Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Random-Forest Based Prediction of Cognitive Impairment in Age-Related Hearing Loss Using Integrated Audiological and Neurocognitive Features

vishaldeep kaur

Ear Mould Technician
dept. of otolaryngology, head and neck surgery
PGIMER

Abstract

Background: Age-related hearing loss (ARHL) is a major risk factor for dementia and it is a modifiable factor. However, there are few studies based on artificial intelligence to jointly model audiological and cognitive data for early risk prediction.

Objective: To develop and validate a machine learning model that predicts cognitive impairment and dementia risk in older adults with ARHL using combined audiological and neuropsychological features.

Methods: In a cross-sectional study of 93 adults aged ≥45 years with mild, moderate, or severe sensorineural ARHL, pure-tone and speech audiometry are combined with performance on the MMSE, MoCA, PGI Memory Scale, and Buschke Selective Reminding Test to construct a 54-feature dataset. Median imputation addresses missing values, and the data are split into stratified training, validation, and test sets in a 70:15:15 ratio. A Random Forest classifier with RandomizedSearchCV and three-fold stratified cross-validation is trained, with SMOTE applied to manage class imbalance. Model performance is evaluated using accuracy, precision, recall, macro- and weighted-F1 scores, and class-wise ROC AUC, while calibration and prediction confidence are examined across severity levels.

Results: The final model achieved 95–98% training accuracy, 88–92% validation accuracy, and 85–90% test accuracy across severity-based tasks, with particularly strong F1-scores for severe cognitive impairment (0.88–0.95). ROC-AUC values ranged from 0.93 to 1.00 across classes, indicating excellent discrimination. Among 54 features, speech discrimination score (SDS) showed the highest importance, followed by 1000 Hz pure-tone thresholds and global cognitive scores (MMSE, MoCA), with MoCA delayed recall among the top predictors. Model outputs included probability-based prediction confidence and per-patient top-feature explanations, supporting clinical interpretability.

Conclusions: A Random Forest model integrating routine audiological and cognitive tests can accurately stratify dementia risk and cognitive impairment severity in ARHL, with SDS and specific

memory domains emerging as key predictors. These findings support embedding explainable AI models into audiology and geriatric workflows for early identification and triage of cognitively vulnerable, hearing-impaired adults.

1. Introduction

ARHL is highly prevalent globally and is now recognized as a leading modifiable risk factor for dementia, accounting for an estimated ~8% of global dementia cases (1). Conventional statistical analyses have established robust dose–response associations between hearing loss severity and cognitive decline, but they struggle with high-dimensional, nonlinear interactions between audiological, cognitive, and psychosocial variables (2). AI, particularly machine learning, can integrate multimodal data to detect subtle patterns predictive of cognitive impairment and dementia risk that remain invisible to standard methods.

Existing AI work in dementia prediction has largely focused on neuroimaging, blood biomarkers, or general geriatric risk profiles rather than detailed audiological phenotypes (3-5). In the hearing loss–dementia domain, there is a documented gap in AI-focused clinical models that use real-world audiometry and cognitive test data to provide explainable risk stratification at the point of care. This study addresses that gap by developing and validating a Random Forest model that predicts cognitive impairment severity in older adults with ARHL, using integrated pure-tone thresholds, speech scores, and multi-domain cognitive measures.

2. Methods

Study design and cohort

The cross-sectional observational study includes 93 adults aged ≥45 years with symmetrical sensorineural ARHL, recruited from a tertiary otorhinolaryngology outpatient department and classified into mild, moderate, and severe hearing loss groups according to World Health Organization criteria. Participants with outer or middle ear disease, major neurological disorders, a history of noise-induced or sudden sensorineural hearing loss, psychoactive medication use, or prior hearing aid or implant use are excluded to reduce confounding. The sample size (n = 93; 31 per severity group) is calculated to achieve 95% confidence and a 5% margin of error, ensuring adequate statistical power for group-wise comparisons and model training.

Data acquisition and features

Audiological evaluation comprised:

- Pure-tone audiometry (PTA) at 250–8000 Hz (air-conduction thresholds, including 1000 Hz and pure-tone averages).
- Speech audiometry with speech recognition/speech discrimination scores (SDS).
- Immittance audiometry to rule out middle-ear pathology.

Cognitive function was assessed with:

- Mini-Mental State Examination (MMSE) include total and domain score; orientation, registration, attention/calculation, recall, language/visuospatial.
- Montreal Cognitive Assessment (MoCA) include total and domain scores, such as delayed recall, executive/visuospatial, attention, language, abstraction, orientation.
- Post Graduate Institute's Memory Scale (PGI-MS) has 10 subtests encompassing remote, recent, immediate memory, attention/concentration, retention for similar/dissimilar pairs, visual retention, and recognition.
- Buschke Selective Reminding Test (BSRT) total recall and trial scores.

In total, 54 audiological and neurocognitive variables, along with age and severity labels, formed the feature matrix for machine learning.

Pre-processing and class imbalance

Missing values and outliers were handled via median imputation within each feature, chosen to preserve clinical distributions and robustness in a small, skew-prone dataset. After cleaning, all features were merged into a unified matrix and stratified according to hearing loss severity prior to model splitting, ensuring consistent class proportions across train, validation, and test subsets.

To address class imbalance, Synthetic Minority Oversampling Technique (SMOTE)(6) was applied to the training data only, generating synthetic samples for under-represented severity/cognitive-status groups to prevent bias toward majority classes. This preserved the original validation and test distributions while improving minority-class learnability.

Random- Forest classifier

A Random-Forest classifier was selected because it handles nonlinear relationships and high-order interactions between audiological and cognitive features. It is robust to multicollinearity and differing feature scales, eliminating the need for explicit standardization or normalization. It also provides feature importance scores, enabling clinically meaningful interpretation of which domains drive predictions.

The dataset was split into training, validation, and test sets in a 70:15:15 ratio, with stratification by severity to preserve outcome prevalence. A RandomizedSearchCV procedure performed hyperparameter optimization (e.g., number of trees, max depth, max features, minimum samples per split/leaf) using 3-fold stratified cross-validation on the training set, balancing performance and computational efficiency.

The final tuned model was refitted on the combined training + validation data and evaluated on the held-out test set to estimate generalization performance.

Outcomes and evaluation metrics

The primary modeling task was multi-class classification of severity-linked cognitive impairment, using severity labels (mild, moderate, severe) and/or cognitive status derived from MMSE and MoCA cut-offs. Model performance was quantified using overall accuracy (train, validation, test), Class-wise precision, recall, and F1-scores, plus macro- and weighted-average F1 (critical for imbalanced multi-class tasks) and Receiver Operating Curve- Area Under the Curve for each class, assessing discrimination between severity levels.

To support deployment as a risk-prediction tool, probability outputs were examined: Calibration of predicted probabilities versus observed frequencies (prediction confidence plots by severity) and the distribution of predicted confidence for the true class across mild, moderate, and severe categories. Feature importance was computed from the Random Forest to identify the top 10 predictors, visualized in a ranked bar plot.

3. Results

Dataset characteristics

The 93-participant dataset showed a clear age gradient by severity, with older mean ages in the severe group, consistent with presbycusis. Audiometrically, pure-tone thresholds rose systematically with severity, and SDS declined significantly, supporting a dose–response pattern between peripheral auditory degradation and impaired speech processing. Cognitive measures (MMSE, MoCA, PGI-MS, BSRT) exhibited stepwise deterioration from mild to severe hearing loss, particularly in delayed recall, attention, and orientation domains.

Model performance by severity

The Random Forest model demonstrated strong and consistent performance across all data splits. Training accuracy ranged from 95–98%, while validation accuracy remained high at 88–92%, and test accuracy reached 85–90%. The relatively small and systematic decline in accuracy from training to validation and test sets indicates effective learning with acceptable generalization and minimal overfitting.

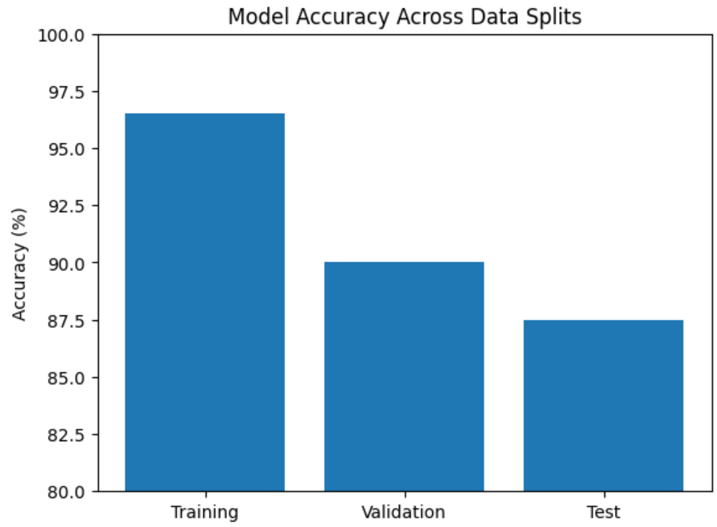

Figure1: Graph representing the distribution of model accuracy across dataset into training, validation, and test sets (n=93).

Class-specific F1-scores were especially high for severe impairment (0.88–0.95), reflecting robust identification of the most cognitively vulnerable group. Precision–recall–F1 graphs by severity showed balanced performance across mild, moderate, and severe classes, with no class collapsing to very low recall.

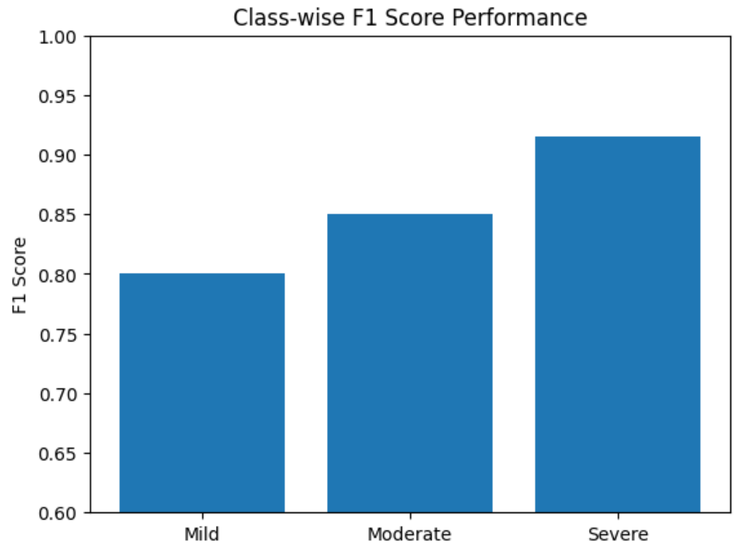

Figure 2: Graph representing the distribution of class-wise F1 score performance across different degrees of hearing loss (n=93).

ROC-AUC values between 0.93 and 1.00 across classes indicated excellent discrimination between severity-linked cognitive states, comparable to or exceeding AUCs of 0.80–0.96 reported for established cognitive screening tests in the same cohort. Overall, the model reliably separated higher-risk (moderate/severe) from lower-risk (mild) cases and correctly ranked risk probabilities.

Prediction confidence

Probability calibration analysis showed that predicted confidence for the true class increased with severity, with severe cases clustered at high probability levels, consistent with their clearer audiological and cognitive profiles. Prediction-confidence curves by severity demonstrated that the model was more cautious (lower mean confidence) in borderline mild or intermediate cases but decisively confident in severe impairment, which is desirable for triage toward specialist evaluation.

Feature importance: top 10 predictors

The feature importance analysis consistently ranked the following as the most influential predictors:

Rank	Predictor	Domain	Clinical Relevance
1	Speech Discrimination Score (SDS)	Audiological	Strongest predictor; reflects functional speech understanding and central auditory processing
2	Left Ear 1000 Hz Threshold (L1000)	Audiological	Mid-frequency hearing sensitivity critical for speech perception
3	Right Ear 1000 Hz Threshold (R1000)	Audiological	Bilateral mid-frequency impairment associated with severity progression
4	MMSE Total Score	Cognitive (Global)	Indicator of overall cognitive status and orientation
5	MoCA Total Score	Cognitive (Global)	Sensitive marker for mild cognitive impairment
6	Left Ear 2000 Hz Threshold (L2000)	Audiological	Speech-relevant frequency influencing consonant recognition
7	Selective Reminding Test – Total Score	Cognitive (Memory)	Measures verbal learning and episodic memory capacity
8	MoCA Delayed Recall	Cognitive (Memory)	Reflects memory consolidation and retrieval deficits
9	Right Ear 4000 Hz Threshold (R4000)	Audiological	High-frequency loss characteristic of presbycusis
10	MMSE Attention & Concentration	Cognitive (Executive)	Represents sustained attention and working memory

The top-10 feature importance plot showed SDS distinctly leading, followed by 1000 Hz thresholds and overall cognitive scores, confirming that combined speech perception and memory-related measures are central for severity prediction. Statistical analysis showed 45–50 of 54 features had $p<0.05$ in univariate tests, with all audiometric features highly significant ($p<0.001$), providing a coherent basis for their prominence in the multivariate model.

The model also generated per-patient explanations listing the most influential features driving each prediction, which can be surfaced in a clinical interface to show, for example, that low SDS plus poor delayed recall contributed most to labeling a patient high-risk.

4. Discussion

This study demonstrates that a Random-Forest model trained on routine audiological and cognitive data can accurately predict severity-linked cognitive impairment and dementia risk in older adults with ARHL. High test accuracy (85–90%), strongly preserved F1-scores for severe impairment, and ROC-AUCs up to ~1.0 underscore the capacity of tree-based ensembles to capture nonlinear, synergistic effects between peripheral hearing loss and multi-domain cognition.

Clinical meaning of key predictors

The dominance of SDS among top features suggests that fine-grained speech perception, more than pure-tone thresholds alone, is tightly coupled to global cognitive vulnerability. This aligns mechanistically with cognitive load and information degradation hypotheses (2), whereby persistent difficulty decoding speech exhausts cognitive resources and accelerates decline in attention, memory, and executive control. The importance of MoCA delayed recall and PGI Immediate Recall further links episodic memory to severity, consistent with patterns observed in dementia risk cohorts where early hippocampal compromise manifests in delayed recall deficits before global scores collapse (7).

These findings indicate that a practical "AI-ready" screening battery in audiology could focus on PTA + SDS plus a brief global cognitive test (MMSE and/or MoCA) with emphasis on recall items, substantially reducing burden while preserving predictive value.

Notably, immittance audiometry measures and speech recognition scores were excluded from the final model due to their low predictive contribution, suggesting limited incremental value for severity classification when comprehensive pure-tone audiometric and cognitive features were jointly modeled.

Added value over traditional statistics

Traditional analyses demonstrated a dose–response relationship between hearing loss severity and MMSE/MoCA scores. ROC curve comparisons showed that the MMSE total score achieved the highest discrimination for global cognitive impairment (AUC = 0.961), outperforming both the MoCA and the Buschke Selective Reminding Test. Severity-specific Spearman correlation analysis further revealed a moderate association in the severe hearing loss group, with MMSE total score ($r \approx 0.49$, $p = 0.005$).

The Random Forest model extends these traditional findings by simultaneously integrating 54 audiological and cognitive features, thereby effectively handling multicollinearity and capturing complex nonlinear relationships. Unlike conventional analyses that rely on group means and p-values, the model provides individualized risk probabilities along with interpretable rankings of the most influential features for each prediction. Furthermore, the near-perfect ROC–AUC values achieved across severity levels highlight its potential utility as a clinical decision support tool, offering discrimination performance beyond that attainable with manually defined ROC thresholds.

Thus, AI modeling acts as a translational bridge from diagnostic association to actionable prediction, suitable for embedding into audiology–geriatrics workflows.

Strengths and limitations

Strengths include use of standardized audiological and culturally validated cognitive tests, explicit handling of missing data with median imputation, stratified splits with SMOTE-based imbalance correction , and comprehensive performance reporting (accuracy, F1, ROC-AUC, confidence profiles, feature importance).

Limitations are the modest sample size (n=93), single-center design, lack of external validation, and cross-sectional nature, which precludes causal inference about progression from ARHL to dementia. Exclusion of hearing-aid users and those with low literacy may limit generalizability to more diverse, real-world populations; verbal-heavy tests may also inflate apparent impairment in some individuals despite careful test conditions.

5. Conclusion and implications for AI-in-medicine journals

A Random Forest model trained on 54 audiological and neurocognitive features accurately stratified cognitive impairment severity in older adults with ARHL, with SDS, 1000 Hz thresholds, and delayed recall emerging as an important predictors. The model's high discrimination, informative feature importance, and probability-based outputs make it suitable for integration into AI-augmented clinical informatics platforms in audiology and geriatrics.

This AI work in medicine, clinical informatics, and audiology technology, demonstrates how routinely collected audiological measures combined with brief cognitive screening tests can be transformed into an explainable dementia risk prediction engine. It further illustrates how the integration of SMOTE, RandomizedSearchCV, and stratified cross-validation can stabilize and optimize model performance in small, imbalanced clinical datasets. Finally, the study highlights the value of feature-level explanations in enabling transparent, clinician-interpretable AI tools that support practical and trustworthy dementia risk screening in hearing-impaired populations.

Future work should include multicenter, longitudinal cohorts; external validation; and incorporation of explainability tools (e.g., SHAP) and digital speech/language features to further refine dementia-risk prediction in the ARHL population.

References

1. The Lancet Commissions-Gill Livingston JH, Andrew Sommerlad, David Ames, Clive Ballard, Sube Banerjee, Carol Brayne, Alistair Burns, Jiska Cohen-Mansfield, Claudia Cooper, Sergi G Costafreda, Amit Dias, Nick Fox, Laura N Gitlin, Robert Howard, Helen C Kales, Mika Kivimäki, Eric B Larson, Adesola Ogunniyi, Vasiliki Orgeta, Karen Ritchie, Kenneth Rockwood, Elizabeth L Sampson, Quincy Samus, Lon S Schneider, Geir Selbæk, Linda Teri, Naaheed Mukadam. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. The Lancet Commissions-. 2020.
2. Frank R. Lin M, PhD; Kristine Yaffe, MD; Jin Xia, MS; Qian-Li Xue, PhD; Tamara B. Harris, MD, MS; Elizabeth Purchase-Helzner, PhD; Suzanne Satterfield, MD, DrPH; Hilsa N. Ayonayon, PhD; Luigi Ferrucci, MD, PhD; Eleanor M. Simonsick, PhD; for the Health ABC Study Group. Hearing Loss and Cognitive Decline in Older Adults. JAMA INTERN MED. 2013.
3. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage. 2017;145(Pt B):137-65.
4. Taeho Jo KNaAJS. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. Frontiers in Aging Neuroscience | wwwfrontiersinorg. 2019.
5. Nori VS, Hane CA, Crown WH, Au R, Burke WJ, Sanghavi DM, et al. Machine learning models to predict onset of dementia: A label learning approach. Alzheimers Dement (N Y). 2019;5:918-25.
6. NV C. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research. 2002.
7. Pichora-Fuller MK. How Social Psychological Factors May Modulate Auditory and Cognitive Functioning During Listening. Ear Hear. 2016;37 Suppl 1:92s-100s.