



Population-Based ADRD Risk Prediction In American Samoans: Comparative Multiclass Modeling And Evaluation Metrics

To’aiva Penelope Hall¹, MPH; Tulimalefoi D. Vaofanua¹, MS, Grace Tuatoo¹, Vaatausili Tofaeono¹, PhD, MBA; James E. Galvin², MD, MPH

¹American Samoa Community Cancer Coalition, PO Box 1716, Pago Pago, American Samoa, 96799

²Comprehensive Center for Brain Health, Department of Neurology, University of Miami Miller School of Medicine, 7700 W Camino Real, Boca Raton, FL, 33433



AMERICAN SAMOA
COMMUNITY CANCER COALITION

INTRODUCTION

Native Hawaiian and Pacific Islander (NHPI) communities experience a disproportionate burden of cardiometabolic disease, limited access to specialty cognitive care, and persistent underrepresentation in dementia research—factors that shape Alzheimer’s disease and related dementias (ADRD) detection and care pathways. The purpose of this study is to develop and evaluate multiclass ADRD risk classification models using routinely collected demographic, clinical, and behavioral variables.

Puipui Malu Manatu ("Preserving Memory") is the first study of its kind in American Samoa, investigating Alzheimer’s disease and related dementias (ADRD) in a probability sample of 1,069 Samoan adults aged 50 and older. Funded by the NIA (RF1AG075901), the study assesses ADRD knowledge, resilience and vulnerability factors, and cognitive status using culturally adapted tools. Gold Standard cognitive evaluations (UDSv3.0) and analyses of genetic (ApoE) and plasma ATN biomarkers (Aβ40, Aβ42, p-tau, NFL, GFAP) are being conducted to estimate ADRD prevalence and compare biological profiles with other populations.



Key Cognitive Assessments

- **MoCA-Blind:** Assesses memory, attention, language, abstraction, and orientation; scores <18 suggest possible impairment
- **Cognivue®:** Computerized assessment evaluating multiple cognitive domains, including memory and executive function
- **Number Symbol Coding Test (NSCT):** Measures processing speed and executive function using timed symbol-number matching

These assessments collectively provide a multidimensional evaluation of cognitive function and are instrumental in identifying individuals at elevated risk for ADRD.



Key Predictive Models

- **Multinomial Logistic Regression:** Statistical model estimating category probabilities
- **Random Forest:** Ensemble machine learning method combining multiple decision trees
- **Neural Network:** Single hidden-layer model capturing nonlinear relationships

METHODOLOGY

Study Design

This study is a retrospective secondary analysis of de-identified data from American Samoa Community Cancer Coalition’s *Puipui Malu Manatu* (PMM) research study (RF1AG075904) on ADRD (n = 819; complete-case analysis).

ADRD Risk Level Classification

ADRD risk was determined using performance across three cognitive assessments (MoCA-Blind, Cognivue®, NSCT).

- **Low risk (per test):**
MoCA-Blind ≥18, Cognivue® ≥71, NSCT ≥34
- **High risk (per test):**
Scores below threshold

Overall ADRD Risk Classification:

- **Low:** Low risk on all three assessments
- **High:** High risk on all three assessments
- **Intermediate:** Mixed low/high risk across assessments

This composite approach was used to stratify participants into ADRD risk categories (see Fig. 1).

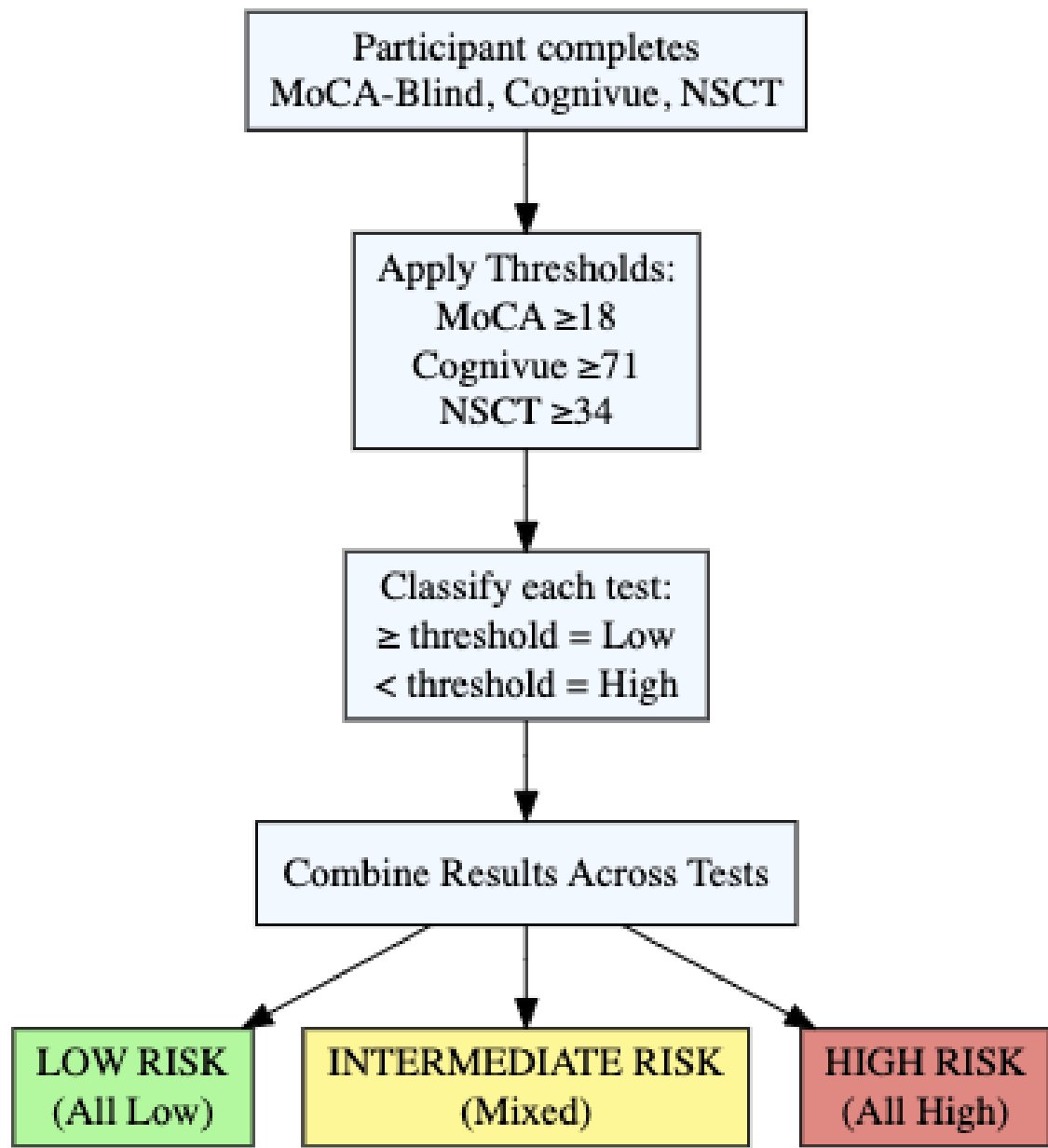


Figure 1. ADRD Risk Level Classification Process

Predictors

Models included routinely collected demographic, clinical, and behavioral variables: **Age (years), Sex, Race, Education (years), BMI, Global Clinical dementia Rating (CDR), Diabetes, Hypertension, tobacco use, and alcohol use.**

Model Development

Data were split into training (80%) and testing (20%) sets using stratified sampling (see Figure 2). Three models were developed and compared:

- Multinomial logistic regression
- Random forest (tuned with class weighting)
- Single hidden-layer neural network

Model performance was evaluated on the test set using accuracy, macro F1 score, multiclass RO

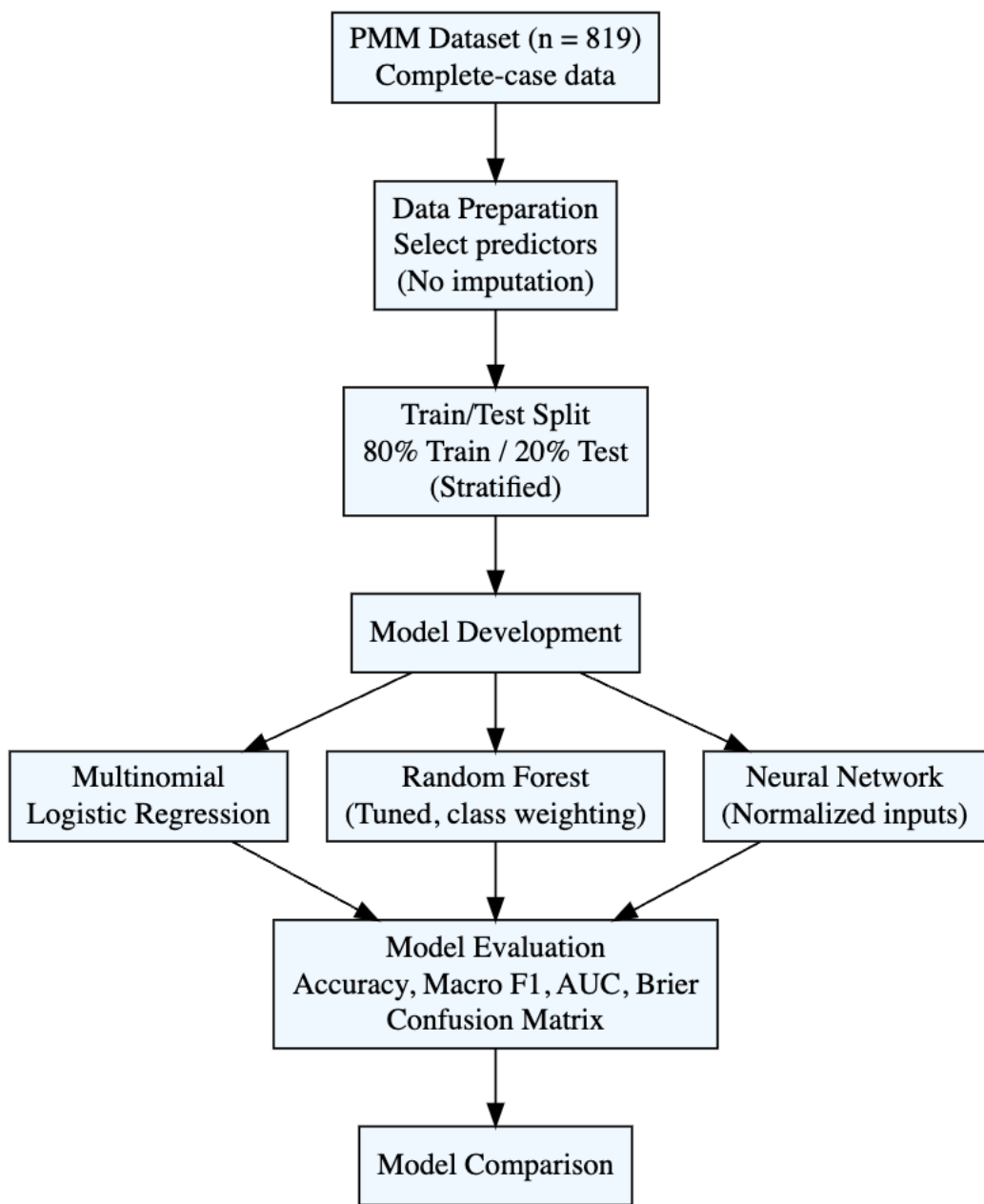


Figure 2. Model Development Process

RESULTS

Participant Characteristics

The sample (n = 819) had a mean Age (years) of 62.6 ± 7.8, with Female (%) comprising 56.2% of participants. Mean Education (years) was 11.4 ± 2.8 and mean BMI was 36.1 ± 8.2. Diabetes (%) and Hypertension (%) prevalence were 40.4% and 56.3%, respectively. ADRD risk distribution was 58.4% High, 33.9% Intermediate, and 7.7% Low. Higher-risk participants had higher Age (years) and lower Education (years), while Low-risk participants had a higher proportion of Female (%) (see Table 1).

Model Evaluation and Performance

Model performance demonstrated moderate discriminative ability across approaches (see Table 2):

- **Multinomial logistic regression**
 - achieved highest overall accuracy (0.636) and AUC (0.726) (see Figure 3), indicating the strongest overall classification and ranking performance
- **Neural network**
 - demonstrated highest macro F1 score (0.534), suggesting improved balance across ADRD risk categories, particularly in identifying the minority Low-risk group
- **Random forest**
 - slightly lower performance across all metrics (accuracy 0.605; AUC 0.692; macro F1 0.424)
 - did not outperform the simpler logistic regression model

Across models, misclassification occurred most frequently between Intermediate and High risk categories, while Low risk individuals remained the most challenging to identify.

Table 1. PMM Participant Characteristics by ADRD Risk Level (n=819)			
Characteristic	High	Intermediate	Low
Age (years)	64 ± 7.6	61 ± 7.8	58.9 ± 7
Female (%)	49.2%	63.7%	76.2%
Education (years)	10.6 ± 2.7	12.3 ± 2.4	14.1 ± 2.9
BMI	35.3 ± 7.9	36.9 ± 8	39.2 ± 9.5
Diabetes (%)	41.4%	38.5%	41.3%
Hypertension (%)	56.5%	55.4%	58.7%

Table 2. Comparison of Model Performance				
Model	Accuracy	Macro F1	AUC	Brier Score
Logistic Regression	0.636	0.498	0.726	0.228
Random Forest	0.605	0.424	0.692	0.494
Neural Network	0.623	0.534	0.682	0.556

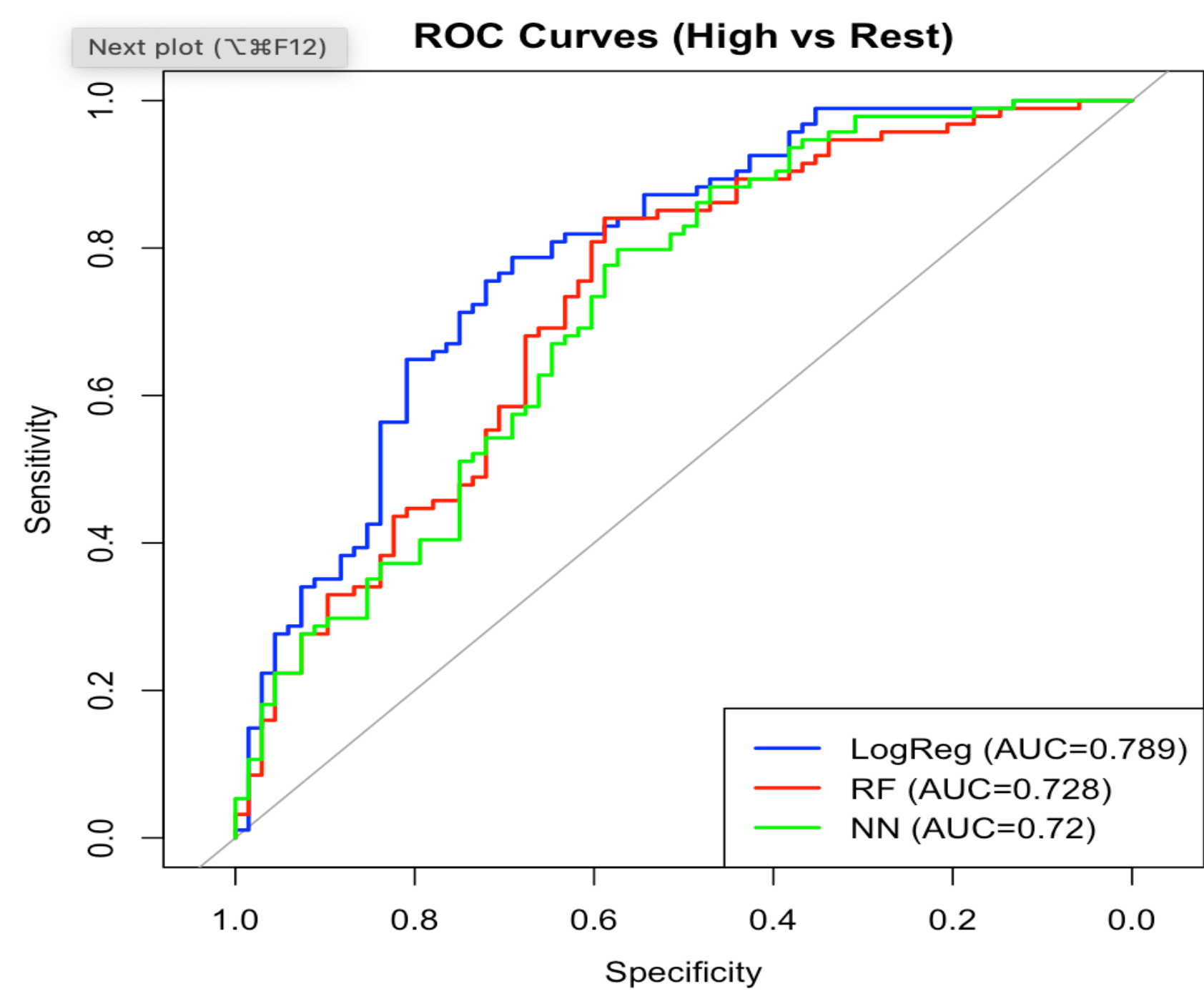


Figure 3. Receiver operating characteristic (ROC) curves comparing multinomial logistic regression, random forest, and neural network models. Curves reflect model performance in distinguishing ADRD risk categories; area under the curve (AUC) values indicate overall discrimination.

DISCUSSION

Overall, routinely available demographic, clinical, and behavioral variables demonstrated *modest* ability to classify three-level ADRD risk. Multinomial logistic regression achieved the strongest overall performance, with the highest accuracy and AUC, suggesting that **simpler, interpretable models can effectively capture risk patterns** in this dataset.

The neural network, while slightly lower in overall discrimination, yielded the highest macro F1 score, indicating improved balance across risk categories and better relative classification of the minority Low-risk group. In contrast, the random forest model did not provide performance gains, highlighting that **increased model complexity did not translate to improved predictive accuracy.**

Across all models, misclassification occurred most frequently between Intermediate and High risk categories, likely reflecting underlying clinical overlap. Low-risk individuals were consistently the most difficult to identify, potentially due to class imbalance and limited distinguishing features using routinely collected variables alone.

Future Directions

This analysis was restricted to complete-case data and a composite cognitive outcome, highlighting opportunities to incorporate more robust approaches to missing data handling and clinical validation in future work. The presence of class imbalance and reliance on a single train-test split further underscore the need for expanded sample sizes and more comprehensive validation strategies, including repeated resampling and external cohorts.

Importantly, biomarker data were available for only a limited subset of participants and were not included in the present analysis. As data collection continues, integrating APOE genotype and plasma ATN biomarkers (e.g., Aβ42/40, p-tau, nFl, GFAP) represents a key opportunity to enhance model performance by capturing underlying disease processes not reflected in clinical and behavioral variables alone.

Additional improvements may be achieved through inclusion of longitudinal cognitive measures and context-specific social and environmental factors relevant to NHPI populations, supporting more accurate, population-specific ADRD risk stratification.

CONCLUSION

This study evaluated the utility of routinely collected demographic, clinical, and behavioral variables for multiclass ADRD risk classification in a population-based cohort of Samoan adults, with multinomial logistic regression achieving the strongest overall performance. These findings support the use of parsimonious, interpretable models for population-level risk stratification in resource-limited settings. Importantly, this work is grounded in an underrepresented NHPI population, demonstrating the feasibility of scalable, data-driven approaches to support earlier and more equitable identification of ADRD risk. While future integration of APOE and plasma ATN biomarkers may enhance predictive performance, these results highlight the potential for population-specific, contextually relevant modeling approaches to advance ADRD risk detection in NHPI communities.