MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190205

Handwriting Features of Multiple Drawing Tests for Early Detection of Alzheimer's Disease: A Preliminary Result

Tatsuya Ishikawa[a], Miyuki Nemoto[b], Kiyotaka Nemoto[b], Tomoko Takeuchi[b], Yuriko Numata[b], Ryohei Watanabe[b], Eriko Tsukada[b], Miho Ota[b], Shinji Higashi[b], Tetsuaki Arai[b], Yasunori Yamada[a]

[a] IBM Research, Tokyo, Japan
[b] University of Tsukuba, Ibaraki, Japan

Abstract

Early detection of Alzheimer's disease (AD) has become increasingly important. Healthy monitoring technology focusing on behavioral changes is a promising approach in this vein. Among such technologies, handwriting features measured by digital tablet devices have attracted attention as potential indicators for detecting AD and mild cognitive impairment (MCI). However, previous studies have mainly investigated features in single tasks, and it remains unclear whether combining the features of multiple tasks could improve the performance of detecting AD and MCI. In this study, we investigated features in five representative tasks used in neuropsychological tests collected from 71 seniors including some diagnosed with MCI and AD. We found that our three-class classification model improved diagnosis accuracy by up to 11.3% by combining features of multiple tasks, for a final accuracy of 74.6%. We also suggested that drawing behaviors during multiple tasks might be useful for estimating disease progression simply by utilizing the labels of disease groups.

Keywords:

Dementia, Handwriting, Classification

Introduction

As the world's elderly population increases, the number of people living with dementia is rising rapidly, making dementia an increasingly serious health and social problem. According to a previous survey, around 47 million people globally were living with dementia as of 2015, corresponding to about 7.6% of the world's over-65-year-olds [1]. At the same time, diagnostic coverage worldwide remains low and dementia is often undiagnosed. Even in high-income countries, only 40–50% of dementia sufferers have received a diagnosis [2, 3]. The low diagnosis coverage makes it difficult for many patients and their families to receive appropriate support and care. In addition, while dementia affects the individuals with the disease, it also affects their supporters—including relatives and the wider society—because people with dementia require constant and costly care for years. In fact, healthcare costs have risen significantly, reaching over $818 billion USD in 2015, and this figure is estimated to rise to $2 trillion USD by 2030 [4]. One strategy to reduce some of this cost is early intervention at the mild cognitive impairment (MCI) or preclinical stages. In fact, longitudinal studies suggest the possibility of early intervention at the MCI stage to reduce the progression to dementia [5]. An intervention that could delay the onset of Alzheimer's disease (AD) by five years is estimated to result in a 57% reduction in the number of AD patients and to reduce 45% of the projected Medicare costs [6].

Health monitoring technology focusing on behavioral features is expected to help improve diagnosis coverage and to detect AD at an earlier stage by expanding opportunities for receiving assessment from the clinical setting to more varied situations including everyday situations. For example, previous lab studies have suggested that behavioral features in gait, speech, and eye movement can be useful indicators for identifying AD and MCI [7]. Being able to infer AD and MCI from these behaviors with better accuracy and wider applicability would be tremendously useful.

One promising behavior to explore is drawing behavior. An advantage of this approach is ease of data collection brought about by the popularization of portable devices such as tablets and smartphones. Drawing behavior assessments such as the Clock Drawing Test (CDT) [8] and the Trail Making Test (TMT) [9] have proven useful for measuring cognitive decline and detecting AD and are commonly used in conventional in-clinic neuropsychological tests. Recent research on drawing behaviors using tablet devices has shown the possibility of automatic detection of patients with cognitive or motor impairments [10–22]. For example, [10] extracted pressure and kinematic features during several tasks including the CDT, while [11] investigated frequencies, velocities, and temporal features during a variant of the TMT. These features have been shown as differentiating healthy subjects and AD patients. Although previous studies have demonstrated how we can build a model for detecting AD and/or MCI by using behavioral features during individual drawing tasks, whether and how we can improve the model performance by combining behavioral features during multiple tasks remains unexplored. In addition, most of these studies focused on developing a classification model for differentiating patients with MCI and AD. Being capable of inferring disease progression on a scalar or ordinal scale defined by in-clinic cognitive assessment scores or biomarkers such as amyloid beta and tau deposition would extend the scope of application, for example, through visualization of the effects of intervention and prevention.

In this study, we investigated drawing behaviors during five representative tasks used in in-clinic neuropsychological tests collected from 71 Japanese seniors including some diganosed with MCI and AD. We extracted a series of drawing behavioral features including pressure, velocity, acceleration, jerk, and in-air and on-screen durations and then built a three-class classification model to distinguish healthy controls (HCs), patients with MCI, and patients with AD. Through

Table 1 – Demographics of participants.

Status	No. of participants (Female)	Mean age (SD)	Mean MMSE score (SD)
HC	36 (21)	70.0 (5.0)	28.3 (1.5)
MCI	25 (15)	75.9 (5.3)	26.8 (3.1)
AD	10 (7)	76.7 (6.0)	18.8 (3.9)

comparison, we found that the model using features in all five tasks could improve accuracy by up to 11.3% by combining the features of multiple tasks, achieving a final accuracy of 74.6% (chance rate 40%). We next investigated using the model for inferring disease progression on a continuous scale. Specifically, we first trained our model to differentiate three groups using only drawing features and disease labels (HC, MCI, and AD), and then we investigated whether the model could estimate in-clinic cognitive assessment scores. The results showed that the scores estimated by the model were significantly correlated with in-clinic cognitive assessment scores, even though we did not use the assessment scores themselves. These results indicate that our approach focusing on drawing behaviors during multiple tasks might be useful for inferring AD progression.

Materials and Methods

Participants

A total of 71 participants were enrolled by the University of Tsukuba Hospital. Ten participants were patients with AD, 25 were patients with MCI, and 36 were HCs. All participants were evaluated with the Mini-Mental State Examination (MMSE), a screening measure of global cognitive functioning [23]. Table 1 shows the number of participants (female), mean age, and mean MMSE score for the HC, MCI, and AD groups.

None of the participants in the HC group were diagnosed as having MCI or dementia before the experiment. The definitions of the MCI and AD groups were based on diagnosis by psychiatrists through medical examinations including structural magnetic resonance imaging, blood tests, and neuropsychological tests. More specifically, the doctors followed the guidelines and criteria in [24] for MCI and [25] for AD. Informed consent was obtained from all participants in accordance with a procedure by the ethics committee, the University of Tsukuba Hospital (H29-65).

Apparatus

A digital drawing tablet (Wacom Cintiq Pro 16) was used to acquire handwriting movements. The detailed specifications of the tablet are as follows: external dimensions (width × depth × height) 410 × 265 × 17.5 mm, spatial resolution 3840 × 2160 dots, pixel size 0.090 × 0.090 mm, temporal resolution 30 ms, and pressure levels 8,192.

The data include 3D coordinates (x, y, z) and the pressure of the pen-tip, altitude and azimuth of the pen, a binary variable (1 for writing state and 2 for erasing state), and timestamp.

When the stylus touches the start button, the software starts acquiring the data, and a black line reproducing the written trace appears. Thus, participants can monitor in real-time what they are writing.

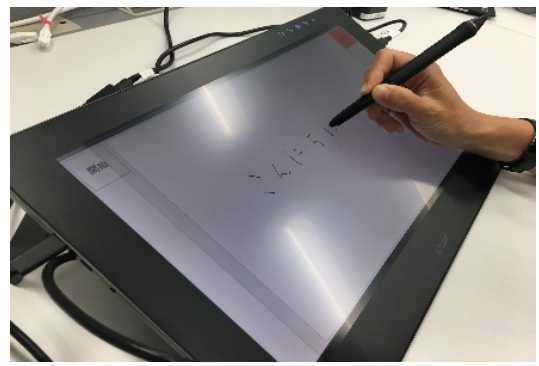


Figure 1 – Example of writing a spontaneous sentence.

Experimental procedure

Participants were seated on a chair with the digital tablet placed on a desk and could freely adjust the position of the device. Each participant was asked to perform five tasks:

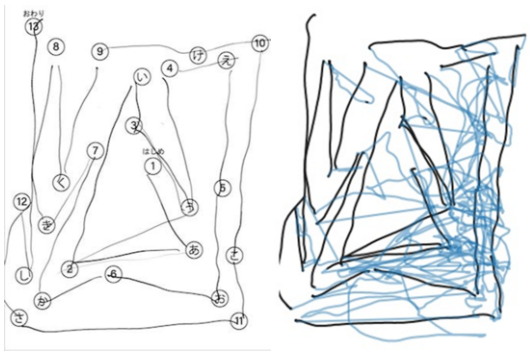
1. Writing spontaneous sentences. This task is included in the Mini-Mental State Examination (MMSE) [23]. Paticipants are asked to write any complete sentences on a sheet.
2. Drawing crossed pentagons. This is also a part of the MMSE. At first, participants see a figure of two intersecting pengagons. Then they are asked to draw the same figure as shown on a sheet.
3. Trail Making Test (TMT) part A. This task requires participants to draw lines connecting consecutive numbers randomly distributed on a sheet [9].
4. TMT part B. This is similar to the TMT part A, but instead of just linking numbers, participants are required to draw lines connecting numbers and letters alternately in their respective sequence [9].
5. Clock Drawing Test. This test asks participants to draw an analog clock-face showing 10 minutes after 10 on a blank sheet [8].

Figure 1 shows an example of writing a spontanous sentence on our tablet device.

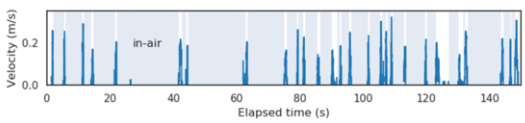
Data analysis

Pressure profiles were obtained from the apparatus as a raw dataset. Numbers of segments were counted to characterize the handwriting behaviors. As kinematic parameters, velocity (m/s), acceleration (m/s^2), and jerk (m/s^3) of the pen-tip movements over the 2D coordinates (x, y) on the tablet surface were computed. The durations (s) of on-screen and in-air stylus pen movements were considered as timing parameters. Figures 2 and 3 show examples of the handwriting behavioral features.

After completion of all five tasks, a three-class classification analysis using the handwriting parameters calculated to distinguish HC, MCI, and AD was performed. To investigate whether the model can be used to estimate in-clinic cognitive assessment scores, correlation analysis between MMSE scores and our model scores was also conducted.

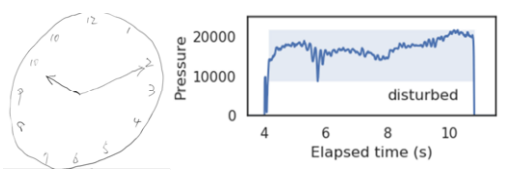


(a) Left: final outcome and right: extracted on-screen (black) and in-air (blue) trajectories

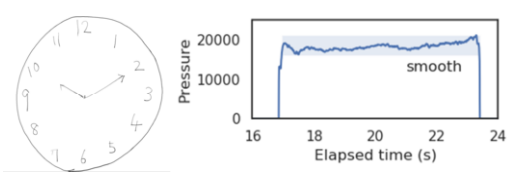


(b) On-screen pen-tip velocity and in-air duration (shaded area)

Figure 2 – Example of handwriting behavioral features of an MCI patient in TMT part A.



(a) Disturbed pressure history of MCI



(b) Smooth pressure history of HC

Figure 3– Example of a difference of pressure features during circle drawing in CDT between MCI and HC regardless of the similarity of the final outcomes.

Results

Three-class classification models

On the basis of the extracted handwriting features, classification analysis using a generalized linear model with a logit link function was performed. In order to differentiate HC, MCI, and AD, we assigned 0, 1, and 2 as targeted values, respectively. Explanatory variables were selected by means of a stepwise method to optimize the Akaike's Information Criterion (AIC). We evaluated three-class classification performances of the models using features of single tasks and all tasks.

Tables 2 and 3 show the accuracy and the selected features of the resultant classification models. As shown, the model that used features of all tasks improved the diagnosis accuracy by up to 11.3% and achieved an accuracy of 74.6% (chance rate 40%). The classification performance of this best model is presented in Table 4.

Table 2 – Classification performance.

Tasks	Accuracy (%)
MMSE sentences	67.6
MMSE pentagons	66.2
TMT part A	69.0
TMT part B	63.3
CDT	67.6
All tasks	74.6

Table 3 – Selected variables in classification models.

Tasks	Selected variables
MMSE sentences	Age, mean velocity, and mean pressure
MMSE pentagons	Age, mean pressure, CV of pressure, in-air duration, and mean velocity
TMT part A	Age, in-air duration, and CV of acceleration
TMT part B	Age, CV of pressure, CV of jerk, and on-screren duration
CDT	Age, number of segments, in-air duration, and CV of jerk
All tasks	Mean velocity and mean pressure (MMSE sentences), mean pressure (MMSE pentagons), on-screen duration (TMT part A), CV of pressure (TMT part B), and number of segments, in-air duration, and mean pressure (CDT)

Table 4 – Classification performance of the model using features from all tasks.

		Actual		
		HC	MCI	AD
Estimated	HC	35	11	1
	MCI	1	10	1
	AD	0	4	8

Estimation of in-clinic cognitive assessment scores from the classification models

We next investigated whether the three-class classification model could estimate in-clinic cognitive assessment scores. Specifically, we explored the relationship between a parameter in the classification model and the MMSE score.

The generalized linear model with the logit link function estimates a continuous and ordinal score that can be interpreted as a probability parameter of a Bernoulli trial. Here, we may regard this as a predicted disease progression score of AD. A correlation analysis was performed between the disease progression scores predicted with the models and the MMSE scores. According to the Pearson's correlation coefficients, our predicted disease progression score obtained from the model combining features of multiple tasks was most strongly correlated with the MMSE scores among all models ($r = -0.70$, Table 5).

Figure 4 shows the best predicted scores according to the labels of the groups. Figure 5 indicates the relationship between the best predicted scores and MMSE scores.

Table 5 – Pearson's correlation coefficients between predicted disease progression scores and MMSE scores.

Tasks	*r*	*p*-value
MMSE sentences	–0.38	$<1.0 \times 10^{-3}$
MMSE pentagons	–0.51	$<1.0 \times 10^{-5}$
TMT part A	–0.59	$<1.0 \times 10^{-7}$
TMT part B	–0.60	$<1.0 \times 10^{-7}$
CDT	–0.65	$<1.0 \times 10^{-9}$
All tasks	–0.70	$<1.0 \times 10^{-11}$

Discussion

As stated in the introduction, healthy monitoring technology focusing on behavioral changes has shown promise for early detection of AD. Among these technologies, handwriting behaviors measured by digital tablet devices are attracting attention as potentially useful indicators. The objectives of the present study were to investigate the effectiveness of combining behavioral features of multiple drawing tasks for detecting AD and MCI and the possibility of estimating disease progression by using such features. We collected and analyzed a series of behavioral features of five representative handwriting tasks used in MMSE, TMT, and CDT from 71 participants including some diagnosed with MCI and AD.

First, although the selected predictive features in our analysis were different depending on the tasks, we found tendencies that features relating to fine motor controls such as velocity, acceleration, and pressure were selected for the MMSE tasks and features relating to attention or cognitive performances such as durations and number of segments were selected for the TMT and the CDT. These results may reflect the characteristics of the different tasks and suggest that exploiting the combined features of multiple tasks may improve the performance of predictive models. We demonstrated that our three-class classification model combining features of multiple tasks improved the accuracy.

Next, we demonstrated that the classification model can also estimate in-clinic cognitive assessment scores via correlation analysis between the model estimated score that can be regarded as a predicted disease progression and the MMSE assessment score. It should be noted that the predicted disease progression scores were obtained just from the labels of disease groups and behavioral features without any progression measures. A more accurate validation of disease progression might require biomarkers such as amyloid beta in the brain and tau in cerebrospinal fluid (CSF), especially when we attempt to build a model for detecting patients at the preclinical AD stage. It would be difficult or impossible to use these biomarkers for training a model, considering the cost and/or invasiveness for assessing positron emission tomography (PET) and CSF. Our approach, which attempts to predict the level of disease progression by using just disease stage information instead of these biomarkers, should reduce the amount of required sample data (e.g., just validation data would be needed).

Some limitations exist in this study. First, since the number of participants was relatively small, generalizability of our results may be limited. Further study with a larger number of participants is our future work. Second, we evaluated the handwriting features using just disease labels and neuropsychological assessment scores. Further study is needed to investigate whether our model for estimating scores can predict known biomarkers for AD such as amyloid beta and tau deposition.

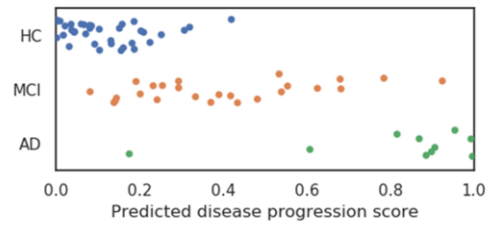


Figure 4 – Strip plot of predicted disease progression scores obtained from model combining multiple tasks.

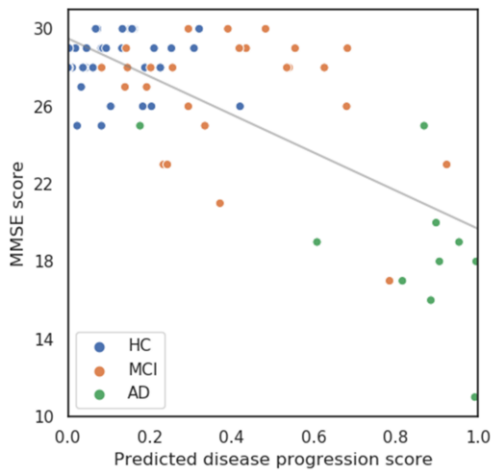


Figure 5 – Scatter plot with the line of best fit between the predicted disease progression scores obtained from the model combining multiple tasks and MMSE score.

References

[1] M. Prince, *et al*., *The global impact of dementia 2013–2050*. Alzheimer's Disease International, 2013.

[2] M. Prince, *et al*., *World Alzheimer report 2016: Improving healthcare for people living with dementia: Coverage, quality, and costs now and in the future*. Alzheimer's Disease International, 2016.

[3] V. Kotagal, *et al*., Factors associated with cognitive evaluations in the United States. *Neurology* **81** (2015), 64–71.

[4] A. Wimo *et al*., The world costs of dementia 2015 and comparisons with 2010. *Alzheimer's Dement*, **13** (2017), 1–7.

[5] T. Horr, *et al*., Systematic review of strengths and limitations of randomized controlled trials for non-pharmacological interventions in mild cognitive impairment: Focus on Alzheimer's disease. *The Journal of Nutrition, Health & Aging,* **19.2** (2015), 141–153.

[6] R.A. Sperling, *et al*. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7.3** (2011), 280–292.

[7] A. Alberdi, A. Aztiria, and A. Basarab. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial Intelligence in Medicine* **71** (2016), 1–29.

[8] M.D. Lezak, D.B. Howieson, and D.W. Loring. *Neuropsychological Assessment (4th ed.)*. Oxford University Press, 2004.

[9] C.R. Bowie, and P.D. Harvey, Administration and interpretation of the Trail Making Test, *Nature Protocols* **1** (2006), 2277–2281.
[10] J. Garre-Olmo *et al*., Kinematic and pressure features of handwriting and drawing: Preliminary results between patients with mild cognitive impairment, Alzheimer's disease, and healthy controls, *Current Alzheimer Research* **14** (2017), 960–968.
[11] P. Gulde, *et al*., Step by Step: Kinematics of the reciprocal trail making task predict slowness of activities of daily living performance in Alzheimer's disease. *Front Neurol* **9** (2018), 140.
[12] C. Kahindo *et al*., Characterizing early-stage Alzheimer's through spatiotemporal dynamics of handwriting. *IEEE Signal Processing Letters* **25** (2018), 1136–1140.
[13] K. Lopez-de-Ipina, *et al*., Automatic analysis of Archimedes' spiral for characterization of genetic essential tremor based on Shannon's entropy and fractal dimension. *Entropy* **20** (2018), 531–545.
[14] G. Pirlo *et al*., Early diagnosis of neurodegenerative diseases by handwriting signature analysis. *Proceedings of International Conference on Image Analysis and Processing 2015*, 290–297.
[15] R. Camicioli *et al*., Handwriting and pre-frailty in the Lausanne cohort 65+ (Lc65+) study. *Archives of Gerontology and Geriatrics* **61** (2015), 8–15.
[16] A. Bisio *et al*., The kinematics of handwriting movements as expression of cognitive and sensorimotor impairments in people with multiple sclerosis. *Scientific Report* **7** (2017), 17730.
[17] M. Mitolo, *et al*., The new qualitative scoring MMSE pentagon test (QSPT) as a valid screening tool between autopsy-confirmed dementia with Lewy bodies and Alzheimer's disease. *J Alzheimer Dis* **39** (2014), 823–832.
[18] J. Dehmen *et al*., An analysis of a digital variant of the trail making test using machine learning techniques. *Technol Health Care* **25** (2017), 251–264.
[19] R. Fellows, *et al*., Multicomponent analysis of a digital trail making test. *Clin Neuropsychol* **31** (2017), 154–167.
[20] J. Heinik, *et al*., Computerized kinematic analysis of the clock drawing task in elderly people with mild major depressive disorder: An exploratory study. *Int Psychogeriatr* **22** (2010), 479–488.
[21] Z. Harbi *et al*., Segmentation of clock drawing based on spatial and temporal features. *Proceedia Computer Science* **60** (2015), 1640–1648.
[22] W. Soullard-Mandar, *et al*., Learning classification models of cognitive conditions from subtle behaviors in the digital clock drawing test. *Mach Learn* **102** (2016), 393–441.
[23] M.F. Folstein, S.E. Folstein, and P.R. McHugh, "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12** (1975), 189–198.
[24] R.C. Petersen *et al*., Aging, memory, and mild cognitive impairment. *International psychogeriatrics* **9** (1997), 65–69.
[25] G. McKhann, *et al*., 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** (2011), 263–269.

Address for correspondence

Tatsuya Ishikawa: ISIKAWA@jp.ibm.com
Yasunori Yamada: YSNR@jp.ibm.com