Automated Video Analysis of Handwashing Behavior as a Potential Marker of Cognitive Health in Older Adults

Ahmed Ashraf and Babak Taati

***Abstract*—The identification of different stages of cognitive impairment can allow older adults to receive timely care and plan for the level of caregiving. People with existing diagnosis of cognitive impairment go through episodic phases of dementia requiring different levels of care at different times. Monitoring the cognitive status of existing patients is, thus, critical to deciding the level of care required by older adults. In this paper, we present a system to assess the cognitive status of older adults by monitoring a common activity of daily living, namely handwashing. Specifically, we extract features from handwashing trials of participants diagnosed with different levels of dementia ranging from cognitively intact to severe cognitive impairment, as assessed by the mini-mental state exam (MMSE). Based on videos of handwashing trials, we extract two classes of features: one characterizing the occupancy of different sink regions by the participant, and the other capturing the path tortuosity of the motion trajectory of participant's hands. We perform correlation analysis to assess univariate capacity of individual features to predict MMSE scores. To assess multivariate performance, we use machine learning methods to train models that predict the cognitive status (aware, mild, moderate, severe), as well as the MMSE scores. We present results demonstrating that features derived from hand washing behavior can be potential surrogate markers of a person's dementia, which can be instrumental in developing automated tools for continuously monitoring the cognitive status of older adults.**

***Index Terms*—Computer vision, dementia, hand washing, mini-mental state exam (MMSE), pervasive health, smart home monitoring, video analysis.**

I. Introduction

DEMENTIA, including Alzheimer's disease, is one of the biggest global public health challenges [1]. As of 2013, over 35 million people worldwide live with the condition, and this number is projected to double by 2030 and more than triple to 135 million by 2050 [1]. Some of the common forms of dementia include Alzheimer's disease, vascular dementia, dementia with Lewy bodies, and frontotemporal dementia [2], [3]. Usual symptoms of dementia include memory loss, cognitive impairment, difficulty in communication, and mood changes among others [1], [2].

Manuscript received September 12, 2014; revised December 9, 2014 and February 3, 2015; accepted February 4, 2015. Date of publication March 16, 2015; date of current version March 3, 2016.

A. Ashraf is with the Toronto Rehabilitation Institute, University Health Network, Toronto, ON M5G 2C4, Canada (e-mail: ahmed.bilal@alumni.cmu.edu).

B. Taati is with the Department of Computer Science, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: Babak.Taati@uhn.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2015.2413358

Primary care clinicians may not recognize cognitive impairment when using routine history and physical examination [4], [5] in as many as 76% of patients with dementia [6]–[8]. Most of these patients are not diagnosed until they are at moderate to severe stages of the disease [9].

Identification of different stages of cognitive impairment can potentially allow patients to receive timely care which could in turn lead to improved quality of life both for older adults and their caregivers who can better plan informal/formal care in advance. Persons with mild cognitive impairment usually continue to live at home, while people with moderate impairment start to need more significant caregiving. As a result, a timely identification of the cognitive status of older adults is of crucial value to allow them to plan the level of care. Persons with dementia typically exhibit mild symptoms in an inconsistent fashion when they are going through a phase of mild cognitive impairment [10]. This makes the diagnosis even more challenging as patients can have normal outcomes in routine exams [4], [5]. Dementia, along with its comorbidities, can adversely affect the patient's ability to carry out activities of daily living (ADL), but the symptoms of the disease may manifest with lesser severity and with inconsistent frequency during early stages of the disease.

Apart from affecting memory and language functions, Alzheimer's disease affects motor planning and hand movement [11], [12]. Specifically, recent studies have noted a correlation between visuomotor impairment affecting eye-hand coordination in early-stage Alzheimer's disease [12]. Moreover, symptoms of apraxia have been reported in patients with senile dementia of Alzheimer's type, which become apparent in patients who otherwise demonstrate good language functions [11]. Handwashing is a common ADL performed routinely requiring both motor planning and eye-hand coordination, and there has been some evidence that it is affected by cognitive impairment [13], [14]. In this paper, we investigate if handwashing behavior can be a potential marker of cognitive disease.

A quantitative representation of the handwashing behavior may be formulated in multiple ways. To establish a proof-of-concept we first tested features derived from manually annotated handwashing trials for predicting cognitive status of the participants. We then explored if automated video analysis could be used for the same purpose. Specifically, we investigated features that encode the following aspects of the handwashing behavior: occupancy of different regions of the sink while completing the handwashing task, and the path tortuosity of the motion trajectory of the participant's hands. The latter class of features

TABLE I
DEMOGRAPHICS OF STUDY PARTICIPANTS

	Aware ($MMSE \geq 25$)	Mild ($MMSE = 19$–24)	Moderate ($MMSE = 10$–18)	Severe ($MMSE \leq 9$)
Number of Participants ($n = 27$)	9 (33%)	3 (11%)	9 (33%)	6 (22%)
Female: 22 (81%)	5 (56%)	3 (100%)	9 (100%)	5 (83%)
Male: 5 (19%)	4 (44%)	0 (0%)	0 (0%)	1 (17%)
Age range: 64–100 years	64–86	82–90	68–95	74–100
Age mean ± Std: 82.4 ± 9.5 years	75.7 ± 9.6	85 ± 4.4	84 ± 8.5	86.3 ± 10.4

is inspired by a recent study which correlates the path tortuosity of walking patterns to the risk of falling in older adults [15]. As part of this study, we first showed that several of the above mentioned features have statistically significant correlations with the dementia level as assessed by the mini-mental state exam (MMSE). We then trained machine learning models to predict the cognitive state categories based on the MMSE score (aware, mild, moderate, severe). We also trained regression models to predict the MMSE scores. Our results show that features encoding the above aspects of handwashing behavior can be potential predictors of cognitive impairment which can facilitate in developing tools for automated continuous monitoring of the cognitive status of older adults through a commonly executed ADL.

The rest of the paper is organized as follows: In Section II, we present a summary of related previous work. Details of the study population and data collection are given in Section III. In Section IV, we mention a proof-of-concept study using manually annotated videos of handwashing. In Section V, we detail the feature extraction methodology for automated video analysis of handwashing behavior. Sections VI and VII present the statistical analysis and results for predicting the MMSE scores and dementia categories respectively. Finally, in Section VIII, we present a discussion on our results and interpretation of top performing features as well as the limitations of the current study and future directions.

II. Previous Work

Smart homes and home automation is increasingly catching traction that can assist in monitoring the health status of older adults [16], [17]. Kaye *et al.* have presented a system that monitors computer usage to assess cognitive impairment of the users based on their usage patterns [17]. In another study, Dodge *et al.* monitored the walking patterns of older adults in their home setting and suggested that walking variability might be a predictor of mild cognitive impairment [16]. Handwashing behavior has been studied previously to provide assistance to older adults while washing their hands. For instance, in a system presented by Hoey *et al.*, hand motion is tracked using a camera while a person is washing hands, and assistance in the form of alerts is given to the user [13], [18]. The prompting system in [18] tracks responsiveness to the prompts, and indirectly estimates the dementia level. Although promising, the study in [18] remained inconclusive whether the estimated levels of dementia correlated with the outcome of a standardized test such as the MMSE score. The focus of our current study is a standalone and passive monitoring system which can assess the cognitive health status of older adults that correlates with MMSE scores. Specifically, we investigate if the level of dementia could be predicted using hand washing behavior as parameterized by the following features: statistics of occupancy of different regions of the sink while handwashing, and the path tortuosity of motion trajectory of participant's hands.

III. Materials and Methods

A. Study Population

We used a subset of the data analyzed in the study that investigated the effect of an operator's familiarity with the faucet on difficulty of operation [14]. All data was collected, stored, and analyzed according to a protocol that had research ethics approval (IRB, University at Buffalo and REB, University of Toronto). A brief review of the data is as follows. The study population consisted of 27 older adults (22 female, five male), age range: 64–100 years (82.4 ± 9.5 years). Each participant was administered the standardized MMSE exam [19] twice; once before the beginning, and once after the completion of the study. The average value of the two scores was taken, although the variation in the two scores was minimal (correlation between pre- and post-study MMSE scores was 0.98, $p < 0.001$). The first MMSE was administered one month before the beginning of the trials. The trials lasted for three months, and the second MMSE was administered immediately after the completion of the trials. MMSE is a 30 point questionnaire test that is used to screen for dementia and cognitive impairment. The exam samples cognitive functions including arithmetic, memory, and orientation [19]. Low to very low MMSE scores have been reported to be correlated with the presence of dementia [20]–[22]. Based on the MMSE score, the cognitive impairment of the test taker can be divided into four categories: severe ($\leq$ 9 points), moderate (10–18 points), mild (19–24 points), aware ($\geq$ 25 points).

B. Hand Washing Trials

Hand washing activity of the study participants was monitored in a designated washroom at a long-term care facility. The participants of the study were asked to wash their hands and videos of the sink area were recorded from an overhead view to enable post-trial analysis. Two researchers were present for all trials; one researcher acted as the caregiver, and the

TABLE II
LIST OF FEATURES EXTRACTED FROM EACH HANDWASHING TRIAL

#	Feature	Feature Description	Corr with MMSE	Corr with Age
1	*Crosshead lever*	1/0 to indicate if the faucet type is Crosshead (c) or not	−0.013	−0.008
2	*Dual lever*	1/0 to indicate if the faucet type is Dual lever (d) or not	−0.018	−0.003
3	*Single lever*	1/0 to indicate if the faucet type is Single lever (s) or not	0.031	0.010
Sink Occupancy Features: (See Fig. 2 for definition of sink regions)				
4	Occ_{Drain}	Percentage occupancy of the Drain region	0.332*	−0.139
5	Occ_{Knob}	Percentage occupancy of the Knob region	0.005	−0.010
6	Occ_{Nozzle}	Percentage occupancy of the Nozzle region	−0.058	0.009
7	$Occ_{\overline{\mathrm{Sink}}}$	Percentage occupancy of the $\overline{\mathrm{Sink}}$ region	0.564*	−0.225
8	Occ_{Sink}	Percentage occupancy of the Sink region	0.554*	−0.221
9	t_{Drain}	Time spent in the Drain region	−0.223*	0.127
10	t_{Knob}	Time spent in the Knob region	−0.319*	0.028
11	t_{Nozzle}	Time spent in the Nozzle region	−0.328*	0.075
12	$t_{\overline{\mathrm{Sink}}}$	Time spent in the $\overline{\mathrm{Sink}}$ region	0.046	−0.049
13	t_{Sink}	Time spent in the Sink region	0.007	−0.042
Path Tortuosity Features:				
14	FD_{Longest}	Fractal D of the longest trajectory (in terms of duration)	−0.410*	0.184
15	$FD_{\max}$	Maximum fractal D	−0.512*	0.197
16	$FD_{\min}$	Minimum fractal D	−0.022	0.017
17	FD_{avg}	Average fractal D	−0.517*	0.225
18	FD_{std}	Standard deviation of fractal D	−0.423*	0.197

The last column represents the Pearson's correlation of the individual feature with MMSE score. A "*" following the correlation represents the statistical significance of the correlation i.e, $p < 0.05$. Last column shows the correlation of the respective feature with age.

other researcher operated study-related equipment. For safety and study uniformity, the caregiver sat the participant in a wheelchair at the beginning of each trial. The caregiver positioned the participant at the sink and asked him or her to wash his or her hands. Each participant was asked to complete about ten hand washing trials for five faucet types, i.e., crosshead, dual lever, single lever, electronic, and plastic wand [14]. In all, 1309 trials were conducted. For the purposes of this study, 741 trials that corresponded to commonly occurring faucet types, i.e., crosshead, dual lever, and single lever, were included. Patient demographics of our study population with break down according to MMSE categories, age, and gender is given in Table I.

IV. PROOF-OF-CONCEPT STUDY

To establish a proof-of-concept, the videos of handwashing trials were manually annotated as follows. A rater scored the number and types of errors in each handwashing trial [14]. To validate data reliability, a second rater scored three randomly selected trials for each participant. Inter-rater agreement between the primary and secondary raters was examined using Cohen's kappa ($\kappa = 0.94$). Following features were computed from the manually annotated data: statistics of time taken to complete the handwashing task and the number and types of errors. A four class random forest classifier [23] was trained to predict the dementia level as aware, mild, moderate, or severe, which gave a classification accuracy of 61%. Prompted by this moderate performance of the classifier using manually annotated videos we explored feature representations which could be extracted automatically, and could also characterize the handwashing behavior with more granularity. We present these features in the following section.

V. FEATURE EXTRACTION FOR AUTOMATED VIDEO ANALYSIS

For automated video analysis of handwashing behavior we explored features characterizing the occupancy of different sink regions by the participant, and path tortuosity of the motion trajectory of participant's hands. Specifically, each handwashing trial was represented as a feature vector as follows. First, the faucet type was encoded as a categorical feature represented by a set of three binary variables. For instance, if the trial involved a crosshead faucet, it was encoded as [1, 0, 0] etc. The three binary variables indicating the faucet type are shown as features 1–3 in Table II. Thereafter, computer vision based features were extracted from each video. This feature extraction is a multistage process as depicted in the block diagram of Fig. 1. Given an input video, the region corresponding to a participant's hands was identified by skin tone detection for every frame. The output of skin tone detection was post processed by a morphological opening step to eliminate small noisy specks [24]. We then computed two types of features described in the following sections.

A. Sink Occupancy Features

Features included in this type capture the occupancy of various regions of the sink area as the participant performs the handwashing task. Specifically, we divided a video frame into subregions as illustrated in Fig. 2. The "Drain" region (shown in blue) constitutes a region centered around the sink drain. The "Knob" region (shown in red) consists of two regions

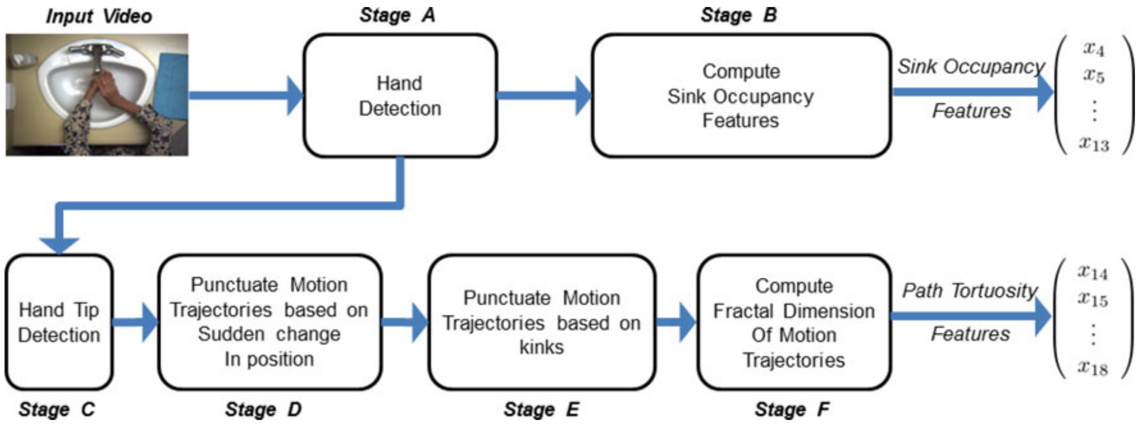


Fig. 1. Block diagram of the feature extraction process.

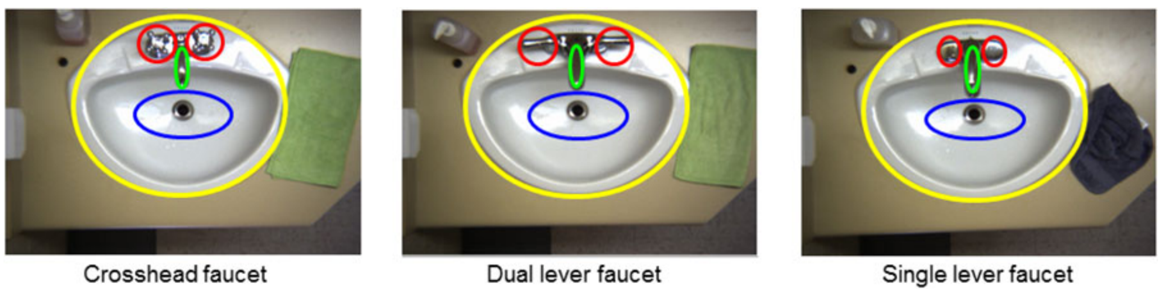


Fig. 2. Definitions of different sink regions. Drain (Blue), Knob (red), Nozzle (green), $\overline{\text{Sink}}$: subset of yellow region excluding all of the above, Sink (yellow).

corresponding to the knobs of the cross-head faucet or the two levers of the dual-lever faucet. There are no knobs/levers in this region for a single-lever faucet, but we still analyze this region for a single lever faucet to potentially account for any confusion a participant might have. The "Nozzle" region (shown in green) corresponds to the top of the faucet where one would find the lever for a singer lever faucet. "$\overline{\text{Sink}}$" region is the area of the sink which does not correspond to any of the above regions. And finally, the "Sink" region (shown in yellow) is the union of all the above regions.

The features that capture the statistics for occupancy of different regions are summarized in Table II. For every region, the percentage area occupied by the hand per video frame was computed, and the average occupancy percentage for each region over the entire video was calculated (features 4–8 in Table II). The time spent in different subregions of the sink was also estimated. To estimate the time, a subregion was declared to be occupied if occupancy percentage exceeded a threshold (10%). The number of frames for which a region was deemed occupied was then divided by the frame rate to get an estimate of the time spent in each region (features 9–13 in Table II).

B. Path Tortuosity Features

Path tortuosity or variability in movement path, as assessed by the fractal dimension of the path, has been previously used to predict the risk of falls in older adults [15]. Motivated by this study, we investigated if the path tortuosity of hand motion while washing could be a helpful feature for assessing cognitive health status.

For computing the path tortuosity, we first detected the tip of the hand based on the output of hand detection (stage A) as shown in the block diagram of Fig. 1. For tracking the tip of the hand, the top-most connected component (as assessed by the y-coordinate of its centroid) in the output of stage A was detected. The centroid of the top 10% pixels within this component was selected as the tip of the hand. It was observed that the skin detection algorithm was occasionally not perfect, and as a result did not allow to consistently detect the hand-tip. To avoid this problem, the motion trajectory was punctuated by points where the change in the position of the tip exceeded a threshold, effectively filtering out sudden motions. This step is carried out in stage D of the block diagram. A sample set of motion trajectories as detected by stage D are illustrated in the first row of Fig. 3. As seen in the figure, there are a number of kinks in the trajectory, which can unnecessarily bias the fractal dimension of the trajectory toward higher values. As further processing, the motion trajectories were also punctuated by the presence of kinks as assessed by a curvature detection algorithm given in [25]. The outcome of this step is shown in the second and third rows of Fig. 3. We should note that shorter segments are predisposed to have smaller fractal dimension. To avoid this bias toward underestimating the fractal dimension, very short motion segments in terms of duration (those that lasted less

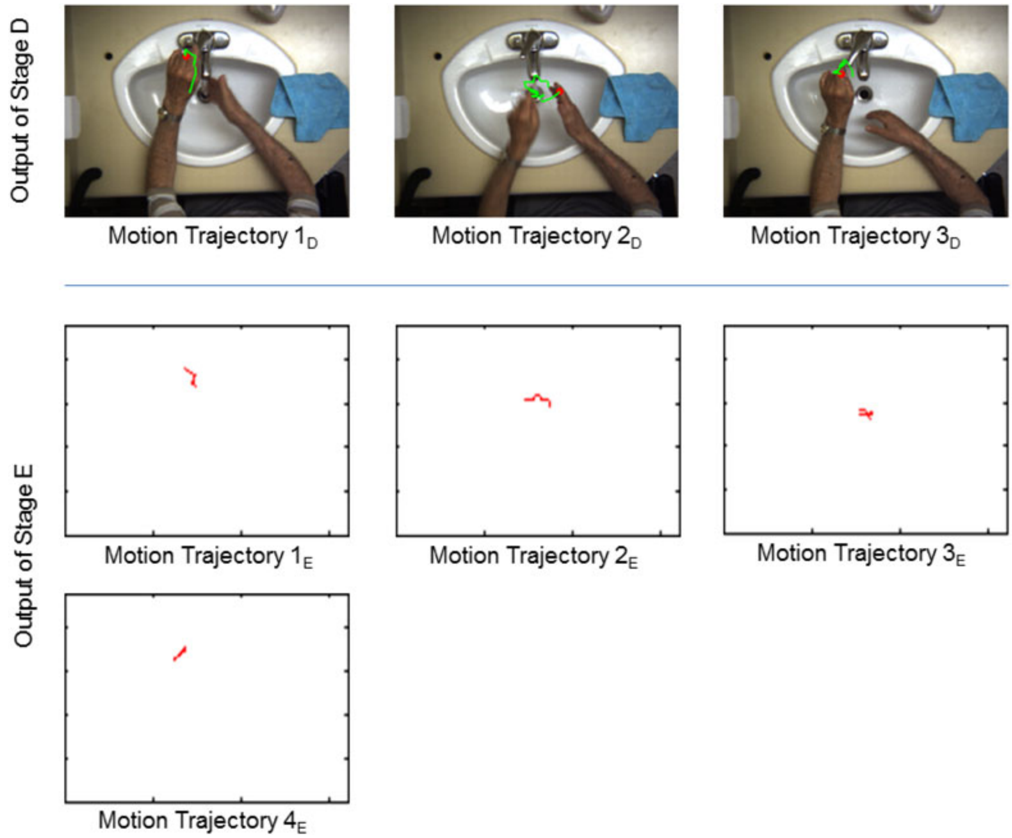


Fig. 3. Detection of motion trajectories. Output of stage D: The first row shows the output of stage D where the motion trajectories (shown in green) are superimposed on the last video frame corresponding to that trajectory, and the red dot represents the hand-tip (Motion trajectory 1_D signifies the first motion trajectory in the output of step D). Output of stage E: Second and third rows show the output of stage E where the trajectories were punctuated by kinks, and very short motion segments were discarded.

than 2 s) were discarded. Starting with the input video, the output of stage E is a collection of motion trajectories executed by the participant during the trial. The fractal dimension of each of these trajectories was computed using the box counting algorithm [26]. Finally, the statistics of the fractal dimension were calculated for multiple motion trajectories executed during a trial as represented by features 14–18 in Table II.

C. Collapsed Features

The features as defined above are extracted from an individual video trial. In order to explore the variability in these features over multiple trials, for every participant all trials corresponding to a particular faucet type were collapsed to a single data point by computing the statistics (up to second order) of the features in Table II. Effectively, all the trials for a particular participant were "*collapsed*" to three trials, each corresponding to a faucet type. For simplicity, from here on we would refer to the features computed at the level of individual videos as "*uncollapsed*" (see Table II), and the statistics of uncollapsed features over multiple video trials as collapsed features. Features computed as part of the collapsed representation are summarized in Table III.

VI. Predicting MMSE Scores

In this section, we explore if the feature representations described above can predict the MMSE scores of participants. To analyze univariate performance of individual features, correlations of features with MMSE scores were evaluated, shown in the last columns of Tables II and III for uncollapsed and collapsed features respectively. Features with statistically significant correlations ($p < 0.05$) are indicated by a "*" after their respective correlations. The last column of Tables II and III shows the correlation between the respective feature and age.

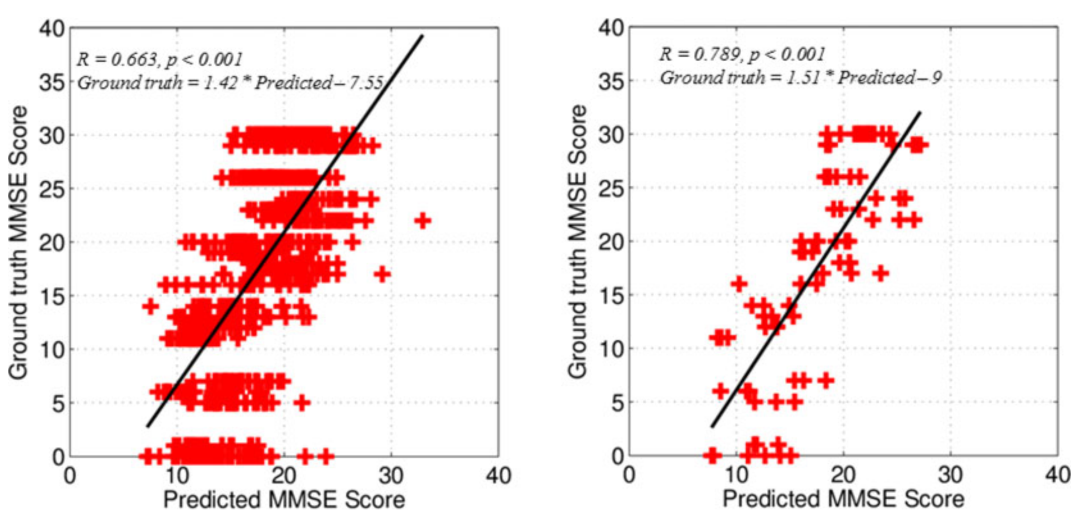
To assess multivariate performance, a regression model was trained using leave-one-subject-out cross validation. Instead of ordinary least squares linear regression, we used regularized regression based on least absolute shrinkage and selection operator (LASSO) [27]. In LASSO regression, the sum of absolute values of the regression weights is constrained to be less than a threshold, which prevents regression weights from taking on very large values, thereby providing robustness against overfitting and inducing sparsity [27]. Specifically, a regression model was trained using the training set corresponding to the current leave-one-out fold, and all trials of the left out subject were tested on this model to give a predicted MMSE score. Pearson's correlation coefficient was computed between the predicted and the actual MMSE score to assess the performance of the model. Regression models were trained for both the uncollapsed and collapsed feature representations.

The predicted MMSE scores using the regression model based on uncollapsed representation had a mild correlation ($R = 0.663,\ p < 0.001$), while the model based on collapsed features

TABLE III
LIST OF COLLAPSED FEATURES

#	Feature	Feature Description	Corr with MMSE	Corr with Age
1	*Crosshead lever*	1/0 to indicate if the faucet type is Crosshead (c) or not	0.000	−0.000
2	*Dual lever*	1/0 to indicate if the faucet type is Dual lever (d) or not	0.000	0.000
3	*Single lever*	1/0 to indicate if the faucet type is Single lever (s) or not	0.000	−0.000
4	$\mu(Occ_{\mathrm{Drain}})$	Mean % occupancy of Drain region	0.486*	−0.190
5	$\sigma(Occ_{\mathrm{Drain}})$	Std deviation in % occupancy of Drain region	−0.105	0.035
6	$\mu(Occ_{\mathrm{Knob}})$	Mean % occupancy of Knob region	0.010	−0.015
7	$\sigma(Occ_{\mathrm{Knob}})$	Std deviation in % occupancy of Knob region	−0.245*	0.043
8	$\mu(Occ_{\mathrm{Nozzle}})$	Mean % occupancy of Nozzle region	−0.090	−0.010
9	$\sigma(Occ_{\mathrm{Nozzle}})$	Std deviation in % occupancy of Nozzle region	−0.210	0.068
10	$\mu(Occ_{\overline{\mathrm{Sink}}})$	Mean % occupancy of $\overline{\mathrm{Sink}}$ region	0.695*	−0.265
11	$\sigma(Occ_{\overline{\mathrm{Sink}}})$	Std deviation in % occupancy of $\overline{\mathrm{Sink}}$ region	0.098	−0.083
12	$\mu(Occ_{\mathrm{Sink}})$	Mean % occupancy of Sink region	0.693*	−0.264
13	$\sigma(Occ_{\mathrm{Sink}})$	Std deviation in % occupancy of Sink region	0.038	−0.104
14	$\mu(t_{\mathrm{Drain}})$	Mean time spent in Drain region	−0.301*	0.186
15	$\sigma(t_{\mathrm{Drain}})$	Std deviation of time spent in Drain region	−0.595*	0.196
16	$\mu(t_{\mathrm{Knob}})$	Mean time spent in Knob region	−0.472*	0.026
17	$\sigma(t_{\mathrm{Knob}})$	Std deviation of time spent in Knob region	−0.507*	0.021
18	$\mu(t_{\mathrm{Nozzle}})$	Mean time spent in Nozzle region	−0.492*	0.092
19	$\sigma(t_{\mathrm{Nozzle}})$	Std deviation of time spent in Nozzle region	−0.578*	0.062
20	$\mu(t_{\overline{\mathrm{Sink}}})$	Mean time spent in $\overline{\mathrm{Sink}}$ region	0.053	−0.070
21	$\sigma(t_{\overline{\mathrm{Sink}}})$	Std deviation of time spent in $\overline{\mathrm{Sink}}$ region	−0.345*	0.025
22	$\mu(t_{\mathrm{Sink}})$	Mean time spent in Sink region	−0.003	−0.064
23	$\sigma(t_{\mathrm{Sink}})$	Std deviation of time spent in Sink region	−0.403*	−0.028
24	$\mu(\mathit{FD\ Longest})$	Mean fractal D of the longest trajectory	−0.631*	0.270
25	$\sigma(\mathit{FD\ Longest})$	Std deviation in fractal D of the longest trajectory	−0.469*	0.138
26	$\mu(FD_{\mathrm{max}})$	Mean maximum fractal D over multiple trials	−0.689*	0.265
27	$\sigma(FD_{\mathrm{max}})$	Std deviation in maximum fractal D over multiple trials	−0.546*	0.037
28	$\mu(FD_{min})$	Mean minimum fractal D over multiple trials	0.011	−0.002
29	$\sigma(FD_{min})$	Std deviation in minimum fractal D over multiple trials	−0.538*	0.144
30	$\mu(FD_{avg})$	Mean of average fractal D over multiple trials	−0.676*	0.285
31	$\sigma(FD_{avg})$	Std dev in average fractal D over multiple trials	−0.522*	0.094
32	$\mu(FD_{std})$	Mean of std dev in fractal D over multiple trials	−0.686*	0.311
33	$\sigma(FD_{std})$	Std dev of std dev in fractal D over multiple trials	−0.267*	−0.033

The last column represents the Pearson's correlation of the individual feature with MMSE score. A "*" following the correlation represents the statistical significance of the correlation i.e, $p < 0{:}05$. Last column represents the correlation of the respective feature with age.



(a) (b)

Fig. 4. Scatter plot between the predicted MMSE and the ground truth MMSE scores based on leave-one-subject-out LASSO regression. (a) Model based on uncollapsed features, (b) Model based on collapsed features.

TABLE IV
CLASSIFICATION RESULTS FOR PREDICTING DEMENTIA CATEGORIES BASED ON LEAVE-ONE-SUBJECT-OUT RANDOM FOREST CLASSIFIER USING UNCOLLAPSED FEATURES

		Predicted				
		Aware	Mild	Moderate	Severe	% correct
True	Aware	149	66	26	5	60%
	Mild	63	1	15	6	1.1%
	Moderate	60	12	146	36	58%
	Severe	5	3	58	90	57.7%
						52.1%

TABLE V
CLASSIFICATION RESULTS FOR PREDICTING DEMENTIA CATEGORIES BASED ON LEAVE-ONE-SUBJECT-OUT RANDOM FOREST CLASSIFIER USING COLLAPSED FEATURES

		Predicted				
		Aware	Mild	Moderate	Severe	% correct
True	Aware	21	5	1	0	77%
	Mild	7	0	2	0	0.0%
	Moderate	3	1	22	1	81.5%
	Severe	0	0	4	14	77%
						70.4%

had a moderate to strong correlation ($R = 0.789,\ p < 0.001$) with the actual MMSE scores. Scatter plots between the predicted and ground truth MMSE scores for the model based on uncollapsed and collapsed features are shown in Fig. 4.

VII. PREDICTING DEMENTIA CATEGORIES

While the performance of regression models was promising, in this section we investigate if the proposed features could also be used for predicting dementia categories as determined by MMSE scores. We followed a similar leave-one-subject-out cross validation scheme as mentioned in the previous section, but instead of regression, we trained random forest classifiers, [23], for predicting the dementia categories, i.e., aware, mild, moderate, and severe. The choice of random forest classifier was motivated by the fact that it has become a standard data analysis tool for analyzing high dimensional datasets and give state of the art performance for a wide variety of applications [28]. Specifically, a leave-one-subject-out cross validation scheme was employed. For every subject, its examples were left out, and the remaining were used to train the classification model. This model was then used to test the unseen examples of the left-out subject. This process was repeated for all the subjects.The leave-one-out classification results in terms of confusion matrices are given in Tables IV and V for uncollapsed and collapsed feature representations, respectively. The overall classification accuracy of the four-class classifier with uncollapsed features was 52.1%, while it was 70.4% for collapsed features.

To rank the features according to their relative importance, the random forest variable importance metric was employed [23]. Specifically, to measure the importance of a feature, the values of this feature are permuted among the training data, and the classification error is again computed on this permuted data. The importance of the feature is computed by averaging the difference in error before and after the permutation over all the trees in the forest. Since we used leave-one-out cross validation, the average rank of features was computed over all cross-validation folds. List of features arranged in ascending order of their rank is given in Table VI.

TABLE VI
AVERAGE RANK OF FEATURES OVER ALL CROSS-VALIDATION LOOPS

#	Feature	Average Feature Rank
1	$\mu(Occ_{\mathrm{Sink}})$	1.72
2	$\mu(Occ_{\overline{\mathrm{Sink}}})$	2.40
3	$\mu(FD_{\mathrm{avg}})$	2.40
4	$\sigma(t_{\mathrm{Drain}})$	3.76
5	$\mu(Occ_{\mathrm{Drain}})$	5.48
6	$\sigma(t_{\mathrm{Knob}})$	6.80
7	$\mu(FD_{\mathrm{max}})$	8.20
8	$\mu(FD_{\mathrm{std}})$	9.64
9	μ*(FD Longest)*	12.12
10	$\sigma(t_{\mathrm{Nozzle}})$	12.20
11	$\sigma(FD_{\mathrm{min}})$	12.60
12	$\sigma(Occ_{\mathrm{Drain}})$	14.28
13	$\sigma(FD_{\mathrm{max}})$	15.20
14	$\mu(FD_{\mathrm{min}})$	16.48
15	$\sigma(FD_{\mathrm{avg}})$	17.20
16	$\sigma(t_{\mathrm{Sink}})$	19.08
17	σ*(FD Longest)*	19.72
18	$\sigma(Occ_{\mathrm{Knob}})$	22.84
19	$\sigma(t_{\overline{\mathrm{Sink}}})$	23.08
20	$\sigma(FD_{\mathrm{std}})$	23.76
21	$\mu(t_{\overline{\mathrm{Sink}}})$	23.80
22	$\sigma(Occ_{\mathrm{Sink}})$	24.92
23	$\mu(t_{\mathrm{Sink}})$	25.60
24	$\mu(Occ_{\mathrm{Nozzle}})$	25.68
25	$\mu(t_{\mathrm{Knob}})$	26.44
26	$\mu(t_{\mathrm{Nozzle}})$	26.88
27	$\mu(t_{\mathrm{Drain}})$	27.08
28	$\sigma(Occ_{\overline{\mathrm{Sink}}})$	27.20
29	$\sigma(Occ_{\mathrm{Nozzle}})$	27.80
30	$\mu(Occ_{Knob})$	27.96
31	*Crosshead lever*	28.40
32	*Single lever*	30.04
33	*Dual lever*	30.52

VIII. DISCUSSION

In this study, we explored if computer vision based features encoding handwashing behavior of older adults can serve as surrogate predictors of their cognitive health. Our results show that the above mentioned features have moderate to strong correlation with the MMSE scores, and can also predict the dementia level as assessed by MMSE score-based categories. The first observation that stems from our results is that collapsed feature representation outperforms the uncollapsed features both for regression ($R_{\mathrm{collapsed}} = 0.789$ versus $R_{\mathrm{uncollapsed}} = 0.663$) and classification tasks (accuracy of 70.2% versus 52.1%). This shows that statistics captured over multiple (~ 10) handwashing trials are more representative of cognitive health as compared to isolated single trials. This is in line with previous findings that about 76% of dementia patients go undiagnosed in routine primary care clinic visits [6]–[8], and highlights the importance

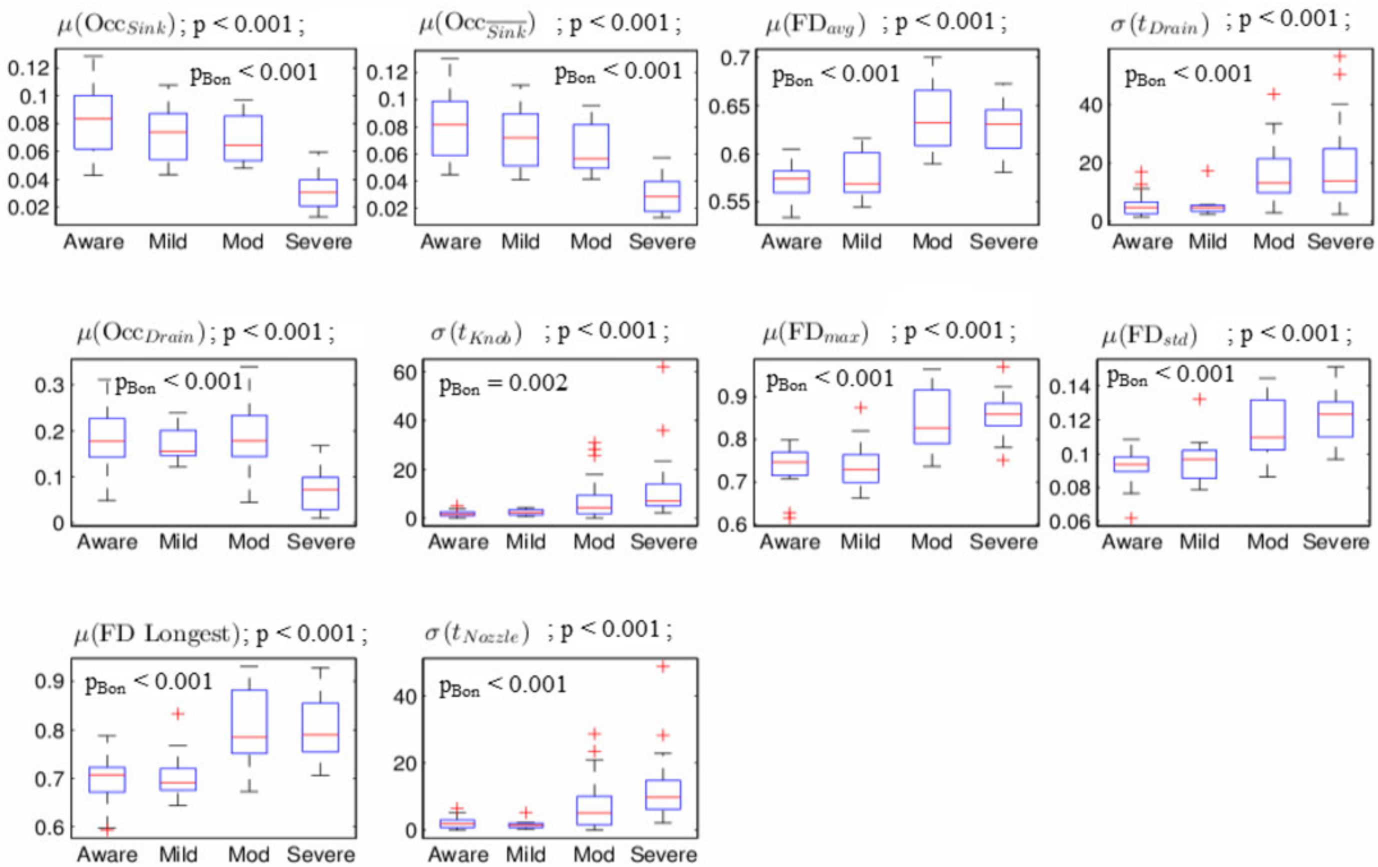


Fig. 5. Box plots of highly ranked features showing the variation of these features across the four dementia categories. On each box, the central red mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as red pluses. Two p-values are shown: p_{Bon} shows the p-value adjusted after Bonferroni correction.

of practical systems capable of monitoring cognitive health on a frequent basis.

In the confusion matrix of Table V, the classifier gives relatively high accuracies of 77%, 81.5%, and 77% to detect aware, moderate, and severe categories, respectively, pointing to the fact that it might be easier to distinguish more extreme categories. All mild cases were misclassified either as aware or moderate. There were only three participants (out of 27) with mild impairment in our dataset. As a result, when videos of a participant with mild dementia were being tested, the leave-one-out model only had *two* participants to learn from. Poor performance in detecting mild cases can in part be because of the underrepresentation of participants with mild impairment in our dataset and needs to be explored further.

Ranking of different features as shown in Table VI needs further attention. Box plots of the variation of top ten features are shown in Fig. 5. To take the effect of multiple comparisons into account, Bonferroni correction was applied to the p-values. As shown on the plots the p-values remain significant after applying the correction. Three main patterns emerge from these box plots. First, more cognitively intact participants tend to have higher percentage occupancy of different sink regions, as can be seen in the box plots of the features $\mu(Occ_{\mathrm{Sink}})$, $\mu(Occ_{\overline{\mathrm{Sink}}})$, and $\mu(Occ_{\mathrm{Drain}})$. More interestingly, participants with moderate and severe level dementia have higher values of path-tortuosity (assessed by fractal dimension of the motion trajectories) as compared to participants who are cognitively aware or have mild dementia. This is evident in the box plots for the following features $\mu(FD_{\mathrm{avg}})$, $\mu(FD_{\mathrm{max}})$, and $\mu(FD_{\mathrm{Longest}})$. This is consistent with a previous study in which the path tortuosity of walking patterns was indicative of cognitive health [15]. Further, from the box plots of $\sigma(t_{\mathrm{Drain}})$ and $\sigma(t_{\mathrm{Nozzle}})$, it can be seen that participants with moderate and severe dementia have higher standard deviations in the time they spend in different subregions of the sink. This potentially points to the fact that inconsistent handwashing behavior in older adults might be representative of different levels of dementia. These findings are in line with [16], which reports variability in walking speed to be an indicator of cognitive impairment.

Some limitations of the study must be noted. Since the data used was initially collected for an earlier study, we did not have control over the study protocol. To assess the generalizability of results, future studies will incorporate cognitive tests other than MMSE scores such as the Montreal Cognitive Assessment, as well as a test for apraxia, and will also control for comorbidities. In the current framework, a participant's hands are tracked by a skin-tone detection algorithm, which is not 100% accurate. In subsequent studies, 3-D depth sensing could be employed for improved hand detection. Since the dataset was not specifically collected for detecting early onset of dementia, participants with mild dementia (based on their MMSE score) are underrepresented. Future studies will work toward analyzing a larger population in which the number of participants from all categories are comparable. One of the aims of this study was to gain insight into the importance of different features for predicting the cognitive status, and if there are clinical interpretations for the top ranked features selected by the model. As a result, all models were trained in the original feature space instead of going in a latent feature subspace. Although random forest classifiers are well suited to avoid overfitting in high dimensional features spaces, future studies will explore the effect of dimensionality reduction techniques. Finally, we are working toward putting together an extended study wherein participants could be monitored on a long term basis as they undergo cognitive decline from cognitively intact to various stages of dementia, which can potentially allow to explore person specific models.

References

[1] (2013). Journey of caring: An analysis of long-term care for dementia. The World Alzheimer Rep., [Online]. Available: http://www.alz.co.uk/research/WorldAlzheimerReport2013.pdf

[2] D. Shub and M. Kunik, "Psychiatric comorbidity in persons with dementia: Assessment and treatment strategies," *Psych. Times*, vol. 26, no. 4, pp. 32–36, Apr. 2009.

[3] "Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. Neuropathology Group of the Medical Research Council Cognitive Function and Aging Study (MRC-CFAS)," *Lancet*, vol. 357, no. 9251, pp. 169–175, 2001.

[4] T. Holsinger, J. Deveau, M. Boustani, and J. W. Williams, "Does this patient have dementia?" *JAMA*, vol. 297, pp. 2391–2404, 2007.

[5] M. Ganguli, E. Rodriguez, B. Mulsant, S. Richards, R. Pandav, and J. V. Bilt, "Detection and management of cognitive impairment in primary care: The steel valley seniors survey," *J. Amer. Geriatrics Soc.*, vol. 52, pp. 1668–1675, 2004.

[6] V. G. Valcour, K. H. Masaki, J. D. Curb, and P. L. Blanchette, "The detection of dementia in the primary care setting," *Archives Internal Med.*, vol. 160, pp. 2964–2968, 2000.

[7] M. Olafsdottir, I. Skoog, and J. Marcusson, "Detection of dementia in primary care: The Linkoping study," *Dement Geriatrics Cogn. Disorders*, vol. 11, pp. 223–229, 2000.

[8] J. Chodosh, D. B. Petitti, M. Elliott, R. D. Hays, V. C. Crooks, and D. B. Reuben, "Physician recognition of cognitive impairment: Evaluating the need for improvement," *J. Amer. Geriatrics Soc.*, vol. 52, pp. 1051–1059, 2004.

[9] A. J. Mitchell, "The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: A meta-analysis," *Int. J. Geriatrics Psych.*, vol. 23, pp. 1191–1202, 2008.

[10] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, "Mild cognitive impairment: Clinical characterization and outcome," *Archives Neurol.*, vol. 56, no. 3, pp. 303–308, 1999.

[11] S. Rapcsak, S. Croswell, and A. Rubens, "Apraxia in Alzheimer's disease," *Neurology*, vol. 39, no. 5, p. 664, May 1989.

[12] S. Verheij, D. Muilwijk, J. Pel, T. J. Cammen, F. U. Mattace-Raso, and J. Steen, "Visuomotor impairment in early-stage Alzheimer's disease: Changes in relative timing of eye and hand movements," *J. Alzheimers Disease*, vol. 30, no. 1, pp. 131–43, 2012.

[13] A. Mihailidis, J. N. Boger, T. Craig, and J. Hoey, "The coach prompting system to assist older adults with dementia through handwashing: An efficacy study," *BMC Geriatrics*, vol. 8, p. 28, Nov. 2008.

[14] J. N. Boger, T. Craig, and A. Mihailidis, "Examining the impact of familiarity on faucet usability for older adults with dementia," *BMC Geriatrics*, vol. 13, no. 63, p. 1, Jun. 2013.

[15] W. D. Kearns, J. L. Fozard, M. Becker, J. M. Jasiewicz, J. D. Craighead, L. Holtsclaw, and C. Dion, "Path tortuosity in everyday movements of elderly persons increases fall prediction beyond knowledge of fall history, medication use, and standardized gait and balance assessments," *J. Amer. Med. Directors Assoc.*, vol. 13, no. 7, p. 665, Sep. 2012.

[16] H. Dodge, N. Mattek, D. Austin, T. Hayes, and J. Kaye, "In-home walking speeds and variability trajectories associated with mild cognitive impairment," *Neurology*, vol. 78, no. 24, pp. 1946–1952, Jun. 2012.

[17] J. Kaye, N. Mattek, H. Dodge, I. Campbell, T. Hayes, D. Austin, W. H. K. Wild, H. Jimison, and M. Pavel, "Unobtrusive measurement of daily computer use to detect mild cognitive impairment," *Alzheimers Dementia*, vol. 10, pp. 10–17, May 2013.

[18] J. Hoey, P. Poupar, A. von Bertoldi, T. Craig, C. Boutilier, and A. Mihailidis, "Automated handwashing assistance for persons with dementia using video and a partially observable Markov decision process," *Comput. Vision Image Understanding*, vol. 114, pp. 503–519, May 2010.

[19] D. Mungas, "In-office mental status testing: A practical guide," *Geriatrics*, vol. 46, no. 7, pp. 54–58, Jul. 1991.

[20] S. Palmqvist, O. Hansson, L. Minthon, and E. Londos, "Practical suggestions on how to differentiate dementia with Lewy bodies from Alzheimer's disease with common cognitive tests," *Int. J. Geriatric Psych.*, vol. 24, no. 12, pp. 1405–1412, Dec. 2009.

[21] A. L. Jefferson, S. A. Cosentino, S. K. Ball, B. Bogdanhoff, N. Leopold, E. Kaplan, and D. J. Libon, "Errors produced on the mini-mental state examination and neuropsychological test performance in Alzheimer's disease, ischemic vascular dementia, and Parkinson's disease," *J. Neuropsych. Clin. Neurosci.*, vol. 14, no. 3, pp. 311–320, Summer 2002.

[22] T. A. Ala, L. F. Hughes, G. A. K. amd M. W. Ghobrial, and R. J. Elble, "The mini-mental state exam may help in the differentiation of dementia with lewy bodies and Alzheimer's disease," *Int. J. Geriatric Psych.*, vol. 17, no. 6, pp. 503–509, Jun. 2002.

[23] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[24] R. C. Gonzalez and R. E. Woods, *Digital Image Processing,* 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.

[25] D. Chetverikov, "A simple and efficient algorithm for detection of high curvature points in planar curves," in *Proc. Comput. Anal. Images Patterns*, (series Lecture Notes in Computer Science), vol. 2756, N. Petkov and M. Westenberg, Eds. Berlin. Germany: Springer, 2003, pp. 746–753.

[26] T. G. Smith, G. D. Lange, and W. B. Marks, "Fractal methods and results in cellular morphology dimensions, lacunarity and multifractals," *J. Neurosci. Methods*, vol. 69, no. 2, pp. 123–136, 1996.

[27] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., Series B*, vol. 58, pp. 267–288, 1994.

[28] A.-L. Boulesteix, S. Janitza, J. Kruppa, and I. R.König, "Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics," *Wiley Interdisciplinary Rev.: Data Min. Knowl. Discovery*, vol. 2, no. 6, pp. 493–507, Nov. 2012.

Authors' photograph and biography not available at the time of publication.