of MCI patients was significantly different in two categories of gait features: pace, such as lower peak gait speed ($p = 0.030$) and shorter step length ($p = 0.040$), and postural control, such as lower maximum toe clearance ($p = 0.020$). Speech of MCI patients showed significant changes in only acoustic features of MFCCs during the picture description tasks ($p \leq 0.015$) and no significant differences in other prosodic and linguistic features. Drawing

Table 2
Examples of behavioral features with statistically significant differences between diagnosis categories of AD, MCI, and CN

		CN (N = 47)	MCI (N = 45)	AD (N = 26)	p
Gait					
Pace	Gait speed [m/s]	1.33 ± 0.16^A	1.24 ± 0.21^A	$1.07 \pm 0.18^{C,M}$	**< 0.0001**
	Step length [m]	$0.68 \pm 0.07^{M,A}$	$0.65 \pm 0.09^{C,A}$	$0.58 \pm 0.07^{C,M}$	**< 0.0001**
Rhythm	Stride time [s]	1.04 ± 0.07^A	1.07 ± 0.08	1.11 ± 0.11^C	**0.0098**
Variability	Stride time variability (SD) [s]	0.06 ± 0.02^A	0.07 ± 0.03	0.08 ± 0.03^C	**0.0292**
Left-right asymmetry	Step time left-right asymmetry [s]	0.04 ± 0.02^A	0.05 ± 0.03	0.07 ± 0.05^C	**0.0182**
Postural control	Maximum toe clearance [m]	$0.13 \pm 0.02^{M,A}$	0.12 ± 0.03^C	0.11 ± 0.02^C	**0.0009**
Speech					
Acoustic	MFCC1 (Picture description)	-496.2 ± 31.9^A	-492.6 ± 36.3	-514.4 ± 35.5^C	**0.0273**
Prosodic	Pitch variability (SD) [Hz] (Counting backwards)	23.3 ± 9.6^A	19.6 ± 9.4	16.8 ± 5.4^C	**0.0148**
	Proportion of pause duration (Semantic verbal fluency)	0.75 ± 0.07^A	0.77 ± 0.05^A	$0.84 \pm 0.07^{C,M}$	**< 0.0001**
Linguistic	Proportion of mistakes (Subtraction)	0.13 ± 0.17^A	0.20 ± 0.23	0.36 ± 0.28^C	**0.0005**
	Honoré's statistics [x10³] (Picture description)	1.93 ± 0.96^A	2.07 ± 0.98^A	$1.34 \pm 0.65^{C,M}$	**0.0061**
	Normalized total number of information units† [words/s] (Picture description)	0.27 ± 0.10^A	0.26 ± 0.12^A	$0.19 \pm 0.09^{C,M}$	**0.0098**
Drawing					
Kinematics	Drawing speed [mm/s] (TMT-B)	102.0 ± 18.5^A	100.5 ± 28.4^A	$79.0 \pm 20.8^{C,M}$	**0.0006**
Pressure-related	Pressure variability (CV) (TMT-B)	$0.077 \pm 0.037^{M,A}$	0.124 ± 0.068^C	0.149 ± 0.074^C	**< 0.0001**
Time-related	Normalized pause duration between drawings‡ [s/mm] (Sentence)	0.016 ± 0.009^A	0.019 ± 0.012	0.025 ± 0.013^C	**0.0154**
	Normalized pause duration between drawings‡ [s/mm] (Pentagon)	0.027 ± 0.015^A	0.032 ± 0.021	0.043 ± 0.029^C	**0.0328**
	Normalized pause duration between drawings‡ [s/mm] (CDT)	0.023 ± 0.009^A	0.031 ± 0.030	0.045 ± 0.039^C	**0.0121**
TMT-specific	Time duration between nodes (TMT-A)	1.10 ± 0.22^A	1.20 ± 0.36	1.41 ± 0.50^C	**0.0107**

Data displayed as mean ± standard deviations were assessed using one-way ANOVAs adjusted for age and sex. Bold values highlight statistically significant differences. Pairwise multiple comparisons (Bonferroni adjusted p values) were performed when comparing individual diagnostic groups. Significant differences between diagnosis categories are marked with C, M, or A (C: Different to CN, M: Different to MCI, A: Different to AD). †Data were normalized by the length of the speech ‡Data were normalized by total stroke length.

features that were significantly different between MCI patients and CN participants were all those during the TMT-A or TMT-B. They consisted of all categories except for kinematics: pressure-related features such as greater pressure variability in TMT-B ($p = 0.004$), time-related features such as longer pause duration between drawings normalized by total stroke length in TMT-B ($p = 0.008$), and TMT-specific features such as longer task duration normalized by the number of answered edges in both TMT-A ($p = 0.037$) and TMT-B ($p = 0.005$).

Classification using gait, speech, and drawing data

We evaluated the model performance for classifying three diagnosis categories of AD, MCI, and CN on the basis of gait, speech, and drawing behavioral features by using iterative ten-fold cross validation. Consequently, combing multiple behavioral modalities improved the model's performance compared with that using a single modality. The model using all three behavioral data achieved the best performance with an accuracy of 93.0% (AuROC of 0.98; Table 3 and Fig. 1). This was higher than an AuROC value of 0.86 calculated as a baseline value by using MMSE scores. The model using all three behavioral data was based on an SVM with a radial basis function (RBF) kernel using 13 gait features, 17 speech features, and 15 drawing features selected by the automatic feature selection procedure. The SVM with an RBF kernel also had the highest accuracies in the other settings using two or single modalities.

We also tested the performance of the classification models in discriminating AD (or MCI) from CN participants. The results of the iterative ten-fold cross validations showed similar trends to those of the three-class classification models: the model using the combined data of all three behavioral modalities achieved more accurate discrimination between AD (or MCI) patients and CN participants, followed by that using two modalities, and that using
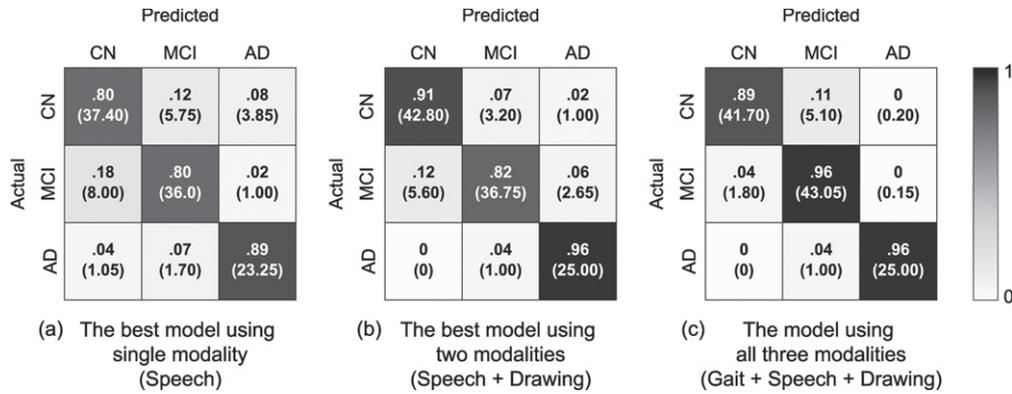
Fig. 1. Normalized confusion matrixes of the three-class classification model for AD, MCI, and CN obtained using 20 iterations of ten-fold cross validation. The number in parentheses indicates the mean number of individuals with CN, MCI, and AD among the 20 iterations. (a) The best model using a single modality (when using speech features) with 81.9% accuracy and AuROC of 0.91. (b) The best model using two different modalities (when using speech and drawing features) with 88.6% accuracy and AuROC of 0.96. (c) The model using all three modalities of gait, speech, and drawing with 93.0% accuracy and AuROC of 0.98.

Table 3
Model performance for classifying three diagnosis categories of AD, MCI, and CN. Values were obtained by 20 iterations of ten-fold cross validation

	Mean [95% CI]	
	Accuracy (%)	AuROC
Gait	75.8 [75.5, 76.1]	0.86 [0.85, 0.86]
Drawing	80.8 [80.3, 81.4]	0.92 [0.91, 0.92]
Speech	81.9 [81.4, 82.4]	0.91 [0.91, 0.91]
Gait + Drawing	86.1 [85.8, 86.4]	0.93 [0.93, 0.94]
Gait + Speech	87.5 [87.3, 87.7]	0.95 [0.95, 0.95]
Speech + Drawing	88.6 [87.9, 89.3]	0.96 [0.96, 0.96]
Gait + Speech + Drawing	93.0 [92.4, 93.6]	0.98 [0.98, 0.98]

a single modality (Tables 4 and 5). Specifically, for discriminating AD from CN participants, the model using gait, speech, and drawing features achieved 100.0% accuracy (100.0% sensitivity, 100.0% specificity, 100.0% F1 score, AuROC of 1.00), while the best accuracy on a single modality was 96.2% when using drawing features (96.9% sensitivity, 95.7% specificity, 94.7% F1 score, AuROC of 0.98). For discriminating MCI from CN participants, the model using gait, speech, and drawing features achieved 89.5% accuracy (94.7% sensitivity, 84.5% specificity, 89.8% F1 score, AuROC of 0.96), while the best accuracy on a single modality was 83.0% when using speech features (82.4% sensitivity, 83.6% specificity, 82.6% F1 score, AuROC of 0.86). Both AuROC values using all three modalities for discriminating AD (or MCI) patients from CN participants were higher than those of 0.98 (0.62) calculated as a baseline value by using MMSE scores. All two-class classification models achieving the best accuracies on

each setting were based on the SVM with an RBF kernel.

Association of behavioral features with cognitive and clinical measures

To better understand how combining multimodal behavioral data could improve classification performance, we investigated the association of behavioral features selected in the model with cognitive and clinical measures. We first applied a principal component analysis to the selected features and obtained the first two PCs for each behavioral modality. They comprised 67.3%, 35.5%, and 54.0% of the total variance in the selected gait, speech, and drawing features, respectively. The estimated factor loadings indicated the characteristics of each principal component as follows: the PC1 of gait features was mainly related to gait acceleration and PC2 was related to stride time variability; the PC1 of speech features was mainly related to pause durations and PC2 was related to variances of the speech spectrum (i.e., variances of MFCCs); and the PC1 of drawing features was mainly related to drawing speed and PC2 was related to pause duration between drawings.

The results of multiple linear regression controlling for age and sex showed that different combinations of behavioral modalities were associated with cognitive and clinical measures (Table 6). Specifically, among the modalities, PCs of only speech features were associated with LM-IIA ($p = 0.003$ for PC1; $p = 0.010$ for PC2), FAB ($p < 0.001$

Table 4
Model performance for classifying diagnosis categories of AD and CN. Values were obtained by 20 iterations of ten-fold cross validation

	Mean [95% CI]			
	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 score (%)
Gait	89.3 [88.3, 90.3]	82.7 [79.7, 85.7]	93.0 [91.4, 94.6]	84.6 [83.1, 86.2]
Speech	92.5 [91.9, 93.0]	90.8 [89.3, 92.2]	93.4 [92.9, 93.9]	89.6 [88.8, 90.3]
Drawing	96.2 [95.4, 96.9]	96.9 [94.8, 99.1]	95.7 [95.7, 95.7]	94.7 [93.6, 95.8]
Gait + Speech	96.7 [96.2, 97.2]	95.0 [93.7, 96.3]	97.7 [97.2, 98.1]	95.4 [94.6, 96.1]
Gait + Drawing	98.4 [97.7, 99.0]	99.2 [97.5, 101.0]	97.9 [97.9, 97.9]	97.7 [96.8, 98.6]
Speech + Drawing	98.6 [98.6, 98.6]	100.0 [100.0, 100.0]	97.9 [97.9, 97.9]	98.1 [98.1, 98.1]
Gait + Speech + Drawing	100.0 [100.0, 100.0]	100.0 [100.0, 100.0]	100.0 [100.0, 100.0]	100.0 [100.0, 100.0]

Table 5
Model performance for classifying diagnosis categories of MCI and CN. Values were obtained by 20 iterations of ten-fold cross validation

	Mean [95% CI]			
	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 score (%)
Gait	77.2 [75.9, 78.4]	84.9 [84.2, 85.6]	69.8 [67.3, 72.3]	78.5 [77.5, 79.4]
Drawing	81.0 [79.1, 82.8]	77.6 [76.0, 79.1]	84.3 [81.3, 87.2]	80.0 [78.2, 81.7]
Speech	83.0 [81.8, 84.3]	82.4 [80.9, 84.0]	83.6 [81.7, 85.5]	82.6 [81.4, 83.9]
Gait + Drawing	87.1 [86.2, 87.9]	80.4 [79.4, 81.4]	93.4 [92.3, 94.5]	85.9 [85.0, 86.8]
Gait + Speech	88.6 [87.7, 89.4]	88.0 [86.3, 89.7]	89.1 [88.7, 89.6]	88.3 [87.3, 89.3]
Speech + Drawing	89.2 [87.3, 91.2]	86.9 [84.0, 89.7]	91.5 [90.1, 92.9]	88.7 [86.6, 90.8]
Gait + Speech + Drawing	89.5 [88.2, 90.7]	94.7 [93.0, 96.4]	84.5 [83.0, 85.9]	89.8 [88.5, 91.0]

for PC1 and PC2), GDS ($p < 0.001$ for PC1), and MTL atrophy ($p < 0.001$ for PC1); PCs of only drawing features were associated with TMT-A ($p < 0.001$ for PC1; $p = 0.008$ for PC2) and CDT ($p < 0.001$ for PC1); PCs of gait and drawing features were associated with ADL ($p < 0.001$ for gait PC2; $p = 0.006$ for drawing PC2); PCs of speech and drawing features were associated with LM-IA ($p = 0.004$ for speech PC1; $p = 0.003$ for speech PC2; $p = 0.041$ for drawing PC1) and MMSE ($p < 0.001$ for speech PC1; $p = 0.006$ for speech PC2; $p = 0.031$ for drawing PC1); PCs of gait, speech, and drawing features were associated with IADL ($p < 0.001$ for gait PC2; $p = 0.004$ for speech PC2; $p = 0.036$ for drawing PC1; $p = 0.033$ for drawing PC2) and TMT-B ($p = 0.037$ for gait PC2; $p = 0.009$ for speech PC1; $p = 0.013$ for speech PC2; $p < 0.001$ for drawing PC1 and PC2).

We also found statistically significant associations of discrete behavioral characteristics with specific cognitive domains reported in previous studies on individual behavioral modalities, even considering the effects of other behavioral modalities. Specifically, they included the associations between stride time variability of gait PC2 and executive function measured by TMT-B ($p = 0.037$); between pause durations of speech PC1 and episodic memory measured by LM-IA ($p = 0.004$) and LM-IIA ($p = 0.003$);

and between drawing speed of drawing PC1 and global cognition measured by MMSE ($p = 0.031$).

DISCUSSION

We investigated multimodal behavioral data of gait, speech, and drawing collected from 118 participants consisting of AD, MCI, and CN participants. Our first main finding was that combining multimodal behavioral data could consistently improve classification performance both for classifying AD, MCI, and CN as well as discriminating AD (or MCI) from CN participants. Combining all three behavioral modalities could consistently achieve more accurate classifications than combining any two modalities, and the model using two modalities could achieve better classification performance than that using a single modality. Our second finding was that each behavioral data may contain different and complementary information for cognitive impairments associated with AD. The regression analysis showed that each behavioral modality was associated with different cognitive and clinical measures for the diagnosis of AD and MCI. In addition, different modalities could achieve the best performance for differentiating AD (or MCI) from CN participants, i.e., drawing for AD versus CN and speech for MCI versus CN. Our findings suggest that gait, speech, and drawing

Table 6
Associations of behavioral features with cognitive and clinical measures

	MMSE[†]	LM-IA[†]	LM-IIA[†]	FAB[†]	TMT-A[†]	TMT-B[†]	CDT[†]	GDS	ADL	IADL[†]	MTLa[†]
Gait											
PC1	0.006	0.027	0.027	0.010	0.008	0.017	0.000	0.012	0.002	0.025	−0.039
	(0.010)	(0.015)	(0.015)	(0.015)	(0.010)	(0.018)	(0.018)	(0.019)	(0.003)	(0.016)	(0.089)
PC2	0.012	0.020	0.018	0.012	0.017	**0.038***	0.009	0.009	**0.010*****	**0.055*****	0.074
	(0.010)	(0.015)	(0.014)	(0.015)	(0.010)	**(0.018)**	(0.017)	(0.018)	**(0.003)**	**(0.016)**	(0.088)
Speech											
PC1	**0.047*****	**0.052****	**0.052****	**0.070*****	0.016	**0.057****	0.021	**0.096*****	−0.003	0.009	**0.448*****
	(0.011)	**(0.018)**	**(0.017)**	**(0.017)**	(0.011)	**(0.021)**	(0.021)	**(0.022)**	(0.003)	(0.019)	**(0.102)**
PC2	**0.034****	**0.059****	**0.049****	**0.070*****	0.018	**0.058***	0.018	−0.022	0.007	**0.059****	0.042
	(0.012)	**(0.019)**	**(0.019)**	**(0.019)**	(0.012)	**(0.023)**	(0.022)	(0.023)	(0.003)	**(0.020)**	(0.113)
Drawing											
PC1	**0.023***	**0.034***	0.028	0.022	**0.059*****	**0.152*****	**0.069*****	−0.028	0.000	**0.037***	0.160
	(0.011)	**(0.017)**	(0.016)	(0.016)	**(0.010)**	**(0.020)**	**(0.019)**	(0.020)	(0.003)	**(0.017)**	(0.096)
PC2	0.016	0.020	0.024	0.003	**−0.025****	**0.085*****	−0.017	−0.019	**0.007****	**0.033***	0.151
	(0.009)	(0.014)	(0.014)	(0.014)	**(0.009)**	**(0.017)**	(0.017)	(0.018)	**(0.003)**	**(0.015)**	(0.084)

Data displayed as beta coefficient (standard error) were assessed using multiple linear regression adjusted for age and sex. The possible score range of all cognitive and clinical measures except for MTLa were rescaled from 0 to 1. In addition, TMT-A, TMT-B, GDS, and MTLa scores were reversed for the purpose of comparison with other variable coefficients, as higher scores indicate better performance. All gait, speech, and drawing variables were standardized by subtracting the mean and dividing by standard deviation. PC, Principal Component; MMSE, Mini-Mental State Examination; LM-IA, immediate recall of the logical memory-story A of the Wechsler memory scale-revised; LM-IIA, delayed recall of the logical memory-story A of the Wechsler memory scale-revised; FAB, Frontal Assessment Battery; TMT-A, Trail making test-part A; TMT-B, Trail making test-part B; CDT, Clock Drawing Test; GDS, Geriatric Depression Scale; ADL, Activity of Daily Living; IADL, Instrumental Activity of Daily Living; MTLa, Medial temporal lobe atrophy. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ [†]Measures with statistically significant differences between diagnosis categories of AD, MCI, and CN assessed by ANOVAs.

behavioral data provides different and complementary information about cognitive impairments useful for the diagnosis of AD and MCI such that clinical diagnostic classification is superior to using either in isolation.

All three behavioral modalities had features that statistically significantly differentiated AD or MCI patients from CN participants and almost all features with significant differences between diagnostic categories showed larger changes from CN in AD compared with MCI. The trends in their changes were consistent with those reported in previous studies on AD or MCI (for gait [7, 8, 27, 33, 42], for speech [19, 31, 43–47], for drawing [22–25]). These behavioral sensitivities to the diagnosis of AD and MCI have been explained as follows: as these behaviors are complex tasks requiring coordination between widespread brain regions related to both motor and cognitive functions [12, 13, 24, 27, 29, 32–36], discrete characteristics of each behavioral modality may reflect neuropathological changes due to AD dementia [12, 13, 27, 56–59]. Our results align with those of previous studies that have shown statistically significant changes in behavioral characteristics from CN individuals and their larger changes in AD compared with earlier stages (i.e., MCI), suggesting the possibility of behavioral markers for detecting AD/MCI and neuropathological changes due to AD. In addi-

tion, we found statistically significant features for differentiating MCI patients from CN participants in specific feature categories (e.g., gait features related to pace and postural control) and specific tasks (e.g., TMTs), while those for differentiating AD patients from CN participants were found both in all categories and in all tasks investigated in this study. This result may help explore behavioral characteristics showing higher sensitivity to changes at early stages of AD in cognitive function and in neuropathological changes.

Our results show that combining gait, speech, and drawing behavioral data can consistently and substantially improve the classification performance of the individual-modality-based classification models. Combining all three behavioral modalities achieved 93.0% accuracy for classifying AD, MCI, and CN, and only 81.9% when using even the best single modality. Because the classification model achieving the best accuracy was based on the SVM with an RBF kernel, nonlinear interactions between behavioral features of different modalities might contribute to the improvement of classification performance by combining multimodal behavioral data. Many studies on biomarkers such as structural MRI, functional imaging, and cerebrospinal fluid have shown that their combinations enhance the classification accuracy of detecting AD or MCI [60, 61]. One of our

contributions lies in providing the first empirical evidence that the benefit of combining different data modalities for detecting AD and MCI can be achieved even with combinations of daily behaviors.

We also suggest that the enhancement of classification performance by combining multimodal behaviors results from their complementary nature regarding 1) cognitive and clinical measures for the diagnosis of AD and MCI as well as 2) discrimination of AD or MCI from CN ones. The selective behavior-cognitive relationships that we showed includes those found in previous studies such as stride time variability and executive function [9–12], pauses during spontaneous speech and episodic memory [21], and drawing speed and global cognition [22]. By demonstrating these associations by using the multimodal behavioral data collected from the same individuals, our results strengthen the notion that these behavioral characteristics in each modality reflect different profiles of cognitive impairments associated with AD [7–13, 18, 21, 22, 27]. In addition, gait, speech, and drawing behaviors each have been suggested to reflect neuropathological changes due to AD [13, 27, 56–59]. Our results suggest that these behavioral modalities may also provide different and complemental information for estimating neuropathological changes. To confirm this suggestion, a further study with validated biomarkers for measuring neuropathological changes is required.

Behavioral data such as gait, speech, and drawing have been proposed as a non-invasive, easy-to-perform means of helping detect AD and MCI in clinical settings [26, 27, 29, 31, 59]. For example, adding a gait test or speech analysis to a neuropsychological assessment has been found to improve the accuracy in detecting people with MCI [31] or a higher probability of developing dementia [29]. In addition, the behavioral data we analyzed in this study may be easily acquired in routine clinical practice. In fact, the speech and drawing data were collected during representative neuropsychological tests for the diagnosis of AD, and the gait data were collected through simple nine-meter walk trials. Although the marker-based motion capture system used in this study might not be readily available in clinical settings, wearable sensors instead might be used for measuring gait characteristics as studies suggested (e.g., [28, 62, 63]). Given the accessibility to behavioral data in clinical practice, our results suggest that combining such behavioral data can help clinicians accurately detect patients with AD and MCI. This would be also useful as an easy-to-perform screen-

ing tool to select individuals who should be further examined with biomarkers both in clinical practice and clinical treatment trials.

Another clinical implication of this study is that our findings might help computerized detection of AD without specialists by improving the performance of computerized screening methods proposed in previous studies [13, 23, 27, 30, 64, 65]. Examples of such methods include automatic assessments of gait characteristics using electronic walkways or wearable sensors [13, 27, 30] and those of drawing characteristics during neuropsychological tests using a digital pen [23]. These methods might benefit from being combined with other behavioral modalities. In addition, when comparing with approaches using computerized assessment tools including digitalized neuropsychological tests, this approach focusing on behavioral data may help promote future efforts towards the development of continuous and passive monitoring tools for the early detection of AD from data collected in everyday life. In fact, recent studies demonstrated the feasibility of using daily walking behavior collected in accelerometer sensors in a free-living setting [28] and daily conversational speech [66–68] for detecting AD or MCI patients. Although further research is needed to investigate the operability and acceptability of real-world data collection, the combination of multimodal behavioral data may be useful for developing an accurate screening tool using behavioral data collected in everyday life.

We acknowledge a number of limitations. First, our analysis did not include validated biomarkers such as cerebrospinal fluid and positron-emission tomography markers of amyloid-β and pathologic tau, even though these biomarkers have been recognized as a standard means of evidencing the biological state of AD [69]. A future study should confirm our findings with a diagnosis based on biomarkers. Second, parts of the cognitive measures we used for the diagnosis and regression analysis were based on the tablet version of drawing tasks instead of the standard paper-and-pencil version; however, given the previous studies showing high correlations between the resulting scores of the tablet- and paper-based tests [70, 71], its effect on our main findings is potentially small. Third, although to our knowledge this is the first study investigating combinations of gait, speech, and drawing data for detecting AD and MCI, the sample size was limited. This might affect the generalizability of our results. Fourth, residual confounding can still exist in addition to age and sex considered in the analysis. Fifth, in terms of statistical

analyses, we did not adjust for multiple comparisons across behavioral features due to the exploratory nature of this investigation. However, our results showed that all three modalities had behavioral features with significant differences at the $p < 0.0001$ level and their trends in their changes were consistent with those reported in previous studies on AD or MCI [7, 8, 19, 22–25, 27, 31, 33, 42–47]. Thus, we believe that our first empirical evidence obtained from the investigation of multimodal behavioral data collected from the same participants could still provide support for the hypothesis that each behavior is sensitive to the diagnosis of AD and MCI, although results need to be replicated on large samples. Finally, all results of the classification performances reported in this work are cross-validation results due to the limited data samples, which is less ideal compared with using a separate validation dataset. To further confirm our results about the improvement of classification performance by combining data of multiple behavioral modalities, future studies should examine the model performance using a separate validation dataset or nested cross-validation procedure.

In conclusion, this pilot study provides initial evidence that multimodal behavioral data of gait, speech, and drawing provide different and complementary information about cognitive impairments such that clinical diagnostic classification for AD and MCI is superior to using either in isolation. Our findings may help develop non-invasive, easy-to-perform screening tools. A future study is needed to better understand the associations of behavioral characteristics with specific profiles of neuropathological changes.

ACKNOWLEDGMENTS

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/21-0684r1).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-210684.

REFERENCES

[1] Alzheimer's Association (2019) Alzheimer's disease facts and figures. *Alzheimers Dement* **15**, 321-387.

[2] Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, Ballard C, Banerjee S, Burns A, Cohen-Mansfield J, Cooper C, Fox N, Gitlin LN, Howard R, Kales HC, Larson E, Ritchie K, Rockwood K, Sampson EL, Samus Q, Schneider LS, Selbæk G, Teri L, Mukadam N (2017) Dementia prevention, intervention, and care. *Lancet* **390**, 2673-2734.

[3] Rasmussen J, Langerman H (2019) Alzheimer's disease-why we need early diagnosis. *Degener Neurol Neuromuscul Dis* **9**, 123.

[4] Petersen RC, Lopez O, Armstrong MJ, Getchius TS, Ganguli M, Gloss D, Gronseth GS, Marson D, Pringsheim T, Day GS, Sager M, Stevens J, Rae-Grant A (2018) Practice guideline update summary: Mild cognitive impairment: Report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. *Neurology* **90**, 126-135.

[5] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, Brayne C, Burns A, Cohen-Mansfield J, Cooper C, Costafreda SG, Dias A, Fox N, Gitlin LN, Howard R, Kales HC, Kivimäki M, Larson EB, Ogunniyi A, Orgeta V, Ritchie K, Rockwood K, Sampson EL, Samus Q, Schneider LS, Selbæk G, Teri L, Mukadam N (2020) Dementia prevention, intervention, and care: 2020 report of the lancet commission. *Lancet* **396**, 413-446.

[6] Prince M, Comas-Herrera A, Knapp M, Guerchet M, Karagiannidou M (2016) *World Alzheimer Report 2016. Improving healthcare for people living with dementia: Coverage, quality and costs now and in the future*. Alzheimer's Disease International, London.

[7] Beauchet O, Allali G, Montero-Odasso M, Sejdić E, Fantino B, Annweiler C (2014) Motor phenotype of decline in cognitive performance among community-dwellers without dementia: Population-based study and meta-analysis. *PloS One* **9**, e99318.

[8] Mielke MM, Roberts RO, Savica R, Cha R, Drubach DI, Christianson T, Pankratz VS, Geda YE, Machulda MM, Ivnik RJ, Knopman DS, Boeve BF, Rocca WA, Petersen RC (2013) Assessing the temporal relationship between cognition and gait: Slow gait predicts cognitive decline in the Mayo Clinic Study of Aging. *J Gerontol A Biol Sci Med Sci* **68**, 929-937.

[9] Hausdorff JM, Yogev G, Springer S, Simon ES, Giladi N (2005) Walking is more like catching than tapping: Gait in the elderly as a complex cognitive task. *Exp Brain Res* **164**, 541-548.

[10] Holtzer R, Wang C, Verghese J (2012) The relationship between attention and gait in aging: Facts and fallacies. *Motor Control* **16**, 64-80.

[11] Beauchet O, Annweiler C, Montero-Odasso M, Fantino B, Herrmann FR, Allali G (2012) Gait control: A specific subdomain of executive function? *J Neuroeng Rehabil* **9**, 12.

[12] Morris R, Lord S, Bunce J, Burn D, Rochester L (2016) Gait and cognition: Mapping the global and discrete relationships in ageing and neurodegenerative disease. *Neurosci Biobehav Rev* **64**, 326-345.

[13] Pieruccini-Faria F, Black SE, Masellis M, Smith EE, Almeida QJ, Li KZ, Bherer L, Camicioli R, Montero-Odasso M (2021) Gait variability across neurodegenerative and cognitive disorders: Results from the Canadian Consortium of Neurodegeneration in Aging (CCNA) and the

Gait and Brain Study. *Alzheimers Dement* **17**, 1317-1328.

[14] Eyigoz E, Mathur S, Santamaria M, Cecchi G, Naylor M (2020) Linguistic markers predict onset of Alzheimer's disease. *E Clinical Medicine* **28**, 100583.

[15] Ahmed S, Haigh A-MF, Jager CA de, Garrard P (2013) Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain* **136**, 3727-3737.

[16] Tóth L, Hoffmann I, Gosztolya G, Vincze V, Szatlóczki G, Bánréti Z, Pákáski M, Kálmán J (2018) A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Curr Alzheimer Res* **15**, 130-138.

[17] Beltrami D, Gagliardi G, Rossini Favretti R, Ghidoni E, Tamburini F, Calzà L (2018) Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline? *Front Aging Neurosci* **10**, 369.

[18] Mueller KD, Koscik RL, Hermann BP, Johnson SC, Turkstra LS (2018) Declines in connected language are associated with very early mild cognitive impairment: Results from the Wisconsin Registry for Alzheimer's Prevention. *Front Aging Neurosci* **9**, 437.

[19] König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, Manera V, Verhey F, Aalten P, Robert PH, Renaud D (2015) Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimers Dement (Amst)* **1**, 112-124.

[20] Pakhomov SV, Hemmy LS (2014) A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the Nun Study. *Cortex* **55**, 97-106.

[21] Pistono A, Jucla M, Barbeau EJ, Saint-Aubert L, Lemesle B, Calvet B, Köpke B, Puel M, Pariente J (2016) Pauses during autobiographical discourse reflect episodic memory processes in early Alzheimer's disease. *J Alzheimers Dis* **50**, 687-698.

[22] Kawa J, Bednorz A, Stepien P, Derejczyk J, Bugdol M (2017) Spatial and dynamical handwriting analysis in mild cognitive impairment. *Comput Biol Med* **82**, 21-28.

[23] Müller 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**, 1-10.

[24] Werner P, Rosenblum S, Bar-On G, Heinik J, Korczyn A (2006) Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment. *J Gerontol B Psychol Sci Soc Sci* **61**, P228-P236.

[25] Yan JH, Rountree S, Massman P, Doody RS, Li H (2008) Alzheimer's disease and mild cognitive impairment deteriorate fine movement control. *J Psychiatr Res* **42**, 1203-1212.

[26] Laske C, Sohrabi HR, Frost SM, Lépez-de-Ipinüa K, Garrard P, Buscema M, Dauwels J, Soekadar SR, Mueller S, Linnemann C, Bridenbaugh SA, Kanagasingam Y, Martins RN, O'Bryant SE (2015) Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimers Dement* **11**, 561-578.

[27] Mc Ardle R, Galna B, Donaghy P, Thomas A, Rochester L (2019) Do Alzheimer's and Lewy body disease have discrete pathological signatures of gait? *Alzheimers Dement* **15**, 1367-1377.

[28] Varma VR, Ghosal R, Hillel I, Volfson D, Weiss J, Urbanek J, Hausdorff JM, Zipunnikov V, Watts A (2021) Continuous gait monitoring discriminates community-dwelling mild Alzheimer's disease from cognitively normal controls. *Alzheimers Dement (N Y)* **7**, e12131.

[29] Grande G, Rizzuto D, Vetrano DL, Marseglia A, Vanacore N, Laukka EJ, Welmer A-K, Fratiglioni L (2020) Cognitive and physical markers of prodromal dementia: A 12-year-long population study. *Alzheimers Dement* **16**, 153-161.

[30] Washabaugh EP, Kalyanaraman T, Adamczyk PG, Claflin ES, Krishnan C (2017) Validity and repeatability of inertial measurement units for measuring gait parameters. *Gait Posture* **55**, 87-93.

[31] Roark B, Mitchell M, Hosom J-P, Hollingshead K, Kaye J (2011) Spoken language derived measures for detecting mild cognitive impairment. *IEEE Trans Audio Speech Lang Process* **19**, 2081-2090.

[32] Tian Q, Chastan N, Bair W-N, Resnick SM, Ferrucci L, Studenski SA (2017) The brain map of gait variability in aging, cognitive impairment and dementia—a systematic review. *Neurosci Biobehav Rev* **74**, 149-162.

[33] Moon Y, Sung J, An R, Hernandez ME, Sosnoff JJ (2016) Gait variability in people with neurological disorders: A systematic review and meta-analysis. *Hum Mov Sci* **47**, 197-208.

[34] Szatloczki G, Hoffmann I, Vincze V, Kalman J, Pakaski M (2015) Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. *Front Aging Neurosci* **7**, 195.

[35] Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF (2015) A review of depression and suicide risk assessment using speech analysis. *Speech Commun* **71**, 10-49.

[36] Boschi V, Catricala E, Consonni M, Chesi C, Moro A, Cappa SF (2017) Connected speech in neurodegenerative language disorders: A review. *Front Psychol* **8**, 269.

[37] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263-269.

[38] Petersen RC, Aisen P, Beckett LA, Donohue M, Gamst A, Harvey DJ, Jack C, Jagust W, Shaw L, Toga A, Trojanowski JQ, Weiner MW (2010) Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology* **74**, 201-209.

[39] Benoit JS, Chan W, Piller L, Doody R (2020) Longitudinal sensitivity of Alzheimer's disease severity staging. *Am J Alzheimers Dis Other Demen* **35**, 1533317520918719.

[40] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps GH (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270-279.

[41] Goodglass H, Kaplan E, Weintraub S (2001) *BDAE: The Boston Diagnostic Aphasia Examination*. Lippincott Williams & Wilkins Philadelphia, PA.

[42] Verghese J, Robbins M, Holtzer R, Zimmerman M, Wang C, Xue X, Lipton RB (2008) Gait dysfunction in mild cognitive impairment syndromes. *J Am Geriatr Soc* **56**, 1244-1251.

[43] Fraser KC, Meltzer JA, Rudzicz F (2016) Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis* **49**, 407-422.

[44] Ahmed S, Jager CA de, Haigh A-M, Garrard P (2013) Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed Alzheimer's disease. *Neuropsychology* **27**, 79-85.

[45] Khodabakhsh A, Yesil F, Guner E, Demiroglu C (2015) Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. *EURASIP J Audio Speech Music Process* **2015**, 1-15.

[46] Martínez-Sánchez F, JJ GM, Pérez E, Carro J, Arana JM (2012) Expressive prosodic patterns in individuals with Alzheimer's disease. *Psicothema* **24**, 16-21.

[47] Meilán JJG, Martínez-Sánchez F, Carro J, López DE, Millian-Morell L, Arana JM (2014) Speech in Alzheimer's disease: Can temporal and acoustic parameters discriminate dementia. *Dement Geriatr Cogn Disord* **37**, 327-334.

[48] Honoré A (1979) Some simple measures of richness of vocabulary. *Assoc Literary Linguistic Comput Bull* **7**, 172-177.

[49] Kavé G, Levy Y (2003) Morphology in picture descriptions provided by persons with Alzheimer's disease. *J Speech Lang Hear Res* **46**, 341-352.

[50] Croisile B, Ska B, Brabant M-J, Duchene A, Lepage Y, Aimard G, Trillet M (1996) Comparative study of oral and written picture description in patients with Alzheimer's disease. *Brain Lang* **53**, 1-19.

[51] Hernández-Domínguez L, Ratté S, Sierra-Martínez G, Roche-Bergua A (2018) Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimers Dement (Amst)* **10**, 260-268.

[52] Goldberger J, Hinton GE, Roweis S, Salakhutdinov RR (2004) Neighbourhood components analysis. *Adv Neural Inf Process Syst* **17**, 513-520.

[53] Breiman L (2001) Random forests. *Mach Learn* **45**, 5-32.

[54] Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. *In Proceedings of the fifth annual workshop on computational learning theory* 144-152.

[55] Hand DJ, Till RJ (2001) A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach Learn* **45**, 171-186.

[56] Wennberg AM, Savica R, Mielke MM (2017) Association between various brain pathologies and gait disturbance. *Dement Geriatr Cogn Disord* **43**, 128-143.

[57] Wilson J, Allcock L, Mc Ardle R, Taylor J-P, Rochester L (2019) The neural correlates of discrete gait characteristics in ageing: A structured review. *Neurosci Biobehav Rev* **100**, 344-369.

[58] Verfaillie SCJ, Witteman J, Slot RER, Pruis IJ, Vermaat LEW, Prins ND, Schiller NO, van de Wiel M, Scheltens P, van Berckel BNM, van der Flier WM, Sikkes SAM (2019) High amyloid burden is associated with fewer specific words during spontaneous speech in individuals with subjective cognitive decline. *Neuropsychologia* **131**, 184-192.

[59] Rentz DM, Papp KV, Mayblyum DV, Sanchez JS, Klein H, Souillard-Mandar W, Sperling RA, Johnson KA (2021) Association of digital clock drawing with PET amyloid and tau pathology in normal older adults. *Neurology* **96**, e1844-e1854.

[60] Jack J Clifford R., Lowe VJ, Senjem ML, Weigand SD, Kemp BJ, Shiung MM, Knopman DS, Boeve BF, Klunk WE, Mathis CA, Petersen RC (2008) 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. *Brain* **131**, 665-680.

[61] Zhang D, Wang Y, Zhou L, Yuan H, Shen D, ADNI (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* **55**, 856-867.

[62] Mc Ardle R, Del Din S, Galna B, Thomas A, Rochester L (2020) Differentiating dementia disease subtypes with gait analysis: Feasibility of wearable sensors? *Gait Posture* **76**, 372-376.

[63] Agurto C, Heisig S, Abrami A, Ho BK, Caggiano V (2021) Parkinson's disease medication state and severity assessment based on coordination during walking. *PLoS One* **16**, e0244842.

[64] Hayes TL, Abendroth F, Adami A, Pavel M, Zitzelberger TA, Kaye JA (2008) Unobtrusive assessment of activity patterns associated with mild cognitive impairment. *Alzheimers Dement* **4**, 395-405.

[65] Petti U, Baker S, Korhonen A (2020) A systematic literature review of automatic Alzheimer's disease detection from speech and language. *J Am Med Inform Assoc* **27**, 1784-1797.

[66] Yamada Y, Shinkawa K, Shimmei K (2020) Atypical repetition in daily conversation on different days for detecting Alzheimer disease: Evaluation of phone-call data from a regular monitoring service. *JMIR Mental Health* **7**, e16790.

[67] Yamada Y, Shinkawa K, Kobayashi M, Nishimura M, Nemoto M, Tsukada E, Ota M, Nemoto K, Arai T (2021) Tablet-based automatic assessment for early detection of Alzheimer's disease using speech responses to daily life questions. *Front Digit Health* **3**, 30.

[68] Nasreen S, Rohanian M, Hough J, Purver M (2021) Alzheimer's dementia recognition from spontaneous speech using disfluency and interactional features. *Front Comput Sci* **3**, 640669.

[69] Jack CR, Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, Hampel H, Jagust WJ, Johnson KA, Knopman DS, Petersen RC, Scheltens P, Sperling RA, Dubois B (2016) A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology* **87**, 539-547.

[70] Lunardini F, Luperto M, Daniele K, Basilico N, Damanti S, Abbate C, Mari D, Cesari M, Ferrante S, Borghese NA (2019) Validity of digital trail making test and bells test in elderlies. *In 2019 IEEE EMBS international conference on biomedical & health informatics (BHI)* IEEE, pp. 1-4.

[71] Fellows RP, Dahmen J, Cook D, Schmitter-Edgecombe M (2017) Multicomponent analysis of a digital trail making test. *Clin Neuropsychol* **31**, 154-167.