

Unlocking the Mysteries of Dementia: How Generative AI Helps Decode Patient Symptoms

Background

- Dementia encompasses diverse clinical syndromes affecting cognitive function, such as in Alzheimer's Disease (AD) and behavioral-variant Frontotemporal Dementia (bvFTD).
- The **heterogeneous presentation of symptoms often results in diagnostic challenges**, leading to delays and potential misdiagnoses.
- While electronic health records (EHRs) contain structured billing data, **crucial clinical information remains embedded within unstructured narrative notes**, presenting an opportunity for advanced natural language processing approaches.

Methods

- We leverage **Large Language Models (LLMs)** to analyze clinical data from the UCSF de-identified clinical data warehouse for patients with Memory and Aging Center (MAC) diagnoses.
- For efficiency, a local **LLAMA 3.3 model** was utilized to extract binary indicators of categories (clinical findings, diagnoses, measurements, medications, and medical history) by enforcing a structured JSON output schema via Pydantic (Figure 1A). This is done on notes from neurology, psychology, and related fields.
- GPT4o** via UCSF Versa API was utilized to identify findings for AD and bvFTD patients for this proof of concept (Figure 1B). Further clinical symptom embeddings that preserve semantic similarity were obtained using OpenAI's **text-embedding-3** model. Clusters of concept embeddings were obtained via HDBSCAN to reveal distinctive symptom patterns. UMAP was utilized for visualization in 2 dimensions.
- To enable phenotyping and insight, AD and bvFTD were compared based on both extracted symptom and symptom clusters. Patient embeddings were obtained by averaging across embedding of findings (Figure 1C).

Figures and Tables

Figure 1: Overview of Pipeline from Clinical Notes to Insight

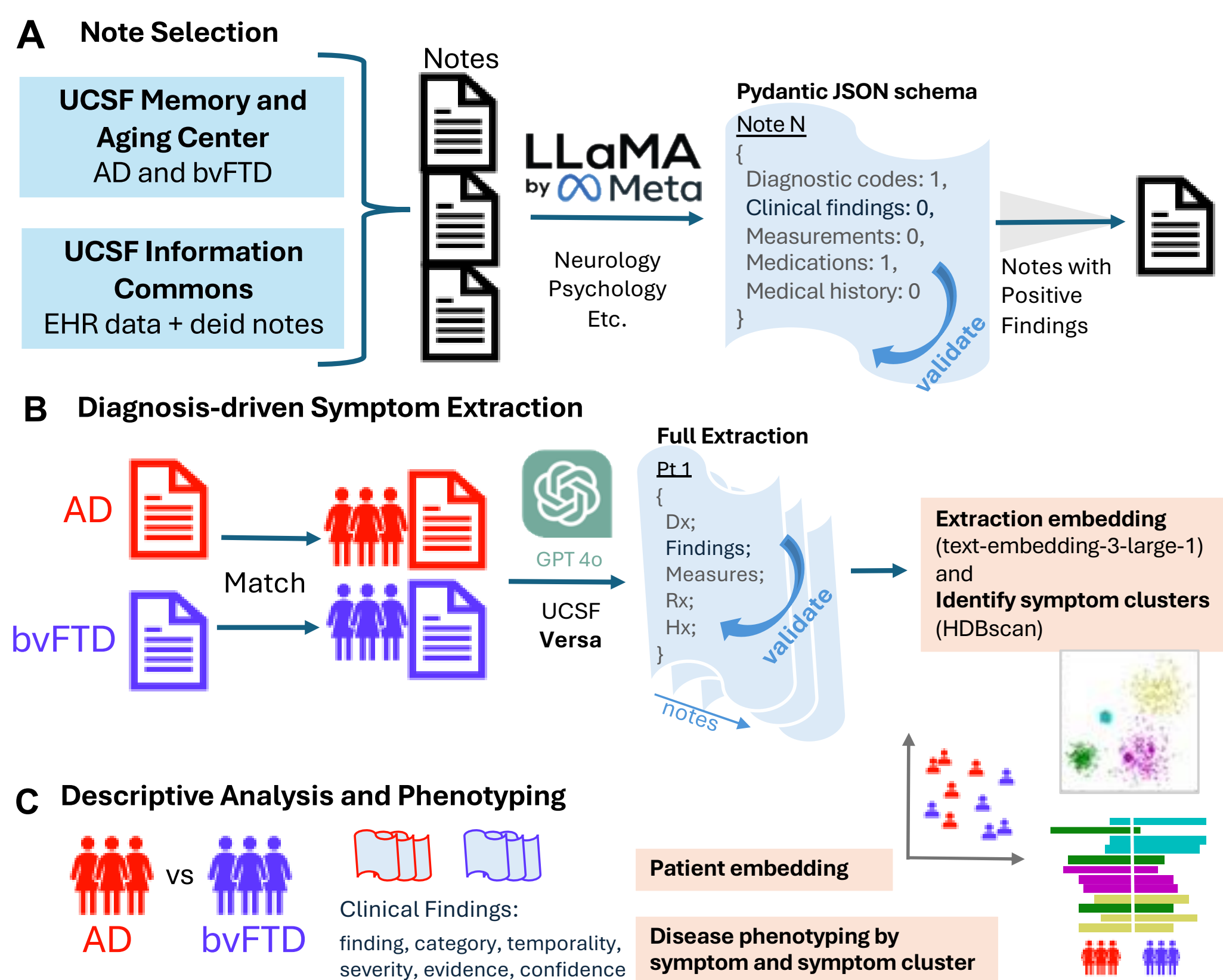


Figure 2. Extraction and Symptom Differences

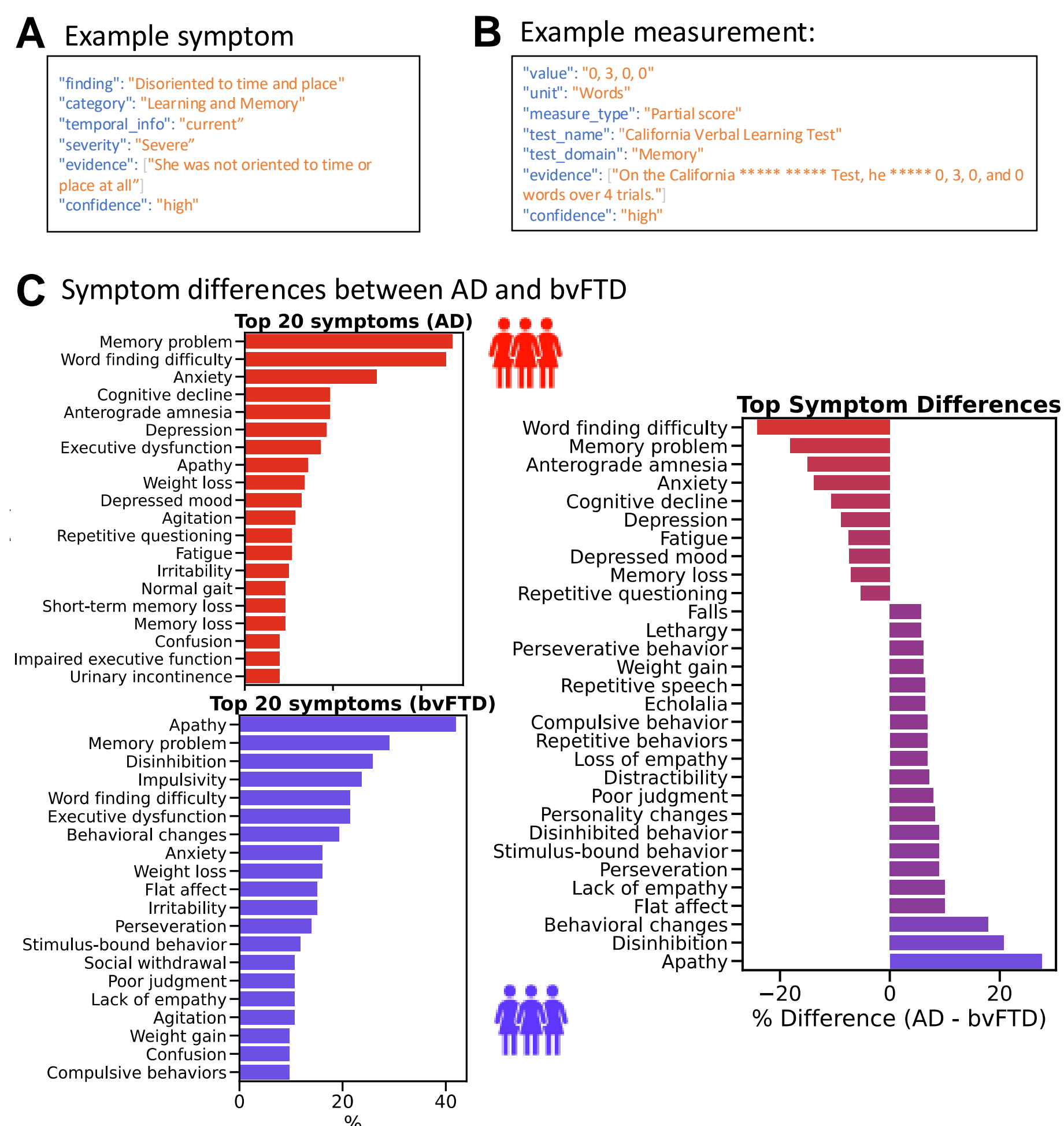
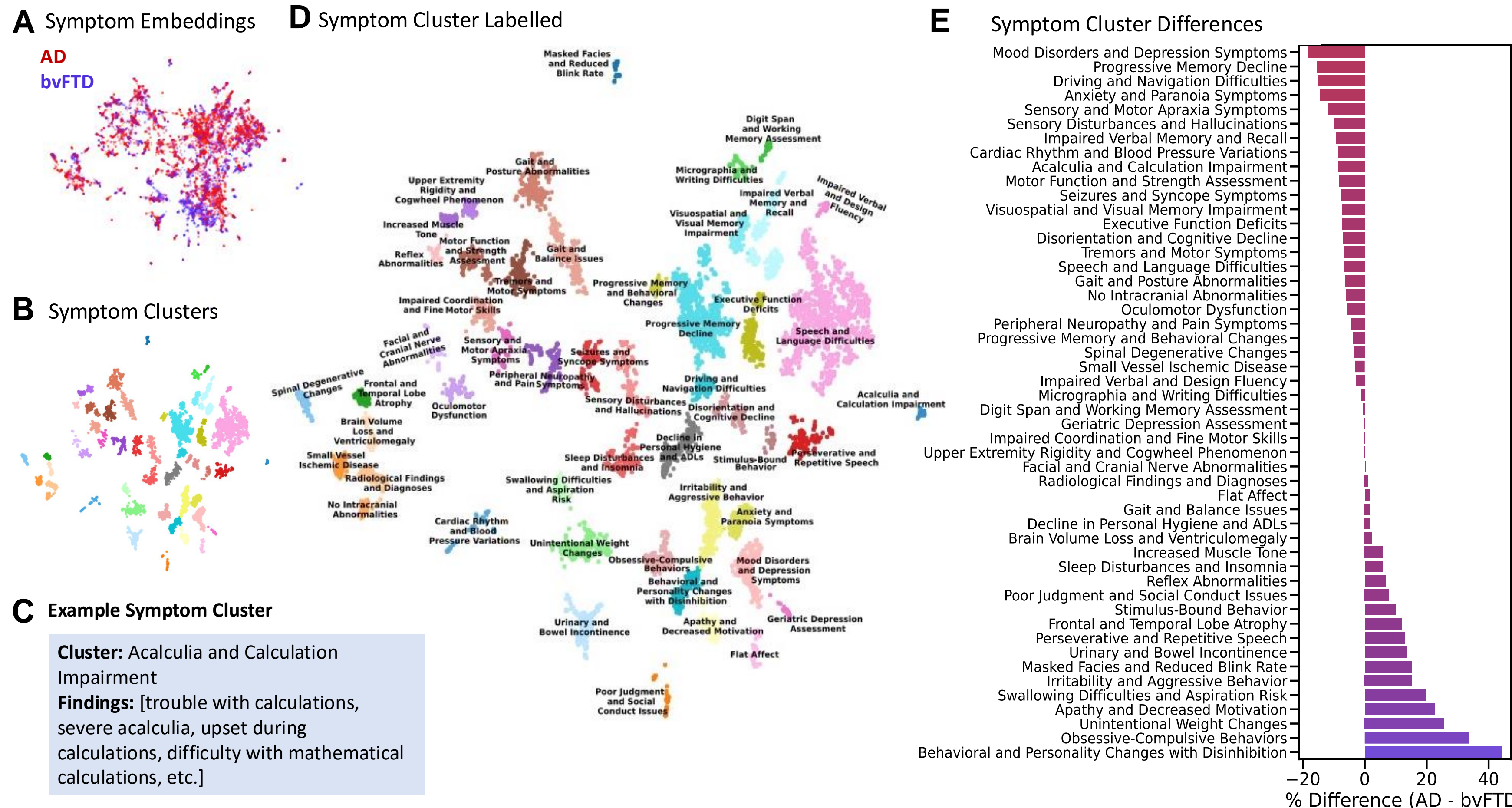


Figure 3. Embeddings preserve semantic meaning and enable comparison with symptom groups



Conclusions

- This novel approach, via LLM-based structured information extraction and traditional data analytics, demonstrates promising results for enhanced symptom characterization and phenotyping in dementia.
- Our findings suggest potential applications in improving diagnostic accuracy, developing prediction models, and optimizing treatment strategies in dementia care.
- We aim to extend this proof of concept to other domains (e.g., measurements, medications) and across dementia syndromes.

Acknowledgements: We would like to acknowledge the UCSF Memory and Aging Center and the UCSF Medical Scientist Training Program for support.

We would like to thank all the patients and providers that contributed to the data. We would further like to thank members of the Sirota Lab for feedback and insights.

Email: alice.tang@ucsf.edu. Twitter: @alicetang_mstp