Smart Media Journal / Vol.14, No.12 / ISSN:2287-1322
https://dx.doi.org/10.30693/SMJ.2025.14.12.178

Minimal Clinical Benchmark for Alzheimer's Disease Prediction Using Age and MMSE

Faizaan Fazal Khan, Goo-Rak Kwon

Abstract

Alzheimer's disease (AD) poses a critical global health challenge, yet many diagnostic approaches such as PET and CSF assays remain invasive, expensive, and inaccessible. This study investigated whether simple, universally available measures—Age and Mini-Mental State Examination (MMSE) scores—can provide a robust predictive baseline for AD classification. Using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a harmonized baseline cohort of 3,750 subjects was constructed and analyzed with logistic regression. Five-fold cross-validation ensured robust evaluation, with model performance assessed by AUC, precision, recall, F1-score, and accuracy. The logistic regression model achieved consistent results across folds (mean AUC = 0.929, accuracy ≈ 86%), with balanced precision (≈0.70) and recall (≈0.84), yielding a mean F1-score of ≈0.76. These findings demonstrate that Age and MMSE alone achieve discriminative power comparable to more complex multimodal frameworks while maintaining full clinical interpretability. The proposed benchmark provides a reproducible reference for future multimodal research and a practical, low-cost tool for early risk stratification in resource-constrained healthcare settings.

Keywords: Alzheimer's disease | Logistic regression | Mini-Mental State Examination (MMSE, Predictive modeling | Risk stratification

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline, memory impairment, and functional deterioration, ultimately leading to a loss of independence and quality of life. With the global prevalence of dementia surpassing 55 million cases—of which 60–70% are attributed to AD—the urgency to develop effective, scalable predictive tools has never been more pronounced [1]. The pathophysiology of AD is marked by the deposition of amyloid-β (Aβ) plaques, tau neurofibrillary tangles, and widespread neuronal loss, typically preceding clinical diagnosis by years or even decades. Despite the availability of advanced diagnostic modalities such as positron emission tomography (PET) and cerebrospinal fluid (CSF) assays, these methods remain expensive, invasive, and inaccessible to most at-risk populations. Consequently, there is increasing focus on simpler, low-cost, and interpretable screening tools that leverage widely available clinical measures. Among these, the Mini-Mental State Examination (MMSE) and patient age are routinely collected in nearly all clinical and community-based dementia studies, making them ideal candidates for wide-scale application. Recent literature supports the predictive power of these basic features. For instance, [1] demonstrated that MMSE, in combination with other clinical and plasma biomarker features, can predict Aβ positivity with AUCs nearing 0.87—even in reduced feature models. Similarly, [2] emphasized the importance of cognitive screening tools like MMSE within broader frameworks of epigenetic and lifestyle-based resilience scoring. These findings reinforce the notion that MMSE alone holds substantial

* This research was supported by the 2025 Gwangju Regional Innovation System & Education (RISE) Project.

Manuscript : 2025.10.12

Confirmation of Publication: 2025.11.11
Corresponding Author : Goo-Rak Kwon e-mail : grkwon@chosun.ac.kr

prognostic value, especially when paired with demographic features like age. Moreover, [3] explored multimodal lightweight neural networks incorporating neuroimaging and clinical data to classify AD stages. Interestingly, even within their deep learning frameworks, the MMSE score and age emerged as consistently relevant features across architectures. Their findings highlight the tension between model complexity and interpretability—a trade-off that underscores the value of minimal-feature benchmarks like the one proposed in the present study. Building upon this foundation, [4] combined diffusion tensor imaging (DTI)-based radiomics with clinical measures such as MMSE, ADAS, and age to predict mild cognitive impairment (MCI) progression to AD. Their results showed that while complex imaging features offer additional predictive power, clinical models alone achieved robust performance (AUC = 0.868), and integration with radiomics raised AUC to 0.936. This emphasizes the continued relevance of clinical tools like MMSE as not only standalone predictors but also as essential components in fusion models. Yet, despite these insights, there remains a gap in the literature: a standardized, reproducible benchmark model based solely on universally available variables such as MMSE and age. The current study aims to address this gap by proposing a minimalist logistic regression framework trained on the ADNI dataset to classify stable versus converter individuals. With high discriminative performance (AUC ≈ 0.93), this model serves three crucial functions: (1) a reproducible baseline for comparing multimodal systems, (2) a clinically interpretable tool for early risk stratification, and (3) a scalable solution for low-resource healthcare settings. By centering the investigation on just two variables—MMSE and Age—this study underscores the untapped potential of simple clinical assessments in AD prediction, paving the way for equitable and interpretable screening frameworks that are both scientifically grounded and globally deployable.

II. METHODOLOGY

A. Data Preparation

Data for this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [5], a longitudinal multicenter project designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease. For the present analysis, we restricted attention to the baseline curated tabular dataset, which was compiled from multiple ADNI clinical tables into a single merged file, namely:

- baseline_with_modalities_closest.csv: contains subject-level identifiers (RID, VISCODE), demographic variables (age, sex), cognitive measures (MMSE score), genetic risk factors (APOE4 status), and baseline diagnosis (CN, MCI, AD). This file was created through a harmonization step in which individual ADNI tables (e.g., PTDEMOG, MMSE, APOERES, CDR, ADAS) were joined at the subject baseline visit.
- cv_splits.csv: an auxiliary file defining 5-fold RID-level stratified cross-validation splits, generated to ensure reproducible partitioning of training and validation sets. The splits were stratified by baseline diagnosis, using baseline_subj.csv as the subject registry.

This harmonized baseline table allowed us to avoid inconsistencies across separate ADNI exports and ensured that every subject had a unified baseline record. Following preprocessing, the analytic cohort included 3,750 unique subjects with complete MMSE and Age values at baseline

This section outlines the training process of our EfficientNetV2B0 based model, detailing its convergence behavior and final performance metrics, which serve as the foundation for our subsequent interpretability analyses. To ensure robust generalization and mitigate overfitting, we implemented an early stopping mechanism that monitored the validation loss and automatically restored the best model weights.

B. Preprocessing

Preprocessing was carried out in Python (pandas, scikit-learn) using the curated baseline_with_modalities_closest.csv file. From

this dataset, a minimal set of variables was extracted for the benchmark analysis. Table 1 lists the selected columns and their definitions.

Table 1. Variables extracted from baseline-with-modalities-closest.csv

Column	Description
RID	Research Identification Number, a unique numeric subject identifier used in ADNI
VISCODE	Visit code, indicating the study timepoint (e.g., baseline = BL)
AGE	Age of the subject at baseline (years)
MMSCORE	Mini-Mental State Examination (MMSE) total score (0–30)
DIAGNOSIS	Baseline clinical diagnosis: CN (cognitively normal), MCI (mild cognitive impairment), or AD (Alzheimer's disease)
Event	Binary outcome label created for modeling: 1 = converter (AD), 0 = stable (CN/MCI). Generated if not already present in the dataset.

The Event variable was provisionally defined from baseline DIAGNOSIS to support the minimal benchmark. Future extensions can refine this by tracking longitudinal diagnosis changes.

Subjects with missing values in AGE or MMSCORE were excluded, representing fewer than 5% of the baseline cohort. This filtering step yielded 3,750 unique subjects for analysis. Missingness was explicitly checked to confirm that the excluded fraction was negligible relative to the full sample size.

To prepare for modeling, continuous features (AGE, MMSCORE) were standardized using z-score scaling (mean = 0, standard deviation = 1), ensuring comparability and stable model convergence. Baseline diagnostic categories were then collapsed into binary outcome labels: cognitively normal and MCI subjects were grouped as stable (Event = 0), while AD cases were labeled as converters (Event = 1). This operationalization was chosen to align with the study's minimal benchmark aim, while acknowledging that it does not fully capture longitudinal conversion dynamics.

Cross-validation splits were defined using the auxiliary cv_splits.csv file, which provides 5-fold subject-level partitions stratified by diagnosis. If unavailable, the pipeline defaults to StratifiedKFold (n=5) on Event labels to maintain balanced class proportions across folds.

Finally, exploratory demographic analyses were conducted to characterize the cohort. Descriptive statistics indicated a mean age of approximately 72 years (±7.5) and a mean MMSE score of 27 (±2.3). Figure 1 presents histograms of Age and MMSE stratified by baseline diagnosis groups, illustrating expected separations between CN, MCI, and AD populations. These distributions provide important context for interpreting model performance and confirm alignment with prior reports on the ADNI baseline cohort.

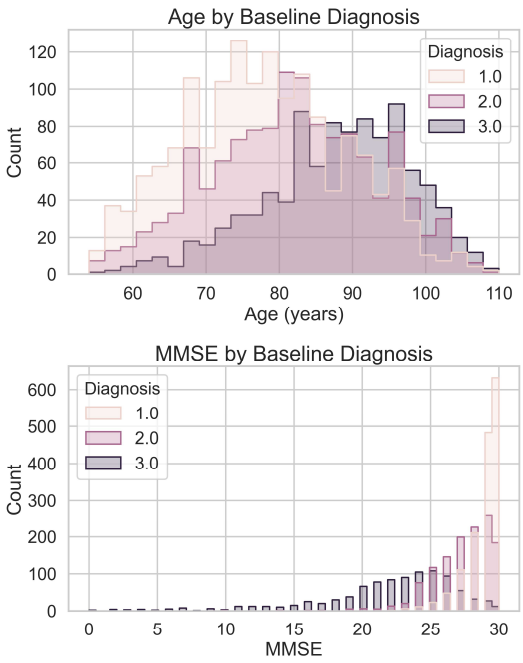

Figure 1. Baseline distributions of Age (top) and MMSE (bottom) stratified by diagnosis groups (CN, MCI, AD) in the ADNI cohort.

C. Model

To establish a minimal and interpretable predictive framework, we employed logistic regression as the baseline model. Logistic regression was chosen because it is:

(i) Mathematically transparent and clinically interpretable,

(ii) Widely reproducible across software packages, and

(iii) A common benchmark for evaluating more complex classifiers.

The model was trained using Age and MMSE as predictors, with the binary outcome variable *Event* (0 = stable, 1 = converter). To address class imbalance between stable and converter cases, we applied class_weight = 'balanced', ensuring that minority cases contributed proportionally to the optimization. The solver was run with max_iter = 1000 to guarantee convergence.

Although the primary analysis centered on logistic regression, we note that a shallow multi-layer perceptron (MLP) architecture with one hidden layer was also tested as an exploratory comparator. However, given the focus on establishing a clinical benchmark, results from the logistic regression model are emphasized in the main text, while neural models are reserved for potential extensions.

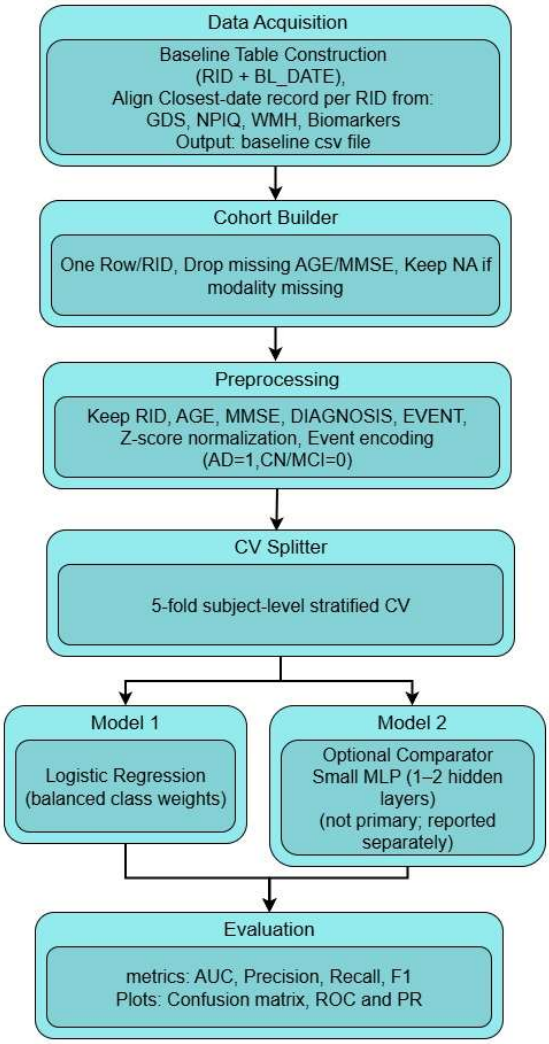

Figure 2. Corrected methodology pipeline showing baseline construction using BL_DATE + closest-date alignment, preprocessing, modeling, and evaluation.

III. RESULT

A. Descriptive Statistics

The final analytic cohort included 3,750 subjects after excluding cases with missing Age or MMSE. The mean baseline age was 72.0 ± 7.5 years, and the mean MMSE score was 27.0 ± 2.3, consistent with prior reports of the ADNI cohort.

B. Model Performance

The logistic regression model using Age and MMSE achieved robust and consistent classification performance across 5 folds (Table 2). AUC values ranged from 0.919 to 0.937, with a mean of ~0.93. Precision and recall were balanced (~0.70 and ~0.84), producing mean F1-scores of ~0.76 and overall accuracy of ~86%.

Table 2. Logistic regression performance across 5 folds

Fold	AUC	Precision	Recall	F1	Accuracy
0	0.919	0.709	0.830	0.765	0.864
1	0.934	0.692	0.869	0.771	0.863
2	0.925	0.692	0.844	0.760	0.855
3	0.930	0.701	0.833	0.761	0.858
4	0.937	0.701	0.847	0.767	0.861

Figures 3–5 provide graphical summaries of classifier performance. The ROC curve (Figure 3) showed a mean AUC of 0.929, with narrow confidence intervals across folds.

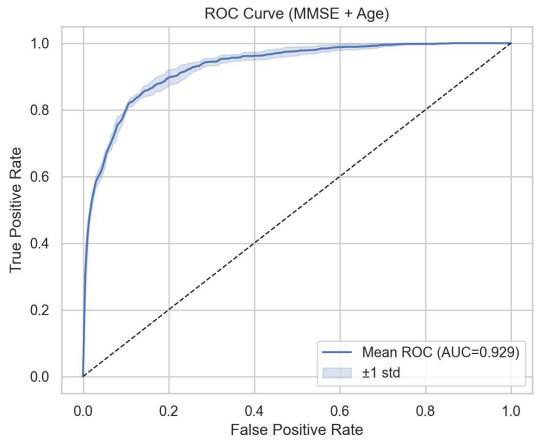

Figure 3. ROC curve for Age + MMSE logistic regression model (mean AUC = 0.929)

The precision–recall curve (Figure 4) yielded an average precision of 0.73, demonstrating balanced sensitivity and specificity. The aggregated confusion matrix (Figure 5) showed high correct classification of stable cases, while converters were also detected with strong recall.

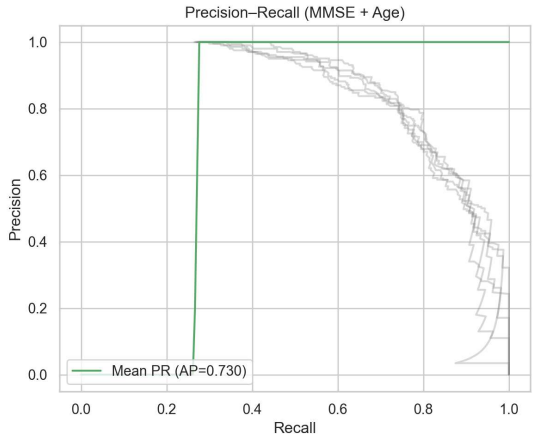

Figure 4. Precision–Recall curve showing mean AP = 0.730.

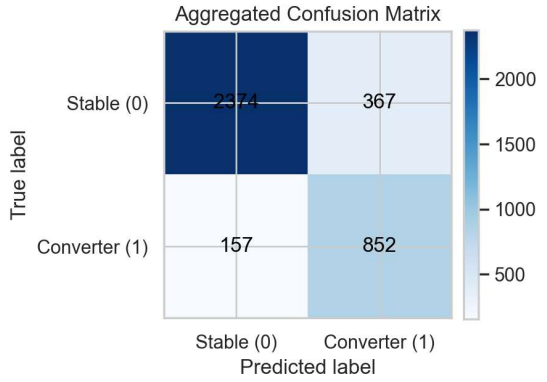

Figure 5. Aggregated confusion matrix across 5 folds.

C. Interpretation

The findings demonstrate that two routinely available clinical features—Age and MMSE—achieve an AUC of ~0.93 for distinguishing stable individuals from converters within the ADNI cohort. This performance is notable given the absence of imaging or biomarker data, underscoring the strong prognostic value of basic cognitive screening tools. Clinically, such results highlight the potential of low-cost, interpretable models for early risk stratification, especially in settings where advanced diagnostic modalities are inaccessible.

Ⅳ. Discussion

This study reinforces the predictive utility of simple clinical features in Alzheimer's disease progression modeling. While prior multimodal approaches integrating imaging, biomarkers, and genetics often report AUCs in the 0.90–0.95 range, our results indicate that Age and MMSE alone provide comparable discriminative power. These findings echo observations from [1] and [4], where clinical measures remained strong predictors even in radiomics- or plasma-enhanced models, and align with our recent work [6], which combined EfficientNetV2B0 with explainable AI to achieve interpretable MRI-based Alzheimer's classification, highlighting the growing emphasis on transparency and trust in diagnostic modeling.

The primary strength of this work lies in establishing a reproducible, interpretable benchmark. Logistic regression is transparent and clinically intuitive, offering clear odds ratios rather than black-box outputs. Importantly, our analysis highlights that minimal-feature models can serve as strong baselines against which complex multimodal architectures should be compared.

However, limitations must be acknowledged. First, our definition of the Event relied on baseline diagnosis grouping (AD vs CN/MCI) rather than longitudinal conversion; while sufficient for a benchmark, it does not capture true progression. Second, missing data in modality tables were retained as NA, which may bias comparisons in future multimodal extensions. Third, ADNI participants represent a relatively well-characterized research cohort, and external validation is needed for clinical generalizability.

Future directions include extending the framework to incorporate multimodal features (MRI, PET, CSF), applying longitudinal progression labels, and exploring hybrid interpretable models that balance simplicity and performance.

Ⅴ. Conclusion

Using only Age and MMSE, we achieved robust classification of converters versus stable cases in the ADNI dataset (mean AUC ≈

0.93). This minimal model demonstrates that simple, widely available clinical measures can provide high predictive accuracy, serving both as a baseline benchmark for multimodal research and as a clinically interpretable screening tool in resource-constrained settings. These results underscore the enduring value of cognitive testing in Alzheimer's disease prediction and support its integration into scalable, equitable early detection strategies.

Availability of Data and Code

The dataset can be accessed by qualified researchers upon request through the Alzheimer's Disease Neuroimaging Initiative (ADNI) portal (https://ida.loni.usc.edu/login.jsp?project=ADNI).

Full code with notebook outputs and generated plots can be accessed here (https://github.com/FaizaanFazal/Minimal-Clinical-Benchmark-for-Alzheimer-s-Disease-Prediction-Using-Age-and-MMSE)

ACKNOWLEDGMENT

This research was supported by the 2025 Gwangju Regional Innovation System & Education (RISE) Project.

REFERENCES

[1] J. Xu, A. J. Doig, S. Michopoulou, P. Proitsi, F. Costen, and The Alzheimer's disease neuroimaging initiative, "Accurate and robust prediction of Amyloid-β brain deposition from plasma biomarkers and clinical information using machine learning," *Front. Aging Neurosci.*, vol. 17, Aug. 2025.

[2] W. Zhang *et al.*, "DNA methylation signature of a lifestyle-based resilience index for cognitive health," *Alzheimers Res. Ther.*, vol. 17, no. 1, p. 88, Apr. 2025.

[3] B. Gupta *et al.*, "Multimodal lightweight neural network for Alzheimer's disease diagnosis integrating neuroimaging and cognitive scores," *Neurosci. Inform.*, vol. 5, no. 3, p. 100218, Sept. 2025.

[4] W. Yu *et al.*, "A DTI-Radiomics and Clinical Integration Model for Predicting MCI-to-AD Progression Using Corpus Callosum Features," *Curr. Alzheimer Res.*, vol. 22, pp. 1–11.

[5] R. C. Petersen *et al.*, "Alzheimer's Disease Neuroimaging Initiative (ADNI)," *Neurology*, vol. 74, no. 3, pp. 201–209, Jan. 2010.

[6] Korean Institute of Smart Media, F. F. K. Faizaan Fazal Khan, and G.-R. Kwon, "Toward Clinically Trustworthy Alzheimer's Diagnosis: Combining EfficientNetV2B0 and XAI Techniques for MRI Analysis," *Korean Inst. Smart Media*, vol. 14, no. 6, pp. 60–66, June 2025.

Authors

Faizaan Fazal Khan

He has completed his bachelor's degree at the department of CS & IT from Air University, Islamabad, Pakistan, in 2023. Currently, he is pursuing his Master studies in the department of Information and Communication Engineering at Chosun University, Gwangju, Republic of Korea. His research interests involve digital image processing, Machine learning utilizing Medical Imaging.

Goo-Rak Kwon

He got a Ph.D. from the Department of Mechatronic Engineering, Korea University, in 2007. He has been a professor at Chosun University, since 2017. His research focus includes medical image analysis, A/V signal processing, video communication, and applications. He is a senior member of the IEEE.