

Multimodal Detection and Prediction of ADRD Using Electronic Health Record Data

Dr. Korin Reid, PhD
Ellison Laboratories

The AIM-AHEAD program is funded by NIH, Agreement No.: 1OT2OD032581-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This study utilized data provided by the Data Bridge from MedStar Health (AADB) formed under The National Institutes of Health (NIH)'s AIM-AHEAD Program. We acknowledge the contributions of the AADB team in ensuring the integrity and accessibility of the data, which was critical to the analyses conducted in this research.

INTRODUCTION	METHODS	RESULTS	CONCLUSIONS	REFERENCES						
Alzheimer’s disease and related dementias (ADRD) impact more than 7 million Americans. ¹ This is expected to reach 13 million by 2060, representing an economic impact of more than \$1 trillion annually. ¹ Despite this significant disease burden, an estimated 61% of ADRD cases remain undiagnosed ² . In addition, FDA-approved disease-modifying treatments are most effective in the early stages of disease. Thus, early diagnosis is critical. However, various diagnostic instruments (MRI, PET, cognitive batteries, cerebral spinal fluid) are expensive, which limits their use at a population scale. Machine learning offers a scalable alternative. That is, models can mine large volumes of electronic health record (EHR) data to automatically flag patients at risk of ADRD. The patients can then be considered for further evaluation using standard diagnostic methods.	Data. Our study uses structured (diagnosis codes, procedure codes, vitals, medications, labs, problem lists, etc.), and unstructured (clinical notes) EHR data in addition to social determinants of health from the MedStar Health system, which includes 10 hospitals and more than 300 outpatient facilities. Cohort. Our cohort includes more than 140,000 patients. This includes 33% rural hospital representation, and a 43% Black patient population. Tasks. Our machine learning model tasks included detection of existing ADRD, and prediction of future ADRD up to 5 years in advance. Interpretability. We used Shapley Additive Explanations (SHAP) analysis to identify which clinical features most influence model predictions.	<table><tr><th>Task</th><th>AUROC</th></tr><tr><td>Detecting existing ADRD</td><td>0.861</td></tr><tr><td>Predicting ADRD, 5 years in advance</td><td>0.733</td></tr></table> Feature ablation analysis demonstrated that including unstructured clinical notes and social determinants of health improved model performance relative to inclusion of structured EHR data alone. In addition, SHAP analysis identified clinically meaningful drivers of ADRD risk. <i>Full model details and feature-level findings appear in our forthcoming publications.</i>	Task	AUROC	Detecting existing ADRD	0.861	Predicting ADRD, 5 years in advance	0.733	Machine learning models can be used to both detect ADRD and predict it up to 5 years prior to diagnosis using a large repository of EHR data. Incorporating unstructured clinical notes and social determinants of health improves model performance. While biologically based biomarkers and robust evaluation are critical to the ADRD diagnosis process, machine learning could offer a scalable and cost-effective means of triaging large volumes of patients for potential follow-up evaluation using standard diagnostic means. This could lead to earlier diagnosis, which could enable more effective treatment options for patients.	1. Alzheimer’s Association. Alzheimer’s Disease Facts and Figures. alz.org/alzheimers-dementia/facts-figures 2. Lang L, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open. 2017;7:e011146. Contact Korin Reid CEO, Ellison Laboratories Korin.reid@ellisonlabs.com ALZHEIMER'S ASSOCIATION 
Task	AUROC									
Detecting existing ADRD	0.861									
Predicting ADRD, 5 years in advance	0.733									