Early Prediction of Alzheimer's Disease via Longitudinal Clinical Informatics

Suvarna Joshi[1], Prasenjit Chatterjee[2,3]

[1] Postdoctoral Researcher, Lincoln University College, Petaling Jaya, Selangor Darul Ehsan Malaysia ; [2] Lincoln University College, Petaling Jaya, Selangor Darul Ehsan, Malaysia. [3] Department of Mechanical Engineering, MCKV Institute of Engineering, Howrah, West Bengal - 711204, India

Email ID : pdf.suvarna@lincoln.edu.my

Abstract Alzheimer's disease (AD) is a progressive neurodegenerative disease in which cognitive and functional decline may begin years before the onset of dementia. Therefore, early identification of high-risk individuals is essential for prompt monitoring and intervention. However, advanced diagnostic approaches based on MRI, PET, and cerebrospinal fluid biomarkers are expensive, infrastructure-intensive, and difficult to implement for large-scale screening. This paper proposes ClinicoFusion-Net, a non-invasive longitudinal machine-learning framework for early AD risk prediction using routinely available clinical and cognitive data. The framework integrates cognitive assessments, functional measures, demographic characteristics, medical history, and longitudinal changes across clinical visits. A multi-domain feature representation is processed through a Bidirectional Gated Recurrent Unit (Bi-GRU) with temporal attention to capture non-linear patterns of cognitive and functional decline. SHAP-based explainability is incorporated to identify the clinical factors contributing to individual risk predictions. The proposed framework aims to provide an accessible and interpretable approach for risk stratification across Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and AD, while supporting future progression-risk estimation.

Keywords: Alzheimer's disease; early risk prediction,; longitudinal clinical data; Mild Cognitive Impairment

---

Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disease and a major cause of age-related cognitive decline. Pathological changes can develop years before the clinical onset of dementia, presenting important opportunities for early detection and intervention [1]. Although the progression of Alzheimer's disease is highly heterogeneous, mild cognitive impairment (MCI) represents an important step on this continuum [2]. In daily medical practice, cognitive assessments such as MMSE, MoCA, ADAS-Cog, CDR-SB, and FAQ allow for relatively inexpensive and non-invasive measurements of cognitive status. However, due to interindividual variability, learning differences, cognitive reserve, and demographics, a single score may not fully reflect disease progression. Nonlinear pathways can also be observed in early cognitive decline.

MRI, PET, and CSF biomarkers are examples of advanced diagnostic techniques that provide useful information, but may not be suitable for widespread screening due to infrastructure, cost, and availability issues [3 , 4].Consequently, longitudinal clinical data provide a promising foundation for scalable AD risk

prediction. Instead of analyzing only the current clinical state, longitudinal models can learn changes in cognitive and functional measures over time..

Related work

Machine learning has increasingly been applied to AD and MCI progression prediction. Existing studies have used clinical, cognitive, neuroimaging, genetic, and biomarker variables. However, a significant proportion of published research relies on MRI, PET, or multimodal datasets, which can reduce practical applicability in primary healthcare [5].

Clinical and cognitive measures have demonstrated considerable predictive value. A systematic review of machine-learning approaches for MCI progression reported that cognitive variables were among the most informative predictors of conversion to AD [5]. Measures such as MMSE, ADAS-Cog, CDR-SB, and functional assessments provide information directly related to cognitive and functional deterioration.

An important limitation of conventional clinical models is their reliance on cross-sectional or baseline observations. Longitudinal studies have demonstrated that incorporating repeated clinical observations can improve prediction. An ADNI-based study using longitudinal observations reported improved MCI-to-AD prediction compared with a cross-sectional approach, demonstrating the value of temporal information [6].

Deep-learning models such as LSTM and GRU are particularly suitable for longitudinal data because they can model dependencies across multiple clinical visits. Recent research has therefore explored recurrent architectures for predicting progression from MCI to AD [6,7].

Another important requirement is interpretability. Clinical users need to understand which variables contribute to a predicted risk. Explainable AI methods such as SHAP can identify influential features and provide patient-specific explanations.

Area	Existing Limitation	Proposed Approach
Clinical assessment	Often cross-sectional	Longitudinal modeling
Imaging-based models	High cost	Clinical-only inputs
Feature representation	Single-domain	Multi-domain fusion
Interpretability	Black-box prediction	SHAP explanations

3. Proposed Methodology

3.1 Clinical Data and Preprocessing

The proposed ClinicoFusion-Net uses longitudinal, non-invasive clinical data for early Alzheimer's disease risk prediction. The input features include cognitive measures such as MMSE, MoCA, and ADAS-Cog13, functional measures including CDR-SB and FAQ, along with demographic characteristics, medical history, comorbidities, and subjective memory complaints. While deliberately excluding out MRI, PET, and CSF biomarkers from the model inputs, the study can make use of longitudinal records from the ADNI database [5].Following the proper imputation of missing information, patient visits are normalized and aligned chronologically. The direction of cognitive and functional decline is represented by changes between subsequent clinical evaluations.

3.2 Multi-Domain Feature Fusion

Cognitive, functional, demographic, medical, and temporal domains comprise the clinical variables. A cohesive picture of each patient's clinical state and history of progression is produced by integrating these domains.

A gradient-boosted feature-selection stage is used to identify the most informative variables and reduce redundant clinical information.

3.3 Temporal Risk Prediction

The selected longitudinal features are processed using a Bidirectional Gated Recurrent Unit (Bi-GRU) with temporal attention. The Bi-GRU captures relationships between observations across successive visits, while the attention mechanism identifies clinically important time points and changes.

The framework provides classification across CN, EMCI, LMCI, and AD, while its primary objective is to estimate the risk of future progression, particularly from MCI toward AD. Longitudinal machine-learning studies have demonstrated the value of repeated clinical observations for improving progression prediction compared with cross-sectional models [6,7].

3.4 Explainability

SHAP (SHapley Additive exPlanations) is incorporated to identify the clinical features contributing to individual risk predictions. This enables the model to provide both a risk estimate and an interpretable explanation, such as declining cognitive scores or worsening functional measures.

4. Conclusion

This paper proposes ClinicoFusion-Net, an explainable longitudinal framework for early-stage Alzheimer's disease risk prediction using non-invasive clinical and cognitive data. The literature indicates that cognitive measures are important predictors of MCI progression and that longitudinal observations can provide additional information compared with isolated baseline assessments. The proposed paradigm uses Bi-GRU with temporal attention to identify patterns of long-term decline by integrating cognitive, functional, demographic, medical, and temporal characteristics. SHAP-based explainability highlights clinical variables that influence specific predictions. In medical settings where advanced imaging and biomarkers are not available, the proposed strategy represents a potentially accessible, scalable, and interpretable alternative for early assessment of Alzheimer's disease risk. Future studies should validate the model using independent cohorts and longitudinal ADNI data, focusing on predicting progression from MCI to AD at 12, 24, and 36 months.

References

1. Jack, C. R., Jr., et al. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia, 14(4), 535–562.
2. Albert, M. S., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease. Alzheimer's & Dementia, 7(3), 270–279.
3. Scheltens, P., et al. (2021). Alzheimer's disease. The Lancet, 397(10284), 1577–1590.
4. Hampel, H., et al. (2023). Blood-based biomarkers for Alzheimer's disease. Nature Reviews Neurology, 19, 355–376.
5. Lee, S. W., et al. (2020). Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review. Medical Image Analysis, 67, 101810.
6. Prediction of Progression from Mild Cognitive Impairment to Alzheimer's Disease with Longitudinal and Multimodal Data. (2024). Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring.

7. Gelir, F., Akan, T., Alp, S., et al. (2025). Machine learning approaches for predicting progression to Alzheimer's disease in patients with mild cognitive impairment. Journal of Medical Systems.
8. Weiner, M. W., et al. (2017). The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. Alzheimer's & Dementia, 13(5), 561–571.
9. Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. Journal of Internal Medicine, 256(3), 183–194.
10. Livingston, G., et al. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. The Lancet, 396, 413–446.