Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting Non-successful Ageing in Rural South Africa: A Prospective Cohort Study using Machine-Learning Analysis of the HAALSI Cohort

Lawrence Ejike Ugwu
Centre for Applied Psychology and Public Health Research in Africa (CAPPHRA)
Bruno Basil
basilb@dufuhs.edu.ng
International Institute for Pathology and Forensic Science Research, David Umahi Federal University of Health Sciences
Uzoamaka Francisca Ugwoke
Enugu State University of Science and Technology
Ekenna Chiedozie Okonkwo
Royal Bournemouth Hospital

Research Article

Keywords: Non-successful ageing, Rural South Africa, HAALSI, Interpretable Machine Learning, SHAP (SHapley Additive exPlanations).

Posted Date: April 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9524228/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Later-life risk in rural South Africa unfolds under conditions of multimorbidity, socioeconomic inequality and uneven access to care, making it difficult to identify who is most vulnerable to adverse ageing trajectories. Using three waves of Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community (HAALSI) data and a cleaned survivor subset of 1,435 adults, we developed an interpretable machine-learning framework to predict non-successful ageing, defined by functional limitation or psychological distress at Wave 3. In repeated 10-fold cross-validation (5 repeats), discrimination increased from 0.774 ± 0.038 in a baseline model to 0.846 ± 0.031 with longitudinal change scores and 0.860 ± 0.029 in the full proximal model; a constrained artificial neural network (ANN) benchmark achieved 0.841 ± 0.042. SHapley Additive exPlanations identified baseline household assets and BMI change as the strongest contributors to prediction, with age and Wave 2 depressive symptoms contributing secondarily. Although rank discrimination was strong, calibration slope and intercept (3.661 and − 1.242) indicated compressed predicted probabilities. These findings suggest that later-life risk in rural South Africa is better captured by multidomain longitudinal information than by baseline characteristics alone and that community health screenings should prioritize monitoring household asset stability and BMI trends as early warning indicators of declining functional health, and that interpretable machine learning may support relative risk stratification, although absolute risk estimation and generalisability remain limited.

Introduction

South Africa provides a particularly important setting for studying later-life vulnerability in Sub-Saharan Africa because population ageing is unfolding alongside persistent socioeconomic inequality, multimorbidity and uneven access to supportive health and social systems [1, 2]. Findings from the Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI) indicate that ageing outcomes in rural South Africa are shaped not only by disease burden, but also by the combined influence of functional health, mental health and material conditions across the life course [1, 3]. As a result, older adults in this setting do not follow a single or stable ageing trajectory, and approaches derived primarily from high-income contexts may miss important variation in risk. This makes rural South Africa a critical context for developing locally-adaptable tools to identify individuals at elevated risk of poorer ageing outcomes over time.

This study is primarily informed by cumulative disadvantage theory, which holds that inequalities in social conditions, material resources and health exposures accumulate across the life course and contribute to divergence in later-life outcomes [4, 5]. This perspective is especially relevant in South Africa, where many older adults age under conditions shaped by long-term structural inequality, household vulnerability and uneven access to protective resources. Within this broader explanatory framework, successful ageing is used as the outcome-oriented gerontological lens for distinguishing more favourable from less favourable ageing statuses in later life [6].

Successful ageing has long been influential in gerontology, commonly emphasizing the avoidance of disease and disability, maintenance of physical and cognitive function and continued engagement with life [6]. However, this framework has also been criticized for relying heavily on evidence from Western, Educated, Industrialized, Rich and Democratic populations and for insufficiently accounting for the social and structural conditions that shape ageing across the life course [7, 8]. Recent meta-analytic and conceptual work further suggests that successful ageing remains relatively uncommon globally and that contemporary frameworks increasingly emphasize subjective and context-sensitive dimensions of ageing alongside objective functioning [9, 10]. In settings such as rural South Africa, later-life well-being may therefore be less accurately captured by idealized models of ageing developed elsewhere. The present study adopts a pragmatic operationalization of later-life outcomes for longitudinal risk prediction while interpreting those outcomes within a cumulative disadvantage perspective.

Despite the growing availability of longitudinal ageing data in Africa, much of the existing literature remains cross-sectional, and recent reviews have highlighted the relatively limited use of longitudinal ageing studies across Sub-Saharan Africa [11]. This limitation is especially relevant in the present setting, where ageing outcomes are likely shaped by the interplay of functional capacity, depressive symptoms, body composition and material resources across time. Machine-learning methods offer a useful complement because they can model complex, non-linear patterns and higher-order interactions without requiring strong a priori functional assumptions [12, 13]. Recent ageing-focused reviews further suggest that explainable machine-learning approaches may be particularly useful when prediction models seek to capture multidomain later-life vulnerability while retaining interpretability [14, 15].

However, predictive performance alone is insufficient if model behaviour remains opaque. SHapley Additive exPlanations (SHAP) provide a framework for examining how individual predictors contribute to both overall model behaviour and person-level predictions, thereby supporting interpretable risk modelling [16]. Using three waves of HAALSI data, the present study evaluates an interpretable machine-learning framework for predicting non-successful ageing in rural South Africa. Specifically, the study aims to quantify the incremental predictive gain associated with adding longitudinal change and proximal follow-up indicators beyond baseline characteristics alone, identify the most influential functional, psychological and socioeconomic predictors of non-successful ageing using SHAP-based interpretation, and assess whether a parsimonious artificial neural network (ANN) benchmark retains meaningful predictive signal in a resource-constrained setting.

Materials and Methods

Study design and data source

This study was a prospective cohort analysis using data from the Health and Ageing in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI), a population-based cohort of adults aged 40 years and older residing in the Agincourt sub-district of Mpumalanga Province, South Africa. Data from three waves were used: Wave 1 (2014–2015), Wave 2 (2018–2019), and Wave 3

(2021–2022). The present analysis examined whether baseline characteristics, longitudinal change, and proximal follow-up indicators improved prediction of non-successful ageing at Wave 3. The reporting of this study follows the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis–Artificial Intelligence (TRIPOD + AI) statement.

Participants and analytical sample

The initial Wave 1 cohort comprised 5,059 participants. For inclusion in the present analysis, participants were required to contribute a valid Wave 3 ageing outcome and to have complete data on the principal Wave 1 to Wave 2 predictors used to estimate longitudinal change. A staged exclusion process was applied. First, 1,731 participants were excluded because of attrition or mortality before Wave 3. Second, 962 participants were excluded because of missing data on key Wave 1→Wave 2 predictors, principally body mass index (BMI) and gait speed. Third, 931 participants were excluded during data hygiene screening because of physiologically implausible changes in BMI or mental-health scores between Waves 1 and 2 exceeding ± 15 units. These values were considered likely to reflect recording artifacts or placeholder values rather than plausible biological change. The final analytical sample therefore comprised 1,435 participants.

Comparative analysis showed that excluded participants were significantly older and more functionally limited at baseline ($p < .01$), indicating potential healthy-survivor bias. Accordingly, the final sample should be interpreted as a cleaned survivor subset rather than as fully representative of the baseline cohort. Complete-case analysis was used because the primary objective was to estimate interpretable longitudinal change metrics, and key repeated-measure predictors were unavailable for a substantial proportion of excluded cases.

Outcome definition

The primary outcome was non-successful ageing at Wave 3, coded as a binary variable (1 = non-successful ageing, 0 = successful ageing). Successful ageing was defined as the simultaneous absence of both functional limitation and psychological distress. Non-successful ageing was coded when either criterion was present. Functional limitation was defined as gait speed below 0.6 m/s or self-reported walking difficulty. Psychological distress was defined as a Center for Epidemiologic Studies Depression Scale (CES-D) score of 9 or higher.

Predictor definition

Predictors were grouped into three nested tiers to evaluate the incremental contribution of baseline characteristics, longitudinal change, and proximal status indicators. Tier 1 comprised baseline variables (age, sex, and Wave 1 household asset index). Tier 2 comprised longitudinal trajectories, operationalized

as Wave 1→Wave 2 change scores for BMI and CES-D. Tier 3 comprised proximal Wave 2 status indicators for depressive symptoms and functional difficulty.

Data cleaning and missing data

To improve interpretability of longitudinal change estimation, data were screened for implausible transitions between Waves 1 and 2. BMI and mental-health shifts greater than ± 15 units were excluded as likely artifacts. No multiple imputation was undertaken because the key repeated-measure predictors required for velocity estimation were unavailable for a substantial proportion of excluded observations. For model fitting, predictors entered into XGBoost were analysed on their original scale, whereas ANN inputs were standardized as appropriate for neural-network training.

Statistical analysis

XGBoost was used as the primary learner because of its capacity to model non-linear relationships and higher-order interactions. A constrained artificial neural network (ANN), comprising a single hidden layer with five neurons and ReLU activation, was retained as a secondary benchmark. To assess potential temporal overlap and proximity bias, three nested model architectures were evaluated: (1) a baseline model containing age, sex, and Wave 1 household asset index; (2) a longitudinal model adding Wave 1→Wave 2 changes in BMI and CES-D; and (3) a full proximal model adding Wave 2 depressive symptoms and functional difficulty.

Internal validation was conducted using repeated 10-fold cross-validation with five repeats. Mean area under the receiver operating characteristic curve (AUC) and standard deviation were reported across folds to quantify performance stability. Calibration was assessed using slope and intercept estimates derived from logistic regression of observed outcomes on predicted log-odds. SHapley Additive exPlanations (SHAP) were used to summarize the relative contribution of predictors to model classification. Model development, validation, and reporting followed established best practices for multivariable prediction modelling, including transparent interpretation of machine-learning outputs [17].

Results

Attrition Flow and Final Analytical Sample

Of the initial 5,059 participants in the Wave 1 cohort, 1,731 were excluded due to attrition or mortality by Wave 3. Missingness in W1 → W2 predictors resulted in the removal of another 962 individuals. Following the data hygiene screening for implausible physiological transitions, a further 931 participants were excluded.

The final analytical sample comprised a cleaned survivor subset of 1,435 adults (Table 1). Comparative analysis revealed that excluded participants were significantly older and more functionally limited at

baseline compared to those retained in the final subset ($p < .01$). These differences indicate a potential healthy-survivor bias within the analytical sample.

Table 1
Sample selection and staged exclusions.

Stage	N excluded	Remaining sample	Reason
Initial HAALSI cohort	—	5,059	Wave 1 participants
Loss to follow-up / mortality	1,731	3,328	Did not contribute Wave 3 outcome
Missing W1→W2 predictors	962	2,366	Missing BMI and/or gait-speed predictors
Data hygiene screening	931	1,435	Implausible BMI or mental-health changes (> ± 15 units)
Final analytical subset	—	1,435	Cleaned survivor subset used for modelling

Nested Model Performance and Proximity Bias

Cross-validated performance estimates from repeated 10-fold cross-validation (5 repeats) showed stable discrimination across the nested models (Fig. 1). The Baseline Model achieved an AUC of 0.774 ± 0.038, the Longitudinal Model achieved 0.846 ± 0.031, and the Full Proximal Model achieved 0.860 ± 0.029. The modest gain of + 0.014 AUC units from adding proximal Wave 2 status indicators suggests that predictive performance was driven primarily by longitudinal trajectories rather than by near-term status indicators that overlap conceptually with the Wave 3 outcome. The constrained ANN benchmark performed similarly to the XGBoost model (AUC = 0.841 ± 0.042), indicating that a meaningful portion of the predictive signal was retained across different algorithmic architectures.

Table 2
Cross-validated nested model performance.

Model	Feature set	Mean AUC	SD
Baseline	Age, sex, Wave 1 asset index	0.774	0.038
Longitudinal	Baseline + BMI change + CES-D change	0.846	0.031
Full proximal	Longitudinal + Wave 2 depressive symptoms + functional difficulty	0.860	0.029
ANN benchmark	Reduced four-feature model	0.841	0.042

Feature Contribution

SHAP analysis (Fig. 2) identified baseline household assets and BMI change as the strongest contributors to the predictive risk score. Lower baseline asset levels and sharper declines in BMI between Waves 1 and 2 were associated with increased predicted risk of non-successful ageing. Age and Wave 2 depressive symptoms also contributed meaningfully, whereas sex and proximal functional status contributed comparatively less.

Calibration and Probability Alignment

Although the model demonstrated strong rank discrimination (AUC = 0.860), calibration analysis yielded a slope of 3.661 and an intercept of − 1.242, indicating substantial probability compression. In practical terms, this means that while the model effectively distinguished higher- from lower-risk individuals, predicted probabilities were insufficiently spread relative to observed outcome frequencies. The model is therefore better suited to relative risk stratification than to precise absolute risk estimation.

Discussion

This study evaluated whether later-life risk in rural South Africa is better predicted by baseline characteristics alone or by a combination of baseline conditions, longitudinal change and proximal follow-up information. Three findings are central. First, discrimination improved substantially when longitudinal trajectories were added to the baseline model, indicating that repeated observation captured important risk information not available from baseline status alone. Second, the additional gain from proximal Wave 2 status indicators was comparatively modest, suggesting that the predictive signal was driven primarily by longer-term trajectories rather than by short-term temporal overlap with the Wave 3 outcome. Third, baseline household assets and BMI change emerged as the strongest contributors to prediction, whereas proximal functional status played a smaller role than expected.

The overall pattern of findings is interpretable through a cumulative disadvantage lens. The strongest contributors to model classification were not confined to a single biomedical marker; instead, they reflected the joint importance of material conditions, bodily change and psychological status. This is consistent with the view that later-life vulnerability in South Africa is shaped by the cumulative consequences of unequal resources and uneven resilience across the life course rather than by one dominant risk factor alone [4, 5]. At the same time, the present results should be read as compatible with cumulative disadvantage rather than as a direct empirical test of that theory. The analysis is predictive rather than causal, and the outcome was deliberately operationalized for risk classification.

A second interpretive point concerns the meaning of non-successful ageing in this study. The outcome was operationalized pragmatically as the presence of either functional limitation or psychological distress at Wave 3. This approach is useful for prediction and stratification, but it does not capture the full multidimensional range often associated with broader successful-ageing frameworks, including cognition, social engagement and subjective well-being [6, 8–10]. The present model therefore classifies a specific adverse ageing profile rather than the entirety of later-life well-being.

The nested analysis is especially important for interpreting reviewer concerns about temporal overlap. The increase from 0.774 to 0.846 when longitudinal trajectories were added suggests that dynamic change between Waves 1 and 2 contains substantial information about later non-successful ageing. By contrast, the smaller increment from 0.846 to 0.860 after the addition of proximal Wave 2 status indicators suggests that the model's performance is not driven mainly by near-term carryover from variables that mirror the Wave 3 outcome. In that sense, the results support the interpretation that longer-term change processes (not only near-term status) carry the dominant predictive signal in this cleaned survivor subset.

The prominence of household assets is consistent with prior South African work showing persistent socioeconomic gradients in later-life health and functional vulnerability [18]. In the present study, household assets likely capture more than material wealth alone; they may also index buffering capacity, access to care and the cumulative effects of structural inequality. Likewise, the strong role of BMI change suggests that bodily trajectories may function as a sensitive marker of nutritional, metabolic and frailty-related processes that are difficult to capture through cross-sectional snapshots alone. The emergence of BMI decline as a leading contributor indicates that change itself may be more informative than static body size in this setting.

Wave 2 depressive symptoms also contributed meaningfully, although they were not the dominant predictor once assets and BMI change were taken into account. This is still substantively important. Prior studies in South Africa have shown that depressive symptoms in older adults are closely linked to disability, functional limitation and wider social vulnerability [19, 20]. The present findings extend that literature by showing that depressive symptoms remain relevant even in a multivariable longitudinal model, but that their contribution is embedded within a broader configuration of material and bodily vulnerability rather than acting as a singular risk determinant.

The constrained ANN benchmark performed similarly to the XGBoost model, indicating that a meaningful portion of the predictive signal was retained across different algorithmic architectures. However, this comparison should be interpreted narrowly. The ANN was intentionally simplified and included as a benchmark rather than as a fully optimized alternative. Its performance therefore suggests that the observed signal was not unique to the boosting framework, but it does not establish equivalence between model classes or imply that neural-network methods were exhaustively tested in this application.

From a methodological perspective, the study demonstrates both the value and the limits of interpretable machine learning in ageing research. On the one hand, the model achieved strong rank discrimination after rigorous data cleaning, and SHAP helped clarify that the principal contributors to classification were baseline socioeconomic conditions and longitudinal BMI change. On the other hand, calibration remained poor despite strong ranking performance. The slope and intercept indicate that predicted risks were compressed and not well aligned with observed outcome frequencies. Accordingly, the framework is more appropriate for relative risk stratification than for precise individual-level

probability estimation. This interpretation is consistent with emerging work showing that explainable machine-learning approaches can improve classification of healthy-ageing outcomes, while recent reviews also stress the need for cautious interpretation, transparent reporting and further validation in older-adult populations [14, 15].

These results also have practical implications. Later-life risk assessment in rural South Africa may be better approached as a multidomain task that integrates material vulnerability, bodily decline and psychological status rather than focusing narrowly on disease burden alone. In particular, the findings suggest that mental health should not be treated as secondary, but they also indicate that longer-term socioeconomic and nutritional trajectories may be at least as important for identifying who is at elevated risk. This implies that multidomain surveillance approaches may be more useful than single-indicator screening in resource-constrained settings.

Some limitations should be acknowledged. First, the final analytical sample was heavily restricted through attrition, missing-data exclusion and outlier screening, resulting in a cleaned survivor subset that was younger and functionally healthier than those excluded. This likely limit generalisability and may inflate observed performance relative to the full cohort. Second, the outlier-removal strategy, although justified by implausible transitions, may also have excluded some true but extreme changes. Third, the operationalization of non-successful ageing was intentionally narrow and did not include cognition or social engagement. Finally, despite strong discrimination, the model has not been externally validated and calibration remained weak for absolute probability estimation.

Conclusion

This analysis suggests that non-successful ageing in rural South Africa is better captured by multidomain longitudinal information than by baseline characteristics alone. The strongest predictive signal was derived from baseline household assets and BMI change, with only modest additional gain from proximal Wave 2 indicators. These findings suggest that community health screenings in rural South Africa should prioritize monitoring household asset stability and BMI trends as early warning signs of declining functional health. Collectively, the results support the use of interpretable machine-learning approaches for relative risk stratification in ageing research, while also underscoring the importance of sample selection, data hygiene, and calibration in evaluating model performance.

Declarations

Ethical Approval and Consent to Participate:

This study is a secondary analysis of de-identified data from the HAALSI cohort. The original HAALSI study was approved by the University of the Witwatersrand Human Research Ethics Committee (Ref: M141159), the Harvard T.H. Chan School of Public Health Office of Human Research Administration (Ref: C13-1608-02), and the Mpumalanga Provincial Research and Ethics Committee. Informed consent was

obtained from all participants at each wave of data collection. The present analysis was conducted in accordance with the data use agreement provided by the HAALSI project.

Consent for Publication:

Not applicable.

Data availability:

The data that support the findings of this study are available from the Harvard Center for Population and Development Studies / MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) HAALSI study, but restrictions apply to the availability of these data, which were used under license for the current study and are therefore not publicly available. Data are, however, available to researchers upon reasonable request and with permission from the HAALSI investigators. Interested researchers can apply for access through the HAALSI data access process via the study website: https://haalsi.org.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions:

LEU and UFU conceived and designed the study. LEU and BB performed the machine-learning analysis and model validation. LEU, BB and ECO drafted the manuscript. All authors have read and approved the final version of the manuscript.

Patient and Public Involvement:

Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of our research.

Acknowledgements:

The authors acknowledge the Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community (HAALSI) team for providing access to the dataset used in this study. We are grateful to the study participants and the field and data management staff of the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) for their invaluable contributions. We also recognize the support of the Harvard Center for Population and Development Studies and collaborating institutions involved in the design, implementation, and maintenance of the HAALSI cohort.

References

1. Gómez-Olivé FX, Montana L, Wagner RG, Kabudula CW, Rohr JK, Kahn K, et al. Cohort profile: Health and Ageing in Africa: a longitudinal study of an INDEPTH community in South Africa (HAALSI). Int J Epidemiol. 2018;47(3):689–j690. 10.1093/ije/dyx247.
2. Payne CF, Houle B, Chinogurei C, Riumallo-Herl C, Kabudula CW, Kobayashi LC, et al. Differences in healthy longevity by HIV status and viral load among older South African adults: an observational cohort modelling study. Lancet HIV. 2022;9(10):e709–16. 10.1016/S2352-3018(22)00198-9.
3. Bassil DT, Farrell MT, Wagner RG, Brickman AM, Glymour MM, Langa KM, et al. Cohort profile update: cognition and dementia in the Health and Aging in Africa Longitudinal Study of an INDEPTH community in South Africa (HAALSI dementia). Int J Epidemiol. 2022;51(4):e217–26. 10.1093/ije/dyab250.
4. Dannefer D. Cumulative advantage/disadvantage and the life course: cross-fertilizing age and social science theory. J Gerontol B Psychol Sci Soc Sci. 2003;58(6):S327–37. 10.1093/geronb/58.6.S327.
5. Ferraro KF, Shippee TP. Aging and cumulative inequality: how does inequality get under the skin? Gerontologist. 2009;49(3):333–43. 10.1093/geront/gnp034.
6. Rowe JW, Kahn RL. Successful aging. Gerontologist. 1997;37(4):433–40. 10.1093/geront/37.4.433.
7. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci. 2010;33(2–3):61–83. 10.1017/S0140525X0999152X.
8. Stowe JD, Cooney TM. Examining Rowe and Kahn's concept of successful aging: importance of taking a life course perspective. Gerontologist. 2015;55(1):43–50. 10.1093/geront/gnu055.
9. Zhou Y, Sun Y, Pan Y, Dai Y, Xiao Y, Yu Y. Prevalence of successful aging in older adults: a systematic review and meta-analysis. Arch Gerontol Geriatr. 2025;128:105604. 10.1016/j.archger.2024.105604.
10. Sabatini S, Rupprecht F, Kaspar R, Wahl H-W, Diehl M, Clare L, et al. Successful aging and subjective aging: toward a framework to research a neglected connection. Gerontologist. 2025;65(1):gnae051. 10.1093/geront/gnae051.
11. Akinrolie O, Iwuagwu AO, Kalu ME, Rayner D, Oyinlola O, Ezulike CD, et al. Longitudinal studies of aging in Sub-Saharan Africa: review, limitations, and recommendations in preparation of projected aging population. Innov Aging. 2024;8(4):igae002. 10.1093/geroni/igae002.
12. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56. 10.1038/s41591-018-0300-7.
13. Wiemken TL, Kelley RR. Machine learning in epidemiology and health outcomes research. Annu Rev Public Health. 2020;41:21–36. 10.1146/annurev-publhealth-040119-094437.
14. Gašperlin Stepančič K, Ramovš A, Ramovš J, Košir A. A novel explainable machine learning-based healthy ageing scale. BMC Med Inf Decis Mak. 2024;24:317. 10.1186/s12911-024-02714-w.
15. Nopour R. Machine learning models in enhancing prediction of health-related indices among older adults: a scoping review. Heliyon. 2025;11(12):e43510. 10.1016/j.heliyon.2025.e43510.

16. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30. Red Hook (NY): Curran Associates; 2017. pp. 4765–74.
17. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. 10.1136/bmj-2023-078378.
18. Riumallo-Herl C, Canning D, Kabudula CW. Health inequalities in the South African elderly: the importance of the measure of social-economic status. J Econ Ageing. 2019;14:100191. 10.1016/j.jeoa.2019.01.005.
19. Peltzer K, Phaswana-Mafuya N. Depression and associated factors in older adults in South Africa. Glob Health Action. 2013;6(1):18871. 10.3402/gha.v6i0.18871.
20. Tomita A, Burns JK. Depression, disability and functional status among community-dwelling older adults in South Africa: evidence from the first South African National Income Dynamics Study. Int J Geriatr Psychiatry. 2013;28(12):1270–9. 10.1002/gps.3954.

Figures

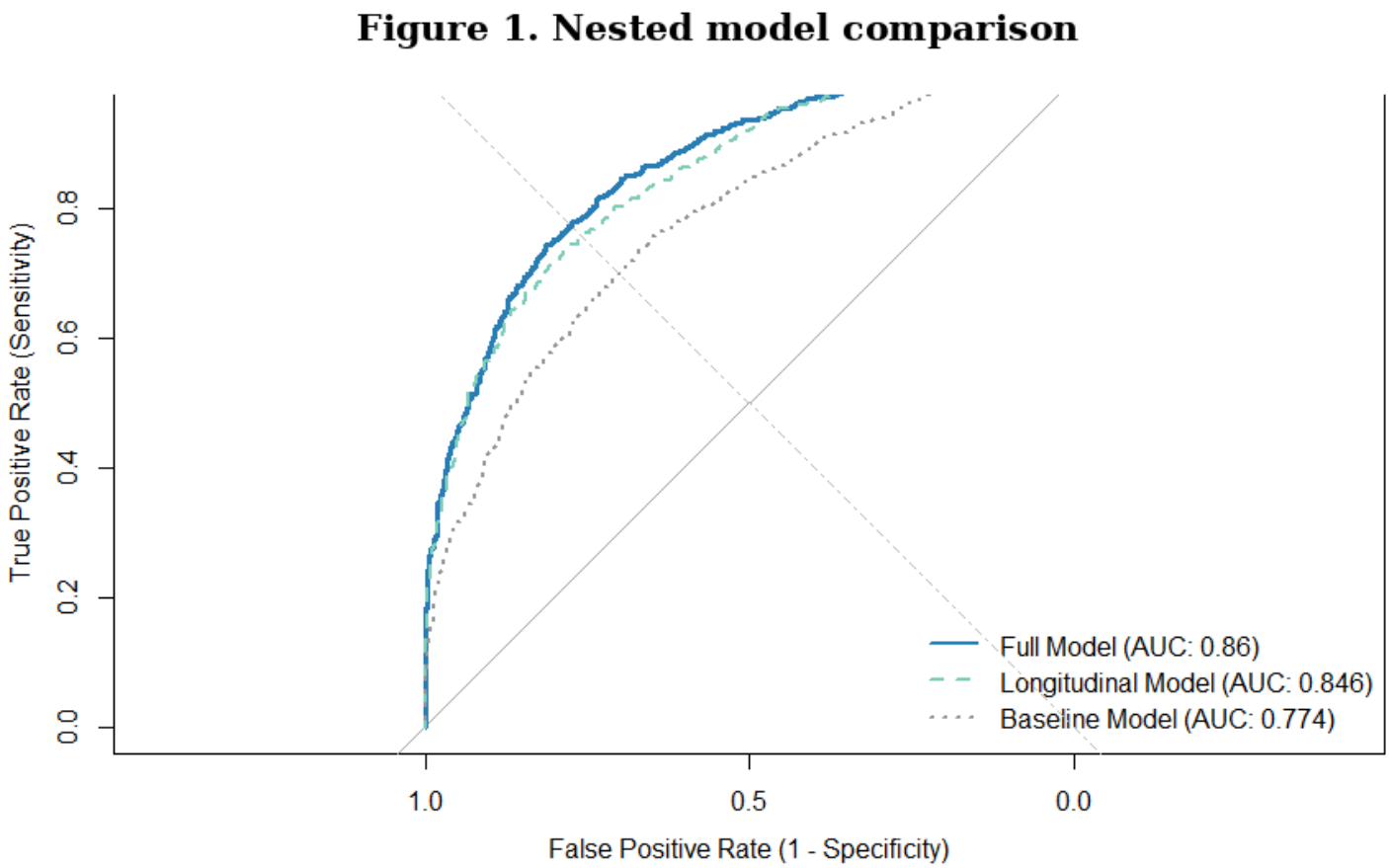


Figure 1

Nested model ROC comparison showing discrimination for baseline, longitudinal and full proximal models.

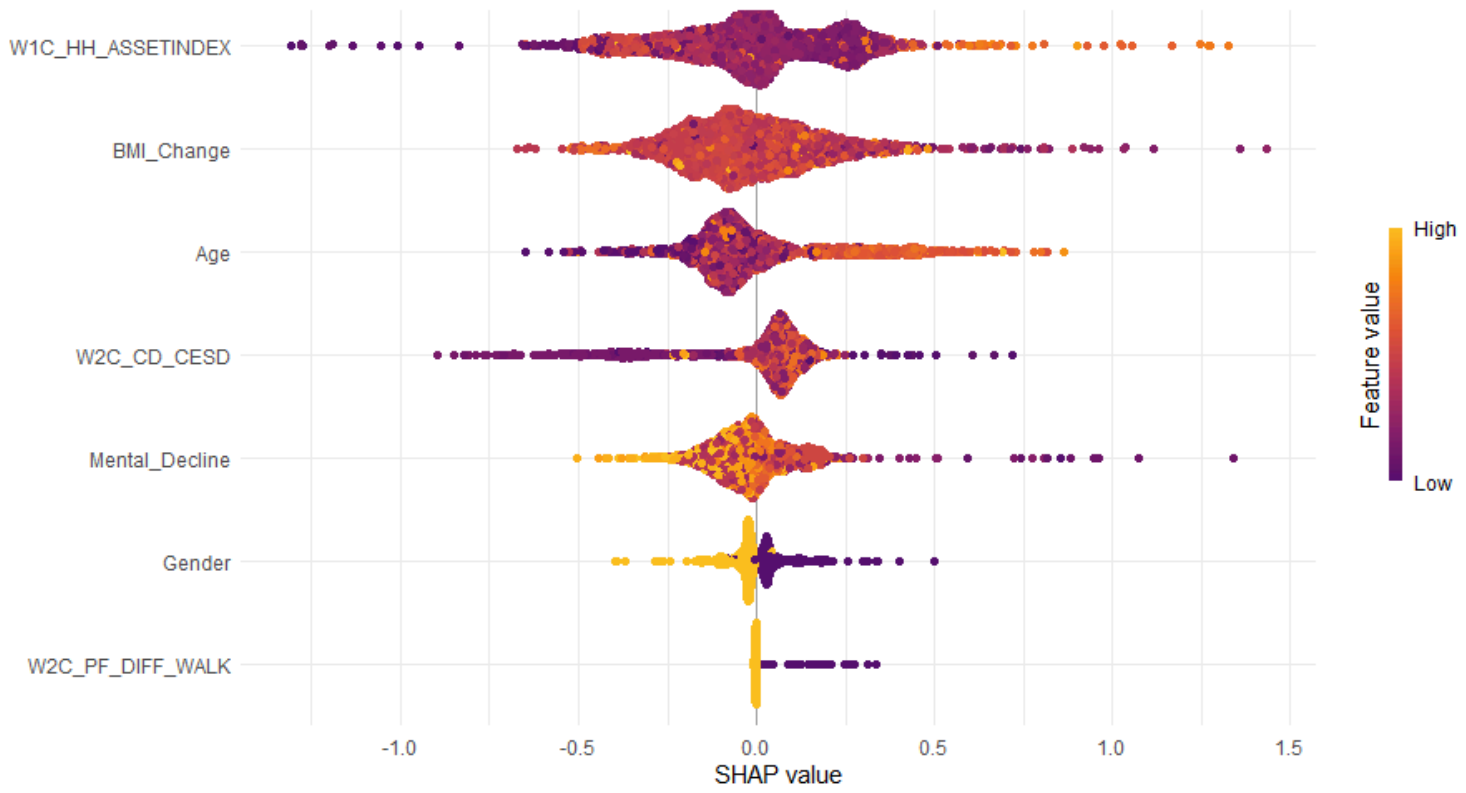


Figure 2

SHAP beeswarm plot showing the relative contribution and direction of major predictors in the final model.

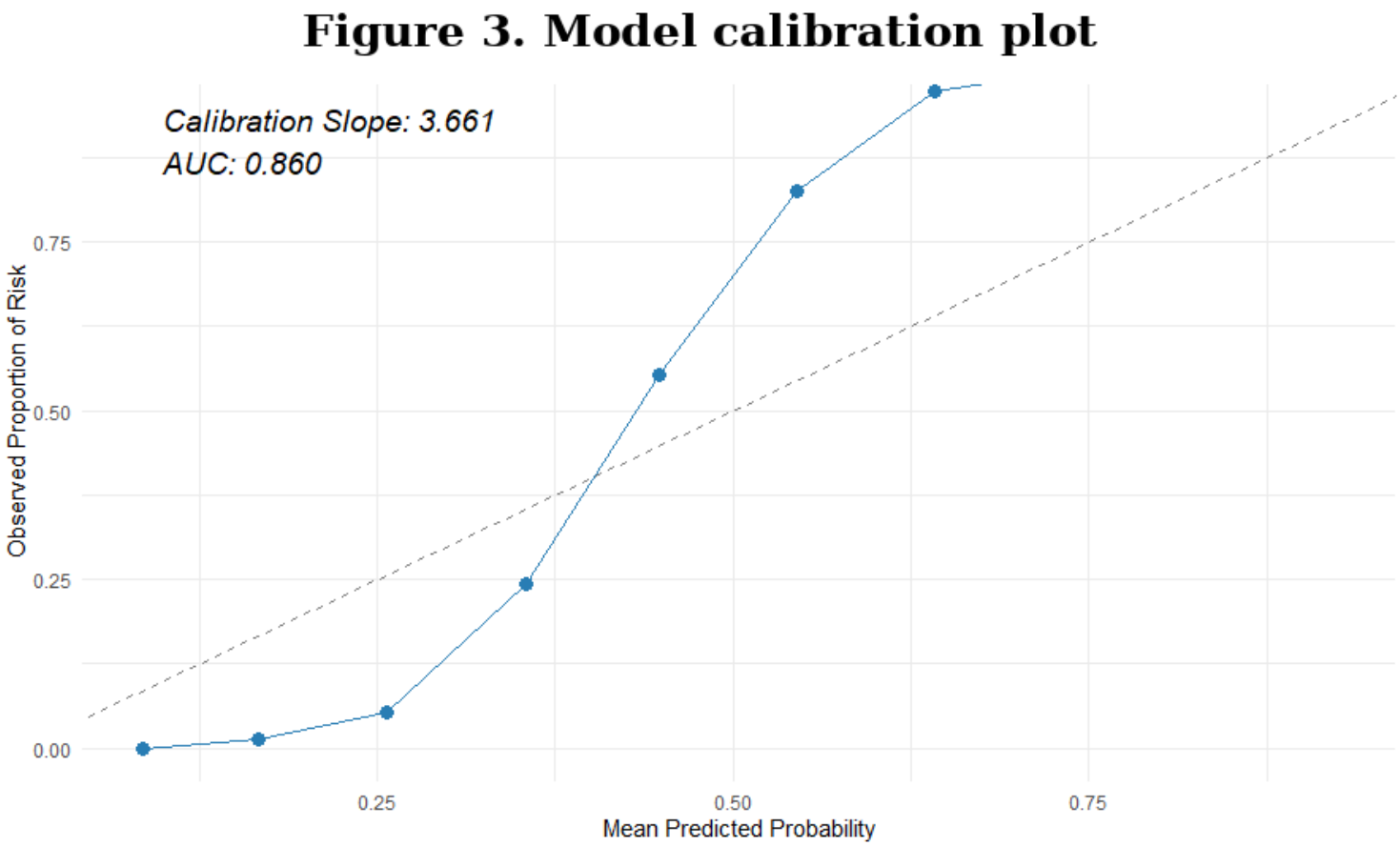


Figure 3

Calibration plot for the final model. Strong ranking performance is accompanied by compressed probability estimates.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TRIPODAIChecklist.docx