Dement Neurocogn Disord. 2026 Apr;25(2):91-102
https://doi.org/10.12779/dnd.2026.25.2.91
pISSN 1738-1495·eISSN 2384-0757

DND Dementia and Neurocognitive Disorder

Original Article

AI-Based Dementia Prediction From Low-Dose Brain CT in Mild Cognitive Impairments

Seongbeom Park,[1] **Jehyun Ahn**,[2,3] **Duk L. Na**,[2,3] **Kichang Kwak**,[1] **Jun Pyo Kim**,[2,3] **On behalf of the K-ROAD study**

[1]BeauBrain Healthcare, Inc., Seoul, Korea
[2]Alzheimer's Disease Convergence Research Center, Samsung Medical Center, Seoul, Korea
[3]Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Received: Feb 23, 2026
Revised: Mar 31, 2026
Accepted: Apr 1, 2026
Published online: Apr 16, 2026

Correspondence to
Kichang Kwak
BeauBrain Healthcare, Inc., 314 Hakdong-ro, Gangnam-gu, Seoul 06098, Korea.
Email: kichang.kwak@beaubrain.bio

Jun Pyo Kim
Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea.
Email: torch0703@gmail.com

ORCID iDs
Seongbeom Park
https://orcid.org/0000-0002-0759-6826
Jehyun Ahn
https://orcid.org/0009-0008-0927-6069
Duk L. Na
https://orcid.org/0000-0002-0098-7592
Kichang Kwak
https://orcid.org/0000-0003-4542-0001

ABSTRACT

Background and Purpose: Magnetic resonance imaging (MRI) is often limited by safety concerns and long acquisition times in elderly patients. We evaluated the potential clinical utility of an artificial intelligence (AI) model, Morph Computed Tomography (CT) Plus, to predict three-year dementia conversion in mild cognitive impairment (MCI) using rapid low-dose CT scans, prioritizing its role as a safe and efficient triage tool.
Methods: We analyzed 896 MCI (Clinical Dementia Rating [CDR]=0.5) patients from the Korea-Registries to Overcome Dementia and Accelerate Dementia Research cohort. A deep learning model quantified regional brain atrophy. We implemented a dual-cutoff risk-stratification strategy to optimize clinical workflow, comparing 90% and 95% sensitivity/specificity operating targets.
Results: The model achieved an area under the curve of 0.870. The 90% strategy was superior for clinical triage, limiting the intermediate-risk group to 27.3% (n=99) compared to 49.6% (n=180) in the 95% strategy. For the combined high- and low-risk groups, the 90% strategy yielded an overall accuracy of 87.5%, with a positive predictive value of 83.3% and a negative predictive value of 90.4%. High-risk patients demonstrated a 14.13-fold higher hazard of conversion (p<0.001).
Conclusions: AI-driven CT analysis provides a high-precision, inclusive triage tool. The 90% strategy offers a scalable solution for early dementia management, especially for patients with MRI contraindications, demonstrating significant potential clinical utility.

Keywords: Mild Cognitive Impairment; Dementia; Artificial Intelligence; Tomography, X-ray Computed; Health Technology Assessment

INTRODUCTION

As South Korea rapidly transitions toward a super-aged society, projected to occur by 2025, the socioeconomic burden of dementia is becoming the nation's most critical public health challenge.[1,2] The current paradigm for dementia screening relies heavily on magnetic

Jun Pyo Kim
https://orcid.org/0000-0003-4376-3107

Conflict of Interest
Seongbeom Park and Kichang Kwak were employed by BeauBrain Healthcare Inc. The other authors declare that they have no competing interests.

Autor Contributions
Conceptualization: Park S, Ahn J, Na DL, Kim JP, Kwak K; Data curation; Ahn J, Na DL, Kim JP; Investigation: Ahn J; Methodology: Park S; Supervision: Kwak K; Writing - original draft: Park S; Writing - review & editing: Kim JP, Kwak K.

resonance imaging (MRI) and amyloid positron emission tomography (PET). However, the clinical application of MRI in the elderly population—the demographic at the highest risk—is frequently hindered by a variety of logistical and physiological factors. Standard MRI protocols often require 20 to 40 minutes of total stillness, a requirement that is often intolerable for frail or cognitively impaired geriatric patients. The loud noise and enclosed space of MRI units frequently trigger severe agitation, claustrophobia, or panic, leading to motion artifacts that compromise image quality or result in complete scan failures.[3]

Furthermore, safety contraindications create a significant "diagnostic void." A substantial portion of the elderly population has implantable medical devices, such as cardiac pacemakers, neurostimulators, or metallic fragments, which completely preclude MRI screening.[4] This means that the most vulnerable individuals are often left without access to early prognostic insights. While amyloid PET offers gold-standard pathology data, its high cost and limited geographic accessibility in primary care and community-based settings further exacerbate the "diagnostic gap," preventing timely intervention for millions.[5,6]

Computed tomography (CT) offers an inclusive, rapid, and universally available alternative. A typical brain CT scan can be completed in less than five minutes and is generally better tolerated by patients with mobility issues or respiratory distress.[7,8] Despite this ubiquity, CT has been historically underutilized for neurodegenerative prognosis because the human eye often fails to detect the subtle, regional brain atrophy patterns characteristic of early-stage dementia.[9] Recognizing this limitation, the Health Insurance Review and Assessment Service (HIRA) in Korea has recently refined its evaluation guidelines. Specifically, HIRA has identified artificial intelligence (AI)-based imaging technologies that derive novel prognostic biomarkers—information that is clinically essential but cannot be identified through traditional human interpretation—as a representative example of Category D medical technology.[10]

This classification underscores the national policy importance of developing AI solutions that can repurpose accessible imaging data into high-level evidence for disease progression. One such example highlighted by regulatory authorities is the use of brain CT to accurately predict the prognosis of neurological diseases.[10] Such technologies are designated as suitable candidates for New Health Technology Assessment due to their potential to provide information beyond the general role of the original diagnostic test.

The primary objective of this study is to validate the potential clinical utility of Morph CT Plus, an AI-based software developed using the Korea-Registries to Overcome Dementia and Accelerate Dementia Research (K-ROAD) cohort.[11] By focusing on patients with mild cognitive impairment (MCI), the critical window for intervention, we demonstrate how AI-driven analysis of low-dose CT scans can serve as a safe, inclusive, and high-precision triage tool to support the South Korean "National Responsibility for Dementia."[12,13]

METHODS

Study participants

This retrospective prognostic cohort study included patients with MCI, operationally defined as having a baseline Clinical Dementia Rating (CDR) score of 0.5.[14] Participants were recruited from the K-ROAD project between January 2015 and December 2023.[11] Eligible participants met the following criteria: 1) age ≥50 years; 2) baseline CDR score of

0.5; and 3) available follow-up for 3 years (±6 months). Dementia conversion was defined as progression to a CDR score ≥1.0. Exclusion criteria included structural brain abnormalities (e.g., stroke, tumors) or severe psychiatric disorders.[14]

All participants provided written informed consent, and the study was approved by the Institutional Review Board of Samsung Medical Center (No. 2021-02-135).

Acquisition of CT images

The CT images analyzed in this study were low-dose scans acquired as part of amyloid PET-CT protocols, primarily for attenuation correction, rather than standalone CT examinations performed for prognostic assessment. All CT scans were acquired using the same PET/CT scanner model (Discovery STe, GE Medical Systems, Milwaukee, WI, USA) in three-dimensional mode at a single center, ensuring consistency in hardware and reconstruction algorithms across participants. Common reconstruction settings were maintained throughout the study period, including a slice thickness of 3.75 mm, a STANDARD convolution kernel, and a 512×512 matrix. Depending on the PET tracer protocol, tube voltage and tube current ranged from 120–130 kVp and 50–95 mA, respectively. The pixel size ranged from 0.488 to 0.500 mm, yielding voxels of 0.488–0.500×0.488–0.500×3.75 mm^3.

Software development

We developed Morph CT Plus (BeauBrain Healthcare Inc., Seoul, Korea), an AI-based software as a medical device approved by the Ministry of Food and Drug Safety. The software is designed to predict three-years dementia conversion risk in patients with MCI (CDR=0.5) who have not been diagnosed with dementia. The software integrates non-contrast brain CT images with clinical variables, including age, sex, years of education, and Clinical Dementia Rating-sum of boxes (CDR-SB). The model generates both a continuous risk score ranging from 0 to 100 and a categorical risk stratification (low, intermediate, high) to support clinical decision-making.

Image processing pipeline

The software employs a deep learning-based segmentation algorithm, previously validated against 3D T1-weighted MRI, to automatically segment 14 predefined anatomical brain regions and quantify cerebrospinal fluid (CSF) and lateral ventricle volumes.[8] To account for demographic and modality-related variability, W-scores were calculated as standardized residuals from linear regression model fitted in the K-ROAD cohort, in which regional brain volumes were modeled as functions of age, sex, and imaging modality. Both regional and a global composite W-scores (calculated as the sum of across 14 regions) are incorporated as imaging-derived features. These features, together with clinical input variables are used in the dementia conversion prediction model.

Model development and test

The dataset (n=896) was split into a training/validation set (n=533, 59.5%) and an independent test set (n=363, 40.5%) using stratified sampling to preserve the class distribution. The independent test set was completely held out from all stages of model development, including preprocessing, model training, and hyperparameter optimization. Candidate predictors included demographic and clinical variables (age, sex, years of education, and CDR-SB) and imaging-derived metrics comprising regional W-scores from 14 predefined brain regions and a global composite W-score, yielding 19 candidate predictors. Continuous variables (age, years of education, and W-scores) were standardized

https://doi.org/10.12779/dnd.2026.25.2.91

using z-score normalization, and CDR-SB was scaled using a robust scaler to reduce sensitivity to outliers. Sex was encoded as a binary variable. All preprocessing parameters were estimated exclusively from the training data and subsequently applied to validation and test cohorts to prevent information leakage.

Six supervised machine learning algorithms—logistic regression, random forest, gradient boosting, support vector machine, extreme gradient boosting, and multilayer perceptron—were evaluated within the training/validation cohort using stratified five-fold cross-validation. Hyperparameters were optimized via grid search with cross-validated area under the curve (AUC) as the selection criterion (**Supplementary Table 1**).[15-20] No resampling techniques were applied to address class imbalance. The final logistic regression model was trained on the full training/validation set (n=533) and evaluated on the independent test set (n=363).

To enhance interpretability, SHapley Additive exPlanations (SHAP) were applied to quantify both global feature importance and individual-level contributions to predicted dementia conversion risk. The classifier generated a predicted probability of conversion for each participant, which was subsequently used for clinically oriented risk stratification. A two cutoffs strategy was implemented to define low-, intermediate-, and high-risk groups. For each predefined operating target (90% and 95% sensitivity or specificity), a lower cutoff was selected to achieve the target sensitivity (minimizing false negatives) and an upper cutoff was selected to achieve the target specificity (minimizing false positives). Participants with predicted probabilities below the lower cutoff were classified as low risk, those above the upper cutoff as high risk, and those between the two thresholds as intermediate risk. These thresholds were fixed and applied without recalibration to the independent test set. Positive predictive values (PPVs) and negative predictive value (NPV) were calculated to assess clinical operating performance of the resulting risk strata.

Statistical analysis

All statistical analyses were performed using Python (version 3.12) with scikit-learn pandas, and statsmodels libraries for model development, and R (R Studio version 2023.12.1+402; Boston, MA, USA) for statistical testing and visualization. Baseline participants characteristics were compared between the development and test sets using Welch's two-sample *t*-test for continuous variables and Pearson's χ^2 tests for categorical variables. Differences across risk stratification groups were evaluated using one-way analysis of variance for continuous variables, followed by Tukey's *post hoc* tests for pairwise comparisons. Categorical variables were compared using χ^2 tests, with Bonferroni correction employed for multiple pairwise comparisons.

Survival analysis was performed to evaluate time-to-dementia conversion within the three-year follow-up period. Kaplan-Meier survival curves were constructed for each risk group, and differences among groups were assessed using the log-rank test. Cox proportional hazards regression was performed to estimate hazard ratios for dementia conversion across risk groups, with the low-risk group serving as the reference category.

SHAP values were calculated using the shap Python library (version 0.50.0) to assess feature importance. Statistical significance was set at $p<0.05$ (two-sided) for all analyses.

To evaluate the incremental value of CT-derived imaging features, the performance of a clinical-only model (age, sex, years of education, and CDR-SB) was compared with

https://doi.org/10.12779/dnd.2026.25.2.91

that of the combined model using AUC, PPV, NPV, and the proportion of intermediate-risk classifications.

RESULTS

Baseline participant characteristics

The baseline demographic and clinical characteristics of the study participants are summarized in **Table 1**. The study included a total of 896 patients with MCI (mean age 72.0 years, 57.1% female, mean education 11.7 years). The development set (n=533) and test set (n=363) showed no significant differences in demographic or clinical characteristics, including age, sex, education, baseline CDR-SB, or dementia conversion rate. The mean follow-up duration was slightly longer in the development set compared with the test set (38.8±18.2 months vs. 35.9±14.2 months, respectively, p=0.008).

Model selection and classification performance

We compared six classifiers using five-fold cross-validation on the development dataset (**Supplementary Table 1**). Logistic regression achieved the highest cross-validation AUC (0.861±0.012) and was selected for subsequent analyses (**Supplementary Fig. 1A**). On the independent test set (n=363), the logistic regression model achieved an AUC of 0.870, with accuracy of 0.793, sensitivity of 0.701, and specificity of 0.861 (**Supplementary Fig. 1B**). During cross-validation on the training/validation set (n=533), fold-wise AUCs ranged from 0.841 to 0.870 (**Supplementary Fig. 1A**). The distribution of predicted probabilities showed clear separation between converters and non-converters in both development and test datasets (**Fig. 1**).

Feature importance

SHAP analysis identified CDR-SB as the dominant predictor, with SHAP values ranging from −0.28 to +0.52 (**Supplementary Fig. 1C**). Among imaging features, parietal and temporal CSF W-scores showed the greatest contribution to model prediction of dementia conversion (right parietal: −0.32 to +0.32; left parietal: −0.31 to +0.35), followed by right temporal CSF W-score (−0.21 to +0.23). As W-scores reflect the degree of regional brain atrophy, greater atrophy in these regions was associated with higher conversion probability, consistent with known patterns of early neurodegeneration in Alzheimer's disease. Sex contributed moderately (SHAP −0.08 to +0.06), while age (−0.05 to +0.05) and education (−0.09 to +0.06) had minimal influence.

Table 1. Baseline participant characteristics

Characteristics	Total (n=896)	Development set (n=533)	Test set (n=363)	p-value
Age (yr)	72.0±7.7	71.7±7.5	72.4±8.0	0.152
Sex (female)	512 (57.1)	310 (58.2)	202 (55.6)	0.498
Education (yr)	11.7±4.7	11.5±4.6	11.9±4.9	0.267
CDR-SB	1.8±1.1	1.8±1.1	1.8±1.1	0.930
Dementia converters	371 (41.4)	217 (40.7)	154 (42.4)	0.659
Follow-up duration (mon)	37.7±16.7	38.8±18.2	35.9±14.2	0.008*

Data presented as mean ± standard deviation for continuous variables and number (%) for categorical variables.
Between group comparisons were performed using Welch two sample t-tests for continuous variables or Pearson's χ^2 tests for categorical variables.
CDR-SB: Clinical Dementia Rating-sum of boxes.
*p<0.05.

https://doi.org/10.12779/dnd.2026.25.2.91

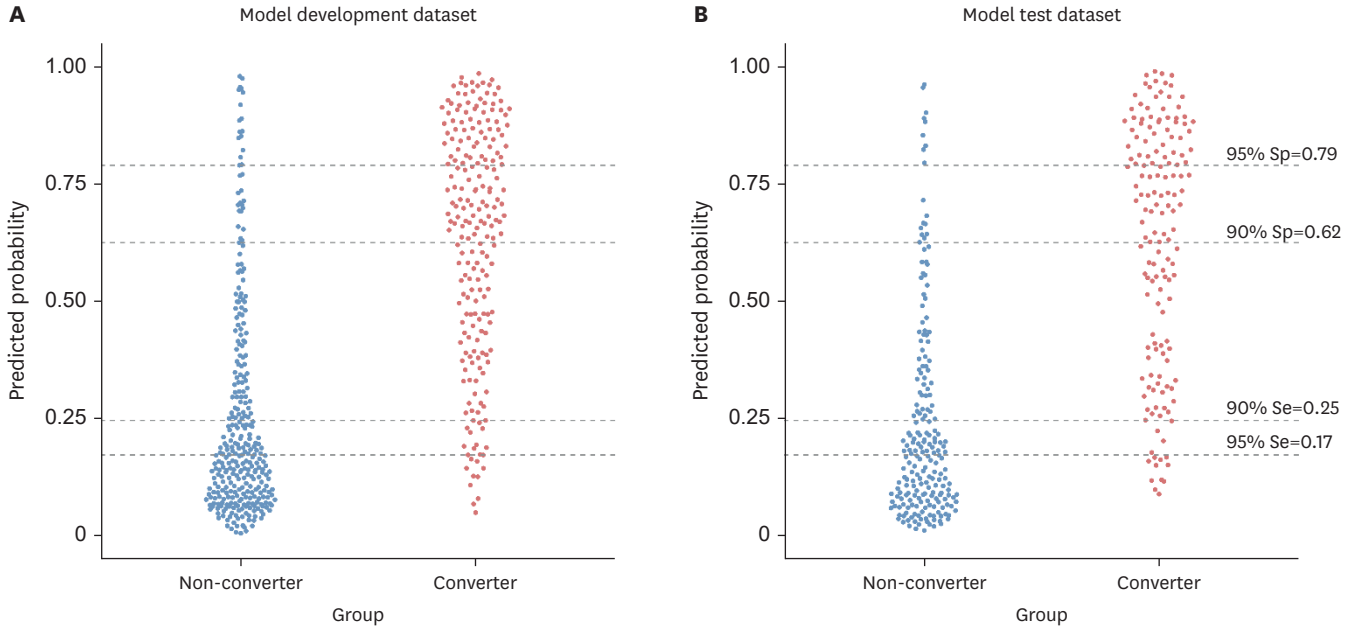

Fig. 1. Distribution of predicted probabilities of conversion based on the logistic regression model. (A) Distribution of predicted conversion probabilities for the model development dataset and (B) the model test dataset. Blue dots represent non-converters and red dots represent converters. On the right y-axis, the probability values corresponding to the evaluated risk thresholds are demonstrated and accompanied by the metric used to define them (90% and 95% Se or 90% and 95% Sp). The dashed lines indicate where these thresholds fall on the probability distribution, with the lower lines corresponding to the Se-based thresholds (0.25 for 90% Se, 0.17 for 95% Se) and the upper lines corresponding to the Sp-based thresholds (0.62 for 90% Sp and 0.79 for 9% Sp). These thresholds define the boundaries for low-, intermediate-, and high-risk groups.
Se: sensitivity, Sp: specificity.

Risk stratification performance

The logistic regression model stratified patients into low-, intermediate-, and high-risk groups for dementia conversion using sensitivity/specificity-based probability cut-offs (**Fig. 1**). We stratified patients into three risk groups using two cut-off strategies (**Fig. 2**). The 90% sensitivity/specificity cut-off strategy (lower cut-off=0.25; higher cut-off=0.62) classified 156 patients (43.0%) as low-risk, 99 (27.3%) as intermediate-risk, and 108 (29.8%) as high-risk. The low-risk group achieved 90.4% (95% confidence interval [CI], 84.6%–94.5%) NPV and the high-risk group achieved 83.3% (95% CI, 74.9%–89.8%) PPV. This strategy limited the intermediate-risk group to 27.3% of patients, enabling definitive risk classification for 72.7% of the cohort with 87.5% overall accuracy (**Table 2**). The 95% sensitivity/specificity strategy (lower cut-off=0.17; higher cut-off=0.79) classified 118 patients (32.5%) as low-risk, 180 (49.6%) as intermediate-risk, and 65 (17.9%) as high-risk, with improved NPV of 90.7% (95% CI, 83.9%–95.3%) and PPV of 86.2% (95% CI, 75.3%–93.5%), achieving 89.1% accuracy for definitive classifications, though nearly half of patients (49.6%) required additional testing.

Baseline characteristics using the 90% strategy differed significantly across risk group (**Table 3**). The high-risk group was older (73.8±7.8 vs. 71.3±7.6 years), had higher CDR-SB scores (3.2±0.8 vs. 0.8±0.4), and substantially higher conversion rate (83.3% vs.

Table 2. Performance metrics for definitive risk classifications

Strategy	Number (low + high)	Accuracy	PPV	NPV	Intermediate %
90% Se/Sp	264	87.5%	83.3%	90.4%	27.3%
95% Se/Sp	183	89.1%	86.2%	90.7%	49.6%

Metrics calculated for patients classified as low-risk or high-risk only (excluding intermediate-risk group).
PPV: positive predictive value, NPV: negative predictive value, Se: sensitivity, Sp: specificity.

https://doi.org/10.12779/dnd.2026.25.2.91

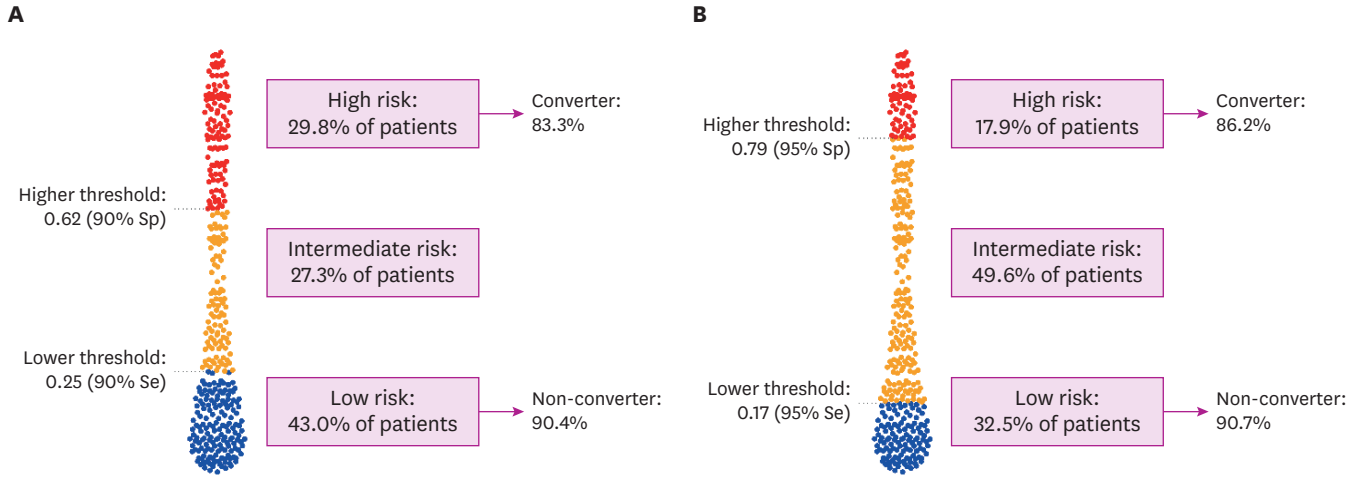

Fig. 2. Risk stratification based on model-predicted conversion probabilities using dual cut-off strategies. The panels illustrate risk stratification workflows based on (A) 90% Se and 90% Sp thresholds strategy and (B) 95% Se and 95% Sp threshold strategy. Each panel shows the distribution of predicted probabilities with individuals categorized into high-risk (red), intermediate-risk (orange), and low-risk (blue) groups. The lower threshold (0.25 for 90% Se; 0.17 for 95% Se) identifies subjects at low risk of conversion with high Se, while the higher threshold (0.62 for 90% Sp; 0.79 for 95% Sp) identifies subjects at high risk with high Sp. The predictive accuracy of the stratification is demonstrated by arrows showing the positive predictive value (percentage of converters in the high-risk group: 83.3% in [A] 86.2% in [B]) and negative predictive value (percentage of non-converters in the low-risk group: 90.4% in [A], 90.7% in [B]).
Se: sensitivity, Sp: specificity.

Table 3. Baseline characteristics by risk group (90% sensitivity/specificity)

Characteristics	Low risk (n=156)	Intermediate risk (n=99)	High risk (n=108)
Age (yr)	71.3±7.6	72.8±8.5	73.8±7.8*
Sex (female)	76 (48.7)	60 (60.6)	66 (61.1)
Education (yr)	12.2±4.8	12.0±4.7	11.4±5.3
CDR-SB	0.8±0.4	1.8±0.7*	3.2±0.8*†
Dementia converters	15 (9.6)	49 (49.5)*	90 (83.3)*†
Follow-up duration (mon)	42.2±14.3	35.0±12.3*	27.8±10.8*†
W-scores			
Left frontal CSF	0.1±0.9	−0.5±1.0*	−0.9±1.1*†
Left occipital CSF	0.1±1.0	−0.5±1.1*	−0.8±1.1*
Left parietal CSF	0.2±0.9	−0.4±1.1*	−1.0±1.1*†
Left temporal CSF	0.3±0.9	−0.7±1.0*	−1.3±1.1*†
Left anterior LV	0.0±1.2	−0.6±1.2*	−0.9±1.2*
Left posterior LV	0.1±1.1	−0.5±1.1*	−1.0±1.6*†
Left inferior LV	0.0±1.2	−0.9±1.4*	−1.8±1.9*†
Right frontal CSF	0.1±0.9	−0.5±1.0*	−0.9±1.0*†
Right occipital CSF	0.0±1.0	−0.4±1.0*	−0.5±0.9*
Right parietal CSF	0.2±1.0	−0.4±1.0*	−0.9±1.1*†
Right temporal CSF	0.1±1.0	−0.8±1.1*	−1.3±1.1*†
Right anterior LV	0.0±1.1	−0.5±1.2*	−0.9±1.3*
Right posterior LV	0.0±1.0	−0.6±1.3*	−1.0±1.6*†
Right inferior LV	0.0±1.1	−1.0±1.4*	−1.6±1.7*†
Global	0.1±1.0	−0.8±1.0*	−1.5±1.1*†

Data presented as mean ± standard deviation for continuous variables and number (%) for categorical variables. Statistical comparisons among risk groups were conducted using analysis of variance for continuous variables and χ^2 tests for categorical variables.
CDR-SB: Clinical Dementia Rating-sum of boxes, CSF: cerebrospinal fluid, LV: lateral ventricle.
*Denotes a significant difference compared to the low-risk group. †Indicates a significant difference compared to the intermediate-risk group.
Statistical significance was defined as $p<0.05$.

9.6%) compared to the low-risk group (all $p<0.05$), with the intermediate-risk group showing intermediate values. All W-scores demonstrated progressive atrophy across risk groups (all $p<0.05$), with the global composite showing a stepwise gradient (0.1±1.0, −0.8±1.0, −1.5±1.1 for low, intermediate, and high risk; all pairwise $p<0.05$).

https://doi.org/10.12779/dnd.2026.25.2.91

DND Dementia and Neurocognitive Disorder

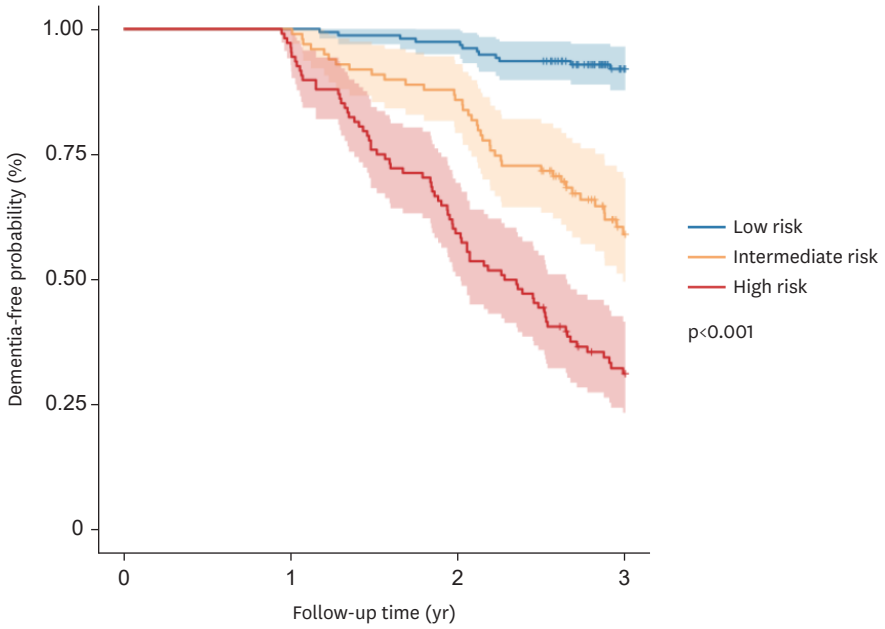

Fig. 3. Kaplan-Meier survival analysis by risk stratification group. Kaplan-Meier survival curves and summary statistics for three risk groups in the test data set. The table shows 3 years dementia-free probability with 95% confidence intervals and hazard ratios for each group. Shaded areas indicate 95% confidence intervals. Log-rank test $p<0.001$.

Kaplan-Meier survival analysis demonstrated significant differences in time-to-dementia conversion across the risk groups (log-rank test, $p<0.001$, **Fig. 3**). Cox proportional hazards regression revealed that compared to the low-risk group, intermediate-risk patients had a 5.97-fold higher hazard of conversion (95% CI, 3.12–11.44; $p<0.001$), and high-risk patients had a 14.13-fold higher hazard (95% CI, 7.66–26.07; $p<0.001$). The three-year dementia-free survival rates were 92.0% for low-risk, 59.1% for intermediate-risk, and 31.3% for high-risk groups (**Supplementary Table 2**).

DISCUSSION

This study demonstrates that a dual-cutoff risk-stratification strategy using AI-based CT analysis provides a robust and actionable prognostic tool for patients with MCI (CDR=0.5). In our independent test set, the 90% sensitivity/specificity strategy successfully limited the intermediate-risk (undecided) group to a manageable 27.3% of the total cohort. This result is particularly significant for clinical practice; for the remaining 72.7% of patients, the model achieved a combined group accuracy of 87.5% and a NPV of 90.4%. These findings indicate that the vast majority of patients can receive a definitive risk assessment immediately, establishing the potential clinical utility required for innovative medical technology adoption by National Evidence-based healthcare Collaborating Agency and HIRA.[10,13] By providing actionable results for nearly three-quarters of the population with such high precision, this technology bridges the technical gap between accessible CT and specialized MRI-based models.[8]

The practical value of Morph CT Plus lies in its ability to optimize the national dementia screening pathway by addressing critical “unmet needs.” In the current healthcare landscape,

https://doi.org/10.12779/dnd.2026.25.2.91

primary care facilities often lack the resources to differentiate stable MCI cases from those at high risk of rapid decline. Our technology provides an efficient "Triage Pathway" that identifies a large low-risk group that can be safely managed with routine monitoring, effectively deferring the need for expensive PET scans or invasive biomarker tests for these individuals.[5,21] This optimizes the allocation of national health insurance resources, prioritizing specialized clinics and high-cost diagnostics for the high-risk group, which in our study exhibited a 14.13-fold increase in conversion hazard. Among individuals classified as intermediate-risk patients, further diagnostic workup would be recommended to enable more definitive risk stratification, including advanced neuroimaging such as amyloid PET or MRI, or emerging fluid biomarkers such as plasma p-tau217. This is consistent with the intended role of Morph CT Plus as a triage tool that prioritizes more costly or invasive diagnostic procedures for those most likely to benefit. Taken together, these findings support the analytical validity and economic potential of the proposed framework and its potential to improve clinical efficiency and resource allocation, aligning with the goals of the National Responsibility for Dementia.[12,22]

We intentionally utilized a dual-cutoff strategy rather than a single-cutoff binary classification to provide a risk-based hierarchy (low, intermediate, high). This approach is strategically designed to support clinical decision-making while avoiding the misconception that the AI is providing a "definitive diagnosis."[23] Concerns are frequently raised regarding AI tools that act as 'black-box' diagnostic replacements; however, by providing a conversion probability and risk stratification, Morph CT Plus serves as a sophisticated prognostic aid that enhances clinician judgment.[24] While a single-cutoff model achieved a baseline accuracy of 79.3%, the dual-cutoff approach significantly enhanced diagnostic confidence for the targeted groups, ensuring that only those with highly certain risk profiles are prioritized for intervention.

In our comparative analysis of operating targets, the 90% strategy proved superior to the 95% threshold for large-scale clinical application. While the 95% strategy provided extreme specificity for the high-risk group, it left 49.6% of patients in the intermediate zone. Such a high rate of "undecided" results would lead to a bottleneck in clinical workflows and reduce the technology's impact on public health. The 90% strategy offers the optimal balance, ensuring high precision while maintaining a low undecided rate. Furthermore, SHAP analysis identified parietal CSF W-scores as a dominant imaging predictor. This aligns with the "retrogenesis" theory, suggesting that association cortices are among the first to atrophy in early-stage Alzheimer's disease.[25,26] This supports the biological validity of the model, suggesting that the AI is capturing real neurodegenerative changes rather than imaging noise.[8]

Although CDR-SB emerged as the dominant predictor in our model, reflecting its central role in capturing cognitive and functional impairment, the incorporation of imaging-derived features provided meaningful complementary value. Rather than substantially altering overall model discrimination, imaging biomarkers contributed to refining risk stratification at clinically relevant decision boundaries. In particular, the combined model demonstrated a significantly improved model fit compared to the clinical-only model, indicating that the inclusion of imaging-derived features provides additional explanatory value beyond clinical variables along (**Supplementary Table 3**). These findings highlight that, even in the presence of strong clinical predictors, imaging features can enhance the practical utility of risk prediction models by supporting more confident and targeted clinical decision-making.

https://doi.org/10.12779/dnd.2026.25.2.91

Strengths of this study include the utilization of the large-scale Korean K-ROAD cohort, ensuring that the model is optimized for the ethnic and anatomical characteristics of the domestic population.[11] Several limitations of this study merit consideration. First, this study was conducted within a single-center cohort (K-ROAD), and the model has not been externally validated in geographically or ethnically diverse populations. Second, the CT images in this study were acquired using a low-dose protocol for PET attenuation correction, which may differ from routine diagnostic non-contrast brain CT in noise level, spatial resolution, and reconstruction parameters. Although the underlying segmentation model was trained with synthetic images to promote robustness across varying acquisition conditions and demonstrated generalizability on external ADNI data,[8] direct validation on routine clinical CT has not been performed. Future multi-center studies incorporating routine diagnostic CT datasets are needed to confirm the broader clinical applicability of this framework. Third, white matter hyperintensities (WMH), and important contributor to cognitive impairment, are difficult to quantify reliably on CT and were not explicitly captured in our model. However, previous studies have shown that when structural brain atrophy and WMH are evaluated simultaneously, the independent contribution of WMH to cognitive decline is often significantly diminished, as its effects are largely mediated through subsequent brain atrophy. Therefore, the prominent regional CSF and ventricular expansions quantified by our model likely capture not only primary neurodegeneration but also the downstream structural consequences of cerebral small vessel disease. Nonetheless, the potential incremental value of explicit WMH quantification remains an area for future investigation. Fourth, individuals with different underlying etiologies may exhibit distinct structural atrophy patterns, and etiology-specific model performance was not evaluated in the current study. Future studies stratified by amyloid status or underlying etiology would help clarify the model's predictive utility across different dementia subtypes.

In conclusion, Morph CT Plus offers a safe and highly efficient triage solution that fulfills the rigorous criteria for New Medical Technology, potentially improving patient outcomes while ensuring the economic sustainability of the national dementia management framework.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
Performance comparison of machine learning classifiers for predicting three-year dementia conversion

Supplementary Table 2
Survival outcomes and HRs by risk stratification group

Supplementary Table 3
Comparison of model performance and incremental value of CT features

Supplementary Fig. 1
Model performance evaluation and feature importance analysis. (A) Five-fold cross-validation performance on the development dataset (individual folds in color, mean in black) and (B) performance on the independent test dataset. (C) The SHAP beeswarm plot displaying the contribution of the nineteen input features to the logistic regression model's output. Features are ranked by their overall impact on model predictions. Each point represents a single subject,

https://doi.org/10.12779/dnd.2026.25.2.91

with the horizontal position indicating the magnitude and direction of the feature's effect on conversion probability, and the color denoting the feature value (red = high, blue = low).

REFERENCES

1. Shon C, Yoon H. Health-economic burden of dementia in South Korea. BMC Geriatr 2021;21:549. PUBMED | CROSSREF
2. Eggleston K, Kim D. Dementia care in a rapidly aging society. Alzheimers Dement 2025;21:e70268. PUBMED | CROSSREF
3. Munn Z, Moola S, Lisy K, Riitano D, Murphy F. Claustrophobia in magnetic resonance imaging: a systematic review and meta-analysis. Radiography 2015;21:e59-e63. CROSSREF
4. Nguyen XV, Tahir S, Bresnahan BW, Andre JB, Lang EV, Mossa-Basha M, et al. Prevalence and financial impact of claustrophobia, anxiety, patient motion, and other patient events in magnetic resonance imaging. Top Magn Reson Imaging 2020;29:125-130. PUBMED | CROSSREF
5. Wittenberg R, Knapp M, Karagiannidou M, Dickson J, Schott J. Economic impacts of introducing diagnostics for mild cognitive impairment Alzheimer's disease patients. Alzheimers Dement (N Y) 2019;5:382-387. PUBMED | CROSSREF
6. Rabinovici GD, Knopman DS, Arbizu J, Benzinger TLS, Donohoe KJ, Hansson O, et al. Updated appropriate use criteria for amyloid and tau PET: a report from the Alzheimer's Association and Society for Nuclear Medicine and Molecular Imaging Workgroup. Alzheimers Dement 2025;21:e14338. PUBMED | CROSSREF
7. Pasi M, Poggesi A, Pantoni L. The use of CT in dementia. Int Psychogeriatr 2011;23 Suppl 2:S6-S12. PUBMED | CROSSREF
8. Park S, Kim K, Lim KY, Na DL, Kim HJ, Jang H, et al. Computed tomography-based nnU-Net for region-specific brain structural changes across the alzheimer's continuum and frontotemporal dementia subtypes. Sci Rep 2025;15:42597. PUBMED | CROSSREF
9. Jang JW, Kim J, Park SW, Kasani PH, Kim Y, Kim S, et al. Machine learning-based automatic estimation of cortical atrophy using brain computed tomography images. Sci Rep 2022;12:14740. PUBMED | CROSSREF
10. Health Insurance Review and Assessment Service (HIRA). Guidelines for reimbursement evaluation of innovative medical technologies: AI-based innovative medical technologies [Internet]. Wonju: HIRA; 2023 [cited 2023 Aug 25]. Available from: https://www.hira.or.kr.
11. Jang H, Shin D, Kim Y, Kim KW, Lee J, Kim JP, et al. Korea-Registries to Overcome Dementia and Accelerate Dementia Research (K-ROAD): a cohort for dementia research and ethnic-specific insights. Dement Neurocogn Disord 2024;23:212-223. PUBMED | CROSSREF
12. Ministry of Health and Welfare (KR). The 5th Comprehensive Plan for Dementia Management (2026-2030) [Internet]. Sejong: Ministry of Health and Welfare (KR); 2026 [cited 2026 Feb 13]. Available from: https://mohw.go.kr.
13. National Evidence-based healthcare Collaborating Agency (NECA). A to Z Guide for New Health Technology Assessment System [Internet]. Sejong: NECA; 2025 [cited 2025 Dec 10]. Available from: https://nhta.neca.re.kr.
14. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:270-279. PUBMED | CROSSREF
15. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 1958;65:386-408. PUBMED | CROSSREF
16. Wiley. Applied Logistic Regression, 3rd Edition [Internet]. Hoboken: Wiley; 2013 [cited 2025 Jul 16]. Available from: https://www.wiley.com/en-us/Applied+Logistic+Regression%2C+3rd+Edition-p-9780470582473.
17. Breiman L. Random forests. Mach Learn 2001;45:5-32. CROSSREF
18. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;29:1189-1232. CROSSREF
19. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273-297. CROSSREF
20. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13–17; San Francisco, CA, USA. New York, NY, USA: Association for Computing Machinery; 2016, 785-794.

21. Palmqvist S, Tideman P, Mattsson-Carlgren N, Schindler SE, Smith R, Ossenkoppele R, et al. Blood biomarkers to detect Alzheimer disease in primary care and secondary care. JAMA 2024;332:1245-1257. PUBMED | CROSSREF

22. Choi H, Kim SH, Lee JH, Lee AY, Park KW, Lee EA, et al. National responsibility policy for dementia care: current and future. J Korean Neurol Assoc 2018;36:152-158. CROSSREF

23. Yim S, Park S, Lim KY, Kang H, Shin D, Jang H, et al. CT-derived brain volumes and plasma p-Tau217 for risk stratification of amyloid positivity in early-stage Alzheimer's disease. Alzheimers Res Ther 2025;17:233. PUBMED | CROSSREF

24. Martin SA, Zhao A, Qu J, Imms P, Irimia A, Barkhof F, et al. Investigating the utility of explainable artificial intelligence for neuroimaging-based dementia diagnosis and prognosis. Hum Brain Mapp 2026;47:e70456. PUBMED | CROSSREF

25. Reisberg B, Franssen EH, Hasan SM, Monteiro I, Boksay I, Souren LEM, et al. Retrogenesis: clinical, physiologic, and pathologic mechanisms in brain aging, Alzheimer's and other dementing processes. Eur Arch Psychiatry Clin Neurosci 1999;249 Suppl 3:S28-S36. PUBMED | CROSSREF

26. Maboudian SA, Willbrand EH, Kelly JP, Jagust WJ, Weiner KS; Alzheimer's Disease Neuroimaging Initiative. Defining overlooked structures reveals new associations between the cortex and cognition in aging and Alzheimer's disease. J Neurosci 2024;44:e1714232024. PUBMED | CROSSREF