

Background

There is growing interest in evaluating the potential of proteomics, plasma biomarkers of Alzheimer's disease and related dementias (ADRD), and their combination to predict incident dementia.

Methods

Data

- We examined women recruited between 1996-1999 across 39 U.S. clinical sites into the Women's Health Initiative Memory Study.
- Participants were ages 65-79 years and cognitively unimpaired at baseline, with follow-up through 2021.
- Our primary outcome was incident dementia. Dementia was ascertained and adjudicated annually based on dementia on *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)* criteria.
- Baseline plasma biomarkers of ADRD (p-tau217, p-tau181, GFAP, NfL, A β 42, and A β 40) were available.
- The Simoa® Human Neurology 4-Plex E (N4PE) was used to measure A β 40, A β 42, NfL, and GFAP and the Simoa® p-tau181 v2 assay was used to measure p-tau181 on the Quanterix platform. P-tau217 was measured using the ALZpath Simoa®pTau-217 v2 assay.
- Proteomics were measured using the SomaScan 11K aptamer-based platform.

Prediction of incident dementia

- We employed a multi-stage analytical **Machine learning** approach combining univariate Cox proportional hazards filtering, Cox regression with elastic net regularization, and survival random forest modeling with competing risk adjustment for all-cause mortality to predict incident dementia.
- Data were randomly partitioned into training and testing sets (80%-20%).
- Model performance was evaluated using the C-index for models trained on proteomics alone, plasma ADRD biomarkers alone, and their combination, with age as a covariate.

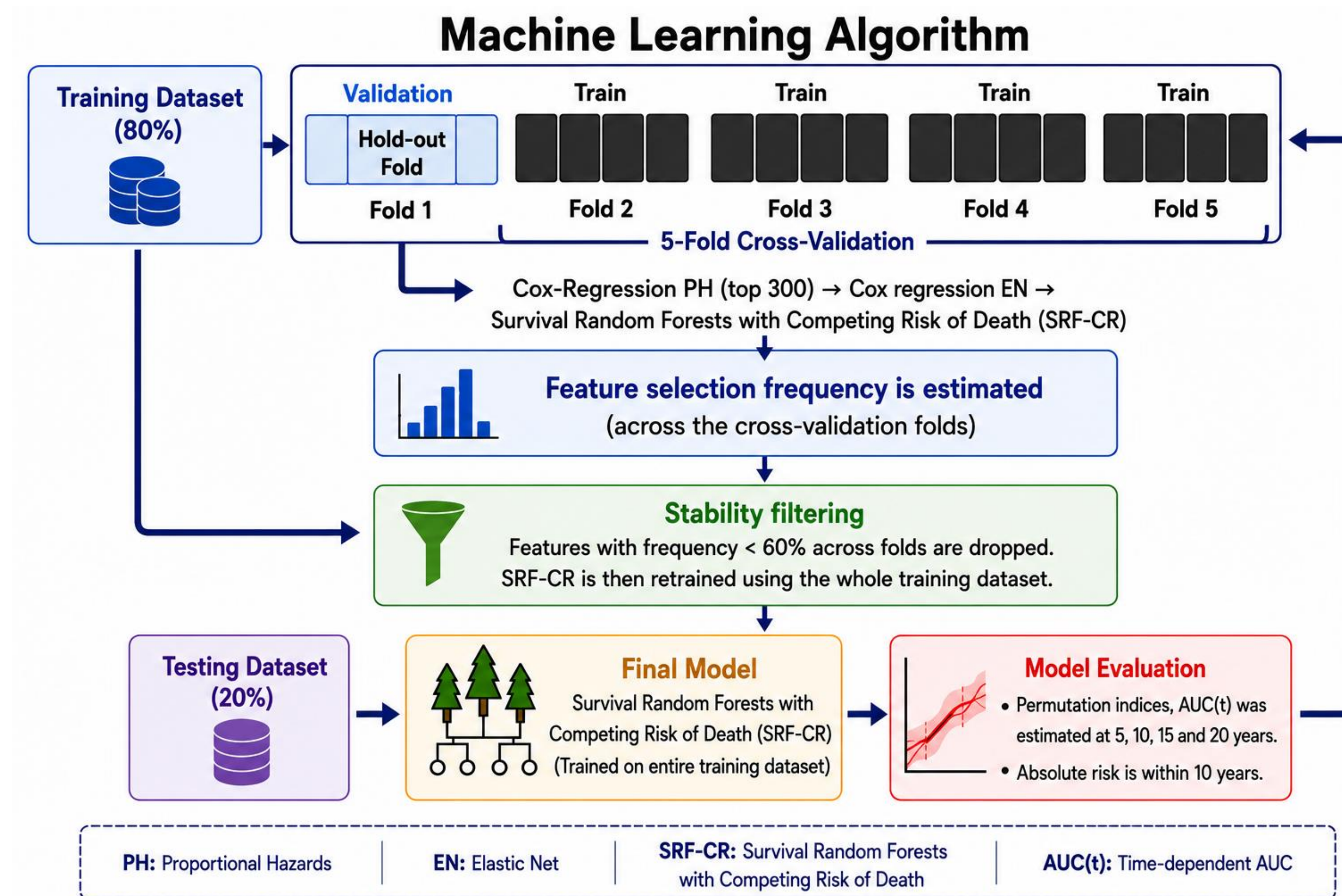


Figure 1. The machine learning algorithm consisted of multiple steps embedded in a cross-validation procedure to reduce dimension and leave in the model the most informative proteins.

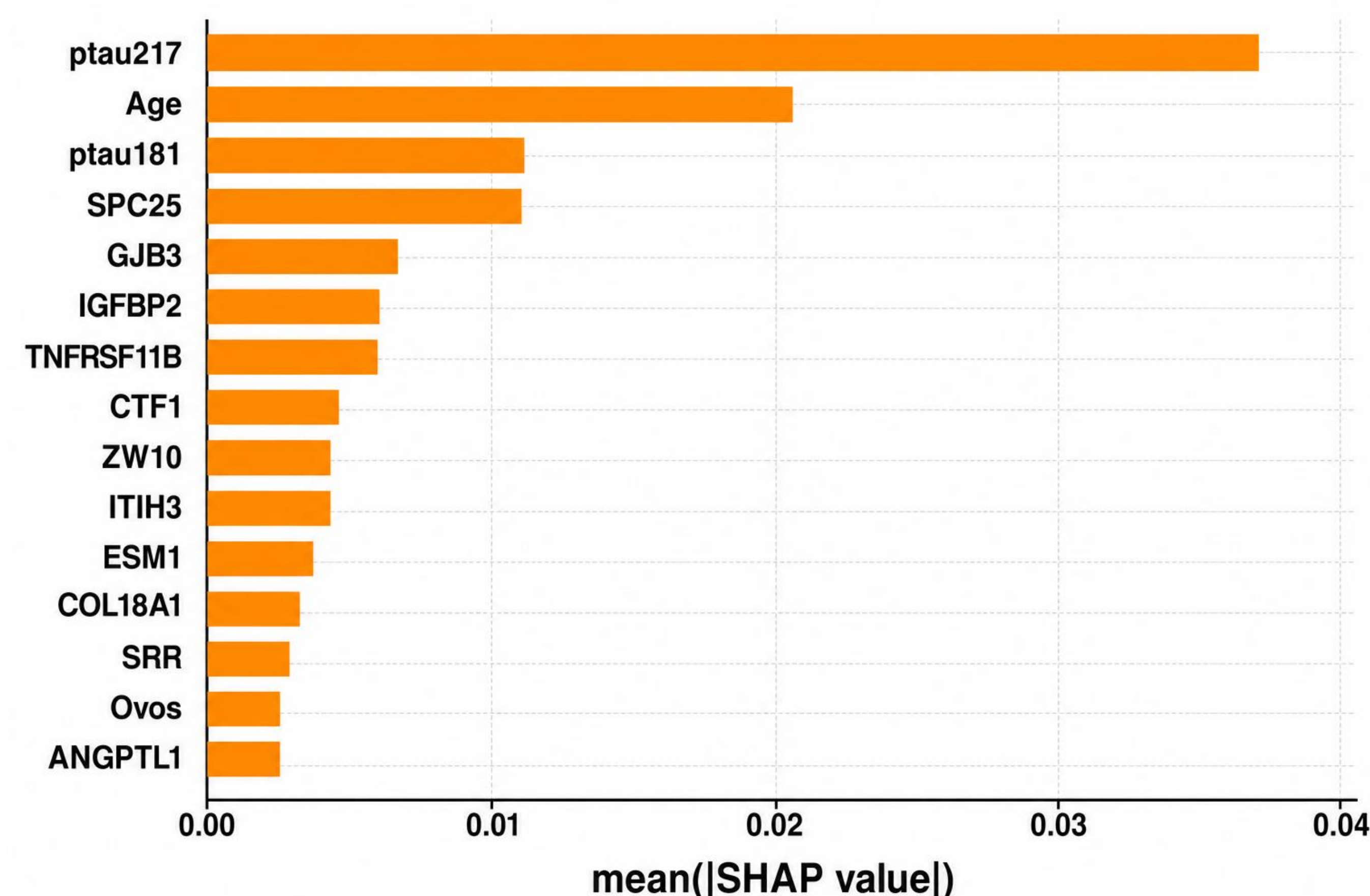


Figure 2. Top 15 predictors selected by the model based on the combination of proteomics, plasma ADRD biomarkers, and age.

- The prediction time horizon was fixed at 10 years of follow-up.
- Shapley values were estimated to identify the most influential predictors within each model.

Results

- Among 2,749 participants (mean baseline age 69.9 [SD-3.8] years), 442 developed incident dementia over a mean follow-up of 8.7 years.
- The C-index was 0.69 for proteomics alone, 0.73 for proteomics with age, 0.79 for plasma ADRD biomarkers alone, 0.81 for plasma ADRD biomarkers with age, and 0.84 for the full combination model.
- In the combined model, p-tau217, age, p-tau181, and 49 plasma proteins were among the 52 selected predictors.
- P-tau217 was the most influential predictor followed by age (Figure 2).
- Three proteins ranked highly across all proteomics-based analyses: 1) kinetochore protein Spc25; 2) Insulin-like Growth Factor Binding Protein 2 and 3) gap junction beta-3 protein.

Conclusions

The combination of plasma biomarkers, proteomics, and age produces a strong predictor of dementia up to 10 years in advance. Ptau217 was found to be the strongest predictor of incident dementia among all biomarkers we investigated. Our results suggest that combining plasma ADRD biomarkers and next-generation proteomics produces synergistic effects on prediction performance.

References

- Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**, 1-22 (2010)
- Ishwaran H & UB., K. Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC). *R package*