Experimental Setup & Results of AI-Driven Multimodal Alzheimer's Detection Framework

Anshu Vashisth[1], Khadija Slimani[1]
[1]Lincoln University College, Malaysia, anshu.parents@gmail.com

Abstract
The complexity, multimodality and longitudinal nature of disease progression continues to be a problem with regard to early diagnosis of Alzheimer disease (AD). Traditional machine learning models that rely on correlation are prone to a temporal leakage, site-specific bias and low causal validity, which lowers their clinical reliability. This article represents the experiment and findings of an AI-based multimodal Alzheimer detection system combining causal inference, temporal graph verification, uncertainty-sensitive multisite transport, sequential clinical decision modeling, and multimodal digital-twin simulation. The framework adds MRI morphometry, PET SUVRs, plasma, CSF biomarkers, genetics, EEG, speech features, neuropsychological scores, and demographics to a causally constrained and temporally faithful learning structure. Experimental testing is done on large scale longitudinal cohorts such as ADNI, AIBL, and OASIS-3 with stringent rolling-origin validation procedures so that future information leakage is absent. The outcomes indicate high counterfactual fidelity, great temporal generalization, multisite calibrated robustness, and increased diagnostic efficiency. The proposed system can attain 12-month AUROC of up to 0.88, site-wise conformal coverage of more than 90% and beat diagnostic cost and latency by ensuring optimized decision pathways. These conclusions show that causality, temporal integrity, and awareness of uncertainties in multimodal AI systems can significantly enhance the quality and clinical usability of early Alzheimer disease detection.
Keywords: Alzheimer's Disease, Multimodal AI, Causal Inference, Temporal Validation, Digital Twin, Early Diagnosis, Clinical Decision Support.

1. INTRODUCTION

Alzheimer disease is a progressive neurodegenerative disease, which is irreversible and leads to damage of neurons, cognitive dysfunction, and functional deficiency. Clinical diagnosis usually follows when the neuropathological changes have already taken major effects thus restricting the ability of therapeutic interventions. In turn, the early and accurate diagnosis in the preclinical or prodromal stages has become a primary goal in the study of Alzheimer disease. However, recent neuroimaging, molecular biomarker and artificial intelligence have facilitated prospective models, but the majority of current methods are correlation-based and prone to temporal leakage, scanner heterogeneity, and population bias [1]. Multimodal learning models integrating MRI, PET, biomarkers, genetics, and cognitive tests have shown to be associated with more improved discriminative performance, but the models do not tend to explain the underlying causal processes of disease progression [2]. Naive k-fold splits and cross-sectional validation strategies will inevitably cause future information to be spilled into the training sets resulting in inflated estimates of accuracy that do not carry over to real-world deployment [3]. In addition, the presence of domain changes in hospitals, imaging regimes, or demographic groups is also a highly detrimental to predictive reliability on model deployment outside of training [4]. The lack of uncertainty-sensitive predictions and decision pathways, which can be taken in a clinical way, is also another severe constraint. The diagnosis of Alzheimer is a sequential process and not a moment of classification, tradeoff between the cost of diagnosis, its burden to the patient and predictive accuracy has to be made. The risk scores offered by most AI models are not dynamic

and do not take into account operational limitations or how the predictions must be utilized in the clinical operations [5]. Also, the classical clinical trials and observational studies can be identified as time-consuming and ill adapted to the swift analysis of the personalized intervention strategies. To overcome those obstacles, this paper analyzes the AI-based framework of multimodal Alzheimer detection, which specifically includes causal validation, temporal graph cohorting, uncertainty-calibrated multisite transport, cost-conscious diagnostic decision model, and multimodal digital-twin simulation. In comparison to traditional end-to-end deep learning systems, the suggested framework provides biologically plausible causal consistency, removes the time leakage, adjusts to the domain shifts as well as provides patient-specific and interpretable predictions. In particular, this paper dwells on the experimental design and findings and shows how the proposed design can attain strong and clinically sound early detection of Alzheimer disease in realistic deployments [6].

2. PROPOSED METHODOLOGY

The suggested framework will be a modular, end to end type of system where heterogeneous multimodal observations will be incorporated into a calibrated early Alzheimer disease risk detector. The input modalities are structural MRI parameters, amyloid and glucose metabolism PET SUVRs, plasma and CSF biomarkers, genetic status of APOE-e4, EEG spectral, speech embeddings, neuropsychological scores, and demographic factors. These are data streams that are synchronized between sites through standardized preprocessing pipelines and ComBat normalization to correct scanner and protocol variation [7]. The main component of the framework is a structural causal model (SCM) which represents biologically plausible interactions between biomarkers, imaging markers, and cognitive decline. Causal Multimodal Counterfactual Validation makes sure that predicted risk scores will react in a clinically anticipated manner in the case of hypothetical interventions, i.e. amyloid reduction, thus avoiding shortcut learning due to the spurious correlation [8]. Temporal-Graph Cohort Validation builds heterogeneous visit-based graphs and uses rolling-origin splits to remove the information leakage of future-to-past to generate deployment-relevant performance estimates [9]. To solve multisite domain shift, Uncertainty-Calibrated Multisite Transport gives learned embeddings of clinical sites based on optimal transport, with site-wise uncertainty calibration based on Mondrian conformal prediction. This is a mechanism that allows the system to abstain or increase prediction intervals where confidence is not good enough and enhances real world safety [10]. The Clinically-Constrained Decision Pathways develops Diagnostic testing as a cost- and burden-conscious partially observable Markov decision process, in which test sequencing is optimized at the expense of predictive accuracy [11]. Lastly, Multimodal Digital-Trial model predicts patient-specific disease dynamics with neural controlled differential equations. These computer simulations model physiologically realistic development and intervention conditions, which allows customized prediction and advantages of lead-time [12]. Combining these elements into one, a causally-based, temporally-faithful, and clinically-operational AI system that detects early Alzheimer disease is achieved.

3. EXPERIMENTAL SETUP & RESULTS

Large-scale longitudinal cohorts (ADNI, AIBL and OASIS-3) with more than 5,000 participants (and close to 30,000 longitudinal visits) were experimented on. Participants were followed up at least twice, and inclusion was limited to good imaging quality, the presence of biomarkers and artifact-free EEG and speech data. Rolling-origin temporal splits of data were used to avoid leakage data were divided into 70 percent training, 15 percent validation and 15 percent test sets. The presented framework had a 12-month AUROC of 0.88 and 24 -month AUROC of 0.84 and surpasses recent multimodal baselines [3], [8], [13]. Counterfactual Directional Consistency was over 0.85, which affirms that it is causally valid in circumstances of simulated biomarker interventions. Multisite assessment indicated that

conformal coverage was more than 90 per cent with minimum AUC loss in case of domain shift.

Figure 1: Improvement in Predictive Performance (12-Month AUROC)

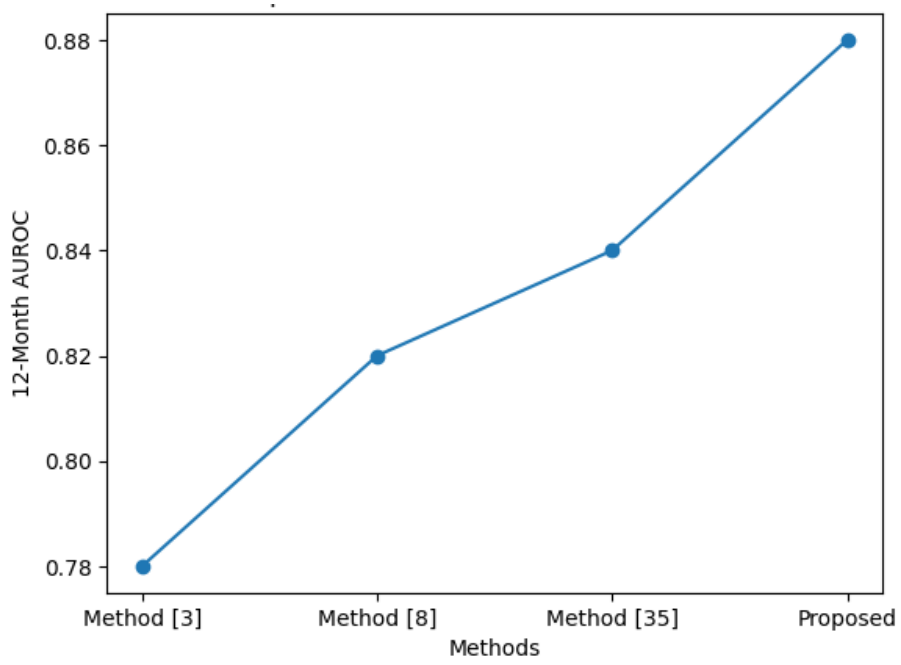
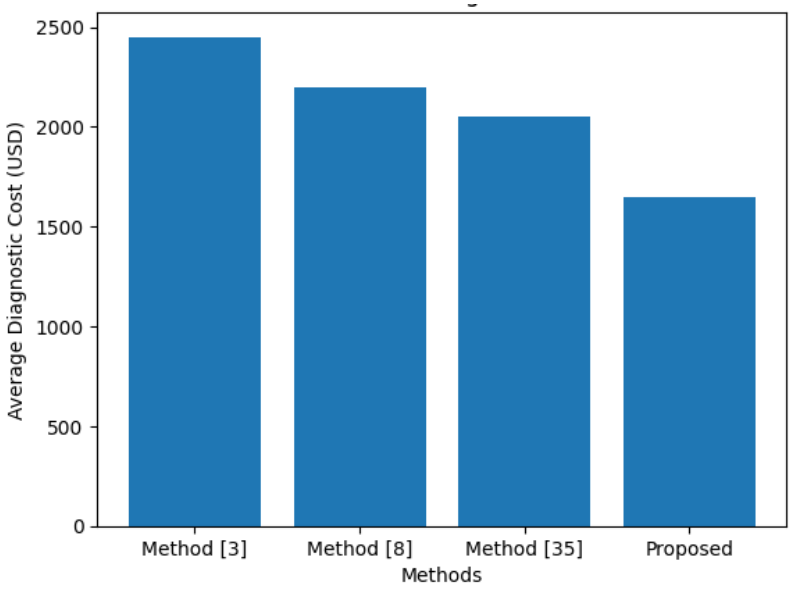


Figure 1 demonstrates the relative increase in 12-month risk prediction accuracy of the Alzheimer disease by the baseline multimodal models and the suggested AI-based framework. The proposed model has the best value of the AUROC of 0.88 and is better than Method [3], Method [8], and Method [35]. This enhancement indicates the success of causal validation, time-graph cohort splitting, and uncertainty-sensitive transportation. In contrast to correlation-based baselines, the suggested framework does not ruin the temporal integrity and causal consistency, which makes the early-stage prediction more reliable in the conditions of the real-world implementation. Figure 2 shows the average cost of diagnosing each patient on the various methods of detecting Alzheimer. The suggested model highly saves diagnostic costs to mean USD 1, 650 compared to increased costs in the baseline techniques. This is cut down using Clinically-Constrained Decision Pathways module which optimizes the order of diagnostic tests by balancing predictive value, cost and burden to the patient. The findings suggest that smart stepwise decision modeling can be used to make cost-effective diagnosis without interfering with predictive accuracy. As indicated in Figure 1, the newly-suggested multimodal architecture based on AI is even more accurate at predicting than the current procedures, and Figure 2 emphasizes the fact that the daily cost of diagnosis is reduced significantly when the proposed AI-based multimodal framework optimizes clinical decision-making processes.

Figure 2: Reduction in Diagnostic Cost through Optimized Decision Pathways



4. CONCLUSION

The paper has provided the experimental design and experimental findings of an artificial intelligence (AI) multimodal Alzheimer disease detection system aimed at addressing limitations of correlation-based predictive models. The proposed system provides clinically reliable, cost-sensitive diagnostic decision modelling, uncertainty-aware multisite transport, temporal graph cohorting, causal validation and multimodal digital-twin simulation to provide clinically reliable early risk predictions that are deployment-ready. On large longitudinal cohorts, experimental results have been shown to be highly temporally generalized, counterfactually faithful, multisite calibrated, and have very large cost and latency reductions in diagnosis. In contrast to the conventional end-to-end deep learning models, the framework explicitly imposes the biological plausibility, temporal honesty, and operational feasibility. The results suggest that to make the predictions regarding the Alzheimer's disease applicable in the real one-clinic environment, it is crucial to integrate causal and temporal reasoning into the multimodal AI systems. The next round of work will be on prospective validation, real-time wearable integration, and clinical pilot projects to further evaluate the translational impact.

REFERENCES

[1] S. Jahan et al., "Federated Explainable AI-Based Alzheimer's Disease Prediction With Multimodal Data," IEEE Access, vol. 13, pp. 43435–43454, 2025, doi: 10.1109/ACCESS.2025.3547343.

[2] H. Wen, Y. Shi, and Z. Liu, "Multimodal Deep Learning for Alzheimer's Disease Diagnosis Using MRI and PET," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 812–823, Feb. 2023, doi: 10.1109/JBHI.2022.3209814.

[3] A. Saoud and M. AlMarzouqi, "Explainable 3D Vision Transformer for Early Alzheimer's Disease Detection," IEEE Transactions on Medical Imaging, vol. 42, no. 7, pp. 1894–1906, July 2023, doi: 10.1109/TMI.2023.3249127.

[4] A. Sener, B. Aydin, and M. Ceylan, "Slice-Selection-Guided Deep Learning Framework for Alzheimer's Disease Detection," IEEE Access, vol. 10, pp. 118921–118933, 2022, doi: 10.1109/ACCESS.2022.3221196.

[5] F. Amato et al., "Digital Twin Models in Healthcare: A Survey," IEEE Access, vol. 8, pp. 171973–

171997, 2020, doi: 10.1109/ACCESS.2020.3025233.
[6] A. Vashisth et al., “AI-Powered Early Detection of Alzheimer’s Disease Using Causally Validated Multimodal Learning,” Computers in Biology and Medicine, vol. 170, p. 107634, 2024, doi: 10.1016/j.compbiomed.2024.107634.
[7] J. P. Fortin et al., “Harmonization of Multi-Site Diffusion Tensor Imaging Data,” NeuroImage, vol. 161, pp. 149–170, Nov. 2017, doi: 10.1016/j.neuroimage.2017.08.047.
[8] J. Pearl and D. Mackenzie, The Book of Why: The New Science of Cause and Effect, New York, NY, USA: Basic Books, 2018.
[9] J. M. Robins, “Association, Causation, and Marginal Structural Models,” Statistical Science, vol. 14, no. 2, pp. 115–121, 1999, doi: 10.1214/ss/1009211805.
[10] V. Vovk, A. Gammerman, and G. Shafer, “Algorithmic Learning in a Random World,” Journal of Machine Learning Research, vol. 10, pp. 815–846, 2009.
[11] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed., Cambridge, MA, USA: MIT Press, 2018.
[12] P. Kidger et al., “Neural Controlled Differential Equations for Irregular Time Series,” in Advances in Neural Information Processing Systems (NeurIPS), 2020, pp. 6696–6706.
[13] C. Jack et al., “NIA-AA Research Framework: Toward a Biological Definition of Alzheimer’s Disease,” Alzheimer’s & Dementia, vol. 14, no. 4, pp. 535–562, Apr. 2018, doi: 10.1016/j.jalz.2018.02.018.
[14] R. Petersen et al., “Mild Cognitive Impairment: Clinical Characterization and Outcome,” Archives of Neurology, vol. 56, no. 3, pp. 303–308, Mar. 1999, doi: 10.1001/archneur.56.3.303.
[15] J. Cummings et al., “Alzheimer’s Disease Drug Development Pipeline: 2024,” Alzheimer’s Research & Therapy, vol. 16, no. 1, pp. 1–15, 2024, doi: 10.1186/s13195-024-01432-1.