International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 1, Jan-Feb-2026, ISSN (Online): 2395-566X

Longitudinal Structural MRI-Based Deep Learn-ing and Radiomics Features for Predicting Alz-heimer's Disease Progression

Diksha Pawar[1], Prof. Jayshree Boaddh[2], Prof. Rahul Patidar[3]
MTech Scholar, Vaishnavi Institute of Technology & Science
Bhopal, MP, India
Department of CSE, Vaishnavi Institute of Tech-nology & Science
Bhopal, MP, India

Abstract- Alzheimer's disease (AD), the leading cause of dementia worldwide, affects more than 55 million individuals and gen-erates annual healthcare costs exceeding two trillion USD [14]. A substantial proportion (30–40% per year) of pa-tients with mild cognitive impairment (MCI) progress to AD [2], making early and accurate prognostication essential for timely intervention, trial enrichment, and resource allocation. This paper presents a comprehensive review of a re-cent longitudinal MRI-based study by Aghajanian et al. [1], which integrates three-dimensional (3D) convolutional neural networks (CNNs), time-aware long short-term memory (T-LSTM) networks with attention mechanisms, and radiomics features to predict MCI-to-AD conversion using structural MRI. The cohort comprises 228 ADNI MCI participants with at least three T1-weighted MRI scans over an 18-month window (684 scans in total) [1]. A 3D Res-Net-18 backbone [9] extracts volumetric features, fed into a T-LSTM incorporating inter-scan intervals and attention mechanisms [10]. The best longitudinal model achieves a concordance index (c-index) of 0.90, with time-specific AUCs of 0.96, 0.94, and 0.89 for 2-, 3-, and 5-year conversion prediction, respectively, and an approximate 11-fold hazard ratio between high- and low-risk groups [1]. This review analyzes the methodology, highlights its strengths and weaknesses, and discusses key implications for clinical translation.

Keywords – Alzheimer's disease, mild cognitive impairment, longitudinal MRI, deep learning, convolutional neural networks, LSTM, radiomics, survival analysis, concordance index.

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common cause of dementia worldwide [14]. It imposes a major societal and economic burden, with estimates exceeding 55 million affected individuals and global annual costs surpassing two trillion USD [14]. Clinically, AD typically evolves through stages: cognitively normal (CN), mild cognitive impairment (MCI), and overt dementia [15]. The MCI stage is particularly critical because patients exhibit measurable cognitive decline but retain functional capacity, making this phase ideal for therapeutic intervention [18].

Epidemiological studies consistently report that approximately 30–40% of MCI patients progress to AD over short time horizons [2]. Accurately identifying which MCI patients are at high risk is therefore central to patient counseling, clinical trial enrichment, and treatment resource allocation [15]. Conventional prognostic tools—cognitive tests such as the Mini-Mental State Examination (MMSE) and Clinical Dementia Rating (CDR), along with fluid and PET biomarkers [18][19]—provide valuable information but often lack sufficient individualized prognostic accuracy.

Structural magnetic resonance imaging (MRI) offers a noninvasive window into brain macrostructure [16]. Longitudinal changes in gray matter (GM) and white matter (WM) volumes, cortical thickness, and shape indices reflect neurodegeneration in AD [16]. However, manual radiological review has limited sensitivity to subtle, spatially distributed patterns evolving over time. This limitation has motivated the use of computational imaging pipelines and deep learning (DL) to automatically extract high-dimensional quantitative features from serial MRI data [2][12].

The longitudinal deep learning framework proposed by Aghajanian et al. [1] represents an important step forward, combining 3D CNNs [9], time-aware LSTMs [10], and radiomics [5] to predict MCI-to-AD progression using structural MRI alone. This review analyzes the methodology, connects the work to existing literature, and discusses key implications for future research and clinical translation.

II. BACKGROUND AND RELATED WORK

A. Evolution of Neuroimaging Biomarkers

The trajectory of neuroimaging biomarkers for AD prediction has progressed from simple region-of-interest (ROI) measures to whole-brain, multimodal, and longitudinal approaches [16]. Early studies focused on hippocampal volume and cortical thickness, achieving modest predictive performance (e.g., AUC 0.65–0.70 for 2–3-year conversion) [16]. Radiomics emerged as a complementary paradigm, systematically extracting high-dimensional descriptors of shape, intensity, and texture from segmented regions [5]. These features capture subtle morphological irregularities beyond gross volumetric atrophy. Parallel advances in deep learning, particularly CNNs [9], enabled data-driven feature learning directly from volumetric MR images, often outperforming hand-crafted feature pipelines [2][12].

B. Longitudinal and Multimodal Deep Learning

More recent work has shifted toward architectures that model temporal dynamics and integrate multiple data types [3][4]. Recurrent neural networks (RNNs), LSTMs, and attention mechanisms [10] capture trajectories of imaging and clinical measures over time, reflecting that disease progression is defined by rate-of-change rather than static morphology [1][3][4]. Multimodal models combine MRI with PET, functional imaging, cerebrospinal fluid biomarkers, and genetic risk scores [13], achieving higher performance but at increased acquisition complexity [3]. A longitudinal, MRI-only deep learning model with survival-based evaluation [8] and interpretability balances performance and feasibility for real-world deployment [1].

C. Comparative Performance Summary

Table I provides a comprehensive comparison of contemporary machine learning approaches for MCI-to-AD prediction [1][2][5][6][12].

Table I: Literature Comparison of Machine Learning Methods for MCI-to-AD Prediction

Method	Modality	Sample	Key Metric	AUC/c-index	Year
SVM	Structural MRI	156	Accuracy	75.4%	2019
Random Forests	Structural MRI	189	Accuracy	76.8%	2020
Single-visit CNN	Structural MRI	228	c-index	0.70	2024
Radiomics (TabNet)	GM features	180	AUC	0.92	2025
CNN-LSTM	Structural MRI	228	c-index	0.90	2025
Multimodal (MRI+PET)	MRI + PET	150	AUC	0.91	2024
MediVision (Hybrid)	Medical imaging	500+	Accuracy	95%+	2025

III. STUDY OVERVIEW AND METHODOLOGY

A. Cohort and Study Design

The reviewed study draws its cohort from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [1]. Inclusion criteria required: (1) MCI diagnosis at baseline, (2) at least three T1-weighted structural MRI scans acquired within 18 months of baseline, and (3) maintenance of MCI status for at least 1.5 years [1]. The final dataset comprised 228 MCI participants: 108 converters to AD (mean follow-up ~3.2 years) and 120 non-converters (~4.2 years follow-up) [1]. Across all participants, there were 684 T1-weighted MRI scans, with serial scans typically obtained at 6–12 month intervals [1].

The dataset was partitioned using stratified random sampling: 70% training, 12.5% validation, and 17.5% testing [1]. Table II summarizes key demographic and acquisition characteristics [1].

Table II: Cohort Demographics and MRI Acquisition Characteristics

Characteristic	Converters (n=108)	Non-Converters (n=120)	Total (n=228)
Age at Baseline (yrs)	74.2 ± 6.8	74.0 ± 7.2	74.1 ± 7.0
MMSE at Baseline	27.6 ± 1.6	27.4 ± 1.8	27.5 ± 1.7
Female (%)	48	52	50
APOE4 Carriers (%)	65	42	53
Follow-up (yrs)	3.21 ± 1.9	4.23 ± 2.5	3.74 ± 2.3

B. Image Preprocessing and Radiomics Feature Extraction

Preprocessing followed established neuroimaging standards [1]. T1-weighted volumes were processed using SPM-based pipelines, including bias field correction, brain extraction, resampling to isotropic 1.5 mm resolution, and registration to common space [1]. For the deep learning backbone, each MRI

was resampled to 128×128×128 voxels and z-score normalized [1].

Radiomics features were computed using PyRadiomics (v3.0.1a1) [1]. A total of 107 features per region of interest were extracted, including shape descriptors (12 features), first-order intensity statistics (18 features), and textural features from GLCM, GLRLM, GLSZM, GLDM, and NGTDM [5]. To mitigate dimensionality, principal component analysis (PCA) was applied, retaining approximately 95% of variance [1]. GM-derived features demonstrated the highest concordance with time-to-conversion [16].

C. Deep Learning Architecture

1) 3D ResNet for Single-Timepoint Prediction

The backbone CNN is a 3D ResNet-18 adapted to volumetric MRI [1][9]. Key design elements include initial 3D convolution (7×7×7 kernel, stride 2, 64 filters), four successive residual stages with increasing channel depth (64, 128, 256, 512), and global average pooling [9]. The network is trained with pairwise ranking loss, suitable for survival-type data with censoring [8]. Single-timepoint performance yields a c-index around 0.70, indicating that purely static imaging has limited prognostic capability [1].

2) Time-Aware LSTM with Attention

CNN-derived embeddings are fed into a time-aware LSTM (T-LSTM) [1]. The T-LSTM explicitly models variable time intervals between observations via a learnable decay applied to the memory cell state [1]:

$C_t^{adj} = (1 - Decay) \odot C_{t-1}$

$Decay = 1 - \log_e(1 - \Delta t)$

where Δt represents elapsed time in years between consecutive scans [1]. An attention mechanism is applied over the sequence of hidden states [10]:

$a_s = softmax(v^T \tanh(W_{att} h_s))$

$z = \Sigma a_s h_s$

This provides improved performance and interpretability, as attention weights indicate which visits contribute most to predicted risk [1].

3) Integration of Radiomics and CNN Features

In some configurations, radiomics features (after PCA) are concatenated with CNN embeddings for each timepoint [1]. This hybrid design aims to combine rich learned features with hand-crafted, clinically interpretable descriptors [5].

IV. RESULTS

A. Survival Prediction Performance

The longitudinal CNN–T-LSTM with attention substantially outperforms the single-timepoint CNN [1]. Table III presents detailed performance metrics across different model configurations [1].

Table III: Model Performance Across Different Architectures

Model Configuration	Training C-Index	Validation C-Index	Test C-Index	Test AUC (2-yr)
ResNet-18 (single-visit)	0.88 ± 0.04	0.69 ± 0.04	0.70 ± 0.04	0.68
T-LSTM + Attention	0.95 ± 0.02	0.85 ± 0.00	0.90 ± 0.01	0.96
T-LSTM + GM Radiomics + Attention	0.94 ± 0.03	0.85 ± 0.01	0.91 ± 0.01	0.96

Key Finding: The longitudinal CNN–T-LSTM with attention achieves a test c-index of 0.90±0.01, compared to 0.70±0.04 for single-timepoint ResNet-18 [1]. This 20-point improvement underscores the critical value of trajectory information.

B. Time-Dependent Discrimination

The model demonstrates strong discrimination at clinically relevant time horizons [1]. Table IV provides comprehensive time-specific classification performance [1].

Table IV: Time-Specific Classification Performance

Prediction Horizon	Training AUC	Validation AUC	Test AUC	95% CI (Test)
2-Year Conversion	0.99	0.99	0.96	0.83–1.00
3-Year Conversion	0.98	0.99	0.94	0.84–1.00
5-Year Conversion	0.98	0.99	0.89	0.73–1.00

Performance degrades with increasing prediction horizon due to accumulating uncertainty, but all time points remain clinically useful (AUC > 0.85 for 2–3 years) [1].

C. Risk Stratification and Feature Importance

When patients are dichotomized into high- and low-risk groups based on median predicted risk, Kaplan–Meier curves show clear separation [1]. The reported hazard ratio is approximately 11 [1]:

Hazard Ratio = 11.1 (95% CI: 6.6–18.6)

This translates into an order-of-magnitude difference in conversion risk, highlighting potential clinical utility for stratifying follow-up intensity and trial eligibility [1][15]. Radiomics analyses converge on several key imaging markers: Surface-to-Volume Ratio (HR 1.50, $p<0.001$), Cortical Elongation (HR 1.42, $p<0.001$), GM Volume Change (HR

0.65, p<0.001), and texture features (GLCM Contrast, GLSZM Zone Entropy) [1].

V. DISCUSSION

A. Methodological Strengths

The reviewed study exhibits several notable strengths [1]:

1. **Longitudinal Modeling:** Incorporation of temporal dynamics via T-LSTM and attention yields a ~20-point c-index improvement compared to static models [1], underscoring the critical value of trajectory information.
2. **Hybrid Feature Strategy:** The combination of CNN-derived embeddings [9] and radiomics [5] fosters both high predictive accuracy and interpretability [1], providing clinically meaningful context [5].
3. **Survival Framework:** Using survival analysis metrics (c-index, time-dependent AUC) and hazard ratios [8] is well-aligned with clinical practice [1].
4. **Explainability:** The use of attention mechanisms [10], Grad-CAM, and feature importance analyses addresses barriers to clinical adoption by making model behavior transparent [1].

B. Limitations and Potential Biases

The study also faces important limitations [1]:

Sample Size and Complexity: With 228 participants and ~11 million CNN parameters, the parameter-to-sample ratio is high (~52,000) [1]. Although regularization mitigates overfitting, performance estimates may still be optimistic [11].

Cohort Homogeneity: ADNI participants are not fully representative of real-world populations, raising concerns about demographic and clinical generalizability [1].

Temporal Leakage: Inclusion of the last MRI before or near conversion may introduce subtle temporal leakage [1].

Lack of External Validation: Without testing on independent datasets (e.g., OASIS-3, AIBL), it is difficult to assess robustness to site, scanner, and population differences [4]. Prior work suggests performance can drop by 5–10 percentage points out-of-distribution [4].

C. Position within the Literature

Compared with earlier machine learning approaches (SVM, random forests), the longitudinal CNN–T-LSTM achieves clearly superior discrimination [1][2][12]. Reported benchmark accuracies for MCI-to-AD prediction in the literature often lie in the 75–80% range [2][12], whereas the reviewed model's 2-year AUC of ~0.96 corresponds to markedly higher balanced accuracy [1].

Radiomics-only models in recent work typically achieve AUCs in the 0.81–0.95 range [5]. Hybrid or multimodal deep learning models (combining MRI with PET and clinical data) have reported AUCs around 0.88–0.92 [3]. The reviewed MRI-only longitudinal model thus approaches or exceeds the performance of more complex multimodal systems while relying on a widely available imaging modality [1].

D. Clinical and Research Implications

From a clinical perspective, such models could [1][15][18]:

Identify high-risk patients: Recognize MCI patients at very high short-term risk for conversion, enabling early initiation of disease-modifying therapies [18].

Enrich clinical trials: Preferentially recruit high-risk individuals, increasing event rates and reducing required sample sizes by 25–50% [15].

Support personalized monitoring: Enable tiered follow-up schedules, with intensive monitoring for high-risk patients [18].

For researchers, the study highlights the importance of modeling temporal dynamics [11], standardized preprocessing and robust radiomics pipelines [5], and survival-based evaluation metrics [8].

VI. ENHANCEMENT FRAMEWORK AND CLINICAL TRANSLATION

The reviewed work, while state-of-the-art on ADNI, can be further improved through systematic enhancement [1]. Table V outlines a proposed three-phase implementation roadmap with expected performance gains.

Table V: Three-Phase Enhancement Roadmap

Phase	Duration	Key Enhancements	Test c-Index	2-Yr AUC	Effort
Current	Baseline	Standard architecture	0.90	0.96	—
Phase 1	W1–2	Mixed precision, 3D augmentation	0.90–0.91	0.96	30 h
Phase 2	W3–4	ResNet-50, multi-head attention	0.93–0.94	0.97	70 h
Phase 3	W5–6	Vision Transformer, Bayesian uncertainty	0.96–0.97	0.98	80 h

VII. CONCLUSIONS

This review has examined a longitudinal structural MRI-based deep learning and radiomics framework for predicting MCI-to-AD progression [1]. The integration of 3D ResNet-18 CNNs [9], time-aware LSTMs with attention [10], and GM radiomics features [5] yields state-of-the-art prognostic performance on the ADNI cohort, with a test c-index around 0.90 and excellent short-term AUCs (0.96 for 2-year prediction) [1].

Key Strengths: Longitudinal modeling, hybrid feature representations, survival-based evaluation [8], and interpretable outputs [1].

Important Limitations: Sample size, cohort representativeness, potential temporal leakage, and lack of external validation [1][4].

Future Priorities:

1. Multi-cohort external validation on OASIS-3, AIBL, and independent datasets [4].
2. Exploration of multimodal extensions (adding PET or CSF biomarkers) [3].
3. Incorporation of uncertainty quantification [11].
4. Prospective clinical studies evaluating real-world impact [15][18].
5. Integration with electronic health record systems [18].

Within the broader field of medical AI, this work exemplifies the shift toward longitudinal, interpretable, and clinically grounded models for precision dementia prediction [1][2][3][4]. As technologies mature and external validations accumulate, such frameworks may become routine components of memory clinic diagnostic pipelines [18].

Acknowledgment

The authors acknowledge the Alzheimer's Disease Neuroimaging Initiative (ADNI) for providing the cohort data used in the reviewed study. This work is supported by institutional resources at Vaishnavi Institute of Technology & Science.

REFERENCES

1. Aghajanian, et al., "Longitudinal structural MRI-based deep learning and radiomics features for predicting Alzheimer's disease progression," Alzheimer's Res. Ther., vol. 17, no. 1, p. 182, 2025.
2. Grueso and R. Viejo-Sobera, "Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: A systematic review," Alzheimer's Res. Ther., vol. 13, no. 1, p. 162, 2021.
3. Li, et al., "Predicting long-term progression of Alzheimer's disease with interpretable deep learning featuring interaction effects and multimodality," Alzheimer's Res. Ther., vol. 16, no. 1, p. 62, 2024.
4. Qiu, et al., "Deep learning prediction of Alzheimer's disease conversion from longitudinal MRI across multi-site cohorts," NeuroImage: Clin., vol. 45, p. 102789, 2025.
5. Yuan, et al., "Radiomics-based Alzheimer's disease classification from structural MRI using TabNet," Med. Image Anal., vol. 92, p. 103030, 2025.
6. Hayat, et al., "MediVision: A hybrid CNN–LSTM–attention framework for medical image classification," IEEE J. Biomed. Health Inform., vol. 29, no. 4, pp. 2103–2116, 2025.
7. Thapaliya, et al., "Graph neural networks for brain connectivity analysis and Alzheimer's disease prediction," Med. Image Anal., vol. 89, p. 102889, 2024.
8. Katzman, et al., "DeepSurv: Personalized treatment recommendations using a deep neural network for survival data," J. Mach. Learn. Res., vol. 18, no. 1, pp. 1–25, 2018.
9. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.
10. Vaswani, et al., "Attention is all you need," in Adv. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, 2017, pp. 5998–6008.
11. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in Proc. Int. Conf. Mach. Learn. (ICML), New York, NY, USA, 2016, pp. 1050–1059.
12. Spasov, L. Passamonti, A. Duggento, P. Lió, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," NeuroImage, vol. 189, pp. 276–287, Apr. 2019.
13. W. Rowe, et al., "Machine learning for Alzheimer's risk prediction using single nucleotide polymorphisms," Brain Commun., vol. 3, no. 4, pp. 1–12, 2021.
14. Brookmeyer, et al., "Forecasting the global burden of Alzheimer's disease," Alzheimer's Dement., vol. 3, no. 3, pp. 186–191, 2007.
15. R. Jack Jr., et al., "NIA–AA Research Framework: Toward a biological definition of Alzheimer's disease," Alzheimer's Dement., vol. 14, no. 4, pp. 535–562, 2018.
16. Wüest, "Structural brain MRI in Alzheimer's disease and mild cognitive impairment: A meta-analysis," Neurobiol. Aging, vol. 45, pp. 1–16, 2016.
17. Pizzagalli, "Frontolimbic dysfunction in depression," NeuroImage, vol. 66, pp. 33–42, 2013.
18. L. Lancet, "Clinical diagnosis and management of Alzheimer's disease," Lancet Neurol., vol. 19, no. 12, pp. 1006–1015, 2020.
19. Blennow, K. Hampel, and H. Zetterberg, "Biomarkers in amnestic mild cognitive impairment and Alzheimer's disease: A review," J. Alzheimers Dis., vol. 19, no. 3, pp. 611–628, 2010.

20. ISO/IEEE 11073-10101:2004, Health informatics -- Point-of-care medical device communication -- Part 10101: Nomenclature, Geneva: International Organization for Standardization, 2004.