Support for a Clinical Diagnosis of Mild Cognitive Impairment Using Photoplethysmography and Gait Sensors

Migyeong Gwak [a], Majid Sarrafzadeh [b], and Ellen Woo [c]
University of California Los Angeles, Los Angeles, CA, USA
E-mail: mgwak@cs.ucla.edu [a]; majid@cs.ucla.edu [b]; ewoo@mednet.ucla.edu [c] Tel: +1-310-825-2273

***Abstract*— Mild Cognitive Impairment (MCI) is known as a transitional state between normal aging and Alzheimer's disease. Early and reliable detection of MCI can prepare progressive dementia. In this study, we propose a novel application area of machine learning in the clinical diagnosis of MCI. Our assessment system includes a commercial smartwatch and a wireless pulse oximeter. We investigate predictors from photoplethysmography (PPG) and gait (accelerometer and gyroscope) sensor data. We also demonstrate a feature selection algorithm for a better classification of MCI from cognitively healthy (CH) using the sensor-derived features. Our classification accuracy result is 82% with the PPG dataset and 86% with the gait dataset from 69 elderly participants (72.45 10.55 years; 34 MCI and 35 CH), which is a higher classification accuracy than only using the administered neuropsychological screening instrument. These serve the criteria of mobility, non-invasiveness and easy assessment, and a reliable data transmission link through a connected smartphone. This study supports that sensor-derived parameters have the potential to support a clinical diagnosis and to reduce the diagnostic burden on healthcare professionals.**

I. Introduction

Dementia is one of the major causes of death among the geriatric population in the United States. There is no complete cure or treatment to stop the most progressive dementia [1]. Currently available treatments assist to temporarily improve symptoms. Cognitive impairment is typically present in the early stage of dementia [2] and is often the only preliminary indicator of the underlying condition. Mild Cognitive Impairment (MCI) is known as a transitional state between normal aging and Alzheimer's disease [2]. MCI is associated with dementia risk and detectable as the earliest clear-cut cognitive deficit [3].

The current detection of cognitive deficits is burdensome to administer and liable to misinterpretation. There is no one test to determine the overall dementia evaluation [1]. The symptoms of dementia vary and gradually appear over a number of years and include the decline in mental ability, such as memory loss and difficulty with reasoning and communication. Memory decline can also be caused by depression, excessive use of alcohol, or vitamin deficiencies [1]. Overlapping symptoms with other diseases make it difficult for doctors to determine the diagnosis immediately. The initial cognitive impairment could be noticed by short-administered neuropsychological tests, but the diagnosis of a specific type of dementia requires a high level of certainty in medical history, laboratory tests, and characteristic changes in mental ability [1]. We need a reliable detection of MCI that enables one to diagnose and then use that information to delay the onset of dementia.

Several sensors can be used for a clinical diagnosis of MCI. A study by Karakostas et al. [4] used a camera, tag sensors, a microphone, and a wristwatch to support clinicians in dementia assessment. They mainly focused on a semantic interpretation of speech analysis and daily activities recognition in a lab environment, but this application is stationary and expensive to install. In our study, we concentrated mobile and wearable devices including photoplethysmography (PPG), accelerometer, and gyroscope sensors to discriminate between MCI and the cognitively healthy (CH). Our assessment system includes a commercial smartwatch and a wireless pulse oximeter. This approach offers mobility, non-invasiveness and easy assessment, and a reliable data transmission link through a connected smartphone. Our proposed system is easily applied to clinical setting or even in the daily life.

We verified that the PPG and gait signals are relevant to detect a person's cognition status through many other studies. PPG uses an optical technique to measure changes in the blood volume in the blood vessel [5]. PPG signals are acquired by absorption of red and infrared wavelengths that passed through the finger-tip or the earlobe [6]. The heart rate variability (HRV), which is the beat-to-beat variations in heart rate, can be measured by PPG. The autonomic nervous system (ANS) influences the human body's major physiological changes including the heart rate, blood pressure, and respiration. HRV is related to the individual's cognitive function including memory performance, attention, and executive function. According to a study by Hansen et al (2003), the higher HRV group has a better performance of more correct responses and faster reaction time on a working memory test and a continuous performance test [7]. A study by Duschek et al. (2009) investigated that on-task peak-to-peak interval and power of mid-frequency band of HRV are negatively correlated with attentional function [8]. Taelman et al (2009) demonstrated that a mental task changes the heart rate and HRV [9]. Shah et al. (2011) had a study that controlled familial and genetic influences and found a statistically positive association between HRV and verbal memory and a learning task [10].

Gait is reliant on cognitive functions and gait abnormalities are common in the early stages of cognitive decline [11]. Gait requires multiple cognitive inputs to maintain upright posture and motor control and complex cognitive functions to shift and avoid obstacles [11]. The study of Verghese et al. (2002) explains that subjects with neurologic gait abnormalities have a higher chance of experiencing non-Alzheimer's dementias [12]. The neurologic gait abnormalities were determined by board-certified neurologists in their study [12]. Many other studies that have examined the gait of older adults demonstrate that gait velocity and stride length decreases during aging [11]. Most smartphone and smartwatch apps provide simple step counts using an accelerometer but do not provide the detailed qualities of a person's gait needed to detect the abnormalities. As above studies declared that HRV and gait quality are highly associated with a person's cognitive level, a system using PPG, accelerometer, and gyroscope sensors can highly support a clinical diagnosis of MCI.

Furthermore, we used machine learning algorithms for a classification of MCI from CH. A novel feature selection algorithm is proposed for the PPG and gait signal derived features. Selecting an optimal subset of relevant features helps to avoid overfitting by a high dimensionality and to improve the learning performance [13]. There are three methods for supervised feature selection: filter, wrapper, and embedded method. The filter method prunes low-scoring features, while the wrapper method searches for an optimal feature subset by evaluating each subset by the pre-determined classifier's predictive accuracy [14]. The filter method only looks at the intrinsic properties of data so it is easy to scale the very high dimensionality [13]. The other hand, the wrapper method includes interaction with the selected classifier but requires a high computational cost to search in the space of possible feature subsets [13]. The embedded method is a hybrid of both filter and wrapper method, which filters the features by the statistical criteria first, and then select candidate subsets with a given cardinality [14]. The embedded method is less computationally intensive than the wrapper method. Several feature selection techniques are suggested through other studies, such as information gain, Euclidean distance, or the weight vector of the support vector machine [13]. However, a single feature selection method did not provide a remarkable classification result with our dataset. We propose a feature selection algorithm (Figure 5) to overcome the existing methods. The objective of this study is the following:

- To investigate reliable makers from PPG and gait signals.
- To validate PPG and gait sensor-derived parameters for classification between MCI and the cognitively healthy.
- To provide an optimal feature selection algorithm to support a clinical diagnosis of MCI.

Our classification accuracy result was 82% with the PPG dataset and 86% with the gait dataset from 69 elderly participants (72.45 ± 10.55 years; 34 MCI and 35 CH), which is a better classification accuracy than only using the administered neuropsychological screening instrument. Our design will support medical professionals being able to get a quick and reliable marker of a person's cognitive status and provide appropriate interventions. Our design will also assist in reducing the diagnostic burden on healthcare professionals.

Fig. 1 System architecture for PPG and gait signal assessment.

II. Data Acquisition

We designed a simple sensor data assessment system using a wireless wearable device and a commercial smartwatch to assess PPG and gait signals.

A. *System Design*

The assessment system architecture is shown in Figure 1. Nonin Wireless Finger Pulse Oximeter (Nonin Onyx II 9560; Nonin Medical, Plymouth, MN) [15] takes several physiological data: heart rate in beats per minute (BPM), peripheral capillary oxygen saturation (SpO2), and PPG. This pulse oximeter is completely non-invasive, wireless, and transmits a PPG signal with a 75 Hz sample rate. Nonin pulse oximeters have been used in several clinical trials [2], [16] because of the portability and convenience. Also, Samsung Smartwatch (Samsung Gear Live; Samsung Electronics, Suwon, South Korea) [17] provides three-axial accelerometer and gyroscope signals. It was used to record gait parameters of each participant and operated by Android Wear. A customized Android app used Nonin software API and Android Wear API [18] to transmit sensor data from the pulse oximeter and smartwatch to an Android phone via Bluetooth. The Android app transferred collected data to a Health Insurance Portability and Accountability Act (HIPAA) secure cloud server via Wi-Fi. The PPG sensor data was stored every minute during the measurement and the smartwatch sensor data was stored as soon as each gait measurement is completed.

B. *Neuropsychological Assessment*

This paper includes sensor data and neuropsychological data from 69 elderly participants who were recruited for the longitudinal aging study from the Department of Neurology, Psychiatry, and Computer Science at University of California, Los Angeles. The longitudinal aging study is a UCLA IRB approved yearly longitudinal-term research to characterize healthy aging, MCI, and dementia in regard to multiple factors for those who are 50 years and above. A series of tests were administered to evaluate the subject's mental abilities including memory, attention, language, visuospatial skills, and mental flexibility during each neuropsychological visit. We have collected PPG signal data during three

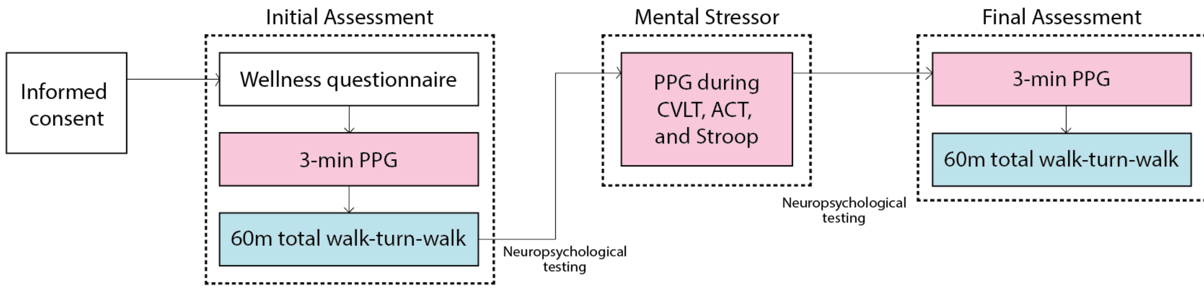

Fig. 2 Measurement protocol during the neuropsychological visit.

neuropsychological tests that measure memory and attention: California Verbal Learning Test (CVLT), Auditory Consonant Trigrams (ACT), and Stroop.

- CVLT measures immediate verbal memory span and level of interaction between verbal memory and conceptual ability. Each of the 16 items in each CVLT list belongs to one of four categories. For example, a CVLT list can be a shopping list containing four names of fruits, clothing, tools, and spices [19].
- ACT measures the level of memory and attention using a distractor task. The examiner gives three letters as a test item and a starting number to count immediately. The examinee should count the starting number backward until signaled to stop, then recall the three letters [19].
- Stroop measures the level of concentration effectiveness because calling out the ink color of the word written in the different colors takes a longer time than reading a word [19].

These tests are selected because they are the most complex cognitive tasks among other neuropsychological tests and may cause some physiological changes during the assessment. These three tests are considered as being mental stressors for PPG assessment.

C. Measurement Protocol

When a participant comes for the neuropsychological study visit, the informed consent form is given. The examiner launches the pre-installed Android app on the smartphone and registers the participant with the corresponding subject ID. The subject completes a wellness questionnaire by choosing one out of five scale numbers. The wellness questionnaire has five criteria: fatigue, mood, stress level, sleep quality of the day before, and sleep duration in hours and minutes. These criteria were modified from the wellness questionnaire for athletes, which was created by McLean et al. (2010) [20].

The participant wears a fingertip pulse oximeter on the index finger of the non-dominant hand. The participant is asked not to talk or move during the measurement to reduce possible artifacts on the signal. The Bluetooth connectivity and functionality of the devices is confirmed by the examiner, then the examiner presses the start button on the app to begin the measurement session. The initial PPG signal data is collected for three minutes (Figure 2). The pulse oximeter is removed after the measurement.

The Samsung smartwatch is placed on the participant's non-dominant wrist. The pre-installed Android Wear app is started by the examiner. The participant walks the 30 meters of hallway beginning at the interior end of the hallway and touching the window of the other end, finishing upon return to the initial location (60m total walk-turn-walk in Figure 2).

The neuropsychological testing begins after collecting the initial sensor data. The examiner asks the participant to again put the pulse oximeter on the index finger right before starting the CVLT, ACT, and Stroop tests. Because the duration of the three tests usually takes less than 30 minutes, the physiological measurement automatically finishes after 30 minutes, unless the pulse oximeter is dismounted manually before the 30 minutes. The participant is allowed to talk, but large hand gestures are restricted during the measurement period. The rest of the neuropsychological testing is continued after the PPG measurement with the mental stressors is completed (Figure 2).

As soon as finishing the entire neuropsychological testing, the PPG and gait measurement occurs again with the same procedure of the initial assessment (Figure 2). The transmitted signal data are stored in the database wirelessly through the smartphone.

D. Datasets

Table I shows the summary of demographics of participants. We formed 63 PPG datasets and 53 gait datasets from 69 participants, less than 69 because of system malfunction during a few of the measurements. Two datasets are independently used for feature extraction, feature selection, and classification. For the result of three neuropsychological tests, we used the percentile rank of Long Delay Free Recall Correct on the CVLT, the percentile rank of total correct across the first two trials on the ACT, and the percentile rank of T-score of the Stroop Color and Word Test. The results from the three neuropsychological tests, wellness questionnaire, and demographics are used in both datasets. The preliminary diagnosis of participants was a target that we want to predict through learning algorithms. PPG and gait datasets had 32 and 27 MCI groups respectively. The MCI group includes all different types of MCI, such as amnestic and non-amnestic, on a single domain or multiple domains, because of the small dataset.

Each feature vector was normalized from 0 to 1 because each feature vector has a different range. For example, the

TABLE I
DEMOGRAPHIC SUMMARY OF PARTICIPANTS.

Variable	Total (n = 69)	PPG Dataset (n = 63)	Gait Dataset (n = 53)
Age, years	72.45 ± 10.55	72.76 ± 10.55	71.58 ± 11.36
Male, %	49.28	49.21	47.18
Education, years	17.06 ± 2.25	17.00 ± 2.23	16.94 ± 2.33
Right-handed, %	86.96	85.71	88.68
MCI diagnosis, %	49.28	50.79	50.94

range of heart rate in beat per minute is from 36 to 130 BPM, but the range of average difference between peaks is from 686.6 to 1524.9 milliseconds. Higher scalar value of feature vector can influence the classification algorithms. Therefore, normalization is an important process to have the same length of the range in feature vectors.

III. FEATURE EXTRACTION

Raw sensor data carries both useful and noisy information. First, we cleaned the noisy information from the raw sensor data. Then, we extracted possible features using digital signal processing and a peak detection algorithm. We also added statistical features of each sensor signal using the sliding window approach. We used Python programming language with Scikit-learn [21] open-source software.

A. *Pre-processing and Peak Detection*

PPG signals using a pulse oximeter are easily corrupted by motion artifacts including respiration and voluntary finger movements [6]. Because we controlled the possible motion artifact during the data acquisition, we simply calibrated each PPG signals by removing the linear trend of time-series (Figure 3). Different patterns and shapes are observed by different PPG signals of a participant.

The Butterworth filter was used to remove noise from the measured gait signals. The Butterworth filter is designed to have a flat frequency response in the passband, while the

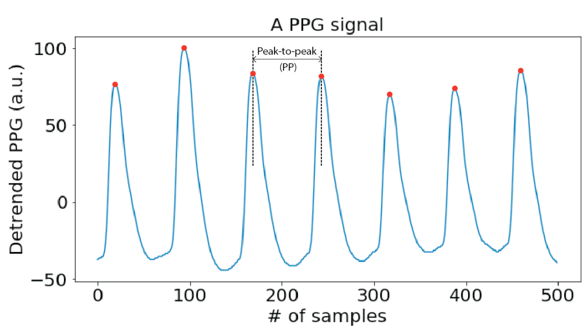

Fig. 3 A sample PPG signal with the peak detection.

frequency response amplitude of the filter rolls off toward zero in the stopband [22]. The Android API provides the ability to set a delay at which the sensor sample is received [23]. Our assessment system collected accelerometer and gyroscope sensor values simultaneously through an Android smartwatch. It recorded one timestamp for the two sensor values and caused irregular delays from 5 milliseconds to 2,145 milliseconds. We averaged different sampling delay (d) and obtained the sampling rate frequency (f_s) using (1) [23].

$$f_s = \frac{1}{N}\sum_{i=0}^{N}\frac{1}{d_i \times 10^{-3}} \quad (1)$$

The N is the number of samples per measured signal. One tenth and one fourth of the sampling rate frequencies were used to set the cut-off of low pass and high pass of the filter respectively. As shown in Figure 4, the Butterworth filter cleans noise and restores missing values due to irregular delays. Because of several long delays, a signal before using the filter mostly has flat peaks and some steps, however, a filtered signal is smoother. Consecutive walking waves are missing in the middle of the signal because it is a turning point during the walk-turn-walk activity. The Butter-worth filter clearly refines peaks of the signal and flattens noise during the turn.

Peaks are significant points to characterize the signal. Peaks on each signal are important for both HRV and gait analysis. The magnitude of the accelerometer signal was calculated by (2) and used to reduce the sensitivity of three-axial smartwatch rotation [24].

$$M_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \quad (2)$$

The *x*, *y*, and *z* represent each orthogonal axis and *i* corresponds to time. The magnitude of the accelerometer signal was mainly used to extract gait variables.

A peak detection algorithm by Duarte [25] was applied to detect peaks of each signal correctly. This peak detection algorithm works by setting up different parameter values: minimum peak height (MPH) and minimum peak distance (MPD) [25]. Both PPG and gait signals are periodic due to continuous heart activity and walking activity. We expected to have an equal distance between peaks for each signal. We found a peak detection with MPH as 0.4 and MPD as 2 samples performs well on the gait signals, while MPH as 10 and MPD as 30 samples performs correctly on the PPG signals.

B. *HRV and Gait Variables*

HRV variables were extracted from the time-domain and the frequency-domain of each PPG signal. We followed guidelines for the HRV from the Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology (1996) [26]. The task force provides HRV analysis using Electrocardiography (ECG) only, but we applied the same guidelines to the PPG since ECG and PPG are highly correlated to each other. The R peak is the highest amplitude on a typical ECG and is an important marker for HRV analysis.

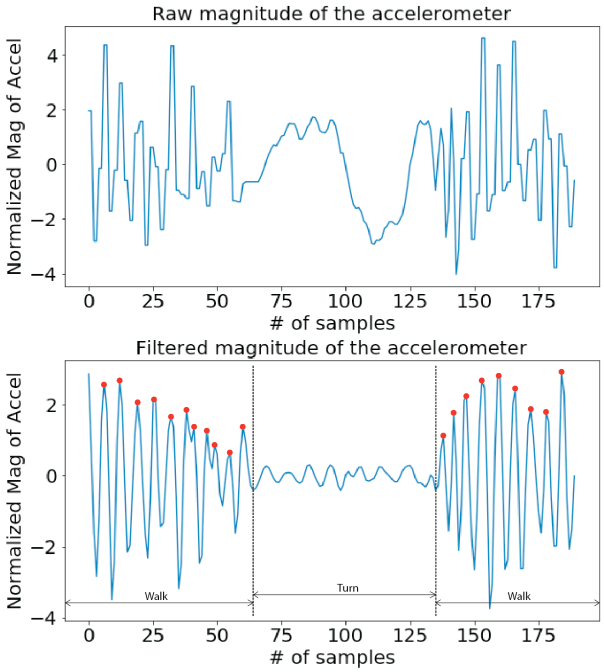

Fig. 4 Magnitude of the accelerometer signal: (top) a noisy raw signal and (bottom) a signal after Butterworth filter with the peak detection.

Several studies have proved that peaks of PPG coincide with the R peaks of ECG [5, 27].

The peak detection algorithm was applied on the time-domain of the PPG signal. Three different states of PPG signals were measured for each participant: initial, mental stress, and final (Figure 2). We calculated the mean (MeanPP) and standard deviation (SDPP) of the peak-to-peak (PP) intervals for each state. The square root of the mean squared differences of successive intervals (RMSSD) was also computed using (3), which is commonly used as a significant metric of HRV.

$$RMSSD = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(PP_{i+1}\text{-}PP_i)^2} \quad (3)$$

The difference between the RMSSD and MeanPP of each state is also extracted as a feature. These differences may represent ANS reaction of participant about the mental stressor. Because the pulse oximeter provides instant heart rate values, the average heart rate for each state was also calculated.

The power spectrum of the PPG signal was computed using a fast Fourier transform (FFT). Since our recordings for PPG were less than 30 minutes in duration and considered therefore as short-term, the power spectrum was divided into three frequency bands: very low frequency (VLF; ≤ 0.04 Hz); low frequency (LF; 0.04 to 0.15 Hz); and high frequency (HF; 0.15 to 0.4 Hz) [26]. The power of each frequency band became a distinct feature. The distribution of power of LF and HF may vary due to the modulations by the ANS. Thus, the ratio of LF to HF was calculated. Total power, which is the highest peak in the frequency domain and within the frequency range of less than 0.4 Hz, also extracted from PPG signals.

Gait variables also derived using peaks from the peak detection algorithm. Each participant walked the same 60m total distance, wearing a smartwatch on the wrist. The number of peaks on the magnitude of the accelerometer is correlated with a step count. (Figure 4). Walking duration is obtained from the time difference between the last step and the first step. The average gait speed is calculated by dividing the walking duration into the walking distance giving the result in meters per second. The average step time is the average time distance between accelerometer data magnitude peaks.

C. *Statistical Features*

Ten statistical characteristics of PPG and gait signals were also considered as possible predictors of classifying MCI against CH. The pulse oximeter provides a single axis of PPG signal, while the smartwatch provides three-axis accelerometer and gyroscope signals. We extracted statistical features from each axis of the sensor output independently for each measurement.

- Average: Average of each measurement.
- Standard Deviation: Standard Deviation of each measurement.
- Max: The maximum value of each measurement.
- Min: The minimum value of each measurement.
- Range: The difference between the maximum and the minimum of each measurement.
- Root mean square: The square root of the mean square of each measurement [23].
- Crest factor: The squared crest factor, which is peak-to-average power ratio (PAPR) of each measurement [23].
- Skewness: The symmetry of probability distribution of each measurement. The skewness of the perfectly symmetrical normal distribution is zero. A positively skewed result has a distribution with an elongated right tail, and a negatively skewed result has a left tail [28].
- Kurtosis: The peakedness and flatness of tail of probability distribution of each measurement [6]. A positive kurtosis represents lighter in the tail, but a negative kurtosis means thicker in the tail than the normal distribution [28].
- Variance: The variance of each measurement.

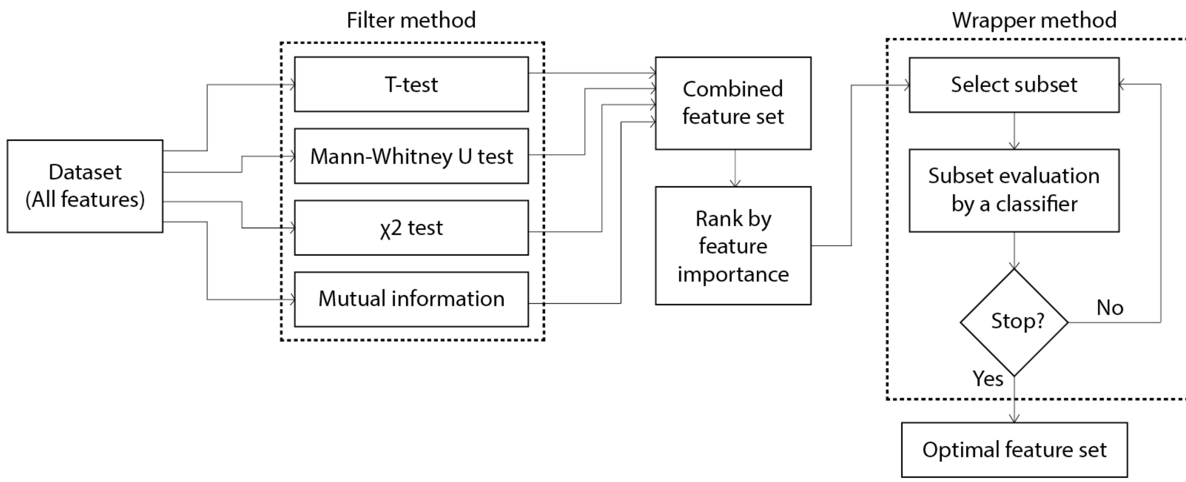

Fig. 5 The proposed feature selection algorithm for MCI classification using sensor-derived features.

D. Sliding Window

Based on the successful gait recognition as an individual biometric by Johnston and Weiss et al. [29], a sliding window approach on our datasets discovers unique patterns or characteristics at the partition of time-series signal values. The PPG time-series sensor data were divided into 20-second non-overlapping windows. Because the PPG data is sampled at 75Hz, a 10-second window has 750 time-series values. The smartwatch sampled both accelerometer and gyroscope sensor data with inconsistence delays, so 50 time-series values of each signal were used for each window. Only the last 200 time-series values of the gait signal were divided into partitions because of various noise sources at the beginning of measurement and when turning around at the end of the hallway. Each window with the time-series values was used to generate the same statistical feature extraction independently.

IV. Feature Selection

We combined the filter and wrapper method to have a better classification of MCI against CH (Figure 5). First, we pruned noisy and redundant features using four statistical criteria: T-test, Mann-Whitney U test, $\chi 2$ test, and mutual information. The T-test helps to analyze the difference between the target group means. The Mann-Whitney U test is like the T-test but more reasonable to use when the distribution of the feature is not normal. The $\chi 2$ test measures the dependency between stochastic variables, while mutual information measures the dependency between two target variables [21]. Our null hypothesis is that extracted features from sensor data does not have any difference between MCI and CH groups. All these approaches are a type of statistical hypothesis testing to prove two numerical data samples significantly differ from one another by deviating the null hypothesis. The subset of each filter was slightly different from each other, and we combined all the subsets into one. In this way, we filtered the extracted features that significantly discriminate the MCI group from the CH group.

The selected features were ranked by feature importance: Gini importance from the ensemble of the decision-tree (Extra Trees) for PPG dataset and coefficient in the logistic regression for the gait dataset. The Gini importance is calculated by total reduction of the Gini impurity brought by that feature [21]. Ranking the features in the subset can reduce the computational cost of the wrapper method by a constant. The wrapper method was used to obtain the optimal subset of the filtered features. Our proposed feature selection algorithm searches the possible subsets starting from the first feature on the ranked list and adding the next feature to create the next subset. The algorithm will move to the second feature and add the next feature to create a new subset, and so on. Each subset was evaluated by the cross-validation (CV) of the selected classifier. We selected the decision-tree classifier for the PPG dataset and the logistic regression classifier for the gait dataset. The algorithm stopped searching for other subsets when it reaches an arbitrary constant. We chose a constant as a one-third of the length of the ranked list because only subsets which include the top feature importance achieves a better classification accuracy than when it is not including the top features.

V. Results and Discussion

The feature subsets from each step of the proposed feature selection algorithm were validated by three learning models: logistic regression, Random Forest, and Extra Trees. The Random Forest model fits several decision-tree models on various sub-samples of the dataset, while the Extra Trees model consists of randomized decision-trees [21]. The classification accuracy of the feature subsets was compared with a classification accuracy when only the CVLT score was trained for the model. The CVLT score was the most significant feature of the classification and highly correlated

TABLE II
CROSS-VALIDATION CLASSIFICATION ACCURACY AND THE STANDARD DEVIATION AFTER EACH FEATURE SELECTION OF THE PPG AND THE GAIT DATASET.

Learning model	PPG Dataset				Gait Dataset			
	All	CVLT	Filter method	Wrapper Method	All	CVLT	Filter method	Wrapper Method
Logistic Regression	0.56 (0.10)	0.74 (0.17)	0.76 (0.15)	0.78 (0.12)	0.57 (0.08)	0.72 (0.14)	0.79 (0.13)	**0.86 (0.08)**
Random Forest	0.50 (0.11)	0.74 (0.08)	0.64 (0.14)	**0.82 (0.07)**	0.51 (0.11)	0.83 (0.06)	0.79 (0.12)	0.70 (0.17)
Extra Trees	0.48 (0.12)	0.76 (0.14)	0.72 (0.18)	**0.82 (0.07)**	0.47 (0.08)	0.83 (0.06)	0.69 (0.15)	0.76 (0.10)

(p-value; 5.7×10^{-7}) with a diagnosis of MCI. The CVLT measures verbal memory and conceptual ability, and this ability may be the main difference between MCI and CH recognition. The 5-fold CV result of each feature selection for PPG and gait dataset is shown in Table II. The classification accuracy using all features is as low as randomly picking one between two targets (Table II). This explains that some of the extracted features from both signals are noisy and redundant. The classification accuracy is improved using feature subset after the filter method of the proposed feature selection algorithm but still not better than using CVLT score alone (Table II). The optimal feature subset after the wrapper method provides the highest classification accuracy, 82% with the PPG dataset and 86% with the gait dataset, which is even better than using CVLT score alone (Table II). These results strongly verify the importance of optimal feature selection and that MCI classification accuracy can be improved when it uses sensor-derived features with some neuropsychological test scores.

The top-ranked features during the feature selection process can be considered as important markers from the PPG and gait signals for the MCI classification. Maximum PPG value during the cognitive task and MeanPP of initial PPG are the significant predictors. These markers imply the amplitude of PPG signals with mental stress and the intervals of PPG peaks are associated with the MCI group. Also, kurtosis and variance of PPG on the specific windows with mental stress are important markers of the MCI group. The average and skewness of the y-axis of the accelerometer signals before mental stress are ranked high among the gait features. These markers explain that the walking characteristics through the y-axis of the accelerometer are different between the MCI and CH group.

The classification accuracy of the randomly split test set for each dataset was evaluated. The PPG test dataset had a 0.93 F1-score with the decision-tree model. The gait test dataset had a 0.55 F1-score with the logistic regression model, but a 0.72 F1-score with the decision-tree model. We observed that the subset generated from the proposed feature selection algorithm is dependent on a specific classifier. The optimal feature subset of the PPG dataset was selected by the Extra Trees classifier, and the classification accuracy increased for the tree-based learning models (Table II). The feature subset of the gait dataset was iterated by the logistic regression classifier, and the maximum classification accuracy is presented on the same model. The proposed feature selection algorithm makes the optimal classifier stronger but has a risk of overfitting small datasets. The classification results from the test set explain that the decision-tree performs well for the wrapper method, but the logistic regression may cause an overfitting problem. Finding the better classifier for the wrapper method on the gait signals will be conducted as a future work.

The proposed feature selection algorithm has reduced a computational intensity of the wrapper method through filtering the high dimensionality first and ranked the selected features by the importance. The feature subset search on the proposed feature selection algorithm needs $O(n^2)$ computational complexity, but the actual computational cost is reduced by dividing the constant value, which stops the subset search.

VI. CONCLUSIONS

In this pilot study, we proposed a novel application area of machine learning in the clinical diagnosis of MCI. We explored predictors from PPG, accelerometer, and gyroscope sensor data to predict MCI. We applied a simple peak detection and a sliding window approach to extract features from the sensor signals. We also demonstrated a new way of using the filter and wrapper methods to find optimal feature subset. We evaluated selected features by several classification algorithms. The classification accuracy using the optimal feature subset was higher than when only using a neuropsychological test score. Therefore, the sensor-derived predictions can support a diagnosis of MCI with neuropsychological tests.

PPG, accelerometer, and gyroscope sensors are mostly available on commercial smartwatches. Other required information, such as simple neuropsychological assessments and demographics, can be served and collected through a mobile app. The current best way to prepare progressive dementia is the early detection of cognitive impairment. This application has a potential to be a reliable detection of MCI using easy-access mobile devices. It will produce social and economic benefits including diagnosing with concrete analysis, providing optimized treatments, decreasing healthcare cost, and reducing the burden on families, caregivers, and doctors. It will produce a positive impact on the geriatric population.

ACKNOWLEDGMENT

We are indebted to all the elderly volunteers for their participation in the longitudinal aging study. We owe special thanks to Dr. Ellen Woo and her neuropsychological research team (Karina Gomez, Jennifer Thompson, Allison T. Diep, and Dr. Valencia Montgomery) who provided professional advice of the neuropsychological fields and on reliable data collection techniques.

REFERENCES

[1] "Dementia | Signs, Symptoms & Diagnosis," Alzheimer's Association. [Online]. Available: https://www.alz.org/what-is-dementia.asp. [Accessed: 05-Nov-2017].

[2] R. C. Petersen, R. Doody, A. Kurz, R. C. Mohs, J. C. Morris, P. V. Rabins, K. Ritchie, M. Rossor, L. Thal, and B. Winblad, "Current Concepts in Mild Cognitive Impairment," Archives of Neurology, vol. 58, no. 12, p. 1985, Jan. 2001.

[3] E. Teng, B. W. Becker, E. Woo, J. L. Cummings, and P. H. Lu, "Subtle Deficits in Instrumental Activities of Daily Living in Subtypes of Mild Cognitive Impairment," Dementia and Geriatric Cognitive Disorders, vol. 30, no. 3, pp. 189–197, 2010.

[4] A. Karakostas, G. Meditskos, T. G. Stavropoulos, I. Kompatsiaris, and M. Tsolaki, "A Sensor-Based Framework to Support Clinicians in Dementia Assessment: The Results of a Pilot Study," Ambient Intelligence - Software and Applications Advances in Intelligent Systems and Computing, pp. 213–221, 2015.

[5] N. Selvaraj, A. Jaryal, J. Santhosh, K. K. Deepak, and S. Anand, "Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography," Journal of Medical Engineering & Technology, vol. 32, no. 6, pp. 479–484, 2008.

[6] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "A Novel Approach for Motion Artifact Reduction in PPG Signals Based on AS-LMS Adaptive Filter," IEEE Transactions on Instrumentation and Measurement, vol. 61, no. 5, pp. 1445–1457, 2012.

[7] A. L. Hansen, B. H. Johnsen, and J. F. Thayer, "Vagal influence on working memory and attention," International Journal of Psychophysiology, vol. 48, no. 3, pp. 263–274, 2003.

[8] S. Duschek, M. Muckenthaler, N. Werner, and G. A. R. D. Paso, "Relationships between features of autonomic cardiovascular control and cognitive performance," Biological Psychology, vol. 81, no. 2, pp. 110–117, 2009.

[9] J. Taelman, S. Vandeput, A. Spaepen, and S. V. Huffel, "Influence of Mental Stress on Heart Rate and Heart Rate Variability," IFMBE Proceedings 4th European Conference of the International Federation for Medical and Biological Engineering, pp. 1366–1369, 2009.

[10] A. J. Shah, S. Su, E. Veledar, J. D. Bremner, F. C. Goldstein, R. Lampert, J. Goldberg, and V. Vaccarino, "Is Heart Rate Variability Related to Memory Performance in Middle-Aged Men?," Psychosomatic Medicine, vol. 73, no. 6, pp. 475–482, 2011.

[11] M. Dorfman, A. Mirelman, J. M. Hausdorff, and N. Giladi, "Gait Disorders in Patients with Cognitive Impairment or Dementia," Movement Disorders in Dementias, pp. 17–44, 2014.

[12] J. Verghese, R. B. Lipton, C. B. Hall, G. Kuslansky, M. J. Katz, and H. Buschke, "Abnormality of Gait as a Predictor of Non-Alzheimers Dementia," New England Journal of Medicine, vol. 347, no. 22, pp. 1761–1768, 2002.

[13] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," Bioinformatics, vol. 23, no. 19, pp. 2507–2517, 2007.

[14] S. Wang, J. Tang, and H. Liu, "Feature Selection," Encyclopedia of Machine Learning and Data Mining, pp. 1–9, 2016.

[15] Nonin Medical Inc., "Nonin Onyx II 9560 ," What is a Pulse Oximeter – Home Pulse Oximeter – Nonin Medical. [Online]. Available: http://www.nonin.com/Onyx9560. [Accessed: 06-May-2018].

[16] D. D. Luxton, R. A. Mccann, N. E. Bush, M. C. Mishkind, and G. M. Reger, "mHealth for mental health: Integrating smartphone technology in behavioral healthcare.," Professional Psychology: Research and Practice, vol. 42, no. 6, pp. 505–512, 2011.

[17] "Samsung Gear Live," Samsung Electronics America. [Online]. Available: https://www.samsung.com/us/support/gear/gear-live-support/. [Accessed: 06-May-2018].

[18] "Wear OS overview | Android Developers," Android Developers. [Online]. Available: https://developer.android.com/training/building-wearables.html. [Accessed: 06-May-2018].

[19] M. D. Lezak, Neuropsychological Assessment, 3rd ed. New York: Oxford Univ. Press, 1995.

[20] B. D. Mclean, A. J. Coutts, V. Kelly, M. R. Mcguigan, and S. J. Cormack, "Neuromuscular, Endocrine, and Perceptual Fatigue Responses during Different Length Between-Match Microcycles in Professional Rugby League Players," International Journal of Sports Physiology and Performance, vol. 5, no. 3, pp. 367–383, 2010.

[21] "scikit-learn." [Online]. Available: http://scikit-learn.org/stable/. [Accessed: 06-May-2018].

[22] S. Hussin, G. Birasamy, and Z. Hamid, "Design of Butterworth Band-Pass Filter," Politeknik & Kolej Komuniti Journal of Engineering and Technology, vol. 1, 2016.

[23] "Deep-Spying: Spying using Smartwatch and Deep Learning." [Online]. Available: https://arxiv.org/pdf/1512.05616v1.pdf. [Accessed: 06-May-2018].

[24] J. K. Urbanek, J. Harezlak, N. W. Glynn, T. Harris, C. Crainiceanu, and V. Zipunnikov, "Stride variability measures derived from wrist- and hip-worn accelerometers," Gait & Posture, vol. 52, pp. 217–223, 2017.

[25] "Detection of peaks in data," Jupyter Notebook Viewer. [Online]. Available: http://nbviewer.jupyter.org/github/demotu/BMC/blob/master/notebooks/DetectPeaks.ipynb. [Accessed: 06-May-2018].

[26] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," Circulation, vol. 93, no. 5, pp. 1043-1065, Mar. 1996.

[27] V. Murthy, S. Ramamoorthy, N. Srinivasan, S. Rajagopal, and M. Rao, "Analysis of photoplethysmographic signals of cardiovascular patients," 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

[28] D. Rindskopf and M. Shiyko, "Measures of Dispersion, Skewness and Kurtosis," Egyptian Journal of Medical Human Genetics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780080448947013440. [Accessed: 06-May-2018].

[29] A. H. Johnston and G. M. Weiss, "Smartwatch-based biometric gait recognition," 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2015.