Alzheimer's & Dementia
**Diagnosis, Assessment
& Disease Monitoring**

RESEARCH ARTICLE

Cognitive screening with functional assessment improves diagnostic accuracy and attenuates bias

David Andrés González[1,2] | **Mitzi M. Gonzales**[1,2] | **Kyle J. Jennette**[3] | **Jason R. Soble**[3,4] | **Bernard Fongang**[2,5]

[1] Department of Neurology, University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA

[2] Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA

[3] Department of Psychiatry, University of Illinois College of Medicine, Chicago, Illinois, USA

[4] Department of Neurology, University of Illinois College of Medicine, Chicago, Illinois, USA

[5] Department of Biochemistry and Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA

Correspondence
David A. González, Medical Arts & Research Center, Neurology 8th floor, 8300 Floyd Curl Dr., San Antonio, TX 78229, USA.
E-mail: GonzalezD15@uthscsa.edu

Abstract

Introduction: Cognitive screening measures often lack sensitivity and are hampered by inequities across ethnoracial groups. A multitrait multimethod (MTMM) classification may attenuate these shortcomings.

Methods: A sample of 7227 participants across the diagnostic spectrum were selected from the National Alzheimer's Coordinating Center cohort. Random forest ensemble methods were used to predict diagnosis across the sample and within Black American ($n = 1025$) and non-Hispanic White groups ($n = 5263$) based on: (1) a demographically corrected Montreal Cognitive Assessment (MoCA), (2) MoCA and Functional Assessment Questionnaire (FAQ), (3) MoCA and FAQ with demographic correction.

Results: The MTMM approach with demographic correction had the highest diagnostic accuracy for the cognitively unimpaired (area under curve [AUC] [95% confidence interval (CI)]: 0.906 [0.892, 0.920]) and mild cognitive impairment (AUC: 0.835 [0.810, 0.860]) groups and reduced racial disparities.

Discussion: With further validation, the MTMM approach combining cognitive screening and functional status assessment may serve to improve diagnostic accuracy and extend opportunities for early intervention with greater equity.

KEYWORDS
cognitive screening, diagnostic accuracy, Functional Activities Questionnaire (FAQ), Montreal Cognitive Assessment (MoCA), random forest

1 | INTRODUCTION

Currently, more than 5.8 million Americans age 65+ are diagnosed with Alzheimer's disease (AD) or another dementia, and new AD diagnoses are expected to double by 2050.[1] This represents a public health threat requiring urgent need for accessible and reliable cognitive screening (CS) tools to identify individuals with underlying neurodegenerative diseases earlier to refer for comprehensive evaluation, provide resources for long-term care planning, establish advanced directives,

improve access to clinical research trials, and avail treatment at earlier disease stages. Further, AD has a clear differential impact on ethnoracial minorities in the United States due to access to and accuracy of CS and assessment as well as course and severity of disease, access to care, and long-term medical/psychological outcomes.[2]

CS is mandated by the Patient Protection and Affordable Care Act of 2010[3 and a] covered preventive service by the US Centers for Medicare and Medicaid Services. Given dementia-associated costs were estimated at more than $300 billion in 2020 with possible annual costs

of more than one trillion by 2050,[1] CS can facilitate early detection and management. Although the Diagnostic and Statistical Manual of Mental Disorders Fifth edition[4] indicates that cognitive impairment (CI) should preferably be documented by standardized neuropsychological testing to support a dementia diagnosis, it is neither practical nor indicated for all individuals to undergo a full evaluation.[5] While CS alone is insufficient to diagnose a neurocognitive disorder, it can minimize health-care costs and unnecessary procedures by identifying those with possible CI who would benefit from referral to a specialist for further confirmatory evaluation.[6,7] CS also can be conducted by a wider array of health-care professionals, which increases access to effective screening by physicians most likely to treat older patients. Along with reducing delayed or misdetected dementia, other CS benefits are that it may facilitate improved care for older adults by identification of potentially reversible CI etiologies, provision of interventions to delay/prevent CI progression, and appropriate family/caregiver education.[6–8] CS also can provide objective information regarding participants' cognitive status in clinical research/trials in which cognition is not the primary outcome.[9]

Several CS tests are available, including the Montreal Cognitive Assessment (MoCA[10]), Mini-Mental State Examination (MMSE[11]), and Mini-Cog (MC[12]). However, each has varying sensitivity/specificity and susceptibility to reduced diagnostic accuracy based on arbitrary cutoff scores. Among the most common CS tests, the MoCA has superior accuracy for detecting mild CI (MCI) and dementia in individuals who otherwise score as cognitively unimpaired (CU) on the MMSE.[10,13] Multiple studies demonstrated the need for different MoCA cut-scores based on age, education/literacy, and ethnoracial background to decrease the likelihood of misclassifying normal individuals as impaired.[14,15] Rossetti et al.[16] established that 80% of community-dwelling Black Americans (BAs) fell below the standard cognitive impairment cutoff using a predominantly non-Hispanic White (NHW) comparison sample. In a large population sample,[14] optimal cut-scores differed by ethnoracial identity and years of education, with substantial improvements in sensitivity/specificity when using demographically corrected cutoffs. These differences are attributed to systemic and multifactorial disparities that have cascading impact on individuals over their lifespans, in addition to cultural biases in test construction, which can impact task performance in a clinical setting. Nonetheless, assessment based on any single test will be limited by the inherent measurement error associated with the tool.

1.1 | Combined approach

One approach for remediating the inherent error from a single measurement is a multitrait, multimethod (MTMM) approach (note: this usage is distinct from Campbell and Fiske's MTMM matrix used to evaluate measure validity), which overcomes limitations of any one measure to assess a complex construct/outcome more comprehensively. Regarding the CU–MCI–dementia syndrome spectrum, a performance-based screener (first method) of cognition (first trait) could be combined with a collateral rating (second method) of activi-

Research in Context

1. Systematic review: The authors reviewed traditional sources (i.e., PubMed) for literature on the diagnostic accuracy of cognitive screeners, finding documented variances related to age, education, and ethnoracial background. Multimodal approaches that combine cognitive screeners with daily functioning assessments may improve accuracy. However, prior studies have been limited by suboptimal measure sensitivity and statistical approaches, as well as lack of guidance for effective clinical translation.

2. Interpretation: Our results, using a multicenter database and ensemble algorithms, showed that integration of multiple traits (i.e., cognition, daily functioning) and methods (i.e., performance-based tasks, collateral ratings) improved cognitive staging classification accuracy compared to lone screening, without adding time or cost to the process. Observed classification disparities between non-Hispanic White and Black participants were attenuated, but not eliminated.

3. Future directions: The article provides a freely accessible online calculator that computes diagnostic probabilities to aid clinical translation. Further validation and extension to broader ethnoracial groups is necessary.

Highlights

- Random forest classification algorithms allow for flexible use of entire scale ranges and multiple scales to validate against a multistage taxonomy (i.e., normal, mild cognitive impairment [MCI], dementia); this is opposed to the static and artificial dichotomization of single scales and taxonomies used in classic receiver operating characteristic curve analyses.

- Functional ratings can be completed by collaterals while a cognitive screener is administered to patients to not increase screening time.

- Inputting both scores into the freely provided calculator allows for improved classification accuracy (i.e., normal, MCI, dementia) and reduces the ethnoracial bias often seen on cognitive screeners.

ties of daily living and functioning (ADLs; second trait). Such a MTMM approach is promising for improving accuracy while not appreciably impacting the cherished efficiency of CS, given a collateral can complete their rating while a professional simultaneously performs CS. A joint approach may also increase the content validity of evaluating the CU–MCI–dementia spectrum compared to CS alone, as cur-

rent criteria for disease syndrome staging considers both cognition and functioning.[17]

Such a MTMM was previously evaluated[18] using classification and regression trees (CART) with the MMSE, MC, and Functional Activities Questionnaire (FAQ; a collateral ADL rating) and found that demographically uncorrected, two-dimensional (2D) classifier models (i.e., cognitive scores are placed on an *x*-axis and functional scores placed on a *y*-axis; an *xy* combination's location is then used to classify that case) had strong overall CU–MCI–dementia accuracy. Although MCI accuracy was the lowest of the three groups, it was substantially higher than prior literature examining MCI accuracy with CS alone. The cutoffs derived from this study were subsequently extended to diverse specialty clinics within the United States[19] and multiple primary care sites outside the United States.[20] Tappen et al.[19] did not use criterion diagnoses to cross-validate accuracy of the cutoffs, but compared 2D classification rates across BAs, Latinxs, and NHWs. They found classification rates across the three groups differed significantly with the MC and MMSE, which was contrasted by the more equitable classifications of the MC-FAQ and MMSE-FAQ 2D-MTMM approach. They also found age and education level impacted 2D classification rates, suggesting correction for these demographics could further increase model accuracy. Kallumpuram et al.[20] used the MC-FAQ across 33 primary care sites, and found ease of implementation across sites. They were able to identify 358 individuals with cognitive impairment (18.1% of those screened) who were not previously identified by standard clinical judgment, and of those who received a specialty evaluation, 94.5% were accurately classified as cognitively impaired.

1.2 | The current study

A simple MTMM approach shows promise in that, compared to CS alone, it may improve overall accuracy of diagnosis without appreciably changing screening time, and reduce disparity in classification across ethnoracial groups. However, the above models are limited by use of a less sensitive CS, use of CART to derive a simple 2D cut-score, and the cross-validations not having criterion diagnoses. CART exploits sample variance in such a way as to overstate accuracy and reduce its reproducibility across samples. Advances in statistical techniques include newer ensemble algorithms, such as random forests (RF), that provide more robust and reproducible results compared to CART.[21] Therefore, the aim of the current study was to extend prior research with inclusions of the MoCA and FAQ using RF within a large, well-characterized multicenter sample.

2 | METHODS

2.1 | Participants

The National Alzheimer's Coordinating Center (NACC) is funded by the National Institute on Aging (NIA) and comprises several Alzheimer's Disease Research Centers (ADRCs) across the United States. The current sample includes data from 31 ADRCs, with data collected between March 2015 and May 2020. A sample of 7227 individuals who completed both the MoCA and FAQ as a part of their first ADRC Uniform Data Set visit were selected from the NACC June 2020 data freeze. Consent was not necessary for this deidentified data.

2.2 | Clinical assessments

All participants were administered the MoCA,[10] which assesses cognition in the domains of visuospatial, language, memory, attention, executive function, and orientation. Total scores range from 0 to 30 with lower scores indicating worse cognition. Participants also had the FAQ,[22] a 10-item measure of ADL functional status, completed by a collateral source. FAQ items are rated on a 4-point ordinal scale, and total scores range from 0 to 30 with higher scores indicating greater functional impairment. The FAQ has robust clinimetric properties and classification accuracy for detecting MCI/dementia.[23] Finally, participants had the Clinical Dementia Rating (CDR Dementia Staging Instrument[24,25]), a valid and reliable semi-structured interview of cognitive and functional status, to stage participants' level of impairment into one of five diagnostic categories (i.e., CU, questionable/MCI, mild dementia, moderate dementia, and severe dementia). The CDR is rated independently from the MoCA and FAQ and, for this study, dementia severities were collapsed consistent with the CU–MCI–dementia staging.

2.3 | Data analysis

See Table 1 for descriptives. Data were evaluated for statistical assumptions. Because all five demographic variables (i.e., age, years of education, sex, ethnoracial group, primary language) had statistically significant relationships with MoCA, FAQ, and CDR classifications, these demographics were included in models below.

We used the RF classification method to build predictors of cognitive status given the participants' MoCAs, FAQs, and demographics. RF is a nonparametric data mining technique that predicts numeric or nominal dependent variables and has undergone substantial developments to improve speed, flexibility, and transportability.[21,26–28] The whole procedure can be described as follows: (1) a collection of bootstrapped samples is computed from the original training set and for each sample, multiple decision trees are generated; (2) at each branch, a predictors-based node is used to determine the best partition of the dataset; (3) the underlying model is represented by the set of all trees and the final classification is simply the majority vote of individual classifications. The out-of-bag (OOB) estimate, which represents the overall error rate of the model, is computed as an aggregate of the predictions of the variables that are not part of the bootstrap samples. The RF model's convergence depends on several parameters, including the number of trees to construct (*ntree*) and the number of variables to select for each tree (*mtry*), which are automatically and optimally selected by the software.[28–30] Statistical analyses were conducted in R and JASP.[31,32] The area under the curve (AUC) is calculated for each

TABLE 1 Sample descriptives

	n (%)	FAQ M (SD)	MoCA M (SD)	Cognitively unimpaired	MCI	Dementia
Total	7227	4.05 (7.44)	22.94 (5.95)	3905	2362	960
Age	–	–	–	67.96 (10.36)	70.43 (9.04)	70.07 (10.15)
Education	–	–	–	16.4 (2.5)	16.1 (2.7)	15.6 (2.8)
Sex	–	–	–	–	–	–
Male	2797 (38.7)	4.95 (7.87)	22.54 (5.84)	1272 (32.57)	1069 (45.26)	456 (47.50)
Female	4430 (61.3)	3.49 (7.09)	23.20 (6.01)	2633 (67.43)	1293 (54.74)	504 (52.50)
Ethnoracial group	–	–	–	–	–	–
Non-Hispanic White	5308 (74.6)	4.51 (7.72)	23.18 (6.12)	2751 (71.69)	1770 (75.99)	787 (83.28)
Black American	1030 (14.5)	2.33 (6.11)	22.29 (5.22)	629 (16.39)	323 (13.86)	78 (8.25)
Hispanic	310 (4.4)	2.99 (6.59)	21.82 (5.85)	179 (4.66)	101 (4.33)	30 (3.17)
Asian	233 (3.3)	4.61 (7.78)	22.06 (6.02)	135 (3.51)	64 (2.74)	34 (3.59)
Other	230 (3.2)	2.14 (5.12)	23.24 (4.81)	143 (3.72)	71 (3.04)	16 (1.69)
Language	–	–	–	–	–	–
English	6811 (94.4)	4.06 (7.46)	23.06 (5.93)	3670 (94.15)	2235 (94.70)	906 (94.57)
Spanish	215 (3.0)	3.36 (6.59)	20.24 (5.96)	119 (3.05)	72 (3.05)	24 (2.50)
Chinese	68 (0.9)	5.12 (8.79)	21.37 (6.55)	48 (1.23)	7 (0.29)	13 (1.35)
Other	122 (1.7)	4.12 (6.80)	21.86 (5.73)	61 (1.56)	46 (1.94)	15 (1.56)
CDR	–	–	–	–	–	–
Cognitively unimpaired	3905 (54.0)	0.18 (1.02)	25.98 (2.95)	–	–	–
MCI	2362 (32.7)	4.11 (5.19)	22.00 (4.58)	–	–	–
Dementia	960 (13.3)	19.68 (6.38)	12.92 (6.28)	–	–	–

Note: Group differences and effect sizes listed in Table S1 in supporting information.
Abbreviations: CDR, Clinical Dementia Rating; FAQ, Functional Activities Questionnaire (FAQ); MCI, mild cognitive impairment; MoCA, Montreal Cognitive Assessment.

class against all other classes, and the 95% confidence intervals are estimated as AUC ± 1.964 its standard errors (SE). AUC interpretive ranges include: .50 to .69 (poor), .70 to .79 (acceptable), .80 to .89 (excellent), and ≥.90 (outstanding). The closer the OOB score is to 1, the better the model can extrapolate predictions to new data.

2.3.1 | Combined models

To assess MoCA and FAQ accuracy for predicting cognitive status based on CDR, we ran models with: (1) MoCA/demographics (corrected-CS), (2) MoCA/FAQ/demographics (corrected-MTMM), and (3) MoCA/FAQ (2D MTMM). The first model was run to provide a comparison for the subsequent two models, as ROC analyses used in the literature do not provide demographic-corrected AUC values. The second determined how well a corrected MTMM model improved classification accuracy. The third provided a more direct comparison to the Steenland et al.[18] 2D-MTMM model.

Of the 7227 individuals with MoCA/FAQ, 7054 had all demographic information completed. To construct the RF model, we split the data as follows: 20% (1410) of the overall data was kept for final testing, and the remaining 80% (5644) was used as training set. The RF parameters *ntree* and *mtry* were optimized from the data as shown in Figure S1

in supporting information to provide a better OOB accuracy for each model.

2.3.2 | Ethnoracial comparisons

To assess the MTMM approach's ability to reduce ethnoracial classification disparities, we divided the sample by ethnoracial groups and compared the first two models: corrected-CS and corrected-MTMM. Ethnoracial classification was determined by self-identification from US Census–consistent options. Given sample size limitations, we were only able to isolate NHWs and BAs.[33]

There were 5305 NHW participants, of which 5263 have age, sex, education, and language information. There were 1030 BA participants, of which 1025 reported their age, sex, language, and education. RF models were built from each ethnoracial group, all adjusting for sex, age, language, and education.

2.3.3 | Probabilistic models

We used the 2D-MTMM RF models in the total sample, and NHW and BA subsamples, to construct probabilistic models. The probability
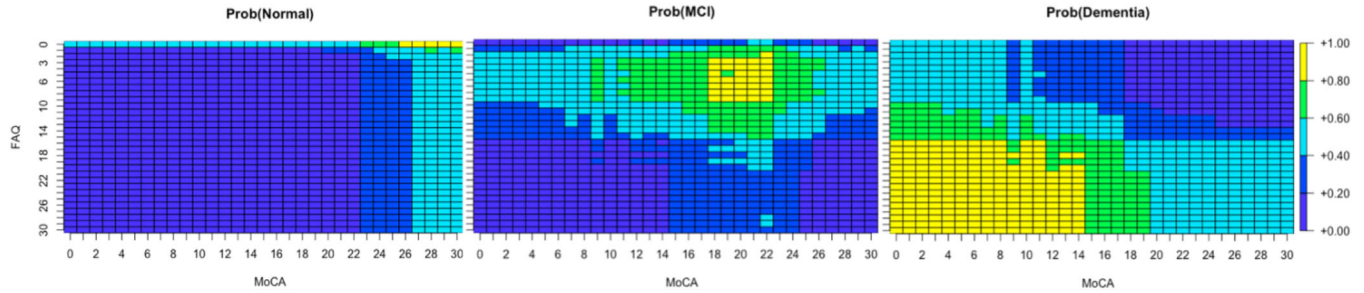
FIGURE 1 Heatmaps of the probability distributions. Visualized probabilities (color coded) of Clinical Dementia Rating (CDR) stages for combinations of Montreal Cognitive Assessment (MoCA; x-axis) and Functional Activities Questionnaire (FAQ; y-axis) values. Provided separately for cognitively unimpaired (left), mild cognitive impairment (MCI; middle), and dementia (right). On rightmost legend, +1.00 (yellow) indicates highest probability and +0.00 (purple) indicates lowest probability
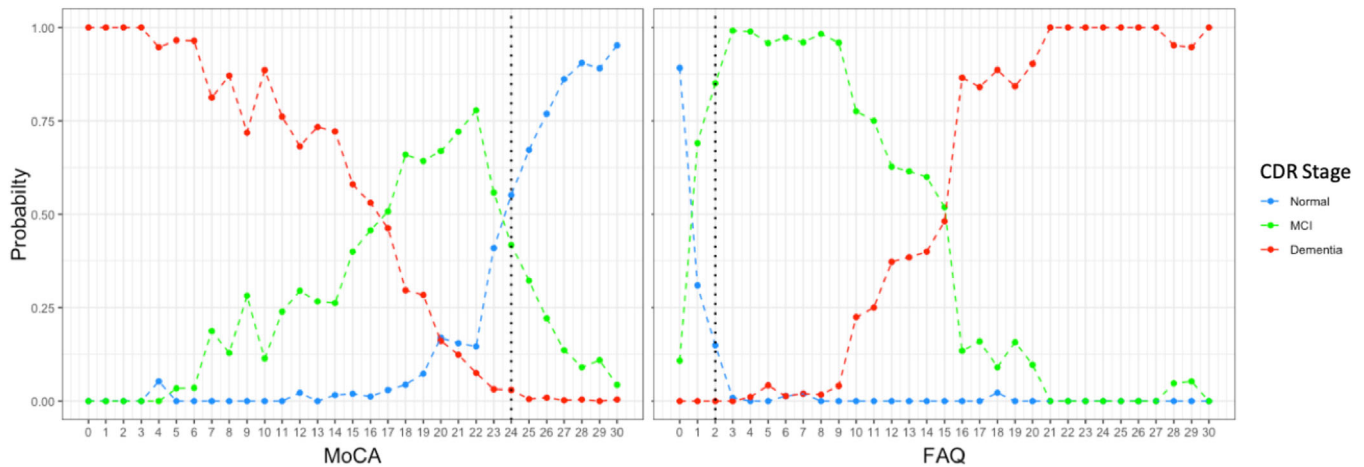

FIGURE 2 Probability distributions for Montreal Cognitive Assessment (MoCA) and Functional Activities Questionnaire (FAQ). Visualized probabilities (y-axis) of each Clinical Dementia Rating (CDR) stage at each measure's score (x-axis). Black dotted lines are derived cutoffs between cognitively unimpaired and cognitively impaired participants for MoCA and FAQ. Provided separately for MoCA (left) and FAQ (right)

$P_t(k, l)$ to have t (t = CU, MCI, or Dementia) as outcome given k value for FAQ and l value for MoCA is given by:

$$P_t(k, l) = \frac{1}{2}\left(\frac{N_t}{N_k} + \frac{N_t}{N_l}\right)$$

where $\frac{N_t}{N_k}$ is the ratio of the number of predicted t cases and the total number of points evaluated for a given value of k.

We used the equation above to compute the joint probability of CU–MCI–dementia classification for pairs of MoCA and FAQ scores. These were positioned into a 2D heatmap (Figure 1) and estimated cognitive impairment cutoffs from cumulative probability distributions (Figure 2). These were used to create a simple 2D nomogram (Figure 3) to determine the most likely classification. The x-axis lays out MoCA raw scores and the y-axis lays out FAQ raw scores. The colors inside the figure indicate a likely diagnosis of CU, MCI, or dementia. The clinician simply needs to find the point for their patient's MoCA and FAQ scores to determine the likely diagnosis. This is also available in an online calculator that provides the likelihood of a cognitive stage with a combination of MoCA and FAQ in the total sample, or NHW–BA sub-samples. While joint probabilities may be more accurate for predicting an individual case, they will not yield probabilities if the original data did not include that score combination, and they are also limited in the number and type of predictors for chained probabilities. Thus, we also offered the unmanipulated RF models for corrected-MTMM that also allow predictions across score combinations and demographic corrections. Both can be found at: https://bernard-fongang.shinyapps.io/moca_faq//

3 | RESULTS

3.1 | Participant characteristics

Despite a wide range, individuals were mostly older ($M = 69.05$, $SD = 9.99$, range = 18–101), NHW (74.6%), female (61.3%), with higher levels of education ($M = 16.01$, $SD = 2.82$, range = 0–26). Other demographics are summarized in Table 1.
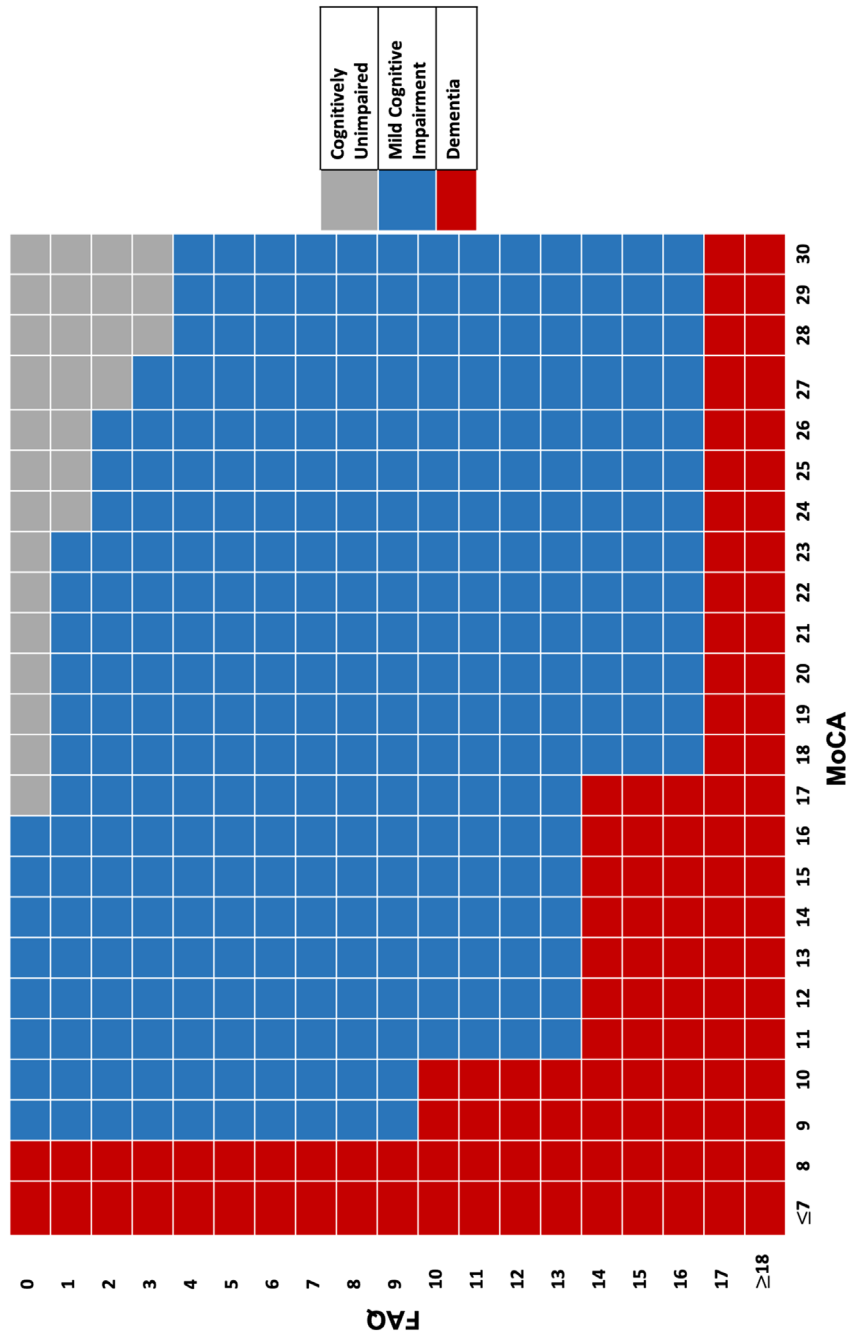
FIGURE 3 2D classification nomogram, Simplified classification aid for determining a cognitive stage based on combinations of Montreal Cognitive Assessment (MoCA; x-axis) and Functional Activities Questionnaire (FAQ; y-axis) values

3.2 | Combined models

3.2.1 | Corrected CS

The variable importance plot (Figure S1) showed ethnoracial group, age, education, sex, and language were marginally important covariates in our prediction of diagnostic group. AUC values are listed in Table 2. The model OOB estimate was .709. Overall, the model was better at classifying dementia than CU, which was better than MCI classification. Per the confusion matrix (Table 3), the model tended to underestimate impairment, predicting people with observed MCI to be CU (n = 217), highlighting the difficulty of distinguishing MCI from CU patients.

3.2.2 | Corrected MTMM

The variable importance plot (Figure S1) indicated ethnoracial group, age, education, sex, and language were marginally important covariates. Overall, the model was similar to corrected CS for predicting dementia; however, there was a drastic improvement in CU and MCI classification, based on non-overlapping confidence intervals. Per the confusion matrix (Table 3), this model still tended to underestimate impairment for those with known MCI, but was markedly more accurate than the first model, which is captured in comparison of averages for model AUCs.

3.2.3 | 2D-MTMM

Comparing the averages of model AUCs, the 2D-MTMM and corrected MTMM were mostly equivalent (Table 2). However, this obscured differences in accuracy for each cognitive stage. The 2D-MTMM was more accurate at classifying dementia, at the cost of reduced accuracy for CU and MCI, compared to the corrected MTMM. However, the 2D-MTMM was still more accurate than the corrected-CS at every cognitive stage.

3.3 | Ethnoracial comparisons of classification accuracy

3.3.1 | Non-Hispanic whites

For corrected CS, the maximum number of trees was 95 and the number of predictors used for each split 2 leading to maximum OOB accuracy of 0.63. For NHWs, the corrected CS model was mostly equivalent overall compared to the model derived from the overall sample, with slightly improved MCI classification (Table 2). Compared to the general sample, the corrected MTMM AUCs were mostly equivalent (Table 2).

3.3.2 | Black Americans

For corrected CS, the maximum number of trees was 19 and the number of predictors used for each split was 2 leading to maximum OOB accuracy of 0.610. The training and the testing datasets had 820 and

205 participants. The corrected CS model demonstrated that classification accuracy was significantly lower for BAs, with MCI and CU being most discrepant compared to NHWs. This was also true compared to the overall sample, which is mostly comprised of NHWs. In contrast, the corrected MTMM model significantly narrowed the classification accuracy gap, with overall model accuracy being quite similar, although MCI and CU classification were still significantly less accurate for BAs compared to NHWs, with dementia classification being much better for BAs.

4 | DISCUSSION

This article aimed to extend prior research with 2D-MTMM to include a more sensitive CS in models using advances in classification algorithms, and evaluate whether such an extension attenuated classification disparities observed with CS alone. 2D-MTMM approaches combining information from the MoCA and FAQ had stronger discriminability of CU and MCI groups relative to the demographically corrected lone CS model. Dementia detection was roughly equivalent across approaches with the highest discriminability for the combined MoCA/FAQ model without demographic correction. We did not anticipate that the demographically unadjusted model would outperform the fully corrected model. These findings may be attributable to the lack of weighting given to demographic factors in the diagnostic criterion measure for dementia (CDR). Consistent with our predictions, the demographically corrected 2D-MTMM approach combining the MoCA and FAQ was superior for correctly classifying the CU and MCI groups relative to the unadjusted 2D-MTMM. These results are salient as MCI classifica-

TABLE 2 Classification accuracy for participants with no cognitive impairment, mild cognitive impairment, and dementia

	MoCA Full Model	MoCA + FAQ Full model	MoCA+ FAQ (No adjustment)
Overall Sample	AUC [95% CI]	AUC [95% CI]	AUC [95% CI]
Cognitively Unimpaired	0.835 [.815, .855]	0.906 [.892, .920]	0.882 [.866, .898]
MCI	0.712 [.676, .748]	0.835 [.810, .860]	0.792 [.763, .821]
Dementia	0.938 [.914, .962]	0.908 [.879, .937]	0.98 [.914, .993]
Average	0.828 [.801, .855]	0.883 [.860, .904]	0.884 [.848, .904]
Non-Hispanic White			
Cognitively unimpaired	0.862 [.841, 883]	0.917 [.902, .932]	–
MCI	0.744 [.706, .782]	0.852 [.825, .879]	–
Dementia	0.937 [.911, .963]	0.919 [.889, .949]	–
Average	0.847 [.819, .876]	0.896 [.872, .920]	–
Black American			
Cognitively Unimpaired	0.731 [.652, .810]	0.849 [.801, .897]	–
MCI	0.614 [.488, .740]	0.751 [.660, .842]	–
Dementia	0.91 [.806, .999]	0.992 [.966, .999]	–
Average	0.751 [.649, .850]	0.864 [.809, .913]	–

Abbreviations: AUC, area under the curve; CI, confidence interval; FAQ, Functional Activities Questionnaire (FAQ) ; MCI, mild cognitive impairment; MoCA, Montreal Cognitive Assessment.

TABLE 3 Confusion matrices

Observed values	Predicted values		
Overall sample-	Normal	MCI	Dementia
Corrected CS			
Normal	656 (87.4%)	94 (12.5%)	1 (0.1%)
MCI	217 (47.9%)	200 (44.2%)	36 (7.9%)
Dementia	8 (3.9%)	66 (32.0%)	132 (64.1%)
Corrected MTMM			
Normal	712 (93.1%)	52 (6.8%)	1 (0.1%)
MCI	147 (30.9%)	295 (62.1%)	33 (7.0%)
Dementia	0 (0.0%)	35 (20.6%)	135 (79.4%)
2D-MTMM			
Normal	732 (95.6%)	34 (4.4%)	0 (0.0%)
MCI	193 (40.5%)	263 (55.1%)	21 (4.4%)
Dementia	1 (0.5%)	45 (22.3%)	156 (77.2%)
Non-Hispanic White-			
Corrected CS			
Normal	457 (86.7%)	67 (12.7%)	3 (0.6%)
MCI	148 (40.7%)	180 (49.4%)	36 (9.9%)
Dementia	5 (3.1%)	40 (24.8%)	116 (72.1%)
Corrected MTMM			
Normal	527 (92.5%)	43 (7.5%)	0 (0.0%)
MCI	92 (27.4%)	229 (68.4%)	14 (4.2%)
Dementia	0 (0.0%)	36 (24.5%)	111 (75.5%)
Black American-			
Corrected CS			
Normal	113 (89.0%)	13 (10.2%)	1 (0.8%)
MCI	45 (70.3%)	17 (26.6%)	2 (3.1%)
Dementia	0 (0.0%)	4 (28.6%)	10 (71.4%)
Corrected MTMM			
Normal	123 (93.9%)	8 (6.1%)	0 (0.0%)
MCI	25 (43.9%)	29 (50.9%)	3 (5.2%)
Dementia	0 (0.0%)	1 (5.9%)	16 (94.1%)

Note: Parenthetical percentages indicate the prediction accuracy between predicted class (column) and actual/observed class (row). Numbers vary depending on number of predictors.
Abbreviations: corrected CS, Model including the Montreal Cognitive Assessment (MoCA) and demographics; corrected MTMM, model including the MoCA, Functional Activities Questionnaire (FAQ), and demographics; 2D MTMM, Model including only the MoCA and FAQ; MCI, mild cognitive impairment.

tion typically lags those of CU and dementia groups.[34] The challenges of detecting MCI are particularly appreciable in the primary care setting due to providers' time limitations and lack of confidence making cognitive diagnoses.[35] Unfortunately, this barrier poses significant limitations to early intervention as primary care providers are often the first point of contact regarding cognitive concerns.

Regarding the question of ethnoracial disparities, we objectively tested the Tappen et al.[19] hypothesis that 2D-MTMM of a CS and a

functional assessment tool may reduce divergences in classification accuracy across diverse ethnoracial groups relative to CS alone. The demographically corrected MTMM approach yielded the most equivalent diagnostic accuracy across NHW and BA groups, which aligns with research demonstrating demographic correction improves classification accuracy based on CS,[14] and clinical staging derived from the FAQ is relatively robust against demographic effects.[23] The ability of two time- and cost-efficient tools to reduce biases in classification accuracy across diverse groups is notable given striking disparities in timely diagnosis and access to care.[2]

With the growing availability of large datasets, there is increasing use of advanced algorithms for data modeling and prediction. Research across chronic diseases indicates that predictive algorithms can improve diagnostic accuracy, reduce errors in decision making, and alleviate workload among providers.[40–42] Numerous studies highlighted the utility of multimodal prediction models for detecting MCI and neurodegenerative diseases.[38,43–45] Unfortunately, these promising research findings have largely failed to influence clinical practice primarily due to integration of expensive measures (i.e., neuroimaging, extensive neuropsychological assessment) that are not widely available and the unreasonable analytical burden in the fast-paced clinical setting.[43] Our approach addressed these barriers by solely basing prediction on a widely used CS, a freely available functional assessment questionnaire, and basic demographics, which are feasible to collect within the clinic. Moreover, we created an online calculator that enables providers to simply enter a patient's basic demographics and MoCA and FAQ scores to generate probability of CU, MCI, and dementia classification. The online tool is freely accessible and was designed with ease of use in mind. A clinician could administer the MoCA while a collateral (e.g., family member) completes the FAQ. Then these scores could be entered into the calculator, with demographic scores prepopulated by staff with information in the medical record. The reported probabilities could then inform treatment plan and consults. The calculator is available on https://bernard-fongang.shinyapps.io/moca_faq//

Several features of our analytical approach enhanced optimization for MCI detection. First, we used a 2D-MTMM to combine a CS and functional status measure. While MCI is conceptualized as a stage without frank functional impairment, there is growing awareness of challenges associated with daily task completion.[36,37] In fact, a study evaluating prediction of clinical conversion within individuals with MCI using a multimodal approach, which incorporated broad factors such as neuropsychological assessment, neuroimaging, and cerebrospinal fluid biomarkers, found the FAQ was the single best predictor.[38] In contrast to prior research,[18] our study used the MoCA, which is more sensitive to MCI than the MMSE.[39] Our fully adjusted model also accounted for factors known to influence CS performance such as age, education, and ethnoracial group.[16] Finally, our study used an RF approach, which derives the ultimate classification decision from numerous classification trees, leading to more robust outcomes.[21,38]

A primary study strength was using a large, multicenter NACC cohort with available standardized clinical assessments and gold standard diagnostic determinations. Nonetheless, the sample skewed female and was more highly educated with a larger proportion of

NHWs relative to the general US population. Due to the more limited representation of diverse ethnoracial groups, we were unable to examine groups beyond NHWs and BAs with the latter more highly underrepresented within the sample. Furthermore, the smaller BA sample size compared to NHWs means that the BA estimates are less reliable, which is reflected in their wider confidence intervals. Taken together, these limitations suggest that the calculator, as is, may not extend to the entire US population or all subpopulations within the United States. This sampling limitation underscores outstanding calls to diversify aging research enrollment. Our study used an RF analytical approach given its ability to derive robust classifications from large datasets; avoid overfitting; and accommodate correlated, nonlinear variables.[21,45] Nonetheless, interpretation of machine learning models can be challenging. Finally, while the factors included in our models have been identified as the strongest predictors for clinical classification of dementia,[38,46] they are unlikely to be able to effectively discriminate diverse etiological causes of dementia.[44] Thus, the value of the approach lies in the enhanced screening performance that can be leveraged to triage individuals more appropriately for comprehensive work-up.

In summary, a combined informatics approach using the MoCA and FAQ enhanced diagnostic accuracy across the CU–MCI–dementia continuum relative to CS alone. With adjustment for demographic factors, we were able to reduce disparities in diagnostic accuracy across NHW and BA groups. Our prediction approach was specifically developed to optimize clinical translation with incorporation of time- and cost-effective variables and the creation of an easy-to-use online tool for clinicians. Together with ongoing research, our findings may serve to reduce diagnostic errors, attenuate diagnostic delays, and enhance opportunities for early intervention more equitably across the population.

CONFLICTS OF INTEREST

Bernard Fongang: No conflicts of interest to declare regarding this manuscript content. Mitzi M. Gonzales: In the last 36 months, has: (1) received grant funding from the National Institute on Aging, Texas Alzheimer's Research and Care Consortium, Alzheimer's Association Part of the Cloud Gates Grant, Alzheimer's Drug Discovery Foundation, Ruby & Perry Stevens Parkinson's Disease Center of Excellence, all paid to institution; and received travel funding from the Alzheimer's Association International Conference 2021 Travel Fellowship, paid to institution. David A. González: In the last 36 months, has received grant funding from Ruby & Perry Stevens Parkinson's Disease Center of Excellence, paid to institution and served/led unpaid roles on professional advisory boards and committees within Epilepsy Foundation of Central and South Texas, American Academy of Neuropsychology Relevance 2050, and National Academy of Neuropsychology. Kyle J. Jennette: No conflicts of interest to declare regarding this manuscript content. Jason R. Soble: No conflicts of interest to declare regarding this manuscript content.

REFERENCES

1. Alzheimer's Association. 2020 Alzheimer's disease facts and figures. Alzheimer's *Dement* 2020;16:391-460. https://doi.org/10.1002/alz.12068.
2. Babulal GM, Quiroz YT, Albensi BCC, et al. Perspectives on ethnic and racial disparities in Alzheimer's disease and related dementias: update and areas of immediate need. Alzheimer's. *Dement J Alzheimer's Assoc*. 2019;15:292-312.
3. Congress 111th. *Patient Protection and Affordable Care Act of 2010*. United States of America: 124; 2010.
4. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. 5th ed. Arlington, VA: American Psychiatric Association; 2013.
5. Block CK, Johnson-Greene D, Pliskin N, Boake C. Discriminating cognitive screening and cognitive testing from neuropsychological assessment: implications for professional practice. *Clin Neuropsychol*. 2017;31:487-500.
6. Cordell CB, Borson S, Boustani M, et al. Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during the Medicare Annual Wellness Visit in a primary care setting. *Alzheimer's Dement*. 2013;9:141-150.
7. Owens DK, Davidson KW, Krist AH, et al. Screening for Cognitive Impairment in Older Adults: uS Preventive Services Task Force Recommendation Statement. *JAMA J Am Med Assoc*. 2020;323:757-763.
8. Wiese LK, Williams CL. Annual cognitive assessment for older adults: update for nurses. *J Community Health Nurs*. 2015;32:187-198.
9. Chou KL, Amick MM, Brandt J, et al. A recommended scale for cognitive screening in clinical trials of Parkinson's disease. *Mov Disord*. 2010;25:2501-2507.
10. Nasreddine ZS, Phillips NA, Bédirian V, et al. MoCA: a brief screening tool for mild cognitive impairment. *The Montreal Cognitive Assessment*. 2005;53:695-699. J Am Geriatr Soc.
11. Folstein MF, Folstein SE, McHugh PR. Mini-Mental State". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.

12. Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The Mini-Cog: a cognitive "vital signs" measure for dementia screening in multi-lingual elderly. *Int J Geriatr Psychiatry*. 2000;15:1021-1027.

13. Ciesielska N, Sokołowski R, Mazur E, Podhorecka M, Polak-Szabela A, Kędziora-Kornatowska K. Czy test Montreal Cognitive Assessment (MoCA) moze być skuteczniejszy od powszechnie stosowanego Mini-Mental State Examination (MMSE) w wykrywaniu łagodnych zaburzeń funkcji poznawczych u osób po 60. roku zycia? Metaanaliza. *Psychiatr Pol*. 2016;50:1039-1052.

14. Milani SA, Marsiske M, Cottler LB, Chen X, Striley CW. Optimal cut-offs for the Montreal Cognitive Assessment vary by race and ethnicity. Alzheimer's Dement Diagnosis. *Assess Dis Monit*. 2018;10:773-781.

15. Malek-Ahmadi M, Powell JJ, Belden CM, et al. Age-and education-adjusted normative data for the Montreal Cognitive Assessment (MoCA) in older adults age 70-99. *Neuropsychol Cogn*. 2015;22:755-761. Aging.

16. Rossetti HC, Lacritz LH, Cullum CM, Weiner MF. Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology*. 2011;77:1272-1275.

17. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. Alzheimer's. *Dement*. 2018;14:535-562.

18. Steenland NK, Auman CM, Patel PM, et al. Development of a rapid screening instrument for mild cognitive impairment and undiagnosed dementia. *J Alzheimer's Dis*. 2008;15:419-427.

19. Tappen RM, Rosselli M, Engstrom G. Use of the MC-FAQ and MMSE-FAQ in cognitive screening of older african americans, hispanic americans, and european americans. *Am J Geriatr Psychiatry*. 2012;20:955-962.

20. Kallumpuram S, Sudhir Kumar C, Khan B, Gavins V, Khan A, Iliffe S. Targeted case finding for dementia in primary care: surrey Downs dementia diagnosis project. *BMJ Qual Improv Reports*. 2015;4:u209827.w4086.

21. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.

22. Pfeffer R, Kurosaki T, Harrah C, Chance J, Filos S. Measurement of functional activities in older adults in the community. *J Gerontol*. 1982;37.

23. González DA, Gonzales MM, Resch ZJ, Sullivan AC, Soble JR. Comprehensive Evaluation of the Functional Activities Questionnaire (FAQ) and Its Reliability and Validity. *Assessment*. 2021.

24. Morris JC, Ernesto C, Schafer K, et al. Clinical dementia rating training and reliability in multicenter studies: the Alzheimer's Disease Cooperative Study experience. *Neurology*. 1997;48:1508-1510.

25. Morris JC. Clinical Dementia Rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *Int Psychogeriatrics*. 1997;9:173-176. Int Psychogeriatr.

26. Fawagreh K, Gaber MM, Elyan E. Random forests: from early developments to recent advancements. *Syst Sci Control Eng*. 2014;2:602-609.

27. Buri M, Hothorn T. Model-based random forests for ordinal regression. *Int J Biostat*. 2020;16.

28. Quadrianto N, Ghahramani Z. A very simple safe-Bayesian random forest. *IEEE Trans Pattern Anal Mach Intell*. 2015;37:1297-1303.

29. Janitza S, Hornung R. On the overestimation of random forest's out-of-bag error. *PLoS One*. 2018;13:e0201904.

30. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*. 2007;8:25.

31. R Core Team. *R: A language and environment for statistical computing 2021*.

32. JASP Team. JASP 2020.

33. Van Der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol*. 2014;14.

34. Kumar A, Singh S, Chaurasiya A, Singh A. Implication of preclinical biomarkers in identification of mild cognitive impairment (MCI) and Alzheimer's disease: promises and future challenges |. *Research Reports Res Reports*. 2018;2:e1-16.

35. Sabbagh MN, Boada M, Borson S, et al. Early Detection of Mild Cognitive Impairment (MCI) in Primary Care. *J Prev Alzheimer's Dis*. 2020;7:165-170.

36. McAlister C, Schmitter-Edgecombe M, Lamb R. Examination of Variables That May Affect the Relationship between Cognition and Functional Status in Individuals with Mild Cognitive Impairment: a Meta-Analysis. *Arch Clin Neuropsychol*. 2015;31:123-147.

37. Puente AN, Terry DP, Faraco CC, Brown CL, Miller LS. Functional impairment in mild cognitive impairment evidenced using performance-based measurement. *J Geriatr Psychiatry Neurol*. 2014;27:253-258.

38. Ritter K, Schumacher J, Weygandt M, Buchert R, Allefeld C, Haynes JD. Multimodal prediction of conversion to Alzheimer's disease based onincomplete biomarkers. Alzheimer's Dement Diagnosis. *Assess Dis Monit*. 2015;1:206-215.

39. Pinto TCC, Machado L, Bulgacov TM, et al. Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly?. *Int Psychogeriatrics*. 2019;31:491-504.

40. Lisboa PJ, Taktak AFG. The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Networks*. 2006;19:408-415.

41. Helldén A, Al-Aieshy F, Bastholm-Rahmner P, et al. Development of a computerised decisions support system for renal risk drugs targeting primary healthcare. *BMJ Open*. 2015;5.

42. Dagliati A, Tibollo V, Sacchi L, et al. Big Data as a Driver for Clinical Decision Support Systems: a Learning Health Systems Perspective. *Front Digit Humanit*. 2018;5:8.

43. Bucholc M, Ding X, Wang H, et al. A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. *Expert Syst Appl*. 2019;130:157-171.

44. Bruun M, Rhodius-Meester HFM, Koikkalainen J, et al. Evaluating combinations of diagnostic tests to discriminate different dementia types. Alzheimer's Dement Diagnosis. *Assess Dis Monit*. 2018;10:509-518.

45. Dauwan M, van der Zande JJ, van Dellen E, et al. Random forest to differentiate dementia with Lewy bodies from Alzheimer's disease. Alzheimer's Dement Diagnosis. *Assess Dis Monit*. 2016;4:99-106.

46. Ewers M, Walsh C, Trojanowski JQ, et al. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging*. 2012;33.

47. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959;56(2), 81.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.