

# Identification Of Factors Associated With Cognitive Status In Older Mexican Adults Using MHAS 2021 Data Through Random Forest

Tuesday-1091

Leonardo Corral-Robles [ [leonardo\\_corral@ciencias.unam.mx](mailto:leonardo_corral@ciencias.unam.mx) ], Claudia Iveth Astudillo-García

This study shows that Random Forest models are Good tools for uncovering the shifting relative importance of multi-domain predictors in cognitive aging, thereby offering crucial, stratified insights to guide targeted public health interventions.

## INTRODUCTION

The prevalence of cognitive impairment and dementia is rising globally and in Mexico, making early screening and intervention essential [1]. To address this, this study aims to identify the demographic, physical health, functional, and social-context variables associated with cognitive status classification in older Mexican adults. By using Random Forest, a supervised learning method that captures complex, non-linear relationships, we highlight the relevance of these multi-domain predictors and determine the relative importance of each factor for the outcome [2]. Additionally, adjusted odds ratios (OR) were calculated to identify the direction of effect for these variables.

## METHODS

This study used data from the 2021 wave of the Mexican Health and Aging Study (MHAS), focusing on adults aged 50 and older. Cognitive function was assessed using the CCCE instrument (covering memory, orientation, and attention), classifying participants into normal cognition or cognitive impairment as the primary outcome. After filtering, a final sample of 11,654 participants was retained. The sample was stratified overall and into four age quartiles to train individual Random Forest classifiers using a 75/25 data split. To mitigate class imbalance and maximize predictive performance, each model was optimized through a randomized hyperparameter search (RandomizedSearchCV) with 5-fold stratified cross-validation focused on the ROC-AUC metric. Finally, test set evaluation was performed by optimizing the decision threshold to maximize balanced accuracy, from which the most important variables were extracted to contrast the relative weight of each predictive factor across the different age groups.

## RESULTS

The Random Forest model identified years of schooling as the most dominant predictor of cognitive status across all age groups, showing its highest relative importance in the youngest quartile (Q1 = 23.1%) and a gradual decline in older cohorts, where psychological and functional markers gained prominent predictive weight. Specifically, depressive symptoms peaked in importance within the third quartile (Q3 = 11.3%), while loneliness, decision-making autonomy, and functional indicators (such as grip strength) exhibited an increased relative contribution in the older groups (Q3 and Q4), suggesting a critical role for functional reserve and frailty dynamics in advanced age. Complementing these machine learning insights, the multivariable regression analyses—fully adjusted for schooling, sex, age, marital status, and urban residence—revealed that high perceived stress was significantly associated with a higher risk of cognitive impairment (OR = 1.79; 95% CI: 1.30–2.48;  $p \approx 0.0004$ ), whereas high levels of depressive symptoms (OR = 1.24) and high feelings of loneliness (OR = 1.25) also trended as risk factors. Conversely, high self-reported health (OR = 0.75), smoking history (OR = 0.79), and medium financial status (OR = 0.84) demonstrated inverse, protective associations against impairment. Notably, severe difficulties in decision-making autonomy (OR = 0.39;  $p \approx 0.002$ ) emerged as a significant protective factor within this adjusted framework, highlighting the complex, non-linear interplay between socioeconomic, mental, and physical health determinants across the aging spectrum.

## CONCLUSIONS

Overall, model performance and the consistency of top-ranked predictors support a central role of socio-educational and psychosocial factors in the outcome, with age-related variation. Random Forest consistently prioritizes schooling, depressive symptoms, sleep quality, and life satisfaction, while highlighting loneliness and functional status in older adults. Stratified regression confirms robust signals for perceived stress and schooling. Notably, unexpected inverse associations—such as a smoking history appearing protective—likely reflect a selective survival effect inherent to older cross-sectional cohorts. Thus, Random Forest emerges as a useful tool to prioritize factors and support more evidence-informed decision-making in public health.

## REFERENCES

- 1) Mejia-Arango, S., Arce Rentería, M., Saenz, J. L., & Wong, R. (2026). Prevalence of dementia and mild cognitive impairment in the Mexican population: Evidence from Mex-Cog, a harmonized cognitive assessment protocol study. *The Journals of Gerontology: Series B*, 81(3). <https://doi.org/10.1093/geronb/gbaf268>
- 2) 2) Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. <https://doi.org/10.58496/BJML/2024/007>

