

Machine Learning Identifies Key IADL Predictors Of Cognitive and Functional outcome In the Oldest-old

Eojin LEE^{1,3}, Hyun Sook KIM^{2,3*}



¹Graduate School of CHA University, Republic of Korea, ²Department of Neurology, CHA Bundang Medical Center, Republic of Korea, ³Memory Center, CHA Bundang Medical Center, Republic of Korea



Wed-0821

- ✓ CDR-SOB reflects dementia severity better than K-MMSE.
- ✓ Telephone Use and Shopping were the strongest IADL predictors in the oldest-old.
- ✓ Two IADL items retained >90% of the explanatory power of the full 33-predictor model.

INTRODUCTION

- Given the high prevalence and diverse manifestations of cognitive and functional impairment among the oldest-old, it is clinically important to identify functional indicators that are simple and sensitive yet reflect cognitive function.
- Instrumental activities of daily living (IADL) such as financial/medication management, shopping, and telephone use are established predictors of cognitive decline in younger-old adults (65–85 years), but it is unclear whether these predictors can be generalized to the rapidly growing population of the oldest-old (≥90 years).
- Machine-learning methods such as random forest (bagging) and gradient boosting (boosting) allow for the selection of robust predictors by validating feature importance across algorithms with distinct learning paradigms; however, these methods have rarely been applied to the oldest-old.
- This study aimed to identify the functional predictors that most strongly predict cognitive function and dementia severity in the oldest-old using item-level ADL, IADL, and neuropsychiatric measures and to derive a simplified set of predictive indicators suitable for clinical use.

METHODS

- This retrospective, single-center study included patients aged ≥90 years who visited our institution between 2019 and 2025.
- The predictor variables consisted of a total of 33 items, comprising the Barthel Index (10 items), the Korean-IADL (K-IADL, 11 items), and Neuropsychiatric Inventory Questionnaire (12 items). Outcome measures were the Korean Mini-Mental State Examination (K-MMSE), Clinical Dementia Rating Sum of Boxes (CDR-SOB), and the Global Deterioration Scale (GDS).
- For statistical analysis, Random Forest and XG Boost models were trained using 5-fold cross-validation. Feature importance was validated across these two algorithms with distinct learning paradigms (bagging vs. boosting).
- Rank-weighted scoring across six model-outcome combinations identified consensus predictors, and sequential variable reduction determined the minimum set maintaining at least 90% of full-model performance across all three outcomes.
- All statistical analyses were performed using IBM SPSS Statistics for Windows, version 27 (IBM Corp, Armonk, NY, USA) and R 4.5.0 software (The R Foundation for Statistical Computing, Vienna, Austria). This study was approved by the Institutional Review Board (IRB No. 2024-12-001).

RESULTS

- A total of 201 patients, 57 (28.4%) were male and 144 (71.6%) were female. The average age was 91.88±1.95 years (range: 90-99 years), with an average duration of education was 7.00±5.86 years (range: 0-20 years).
- In the full random forest model including all predictors, R² values were 0.417 for K-MMSE, 0.611 for CDR-SOB, and 0.484 for GDS. In the full XGBoost model using the same predictors, the corresponding R² values were 0.308 (K-MMSE), 0.538 (CDR-SOB), and 0.383 (GDS). Overall random forest full model explained more variance than the XGBoost full model across all three outcomes (Figure 1).

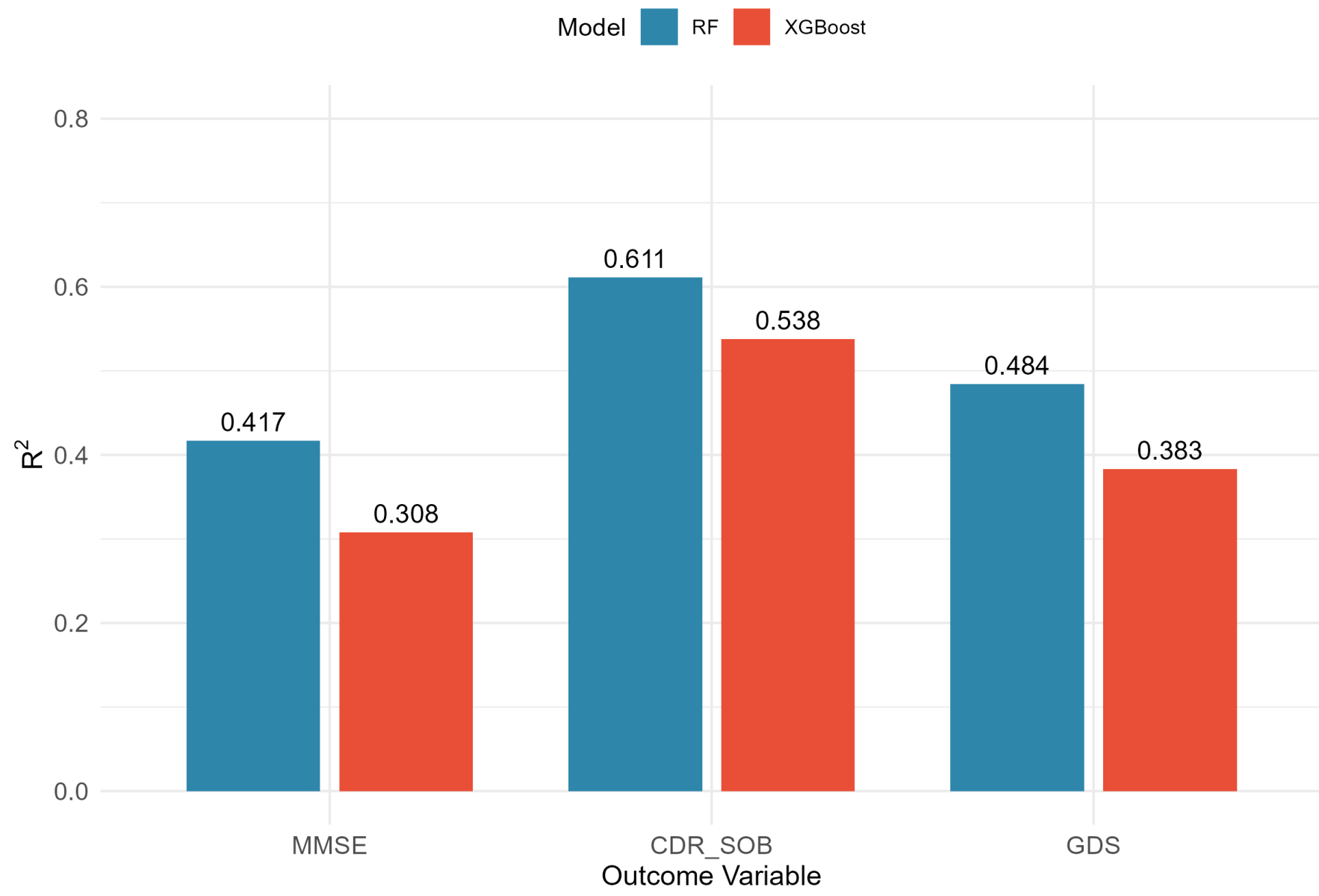


Figure 1. Model performance comparison between random forest and XGBoost.

- Among 33 variables from ADL, IADL, and neuropsychiatric scales, K-IADL items ranked highest in feature importance across both algorithms. Telephone Use (rank-weighted score 52/60) and Shopping (51/60) showed the greatest consistency across model-outcome combinations, while financial management ranked third (45/60) because of its relatively lower importance in predicting GDS (Figure 2).

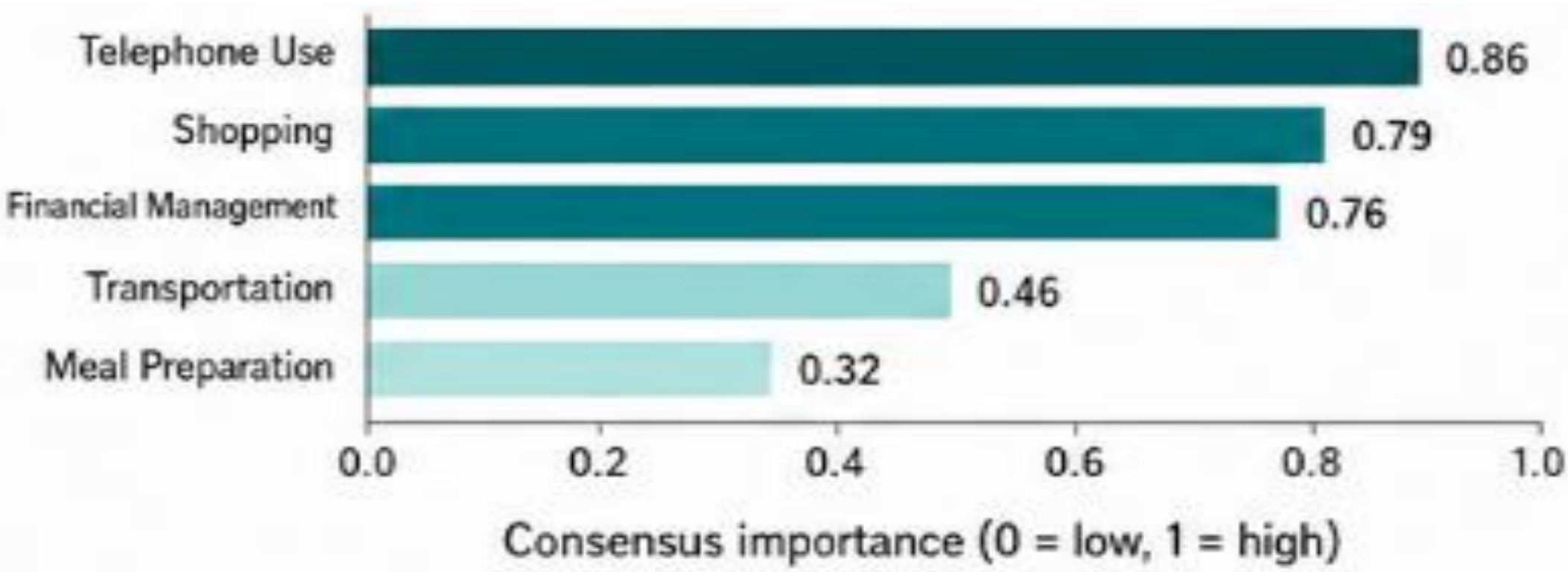


Figure 2. Top 5 predictors across models.

- A simplified model including only Telephone Use and Shopping retained over 90% of the predictive performance of the full 33-variable model across all outcomes, with R² values of 0.422 (101.2% of the full model) for K-MMSE-2, 0.572 (93.6%) for CDR-SOB, and 0.438 (90.5%) for GDS (Figure 3).

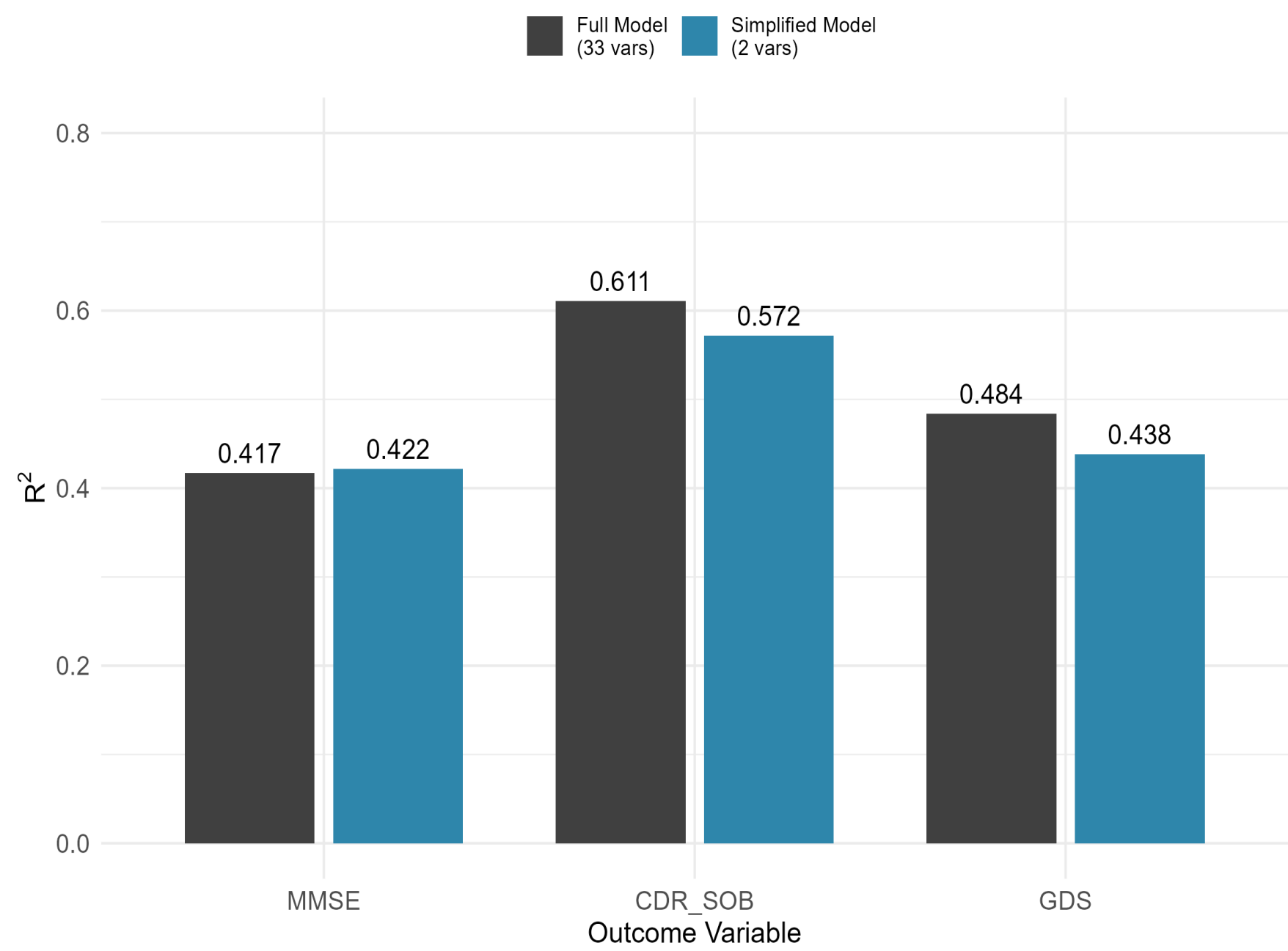


Figure 3. R² of full and two-variable models across outcomes.

CONCLUSIONS

- The stronger overall predictability of CDR-SOB compared with K-MMSE underscores that functional dependence is more tightly aligned with dementia staging than with brief cognitive test scores alone, reinforcing current geriatric practice that emphasizes functional assessment in diagnosis and follow-up.
- Clinically, careful assessment of Telephone Use and Shopping may provide a rapid and practical screening approach to identify oldest-old individuals at increased risk of cognitive impairment or functional decline, particularly in settings where comprehensive neuropsychological testing or full IADL assessment is not feasible. These two items achieved comparable performance to the 33-variable full model.
- To the best of our knowledge, this is the first study to use machine learning to systematically identify the most informative IADL domains for predicting cognitive and functional outcomes in the oldest-old.
- Future studies with larger, independent cohorts are warranted to externally validate these findings and evaluate their generalizability across diverse populations.