Replicating Alzheimer's Research using standardized phenotyping with the OMOP common data model imaging extension

Gabriel Lucca de Oliveira Salvador[1], Jen Park[1,2], Teri Sippel Schmidt[1,2], Blake Dewey[1,2], Paul Nagy[1,2]

1- Division of Biomedical Informatics and Data Science, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

2- OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA

Background

Alzheimer's Disease (AD) is the most common form of dementia, affecting over 55 million people worldwide, with projections estimating 78 million in 2030 and 139 in 2060[1]. Neuroimaging plays a pivotal role in diagnosing and managing AD, complementing clinical assessments by identifying cortical and subcortical changes and enabling early detection through molecular PET imaging and MRI. These and other advanced imaging technologies now support more accurate and quantitative analyses of brain pathology[2].

Healthcare systems and research groups struggle with inconsistent imaging data, poor quality, and challenges in integrating clinical and imaging finding when working with AD data[3]. Variability in image acquisition, segmentation, and feature extraction hampers reproducibility especially in imaging research[3]. Moreover, defining AD research cohorts is complicated by the disease's heterogeneity, which can bias investigations and impede validation across datasets[4]. Lastly, without standards based transparent computational methods, it is unlikely this research can be replicated at other institutions leading to challenges in turning research into effective generalizable clinical evidence. These factors often lead to an inefficient and costly increase throughout the research process, even in observational studies related to AD.

The factors described above reflect some of the main challenges, especially in conducting research related to AD. This study describes the efforts to integrate different data sources encompassing clinical, laboratory, imaging data and volumetric into a single Common Data Model (CDM) to create cohorts of patients with AD, providing the highest amount of clinical information and essential imaging attributes for research to solve the problems above mentioned.

There are three challenges we are addressing that taken together represents a systematic way to approach multi-modal imaging research.

1. Standardizing Inclusion/exclusion criteria: A reproducible approach to defining patient inclusion/exclusion criteria using clinical phenotyping based upon the DICOM acquisition parameters and clinical features coded into the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

2. Reproducible Machine Learning Pipelines: Use of an automated imaging pipeline process to generate and verify results from deep learning segmentation algorithms.

3. Reproducible Analysis of results: Those imaging segmentation features are then submitted as new measurements in the OMOP CDM. This then allows us to use transparent and standardized analytics to generate a table one comparison to replicate the original paper.

Our work aims to leverage the OMOP-CDM and its medical imaging extension to integrate different aspects and create a platform to address the challenges described above.

Methods

Our platform integrates clinical and imaging data using the OMOP CDM and its extension for medical imaging. The clinical dataset follows the OMOP CDM structure, encompassing tables such as Person, Visit Occurrence, Condition Occurrence, Drug Exposure, Measurement, and others. These tables were populated with structured clinical, laboratory data, particularly for patients with AD. Imaging data, originally in DICOM format, were integrated by mapping standard DICOM attributes (DICOM tags) into the OMOP CDM using the Medical Imaging extension (MI-CDM), as proposed by Park et al[5]. Volumetric results from machine learning algorithms were stored in the Measurement table.

To ensure reproducibility in image analysis, we employed OpenMap-T1[6], an open-source deep learning framework optimized for automated segmentation of T1-weighted brain MRIs. It includes steps such as preprocessing, skull stripping, and parcellation into 280 brain structures using convolutional neural networks (CNNs). The algorithm, previously trained using our institutional image database, is highly efficient and suitable for large-scale research, reducing processing time to under 90 seconds. It is particularly robust across MRI scanners and commonly used in dementia research to identify hippocampal atrophy.

For reproducible analysis workflows, we used the OHDSI ATLAS tool to define clinical cohorts. ATLAS-generated queries returned both clinical data and Study Instance Unique Identifiers (UIDs), which were used to retrieve imaging series for analysis with OpenMap-T1. Volumetric data were stored and linked to the originating imaging studies using consistent identifiers.

Our data source was the Alzheimer's Disease Neuroimaging Initiative (ADNI)[7], version 4 (ADNI-4), which includes imaging and clinical data for AD, Mild Cognitive Impairment (MCI), and cognitively normal individuals. The use of ADNI data was pre-approved, and ethical review at our institution was waived.

To validate the platform, we replicated a published study that analyzed hippocampal volume across cognitive groups[8]. We selected a similar cohort from ADNI-4 and ran volumetric analyses using OpenMap-T1. The segmentation pipeline included standard preprocessing steps and measurement of hippocampal volumes. We reproduced comparisons by age group, gender, and diagnosis (AD, MCI, CN), and created summary tables to mirror the original study's structure

Results

We extracted 2,824 DICOM attributes and 5,223 value sets from imaging data, integrating them as custom concepts in the OMOP CDM. The dataset included 545 DICOM studies (4,756 series, 14,816 images) from 289 patients and 4,152 clinical records. Neuropsychiatric scores (88,819 entries) were mapped to the Measurement table. Using OHDSI's ATLAS, we selected volumetric T1-weighted MRI series based on acquisition parameters and clinical criteria. Final hippocampal volume values and summary statistics were visualized in ATLAS.

Volume trends showed progressive hippocampal atrophy with age and dementia severity across control, MCI, and AD groups. Data were stratified by gender and age. Table 1 presents volumetric comparisons between our platform (OpenMap-T1) and the original study (OS). OpenMap-T1 analyzed the entire hippocampal formation (head, body, tail), while the original study focused mainly on the hippocampal head (Figure 1). This broader segmentation may explain volume differences and supports detailed structural analysis relevant to AD research.

	average right hippocampus volume (mm^3) - OpenMap-T1	average right hippocampus volume (mm^3) - OS	average right hippocampus volume (mm^3) - OpenMap-T1	average right hippocampus volume (mm^3) - OS	average left hippocampus volume (mm^3) - Open Map T1	average left hippocampus volume (mm^3) - OS	average left hippocampus volume (mm^3) - Open Map T1	average left hippocampus volume (mm^3) - OS
	Female	Female	Male	Male	Female	Female	Male	Male
CN (147)								
55-65	-	-	3750	2582	-	-	3776	2648
66-79	3451	2518	3838	2570	3437	2447	3815	2521
80-95	3261	2423	3587	2481	3335	2392	3644	2414
MCI (89)								
55-65	3346	2211	4166	2414	3422	2197	3897	2392
66-79	3256	1927	3571	2005	3180	1887	3601	1946
80-95	3093	1863	3319	1726	3057	1833	3255	1731
AD (26)								
55-65	1985	1871	3630	1774	1730	1763	3414	1741
66-79	3198	1474	3468	1506	3094	1390	3441	1378

Table 1– hippocampal volume values in relation to age range and gender, OS- Original Study, CN – normal controls, MCI – mild cognitive impairment and AD – Alzheimer disease, OS – Original Study. The authors of the original study did not evaluate the volumes for females between 55 and 65 years old, so to maintain the replication aspect, we did not evaluate them either.

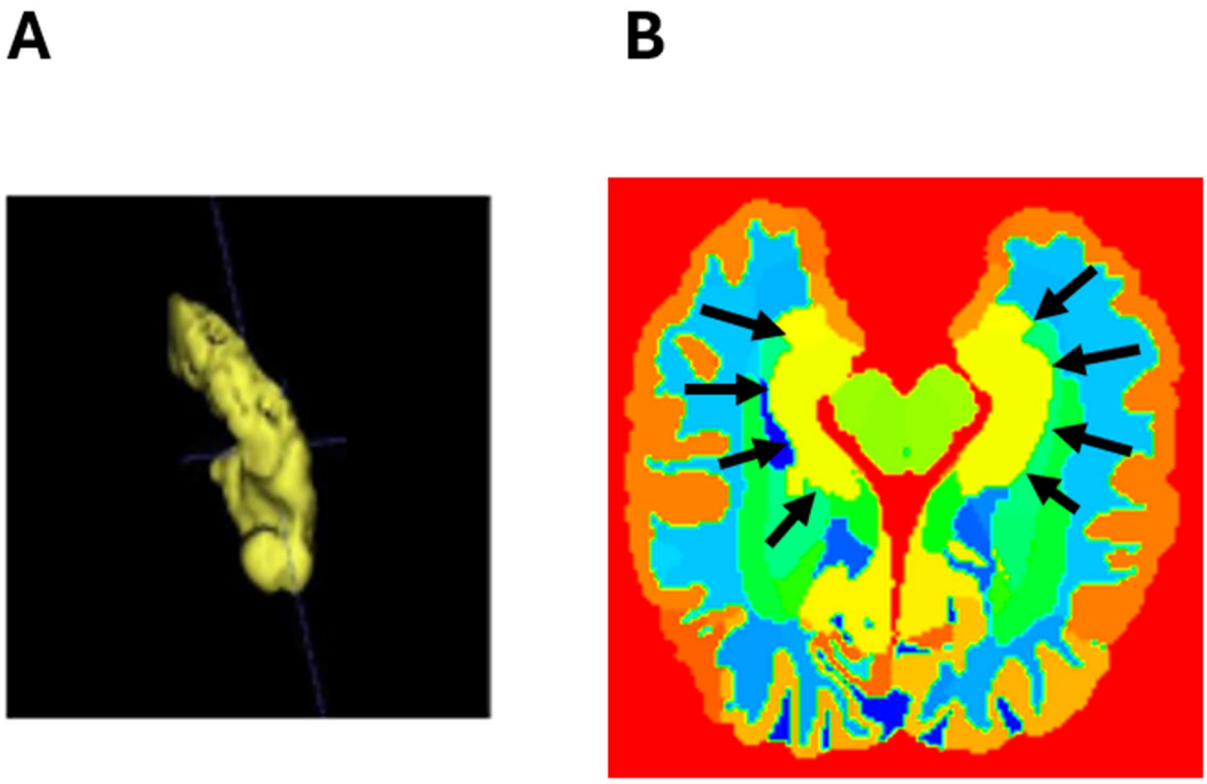

Figure 1 – Differences in segmentation patterns between the original studies and OpenMap-T1. A - Image from the study replicated showing the authors' semi-manual segmentation of one of the hippocampal formations.

B – Segmentation performed by OpenMap-T1 illustrating the head, body, and tail of the hippocampus (black arrows).

Conclusion

This study presents a novel platform integrating clinical and imaging data using the OMOP MI-CDM to accelerate cohort creation and reduce the time and effort required for Alzheimer's disease research. Traditional studies often take months to compile and process data, but our system enables rapid, reproducible cohort generation using ATLAS and automated segmentation via OpenMap-T1, which processes brain MRIs in under a minute. Despite minor differences in hippocampal volume due to segmentation scope and patient selection, the replicated studies confirmed similar data trends. This platform demonstrates strong potential for broader clinical radiology applications.

References

1. 2023 Alzheimer's disease facts and figures. Alzheimers Dement. 2023 Apr;19(4):1598-1695.
2. Whitwell JL. Alzheimer's disease neuroimaging. Curr Opin Neurol. 2018 Aug;31(4):396-404.
3. Dokholyan NV, Mohs RC, Bateman RJ. Challenges and progress in research, diagnostics, and therapeutics in alzheimer's disease and related dementias. Alzheimers Dement (N Y). 2022 Jul 26;8(1):e12330.
4. Dokholyan N, Bateman RJ, Mohs R & Carrillo M. The alzheimer's crisis: examining testing and treatment pipelines and fiscal implications. hearing before the subcommittee on health care of the committee on finance united states senate one hundred sixteenth congress second session. (2020).
5. Park WY, Jeon K, Schmidt TS, Kondylakis H, Alkasab T, Dewey BE, You SC, Nagy P. Development of medical imaging data standardization for imaging-based observational research: OMOP common data model extension. J Imaging Inform Med. 2024 Apr;37(2):899-908.
6. OpenMAP-T1 github page https://github.com/OishiLab/OpenMAP-T1.
7. ADNI website available at https://adni.loni.usc.edu/
8. Uysal G, Ozturk M. Hippocampal atrophy based Alzheimer's disease diagnosis via machine learning methods. J Neurosci Methods. 2020 May 1;337:108669