

Developing A Claims-based Algorithm For Estimating State- And Local-level Dementia Prevalence

Ruijia Chen, ScD¹, Jingxuan Wang, M.S², Maria Glymour, ScD³, John Hsu, MD⁴

¹ Department of Community Health Sciences, UCLA Fielding School of Public Health

² Department of Epidemiology, Harvard T.H. Chan School of Public Health

³ Department of Epidemiology, Boston University School of Public health

⁴ Mongan Institute for Health Policy , Massachusetts General Hospital



Background

- State- and local-level dementia prevention policy planning and monitoring requires reliable dementia surveillance data, which are currently lacking. Although Medicare claims cover most older adults in the U.S., claims-based dementia diagnoses are subject to misdiagnosis bias.
- Population-based cohort studies with detailed neuropsychological assessments tend to provide more accurate dementia ascertainment, but they are often not representative at small geographic levels.
- Using 2006–2020 Health and Retirement Study (HRS)–Medicare linkage, this study aims to develop and validate a dementia prediction algorithm to improve state- and local-level dementia prevalence estimation.

Data and Measure

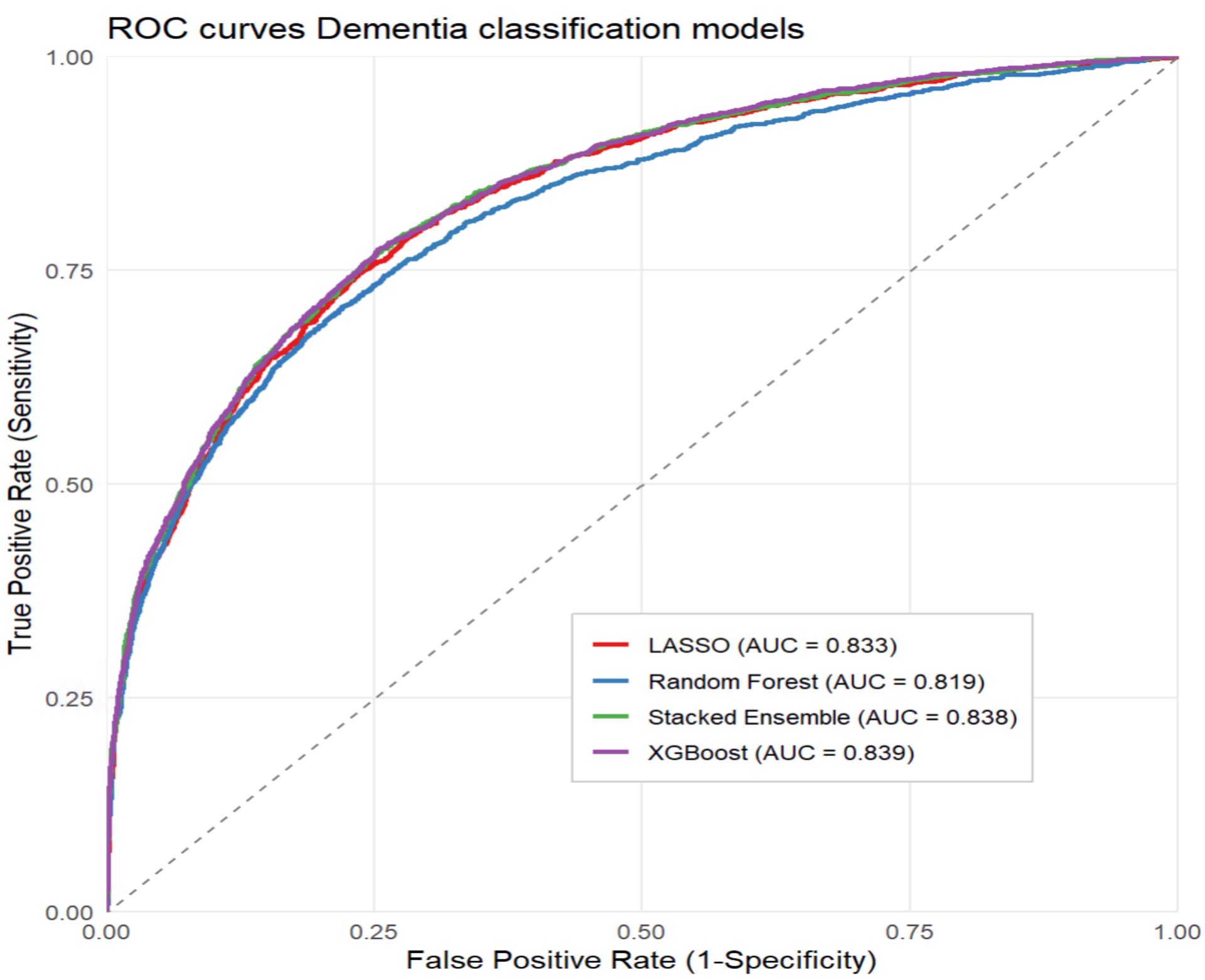
- Dementia status was defined using the Gianattasio-Power algorithm, an established approach based on HRS survey measures and neuropsychological assessments.
- Predictors included age, sex, race and ethnicity, dual Medicare/Medicaid eligibility, state of residence, mid-year Alzheimer’s disease diagnosis, diabetes, hypertension, stroke, chronic obstructive pulmonary disease, obesity, and hearing impairment.
- The analytic sample was partitioned at the person level into a training set (80%) and a test set (20%) using stratified sampling on the outcome. Person-level splitting (rather than person-year splitting) ensured that all observations from a given respondent fell into either the training or test set.

Methods

- We trained three base learners, including a LASSO-penalized logistic regression with the regularization parameter selected by cross-validation, a Random Forest classifier, and XGBoost (gradient-boosted trees).
- Hyperparameters for all base learners were tuned using 5-fold cross-validation within the training set.
- We additionally constructed a stacked ensemble combining the LASSO, Random Forest, and XGBoost models, in which out-of-fold predictions from the 5-fold cross-validation procedure were used as features in a logistic regression meta-learner.
- We evaluated model performance using AUC with 95% confidence intervals obtained via DeLong's method. At the probability threshold maximizing Youden's J statistic, we additionally reported sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and F1 score. Calibration was assessed visually using calibration plots and quantified by the Expected Calibration Error.

Characteristic	Overall (N = 41286) ¹	No dementia (N = 31630) ¹	Dementia (N = 9656) ¹	p-value ²
Age, years	79.5 (6.1)	78.1 (5.2)	84.2 (6.3)	<0.001
Missing	26	15	11	
Sex				<0.001
Male	16,800 (40.7%)	13,287 (42.0%)	3,513 (36.4%)	
Female	24,460 (59.3%)	18,328 (58.0%)	6,132 (63.6%)	
Unknown	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Missing	26	15	11	
Race / ethnicity				<0.001
Black	5,477 (13.3%)	3,716 (11.8%)	1,761 (18.3%)	
Hispanic	3,462 (8.4%)	2,736 (8.7%)	726 (7.5%)	
Other	255 (0.6%)	200 (0.6%)	55 (0.6%)	
White	32,066 (77.7%)	24,963 (79.0%)	7,103 (73.6%)	
Missing	26	15	11	
Dual eligible (Medicare/Medicaid)	5,946 (14.4%)	3,412 (10.8%)	2,534 (26.3%)	<0.001
Missing	26	15	11	
Prior dementia claim (Mattke algorithm)				<0.001
0	38,387 (93.0%)	31,120 (98.4%)	7,267 (75.3%)	
1	2,899 (7.0%)	510 (1.6%)	2,389 (24.7%)	

¹ Mean (SD); n (%). ² Pearson's Chi-squared test; Kruskal–Wallis rank sum test.



Key findings

- Out of 41226 person years and 11576 individual data, 9637 observations had dementia diagnosis from the expert model. The average age is 80 years old.
- Overall, all models demonstrated good discriminative ability, with area under the receiver operating characteristic curve (AUC) values exceeding 0.80. XGBoost achieved the highest performance (AUC = 0.839), closely followed by the stacked ensemble model (AUC = 0.838) and LASSO (AUC = 0.833).
- Random Forest showed slightly lower discrimination (AUC = 0.819).
- Despite these small differences, the ROC curves largely overlapped across models, indicating comparable predictive performance.

Discussion

With further refinement, this algorithm could be applied to full Medicare claims data to improve small-area dementia prevalence estimates and inform state and local resource allocation for dementia prevention.

Funding

This project is supported by NIA 1K99AG088369-01