

Utility of MemTrax and Machine Learning Modeling in Classification of Mild Cognitive Impairment

Michael F. Bergeron^{a,*}, Sara Landset^b, Xianbo Zhou^{c,d}, Tao Ding^e, Taghi M. Khoshgoftaar^b, Feng Zhao^f, Bo Du^c, Xinjie Chen^g, Xuan Wang^c, Lianmei Zhong^g, Xiaolei Liu^{g,*} and J. Wesson Ashford^{h,i}

^a*SIVOTEC Analytics, Boca Raton, FL, USA*

^b*Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, USA*

^c*SJN Biomed LTD, Kunming, Yunnan, China*

^d*Center for Alzheimer's Research, Washington Institute of Clinical Research, Washington, DC, USA*

^e*Department of Rehabilitation Medicine, The First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China*

^f*Department of Neurology, Dehong People's Hospital, Dehong, Yunnan, China*

^g*Department of Neurology, the First Affiliated Hospital of Kunming Medical University, Wuhua District, Kunming, Yunnan Province, China*

^h*War-Related Illness and Injury Study Center, VA Palo Alto Health Care System, Palo Alto, CA, USA*

ⁱ*Department of Psychiatry & Behavioral Sciences, Stanford University School of Medicine, Palo Alto, CA, USA*

Handling Associate Editor: David Loewenstein

Accepted 20 July 2020

Abstract.

Background: The widespread incidence and prevalence of Alzheimer's disease and mild cognitive impairment (MCI) has prompted an urgent call for research to validate early detection cognitive screening and assessment.

Objective: Our primary research aim was to determine if selected MemTrax performance metrics and relevant demographics and health profile characteristics can be effectively utilized in predictive models developed with machine learning to classify cognitive health (normal versus MCI), as would be indicated by the Montreal Cognitive Assessment (MoCA).

Methods: We conducted a cross-sectional study on 259 neurology, memory clinic, and internal medicine adult patients recruited from two hospitals in China. Each patient was given the Chinese-language MoCA and self-administered the continuous recognition MemTrax online episodic memory test on the same day. Predictive classification models were built using machine learning with 10-fold cross validation, and model performance was measured using Area Under the Receiver Operating Characteristic Curve (AUC). Models were built using two MemTrax performance metrics (percent correct, response time), along with the eight common demographic and personal history features.

*Correspondence to: Michael F. Bergeron, PhD, FACS, SIVOTEC Analytics, Boca Raton Innovation Campus, 4800 T-Rex Avenue, Suite 315, Boca Raton, FL 33431, USA. E-mail: mbergeron@sivotecanalytics.com; Xiaolei Liu, MD, Department

of Neurology, First Affiliated Hospital of Kunming Medical University, 295 Xichang Road, Wuhua District, Kunming, Yunnan Province 650032, China. E-mail: ring@vip.163.com.

Results: Comparing the learners across selected combinations of MoCA scores and thresholds, Naïve Bayes was generally the top-performing learner with an overall classification performance of 0.9093. Further, among the top three learners, MemTrax-based classification performance overall was superior using just the top-ranked four features (0.9119) compared to using all 10 common features (0.8999).

Conclusion: MemTrax performance can be effectively utilized in a machine learning classification predictive model screening application for detecting early stage cognitive impairment.

Keywords: Aging, Alzheimer's disease, dementia, mass screening

INTRODUCTION

The recognized (albeit underdiagnosed) widespread incidence and prevalence and parallel escalating medical, social, and public health costs and burden of Alzheimer's disease (AD) and mild cognitive impairment (MCI) are increasingly straining for all stakeholders [1, 2]. This distressing and burgeoning scenario has prompted an urgent call for research to validate early detection cognitive screening and assessment instruments for regular practical utility in personal and clinical settings for older patients across diverse regions and populations [3]. These instruments must also provide for seamless translation of informative results into electronic health records. The benefits will be realized by informing patients and assisting physicians in recognizing significant changes earlier and thus enable more prompt and timely stratification, implementation, and tracking of appropriate individualized and more cost-effective treatment and patient care for those beginning to experience cognitive decline [3, 4].

The computerized MemTrax tool (<http://www.memtrax.com>) is a simple and brief continuous recognition assessment that can be self-administered online to measure challenging timed episodic memory performance where the user responds to repeated images and not to an initial presentation [5, 6]. Recent research and resulting practical implications are beginning to progressively and collectively demonstrate the clinical efficacy of MemTrax in early AD and MCI screening [5–7]. However, direct comparison of clinical utility to existing cognitive health assessment and conventional standards is warranted to inform professional perspective and corroborate MemTrax utility in early detection and diagnostic support. van der Hoek et al. [8] compared selected MemTrax performance metrics (reaction speed and percent correct) to cognitive status as determined by the Montreal Cognitive Assessment (MoCA). However, this study was limited to associating

these performance metrics with characterization of cognitive status (as determined by MoCA) and defining the relative ranges and cutoff values. Accordingly, to expand on this investigation and improve classification performance and efficacy, our primary research question was:

- Can an individual's selected MemTrax performance metrics and relevant demographics and health profile characteristics be effectively utilized in a predictive model developed with machine learning to classify cognitive health dichotomously (normal versus MCI), as would be indicated by one's MoCA score?

Secondary to this, we wanted to know:

- Including the same features, can a MemTrax performance-based machine learning model be effectively applied to a patient to predict severity (mild versus severe) within selected categories of cognitive impairment as would be determined by an independent clinical diagnosis?

The advent and evolving practical application of artificial intelligence and machine learning in screening/detection have already demonstrated distinct practical advantages, with predictive modeling effectively guiding clinicians in the challenging assessment of cognitive/brain health and patient management [7, 9–11]. In our study, we chose a similar approach in MCI classification modeling and cognitive impairment severity discrimination as confirmed by clinical diagnosis from three datasets representing selected volunteer inpatients and outpatients from two hospitals in China. Using machine learning predictive modeling, we identified the top-performing learners from the various dataset/learner combinations and ranked the features to guide us in defining the most clinically practical model applications.

Our hypotheses were that a validated MemTrax-based model can be utilized to classify cognitive health dichotomously (normal or MCI) based on the

MoCA aggregate score threshold criterion, and that a similar MemTrax predictive model can be effectively employed in discriminating severity in selected categories of clinically diagnosed cognitive impairment. Demonstrating the anticipated outcomes would be instrumental in supporting the efficacy of MemTrax as an early detection screen for cognitive decline and cognitive impairment classification. Favorable comparison to an industry purported standard complemented by far greater ease and quickness of utility would be influential in helping clinicians adopt this simple, reliable, and accessible tool as an initial screen in detecting early (including prodromal) stage cognitive deficits. Such an approach and utility could thus prompt more timely and better stratified patient care and intervention. These forward-thinking insights and improved metrics and models could also be helpful in mitigating or stopping dementia progression, including AD and AD-related dementias (ADRD).

MATERIALS AND METHODS

Study population

Between January 2018 and August 2019, cross-sectional research was completed on patients recruited from two hospitals in China. The administration of MemTrax [5] to individuals aged 21 years and over and the collection and analysis of those data were reviewed and approved by and administered in accord with the ethical standards of the Human Subject Protection Committee of Stanford University. MemTrax and all other testing for this overall study were performed according to the Helsinki declaration of 1975 and approved by the Institutional Review Board of the First Affiliated Hospital of Kunming Medical University in Kunming, Yunnan, China. Each user was provided an informed consent form to read/review and then voluntarily agree to participate.

Participants were recruited from the pool of outpatients in the neurology clinic at the Yanhua Hospital (YH sub-dataset) and the memory clinic at the First Affiliated Hospital of Kunming Medical University (XL sub-dataset) in Beijing, China. Participants were also recruited from neurology (XL sub-dataset) and internal medicine (KM sub-dataset) inpatients at the First Affiliated Hospital of Kunming Medical University. Inclusion criteria included 1) men and women at least 21 years old, 2) ability to speak Chinese (Mandarin), and 3) ability to understand verbal and written

directions. Exclusion criteria were vision and motor impairments preventing participants from completing the MemTrax test, as well the inability to understand the specific test instructions.

Chinese version of MemTrax

The online MemTrax test platform was translated into Chinese (URL: <https://www.memtrax.com.cn>) and further adapted to be utilized through WeChat (Shenzhen Tencent Computer Systems Co. LTD., Shenzhen, Guangdong, China) for self-administration. Data were stored on a cloud server (Ali Cloud) located in China and licensed from Alibaba (Alibaba Technology Co. Ltd., Hangzhou, Zhejiang, China) by SJN Biomed LTD (Kunming, Yunnan, China). Specific details on MemTrax and test validity criteria used here have been described previously [6]. The test was provided at no charge to the patients.

Study procedures

For the inpatients and outpatients, a general paper questionnaire for collecting demographic and personal information such as age, sex, years of education, occupation, living alone or with family, and medical history was administered by a member of the study team. Following completion of the questionnaire, the MoCA [12] and MemTrax tests were administered (MoCA first) with no more than 20 minutes between tests. MemTrax percent correct (MTx-%C), mean response time (MTx-RT), and date and time of the testing were recorded on paper by a member of the study team for each participant tested. The completed questionnaire and the results of the MoCA were uploaded into an Excel spreadsheet by the researcher who administered the tests and verified by a colleague before the Excel files were saved for analyses.

MemTrax test

The MemTrax online test included 50 images (25 unique and 25 repeats; 5 sets of 5 images of common scenes or objects) shown in a specific pseudo-random order. The participant would (per instructions) touch the Start button on the screen to commence the test and begin viewing the image series and again touch the image on the screen as quickly as possible whenever a repeated picture appeared. Each image appeared for 3 s or until the image on the screen was

touched, which prompted immediate presentation of the next picture. Using the internal clock of the local device, MTx-RT for each image was determined by the elapsed time from presentation of the image to when the screen was touched by the participant in response to indicating recognition of the image as one that had been already shown during the test. MTx-RT was recorded for every image, with a full 3 s recorded indicating no response. MTx-%C was calculated to indicate the percentage of repeat and initial images to which the user responded correctly (true positive + true negative divided by 50). Additional details of the MemTrax administration and implementation, data reduction, invalid or “no response” data, and primary data analyses are described elsewhere [6].

The MemTrax test was explained in detail and a practice test (with unique images other than those used in the test for recording results) was provided to the participants in the hospital setting. Participants in the YH and KM sub-datasets took the MemTrax test on a smartphone that was loaded with the application on WeChat; whereas a limited number of the XL sub-dataset patients used an iPad and the rest used a smartphone. All participants took the MemTrax test with a study investigator unobtrusively observing.

Montreal cognitive assessment

The Beijing version of the Chinese MoCA (MoCA-BC) [13] was administered and scored by trained researchers according to the official test instructions. Suitably, the MoCA-BC has been shown to be a reliable test for cognitive screening across all education levels in Chinese elderly adults [14]. Each test took about 10 to 30 minutes to administer based on the respective participant’s cognitive abilities.

MoCA classification modeling

There was a total of 29 usable features, including two MemTrax test performance metrics and 27 features related to demographic and health information for each participant. Each patient’s MoCA aggregate test score was used as the cognitive screening “benchmark” to train our predictive models. Accordingly, because MoCA was used to create the class label, we could not use the aggregate score (or any of the MoCA subset scores) as an independent feature. We performed preliminary experiments in which we modeled (classifying cognitive health defined by MoCA) the original three hospital/clinic(s) sub-datasets individually and then combined using all

features. However, all the same data elements were not collected in each of the four clinics representing the three sub-datasets; thus, many of our features in the combined dataset (when using all features) had a high incidence of missing values. We then built models with the combined dataset using only common features which resulted in improved classification performance. This was likely explained by a combination of having more instances to work with by combining the three patient sub-datasets and no features with an undue prevalence of missing values (only one feature in the combined dataset, work type, had any missing values, affecting only three patient instances), because only common features recorded at all three sites were included. Notably, we did not have a specific rejection criterion for each feature that was ultimately not included in the combined dataset. However, in our preliminary combined dataset modeling, we first used all features from each of the three separate patient sub-datasets. This widely resulted in model performance that was measurably lower than the initial preliminary modeling on each individual sub-dataset. Moreover, whereas the classification performance of the models built using all the features was encouraging, across all learners and classification schemes, performance improved for twice as many models when using only common features. In fact, among what ended up being our top learners, all but one model improved upon eliminating non-common features.

The final aggregate dataset (YH, XL, and KM combined) included 259 instances, each representing a unique participant who took both the MemTrax and the MoCA tests. There were 10 shared independent features: MemTrax performance metrics: MTx-%C and mean MTx-RT; demographic and medical history information: age, sex, years of education, work type (blue collar/white collar), social support (whether the test taker lives alone or with family), and yes/no answers as to whether the user had a history of diabetes, hyperlipidemia, or traumatic brain injury. Two additional metrics, MoCA aggregate score and MoCA aggregate score adjusted for years of education [12], were used separately to develop dependent classification labels, thus creating two distinct modeling schemes to be applied to our combined dataset. For each version (adjusted and unadjusted) of the MoCA score, the data were again separately modeled for binary classification using two different criterion thresholds—the initially recommended one [12] and an alternate value used and promoted by others [8, 15]. In the alternate threshold classification scheme,

Table 1

Summary of modeling scheme variations used for MoCA classification (Normal Cognitive Health versus MCI)

Modeling Scheme	Normal Cognitive Health (Negative Class)	MCI (Positive Class)
Adjusted-23 Unfiltered/Filtered	101 (39.0%)	158 (61.0%)
Adjusted-26 Unfiltered/Filtered	49 (18.9%)	210 (81.1%)
Unadjusted-23 Unfiltered/Filtered	92 (35.5%)	167 (64.5%)
Unadjusted-26 Unfiltered/Filtered	42 (16.2%)	217 (83.8%)

Respective number and percent of total patients in each class are differentiated by adjustment of score for education (Adjusted or Unadjusted) and classification threshold (23 or 26), as applied to both feature sets (Unfiltered and Filtered).

a patient was considered to have normal cognitive health if s/he scored ≥ 23 on the MoCA test and having MCI if the score was 22 or lower; whereas, in the initial recommended classification format, the patient had to score a 26 or better on the MoCA to be labeled as having normal cognitive health.

Filtered data for MoCA classification modeling

We further examined MoCA classification using four commonly used feature ranking techniques: Chi-Squared, Gain Ratio, Information Gain, and Symmetrical Uncertainty. For interim perspective, we applied the rankers to the entire combined dataset using each of our four modeling schemes. All rankers agreed on the same top features, i.e., age, number of years of education, and both MemTrax performance metrics (MTx-%C, mean MTx-RT). We then rebuilt the models using each feature selection technique to train the models on only the top four features (see *Feature selection* below).

The resultant final eight variations of the MoCA score classification modeling schemes are presented in Table 1.

MemTrax-based clinical evaluation modeling

Of our three original sub-datasets (YH, XL, KM), only the XL sub-dataset patients were independently clinically diagnosed for cognitive impairment (i.e., their respective MoCA scores were not used in establishing a classification of normal versus impaired). Specifically, the XL patients were diagnosed with either Alzheimer's disease (AD) or vascular dementia (VaD). Within each of these primary diagnosis

Table 2

Summary of modeling scheme variations used for diagnosis severity classification (Mild versus Severe)

Modeling Scheme	Mild (Negative Class)	Severe (Positive Class)
MCI-AD versus AD	12 (17.4%)	57 (82.6%)
MCI-VaD versus VaD	38 (50.0%)	38 (50.0%)

Respective number and percent of total patients in each class are differentiated by primary diagnosis category (AD or VaD).

categories, there was a further designation for MCI. Diagnoses of MCI, dementia, vascular neurocognitive disorder, and neurocognitive disorder due to AD were based on specific and distinctive diagnostic criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders: DSM-5 [16]. Considering these refined diagnoses, two classification modeling schemes were separately applied to the XL sub-dataset to distinguish level of severity (degree of impairment) for each primary diagnosis category. Data utilized in each of these diagnostic modeling schemes (AD and VaD) included demographic and patient history information, as well as MemTrax performance (MTx-%C, mean MTx-RT). Each diagnosis was labeled mild if designated MCI; otherwise, it was considered severe. We initially considered including the MoCA score in the diagnosis models (mild versus severe); but we determined that would defeat the purpose of our secondary predictive modeling scheme. Here the learners would be trained using other patient characteristics readily available to the provider and performance metrics of the simpler MemTrax test (in lieu of the MoCA) against the reference "gold standard", the independent clinical diagnosis. There were 69 instances in the AD diagnosis dataset and 76 instances of VaD (Table 2). In both datasets, there were 12 independent features. In addition to the 10 features included in the MoCA score classification, patient history also included information on history of hypertension and stroke.

Statistics

Comparison of participant characteristics and other numerical features between sub-datasets for each model classification strategy (to predict MoCA cognitive health and diagnosis severity) was performed using Python programming language (version 2.7.1) [17]. The model performance differences were initially determined using a single- or two-factor (as appropriate) ANOVA with a 95% confidence interval and the Tukey honest significant difference (HSD)

test to compare the performance means. This examination of differences between model performances was performed using a combination of Python and R (version 3.5.1) [18]. We employed this (albeit, arguably less than optimal) approach only as a heuristic aid at this early stage for initial model performance comparisons in anticipating potential clinical application. We then utilized the Bayesian signed-rank test using a posterior distribution to determine the probability of model performance differences [19]. For these analyses, we used the interval $-0.01, 0.01$, signifying that if two groups had a performance difference of less than 0.01, they were considered the same (within the region of practical equivalence), or otherwise they were different (one better than the other). To perform the Bayesian comparison of classifiers and calculate these probabilities, we used the baycomp library (version 1.0.2) for Python 3.6.4.

Predictive modeling

We built predictive models using the ten total variations of our modeling schemes to predict (classify) the outcome of each patient's MoCA test or severity of the clinical diagnosis. All learners were applied and the models were built using the open source software platform Weka [20]. For our preliminary analysis, we employed 10 commonly used learning algorithms: 5-Nearest Neighbors, two versions of C4.5 decision tree, Logistic Regression, Multilayer Perceptron, Naïve Bayes, two versions of Random Forest, Radial Basis Function Network, and Support Vector Machine. Key attributes and contrasts of these algorithms have been described elsewhere [21] (see respective Appendix). These were chosen because they represent a variety of different types of learners and because we have demonstrated success using them in previous analyses on similar data. Hyperparameter settings were chosen from our previous research indicating them to be robust on a variety of different data [22]. Based on the results of our preliminary analysis using the same combined dataset with common features that were used subsequently in the full analysis, we identified three learners which provided consistently strong performance across all classifications: Logistic Regression, Naïve Bayes, and Support Vector Machine.

Cross-validation and model performance metric

For all predictive modeling (including the preliminary analyses), each model was built using 10-fold

cross validation, and model performance was measured using Area Under the Receiver Operating Characteristic Curve (AUC). Cross-validation began with randomly dividing each of the 10 modeling scheme datasets into 10 equal segments (folds), using nine of these respective segments to train the model and the remaining segment for testing. This procedure was repeated 10 times, using a different segment as the test set in each iteration. The results were then combined to calculate the final model's result/performance. For each learner/dataset combination, this entire process was repeated 10 times with the data being split differently each time. This last step reduced bias, ensured replicability, and helped in determining the overall model performance. In total (for MoCA score and diagnosis severity classification schemes combined), 6,600 models were built. This included 1,800 unfiltered models (6 modeling schemes applied to the dataset \times 3 learners \times 10 runs \times 10 folds = 1,800 models) and 4,800 filtered models (4 modeling schemes applied to the dataset \times 3 learners \times 4 feature selection techniques \times 10 runs \times 10 folds = 4,800 models).

Feature selection

For the filtered models, feature selection (using the four feature ranking methods) was performed within the cross-validation. For each of the 10 folds, as a different 10% of the dataset was the test data, only the top four selected features for each training dataset (i.e., the other nine folds, or the remaining 90% of the entire dataset) were used to build the models. We were unable to confirm which four features were used in each model, as that information is not stored or made available within the modeling platform we utilized (Weka). However, given the consistency in our initial selection of top features when the rankers were applied to the entire combined dataset and the subsequent similarity in modeling performances, these same features (age, years of education, MTx-%C, and mean MTx-RT) are likely the most prevalent top four used concomitant with the feature selection within the cross-validation process.

RESULTS

Participant numerical characteristics (including MoCA scores and MemTrax performance metrics) of the respective datasets for each model classification strategy to predict MoCA-indicated cognitive health

Table 3
Participant characteristics, MoCA scores, and MemTrax performance for each model classification strategy

Classification Strategy	Age	Education	MoCA Adjusted	MoCA Unadjusted	MTx-%C	MTx-RT
MoCA Category	61.9 y (13.1)	9.6 y (4.6)	19.2 (6.5)	18.4 (6.7)	74.8% (15.0)	1.4 s (0.3)
Diagnosis Severity	65.6 y (12.1)	8.6 y (4.4)	16.7 (6.2)	15.8 (6.3)	68.3% (13.8)	1.5 s (0.3)

Values shown (mean, SD) differentiated by modeling classification strategies are representative of the combined dataset used to predict MoCA-indicated cognitive health (MCI versus normal) and the XL sub-dataset only used to predict diagnosis severity (mild versus severe).

Table 4

Dichotomous MoCA score classification performance (AUC; 0.0–1.0) results for each of the three top-performing learners for all respective modeling schemes

Feature Set Used	MoCA Score	Cutoff Threshold	Logistic Regression	Naïve Bayes	Support Vector Machine
Unfiltered (10 features)	Adjusted	23	0.8862	0.8913	0.8695
		26	0.8971	0.9221	0.9161
	Unadjusted	23	0.9103	0.9085	0.8995
		26	0.8834	0.9153	0.8994
Filtered (4 features)	Adjusted	23	0.8929	0.8954	0.8948
		26	0.9188	0.9247	0.9201
	Unadjusted	23	0.9135	0.9134	0.9122
		26	0.9159	0.9236	0.9177

Utilizing variations of feature set, MoCA score, and MoCA score cutoff threshold, the highest performance for each modeling scheme is shown in **bold** (not necessarily statistically different than all others not in **bold** for the respective model).

(normal versus MCI) and diagnosis severity (mild versus severe) are shown in Table 3.

For each combination of MoCA score (adjusted/unadjusted) and threshold (26/23), there was a statistical difference ($p=0.000$) in each pairwise comparison (normal cognitive health versus MCI) for age, education, and MemTrax performance (MTx-%C and MTx-RT). Each patient sub-dataset in the respective MCI class for each combination was on average about 9 to 15 years older, reported about five fewer years of education, and had less favorable MemTrax performance for both metrics.

Predictive modeling performance results for the MoCA score classifications using the top three learners, Logistic Regression, Naïve Bayes, and Support Vector Machine, are shown in Table 4. These three were chosen based on the most consistently high absolute learner performance across all the various models applied to the datasets for all the modeling schemes. For the unfiltered dataset and modeling, each of the data values in Table 4 indicates the model performance based on the AUC respective mean derived from the 100 models (10 runs \times 10 folds) built for each learner/modeling scheme combination, with the respective highest performing learner indicated in bold. Whereas for the filtered dataset modeling, the results reported in Table 4 reflect the overall average model performances from 400 models for each learner using each of the feature ranking methods (4 feature ranking methods \times 10 runs \times 10 folds).

Comparing the learners across all combinations of MoCA score versions and thresholds (adjusted/unadjusted and 23/26, respectively) in the combined unfiltered dataset (i.e., using the 10 common features), Naïve Bayes was generally the top-performing learner with an overall classification performance of 0.9093. Considering the top three learners, the Bayesian-correlated signed-rank tests indicated that the probability (Pr) of Naïve Bayes outperforming Logistic Regression was 99.9%. Moreover, between Naïve Bayes and Support Vector Machine, a 21.0% probability of practical equivalence in learner performance (thus, a 79.0% probability of Naïve Bayes outperforming Support Vector Machine), coupled with the 0.0% probability of Support Vector Machine performing better, measurably reinforces the performance advantage for Naïve Bayes. Further comparison of MoCA score version across all learners/thresholds suggested a slight performance advantage using unadjusted MoCA scores versus adjusted (0.9027 versus 0.8971, respectively; Pr (unadjusted > adjusted) = 0.988). Similarly, a comparison of cutoff threshold across all learners and MoCA score versions indicated a small classification performance advantage using 26 as the classification threshold versus 23 (0.9056 versus 0.8942, respectively; Pr (26 > 23) = 0.999). Lastly, examining the classification performance for the models utilizing only the filtered results (i.e., top-ranked four features only), Naïve Bayes (0.9143) was

numerically the top-performing learner across all MoCA score versions/thresholds. However, across all feature ranking techniques combined, all the top-performing learners performed similarly. Bayesian signed-rank tests showed 100% probability of practical equivalence between each pair of filtered learners. As with the unfiltered data (using all 10 common features), there was again a performance advantage for the unadjusted version of the MoCA score ($Pr(\text{unadjusted} > \text{adjusted}) = 1.000$), as well as a similarly distinct advantage for the classification threshold of 26 ($Pr(26 > 23) = 1.000$). Notably, the average performance of each of the top three learners across all MoCA score versions/thresholds using only the top-ranked four features exceeded the average performance of any learner on the unfiltered data. Not surprisingly, classification performance of the filtered models (using the top-ranked four features) overall was superior (0.9119) to the unfiltered models (0.8999), regardless of the feature ranking method models that were compared to those respective models using all 10 common features. For each feature selection method, there was 100% probability of a performance advantage over the unfiltered models.

With the patients considered for AD diagnosis severity classification, between-group (MCI-AD versus AD) differences for age ($p = 0.004$), education ($p = 0.028$), MoCA score adjusted/unadjusted ($p = 0.000$), and MTx-%C ($p = 0.008$) were statistically significant; whereas for MTx-RT it was not ($p = 0.097$). With those patients considered for VaD diagnosis severity classification, between-group (MCI-VaD versus VaD) differences for MoCA score adjusted/unadjusted ($p = 0.007$) and MTx-%C ($p = 0.026$) and MTx-RT ($p = 0.001$) were statistically significant; whereas for age ($p = 0.511$) and education ($p = 0.157$) there were no significant between-group differences.

Predictive modeling performance results for the diagnosis severity classifications using the three previously selected learners, Logistic Regression, Naïve Bayes, and Support Vector Machine, are shown in Table 5. Whereas additional examined learners demonstrated slightly stronger performances individually with one of the two clinical diagnosis categories, the three learners we had identified as the most favorable in our previous modeling offered the most consistent performance with both new modeling schemes. Comparing the learners across each of the primary diagnosis categories (AD and VaD), there was no consistent classification performance difference between learners for MCI-VaD versus

Table 5

Dichotomous clinical diagnosis severity classification performance (AUC; 0.0–1.0) results for each of the three top-performing learners for both respective modeling schemes

Modeling Scheme	Logistic Regression	Naïve Bayes	Support Vector Machine
MCI-AD versus AD	0.7465	0.7810	0.7443
MCI-VaD versus VaD	0.8033	0.8044	0.8338

The highest performance for each modeling scheme is shown in **bold** (not necessarily statistically different than others not in **bold**).

VaD, although Support Vector Machine generally performed more prominently. Similarly, there were no significant differences between learners for the MCI-AD versus AD classification, although Naïve Bayes (NB) had a slight performance advantage over Logistic Regression (LR) and just a negligible plurality over Support Vector Machine, with probabilities of 61.4% and 41.7% respectively. Across both datasets, there was an overall performance advantage for Support Vector Machine (SVM), with $Pr(\text{SVM} > \text{LR}) = 0.819$ and $Pr(\text{SVM} > \text{NB}) = 0.934$. Our overall classification performance across all learners in predicting severity of diagnosis in the XL sub-dataset was better in the VaD diagnosis category versus AD ($Pr(\text{VAD} > \text{AD}) = 0.998$).

DISCUSSION

Early detection of changes in cognitive health has important practical utility in personal health management and public health alike. Indeed, it is also very much a high priority in clinical settings for patients worldwide. The shared goal is to alert patients, caregivers, and providers and prompt earlier appropriate and cost-effective treatment and longitudinal care for those beginning to experience cognitive decline [1, 3, 4]. Merging our three hospital/clinic(s) data subsets, we identified three distinctively preferable learners (with one notable standout – Naïve Bayes) to build predictive models utilizing MemTrax performance metrics that could reliably classify cognitive health status dichotomously (normal cognitive health or MCI) as would be indicated by a MoCA aggregate score. Notably, overall classification performance for all three learners improved when our models utilized only the top-ranked four features which principally encompassed these MemTrax performance metrics. Moreover, we revealed the substantiated potential for utilizing the same learners and MemTrax performance metrics in a diagnostic support classification modeling scheme to distinguish severity

of two categories of dementia diagnosis: AD and VaD.

Memory testing is central to early detection of AD [23, 24]. Thus, it is opportune that MemTrax is an acceptable, engaging, and easy-to-implement online screening test for episodic memory in the general population [6]. Recognition accuracy and response times from this continuous performance task are particularly revealing in identifying early and evolving deterioration and consequent deficits in the neuroplastic processes related to learning, memory, and cognition. That is, the models here that are based largely on MemTrax performance metrics are sensitive to and are more likely to readily and with minimal cost reveal biological neuropathologic deficits during the transitional asymptomatic stage well prior to more substantial functional loss [25]. Ashford et al. closely examined the patterns and behaviors of recognition memory accuracy and response time in online users who participated on their own with MemTrax [6]. Respecting that these distributions are critical in optimal modeling and developing valid and effective patient care applications, defining clinically applicable recognition and response time profiles is essential in establishing a valuable foundational reference for clinical and research utility. The practical value of MemTrax in AD screening for early stage cognitive impairment and differential diagnostic support needs to then be more closely examined in the context of a clinical setting where comorbidities and cognitive, sensory, and motor capabilities affecting test performance can be considered. And to inform professional perspective and encourage practical clinical utility, it is first imperative to demonstrate comparison to an established cognitive health assessment test, even though the latter may be recognizably constrained by cumbersome testing logistics, education and language deterrents, and cultural influences [26]. In this regard, the favorable comparison of MemTrax in clinical efficacy to MoCA that is commonly purported as an industry standard is significant, especially when weighing the greater ease of utility and patient acceptance of MemTrax.

Previous exploration comparing MemTrax to MoCA highlights the rationale and preliminary evidence warranting our modeling investigation [8]. However, this prior comparison merely associated the two key MemTrax performance metrics we examined with cognitive status as determined by MoCA and defined respective ranges and cutoff values. We deepened the clinical utility assessment of MemTrax by exploring a predictive modeling-based approach that

would provide a more individualized consideration of other potentially relevant patient-specific parameters. In contrast to others, we did not find an advantage in model performance using an education correction (adjustment) to the MoCA score or in varying the cognitive health discriminating MoCA aggregate score threshold from the originally recommended 26 to 23 [12, 15]. In fact, the classification performance advantage favored using the unadjusted MoCA score and the higher threshold.

Key points in clinical practice

Machine learning is often best utilized and most effectual in predictive modeling when the data are extensive and multi-dimensional, that is, when there are numerous observations and a concomitant wide array of high-value (contributing) attributes. Yet, with these current data, the filtered models with only four select features performed better than those utilizing all 10 common features. This suggests that our aggregate hospital dataset did not have the most clinically appropriate (high value) features to optimally classify the patients in this way. Nevertheless, the feature ranking emphasis on the key MemTrax performance metrics—MTx-%C and MTx-RT—strongly supports building early stage cognitive deficit screening models around this test that is simple, easy to administer, low-cost, and aptly revealing regarding memory performance, at least right now as an initial screen for a binary classification of cognitive health status. Given the ever-mounting strain on providers and healthcare systems, patient screening processes and clinical applications should be suitably developed with an emphasis on collecting, tracking, and modeling those patient characteristics and test metrics that are most useful, advantageous, and proven effective in diagnostic and patient management support.

With the two key MemTrax metrics being central to MCI classification, our top-performing learner (Naïve Bayes) had a very high predictive performance in most models (AUC over 0.90) with a true-positive to false-positive ratio nearing or somewhat exceeding 4:1. A translational clinical application using this learner would thus capture (correctly classify) by far most of those with a cognitive deficit, while minimizing the cost associated with mistakenly classifying someone with normal cognitive health as having a cognitive deficit (false positive) or missing that classification in those who do have a cognitive deficit (false negative). Either one of these scenarios

of misclassification could impose an undue psychosocial burden to the patient and caregivers.

Whereas in the preliminary and full analyses we used all ten learners in each modeling scheme, we focused our results on the three classifiers showing the most consistent strong performance. This was also to highlight, based on these data, the learners that would anticipatedly perform dependably at a high level in a practical clinical application in determining cognitive status classification. Moreover, because this study was intended as an introductory investigation into the utility of machine learning on cognitive screening and these timely clinical challenges, we made the decision to keep the learning techniques simple and generalized, with minimal parameter tuning. We appreciate that this approach may have limited the potential for more narrowly defined patient-specific predictive capabilities. Likewise, whereas training the models using only the top features (filtered approach) informs us further regarding these data (specific to the shortcomings in data collected and highlighting the value in optimizing precious clinical time and resources), we recognize that it is premature to narrow the scope of the models and, therefore, all (and other features) should be considered with future research until we have a more definitive profile of priority features that would be applicable to the broad population. Thus, we also fully recognize that more inclusive and broadly representative data and optimization of these and other models would be necessary before integrating them into an effective clinical application, especially to accommodate comorbidities affecting cognitive performance that would need be considered in further clinical evaluation.

Utility of MemTrax was further edified by the modeling of disease severity based on separate clinical diagnosis. A better overall classification performance in predicting severity of VaD (compared to AD) was not surprising given the patient profile features in the models specific to vascular health and stroke risk, i.e., hypertension, hyperlipidemia, diabetes, and (of course) stroke history. Though it would have been more desirable and fitting to have the same clinical assessment conducted on matched patients with normal cognitive health to train the learners with these more inclusive data. This is especially warranted, as MemTrax is intended to be used primarily for early stage detection of a cognitive deficit and subsequent tracking of individual change. It is also plausible that the more desirable distribution of data in the VaD dataset contributed in part to the comparatively

better modeling performance. The VaD dataset was well-balanced between the two classes, whereas the AD dataset with far fewer MCI patients was not. Particularly in small datasets, even a few additional instances can make a measurable difference. Both perspectives are reasonable arguments underlying the differences in disease severity modeling performance. However, proportionately attributing improved performance to dataset numerical characteristics or the inherent features specific to the clinical presentation under consideration is premature. Nonetheless, this novel demonstrated utility of a MemTrax predictive classification model in the role of clinical diagnostic support provides valuable perspective and affirms pursuit for additional examination with patients across the continuum of MCI.

The implementation and demonstrated utility of MemTrax and these models in China, where the language and culture are drastically different from other regions of established utility (e.g., France, Netherlands, and United States) [7, 8, 27], further underscores the potential for widespread global acceptance and clinical value of a MemTrax-based platform. This is a demonstrable example in striving toward data harmonization and developing practical international norms and modeling resources for cognitive screening that are standardized and easily adapted for use worldwide.

Next steps in cognitive decline modeling and application

Cognitive dysfunction in AD indeed occurs on a continuum, not in discrete stages or steps [28, 29]. However, at this early phase, our goal was to first establish our ability to build a model incorporating MemTrax that can fundamentally distinguish “normal” from “not normal”. More inclusive empirical data (e.g., brain imaging, genetic features, biomarkers, comorbidities, and functional markers of complex activities requiring cognitive control) [30] across varied global regions, populations, and age groups to train and develop more sophisticated (including aptly weighted ensemble) machine learning models will support a greater degree of enhanced classification, that is, the capacity to categorize groups of patients with MCI into smaller and more definitive subsets along the cognitive decline continuum. Moreover, concomitant clinical diagnoses for individuals across regionally diverse patient populations are essential to effectively train these more inclusive and predictably robust models. This will

facilitate more specific stratified case management for those with similar backgrounds, influences, and more narrowly defined characteristic cognitive profiles and thus optimize clinical decision support and patient care.

Much of the relevant clinical research to-date has addressed patients with at least mild dementia; and, in practice, too often patient intervention is only attempted at advanced stages. However, because cognitive decline begins well before clinical criteria for dementia are met, an effectively applied MemTrax-based early screen could encourage appropriate education of individuals about the disease and its progressions and prompt earlier and more timely interventions. Thus, early detection could support suitable involvements ranging from exercise, diet, emotional support, and improved socialization to pharmacological intervention and reinforce patient-related changes in behavior and perception that singly or in aggregate could mitigate or potentially stop dementia progression [31, 32]. Moreover, with effective early screening, individuals and their families may be prompted to consider clinical trials or get counseling and other social services support to help clarify expectations and intentions and manage daily tasks. Further validation and widespread practical utility in these ways could be instrumental in mitigating or stopping the progression of MCI, AD, and ADRD for many individuals.

Indeed, the low end of the patient age range in our study does not represent the population of traditional concern with AD. Nonetheless, the average age for each group utilized in the classification modeling schemes based on the MoCA score/threshold and diagnosis severity (Table 3) underscores a clear majority (over 80%) being at least 50 years old. This distribution is thus very appropriate for generalization, supporting the utility of these models in the population characterizing those typically affected by early onset and burgeoning neurocognitive illness due to AD and VaD. Also, recent evidence and perspective stress those recognized factors (e.g., hypertension, obesity, diabetes, and smoking) potentially contributing to higher early adult and midlife vascular risk scores and consequent subtle vascular brain injury that develops insidiously with evident effects even in young adults [33–35]. Accordingly, the most optimal initial screening opportunity for detecting early stage cognitive deficits and initiating effective prevention and intervention strategies in successfully addressing dementia will emerge from examining contributing factors and antecedent indicators across the age

spectrum, including early adulthood and potentially even childhood (noting the relevance of genetic factors such as apolipoprotein E from early gestation).

In practice, valid clinical diagnoses and costly procedures for advanced imaging, genetic profiling, and measuring promising biomarkers are not always readily available or even feasible for many providers. Thus, in many instances, initial overall cognitive health status classification may have to be derived from models using other simple metrics provided by the patient (e.g., self-reported memory problems, current medications, and routine activity limitations) and common demographic features [7]. Registries such as the University of California Brain Health Registry (<https://www.brainhealthregistry.org/>) [27] and others with an inherent greater breadth of self-reported symptoms, qualitative measures (e.g., sleep and every day cognition), medications, health status, and history, and more detailed demographics will be instrumental in developing and validating the practical application of these more primitive models in the clinic. Further, a test such as MemTrax, which has demonstrated utility in assessing memory function, may in fact provide a substantially better estimate of AD pathology than biological markers. Given that the core feature of AD pathology is disruption of neuroplasticity and an overwhelmingly complex loss of synapses, which is manifest as episodic memory dysfunction, a measure which assesses episodic memory may in fact provide a better estimate of AD pathological burden than biological markers in the living patient [36].

With all predictive models—whether complemented by complex and inclusive data from state-of-the-art technology and refined clinical insights across multiple domains or those limited to more basic and readily available information characteristic of existing patient profiles—the recognized advantage of artificial intelligence and machine learning is that the resultant models can synthesize and inductively “learn” from relevant new data and perspective provided by ongoing application utilization. Following practical technology transfer, as the models here (and to be developed) are applied and enriched with more cases and pertinent data (including patients with comorbidities that could present with ensuing cognitive decline), prediction performance and cognitive health classification will be more robust, resulting in more effective clinical decision support utility. This evolution will be more fully and practically realized with embedding MemTrax into custom (targeted to the available capabilities)

platforms that healthcare providers could utilize in real-time in the clinic.

Imperative to the validation and utility of the MemTrax model for diagnostic support and patient care are highly sought-after meaningful longitudinal data. By observing and recording the concomitant changes (if any) in clinical status across an adequate range of normal through early-stage MCI, the models for appropriate ongoing assessment and classification can be trained and modified as patients age and are treated. That is, repeated utility can assist with longitudinal tracking of mild cognitive changes, intervention effectiveness, and maintaining informed stratified care. This approach aligns more closely with clinical practice and patient and case management.

Limitations

We appreciate the challenge and value in collecting clean clinical data in a controlled clinic/hospital setting. Nonetheless, it would have strengthened our modeling if our datasets included more patients with common features. Moreover, specific to our diagnosis modeling, it would have been more desirable and fitting to have the same clinical assessment conducted on matched patients with normal cognitive health to train the learners. And as underscored by the higher classification performance using the filtered dataset (only the top-ranked four features), more general and cognitive health measures/indicators would likely have improved modeling performance with a greater number of common features across all patients.

Certain participants might have been concomitantly experiencing other illnesses that could have prompted transitory or chronic cognitive deficiencies. Other than the XL sub-dataset where the patients were diagnostically classified as having either AD or VaD, comorbidity data were not collected/reported in the YH patient pool, and the predominant reported comorbidity by far in the KM sub-dataset was diabetes. It is arguable, however, that including patients in our modeling schemes with comorbidities that could prompt or exacerbate a level of cognitive deficiency and a consequent lower MemTrax performance would be more representative of the real-world targeted patient population for this more generalized early cognitive screening and modeling approach. Moving forward, accurate diagnosis of comorbidities potentially affecting cognitive performance is broadly beneficial for optimizing the models and resultant patient care applications.

Lastly, the YH and KM sub-dataset patients used a smartphone to take the MemTrax test, whereas a limited number of the XL sub-dataset patients used an iPad and the rest used a smartphone. This could have introduced a minor device-related difference in MemTrax performance for the MoCA classification modeling. However, differences (if any) in MTx-RT, for example, between devices would likely be negligible, especially with each participant being given a “practice” test just before the recorded test performance. Nevertheless, utility of these two handheld devices potentially compromises direct comparison to and/or integration with other MemTrax results where users responded to repeat pictures by touching the spacebar on a computer keyboard.

Key points on MemTrax predictive modeling utility

- Our top-performing predictive models encompassing selected MemTrax performance metrics could reliably classify cognitive health status (normal cognitive health or MCI) as would be indicated by the widely recognized MoCA test.
- These results support integration of selected MemTrax performance metrics into a classification predictive model screening application for early stage cognitive impairment.
- Our classification modeling also revealed the potential for utilizing MemTrax performance in applications for distinguishing severity of dementia diagnosis.

These novel findings establish definitive evidence supporting the utility of machine learning in building enhanced robust MemTrax-based classification models for diagnostic support in effective clinical case management and patient care for individuals experiencing cognitive impairment.

ACKNOWLEDGMENTS

We recognize the work of J. Wesson Ashford, Curtis B. Ashford, and colleagues for developing and validating the online continuous recognition task and tool (MemTrax) utilized here and we are grateful to the numerous patients with dementia who contributed to the critical foundational research. We also thank Xianbo Zhou and his colleagues at SJN Biomed LTD, his colleagues and collaborators at the hospitals/clinics sites, especially Drs. M. Luo and M. Zhong, who helped with recruitment of participants,

scheduling tests, and collecting, recording, and front-end managing the data, and the volunteer participants who donated their valuable time and made the commitment to taking the tests and providing the valued data for us to evaluate in this study. This study was supported in part by the MD Scientific Research Program of Kunming Medical University (Grant no. 2017BS028 to X.L.) and the Research Program of Yunnan Science and Technology Department (Grant no. 2019FE001 (-222) to X.L.).

J. Wesson Ashford has filed a patent application for the use of the specific continuous recognition paradigm described in this paper for general testing of memory. He also owns the URL for www.memtrax.com.

MemTrax, LLC is a company owned by Curtis Ashford, and this company is managing the memory testing system described in this paper and the www.memtrax.com URL.

Authors' disclosures available online (<https://www.j-alz.com/manuscript-disclosures/19-1340r2>).

REFERENCES

- [1] Alzheimer's Association (2016) 2016 Alzheimer's disease facts and figures. *Alzheimers Dement* **12**, 459-509.
- [2] Gresenz CR, Mitchell JM, Marrone J, Federoff HJ (2019) Effect of early-stage Alzheimer's disease on household financial outcomes. *Health Econ* **29**, 18-29.
- [3] Foster NL, Bondi MW, Das R, Foss M, Hershey LA, Koh S, Logan R, Poole C, Shega JW, Sood A, Thothala N, Wicklund M, Yu M, Bennett A, Wang D (2019) Quality improvement in neurology: Mild cognitive impairment quality measurement set. *Neurology* **93**, 705-713.
- [4] Tong T, Thokala P, McMillan B, Ghosh R, Brazier J (2017) Cost effectiveness of using cognitive screening tests for detecting dementia and mild cognitive impairment in primary care. *Int J Geriatr Psychiatry* **32**, 1392-1400.
- [5] Ashford JW, Gere E, Bayley PJ (2011) Measuring memory in large group settings using a continuous recognition test. *J Alzheimers Dis* **27**, 885-895.
- [6] Ashford JW, Tarpin-Bernard F, Ashford CB, Ashford MT (2019) A computerized continuous-recognition task for measurement of episodic memory. *J Alzheimers Dis* **69**, 385-399.
- [7] Bergeron MF, Landset S, Tarpin-Bernard F, Ashford CB, Khoshgoftaar TM, Ashford JW (2019) Episodic-memory performance in machine learning modeling for predicting cognitive health status classification. *J Alzheimers Dis* **70**, 277-286.
- [8] van der Hoek MD, Nieuwenhuizen A, Keijer J, Ashford JW (2019) The MemTrax test compared to the montreal cognitive assessment estimation of mild cognitive impairment. *J Alzheimers Dis* **67**, 1045-1054.
- [9] Falcone M, Yadav N, Poellabauer C, Flynn P (2013) Using isolated vowel sounds for classification of mild traumatic brain injury. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, pp. 7577-7581.
- [10] Dabek F, Caban JJ (2015) Leveraging big data to model the likelihood of developing psychological conditions after a concussion. *Procedia Comput Sci* **53**, 265-273.
- [11] Climent MT, Pardo J, Munoz-Almaraz FJ, Guerrero MD, Moreno L (2018) Decision tree for early detection of cognitive impairment by community pharmacists. *Front Pharmacol* **9**, 1232.
- [12] Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H (2005) The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* **53**, 695-699.
- [13] Yu J, Li J, Huang X (2012) The Beijing version of the montreal cognitive assessment as a brief screening tool for mild cognitive impairment: A community-based study. *BMC Psychiatry* **12**, 156.
- [14] Chen KL, Xu Y, Chu AQ, Ding D, Liang XN, Nasreddine ZS, Dong Q, Hong Z, Zhao QH, Guo QH (2016) Validation of the Chinese version of Montreal cognitive assessment basic for screening mild cognitive impairment. *J Am Geriatr Soc* **64**, e285-e290.
- [15] Carson N, Leach L, Murphy KJ (2018) A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *Int J Geriatr Psychiatry* **33**, 379-388.
- [16] American Psychiatric Association (2013) *Task Force Diagnostic and statistical manual of mental disorders: DSM-5™*, American Psychiatric Publishing, Inc., Washington, DC.
- [17] Python. Python Software Foundation, <http://www.python.org>, Accessed November 15, 2019.
- [18] R Core Group, R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>, 2018, Accessed November 15, 2019.
- [19] Benavoli A, Corani G, Demšar J, Zaffalon M (2017) Time for a change: A tutorial for comparing multiple classifiers through Bayesian analysis. *J Mach Learn Res* **18**, 1-36.
- [20] Frank E, Hall MA, Witten IH (2016) The WEKA Workbench. In *Data Mining: Practical Machine Learning Tools and Techniques*, Frank E, Hall MA, Witten IH, Pal CJ, eds. Morgan Kaufmann https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf
- [21] Bergeron MF, Landset S, Maugans TA, Williams VB, Collins CL, Wasserman EB, Khoshgoftaar TM (2019) Machine learning in modeling high school sport concussion symptom resolve. *Med Sci Sports Exerc* **51**, 1362-1371.
- [22] Van Hulse J, Khoshgoftaar TM, Napolitano A (2007) Experimental perspectives on learning from imbalanced data. In *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, Oregon, USA, pp. 935-942.
- [23] Ashford JW, Kolm P, Colliver JA, Bekian C, Hsu LN (1989) Alzheimer patient evaluation and the mini-mental state: Item characteristic curve analysis. *J Gerontol* **44**, P139-146.
- [24] Ashford JW, Jarvik L (1985) Alzheimer's disease: Does neuron plasticity predispose to axonal neurofibrillary degeneration? *N Engl J Med* **313**, 388-389.
- [25] Jack CR Jr, Therneau TM, Weigand SD, Wiste HJ, Knopman DS, Vemuri P, Lowe VJ, Mielke MM, Roberts RO, Machulda MM, Graff-Radford J, Jones DT, Schwarz CG, Gunter JL, Senjem ML, Rocca WA, Petersen RC (2019) Prevalence of biologically vs clinically defined Alzheimer spectrum entities using the National Institute on Aging-Alzheimer's Association Research framework. *JAMA Neurol* **76**, 1174-1183.

- [26] Zhou X, Ashford JW (2019) Advances in screening instruments for Alzheimer's disease. *Aging Med* **2**, 88-93.
- [27] Weiner MW, Nosheny R, Camacho M, Truran-Sacrey D, Mackin RS, Flenniken D, Ulbricht A, Insel P, Finley S, Fockler J, Veitch D (2018) The Brain Health Registry: An internet-based platform for recruitment, assessment, and longitudinal monitoring of participants for neuroscience studies. *Alzheimers Dement* **14**, 1063-1076.
- [28] Ashford JW, Schmitt FA (2001) Modeling the time-course of Alzheimer dementia. *Curr Psychiatry Rep* **3**, 20-28.
- [29] Li X, Wang X, Su L, Hu X, Han Y (2019) Sino Longitudinal Study on Cognitive Decline (SILCODE): Protocol for a Chinese longitudinal observational study to develop risk prediction models of conversion to mild cognitive impairment in individuals with subjective cognitive decline. *BMJ Open* **9**, e028188.
- [30] Tarnanas I, Tsolaki A, Wiederhold M, Wiederhold B, Tsolaki M (2015) Five-year biomarker progression variability for Alzheimer's disease dementia prediction: Can a complex instrumental activities of daily living marker fill in the gaps? *Alzheimers Dement (Amst)* **1**, 521-532.
- [31] McGurran H, Glenn JM, Madero EN, Bott NT (2019) Prevention and treatment of Alzheimer's disease: Biological mechanisms of exercise. *J Alzheimers Dis* **69**, 311-338.
- [32] Mendiola-Precoma J, Berumen LC, Padilla K, Garcia-Alcocer G (2016) Therapies for prevention and treatment of Alzheimer's disease. *Biomed Res Int* **2016**, 2589276.
- [33] Lane CA, Barnes J, Nicholas JM, Sudre CH, Cash DM, Malone IB, Parker TD, Keshavan A, Buchanan SM, Keuss SE, James SN, Lu K, Murray-Smith H, Wong A, Gordon E, Coath W, Modat M, Thomas D, Richards M, Fox NC, Schott JM (2020) Associations between vascular risk across adulthood and brain pathology in late life: Evidence from a British birth cohort. *JAMA Neurol* **77**, 175-183.
- [34] Seshadri S (2020) Prevention of dementia-thinking beyond the age and amyloid boxes. *JAMA Neurol* **77**, 160-161.
- [35] Maillard P, Seshadri S, Beiser A, Himali JJ, Au R, Fletcher E, Carmichael O, Wolf PA, DeCarli C (2012) Effects of systolic blood pressure on white-matter integrity in young adults in the Framingham Heart Study: A cross-sectional study. *Lancet Neurol* **11**, 1039-1047.
- [36] Fink HA, Linskens EJ, Silverman PC, McCarten JR, Hemmy LS, Ouellette JM, Greer NL, Wilt TJ, Butler M (2020) Accuracy of biomarker testing for neuropathologically defined Alzheimer disease in older adults with dementia. *Ann Intern Med* **172**, 669-677.