Supplementary Materials for *Towards Early Dementia Detection: Fusing Linguistic and Non-Linguistic Clinical Data*

Appendix A: The ADNI Data Resource

The data used here was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (`adni.loni.usc.edu`). The methods employed here constitute a secondary use of the data for a purpose that is in line with the general goal of identifying dementia markers. The following two paragraphs are included verbatim, as required by the ADNI Data Use Agreement.

The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see `www.adni-info.org`.

Appendix B: Files Used with Text Data

Table 1 on the next page shows the files used with text from the ADNI data resource.

Appendix C: Overview of Modeling Stages

Figure 1 on the next page provides an overview of modeling stages.

File	Content description	Total # entries	# usable entries	# unique subjects
RECMHIST	Recent medical history	30,727	7,153	678
RECADV	Recent adverse events/hospitalizations	16,063	1,375	384
RECBLLOG	Symptoms at initial/baseline visit	12,768	2,156	585
BLCHANGE	Changes since initial/baseline visit	8,571	1,955	635

Table 1: Files used with text data. An *entry* refers to one medical visit. One subject may have multiple visits/entries. The last two columns identify the usable entries and subjects described in the main paper.

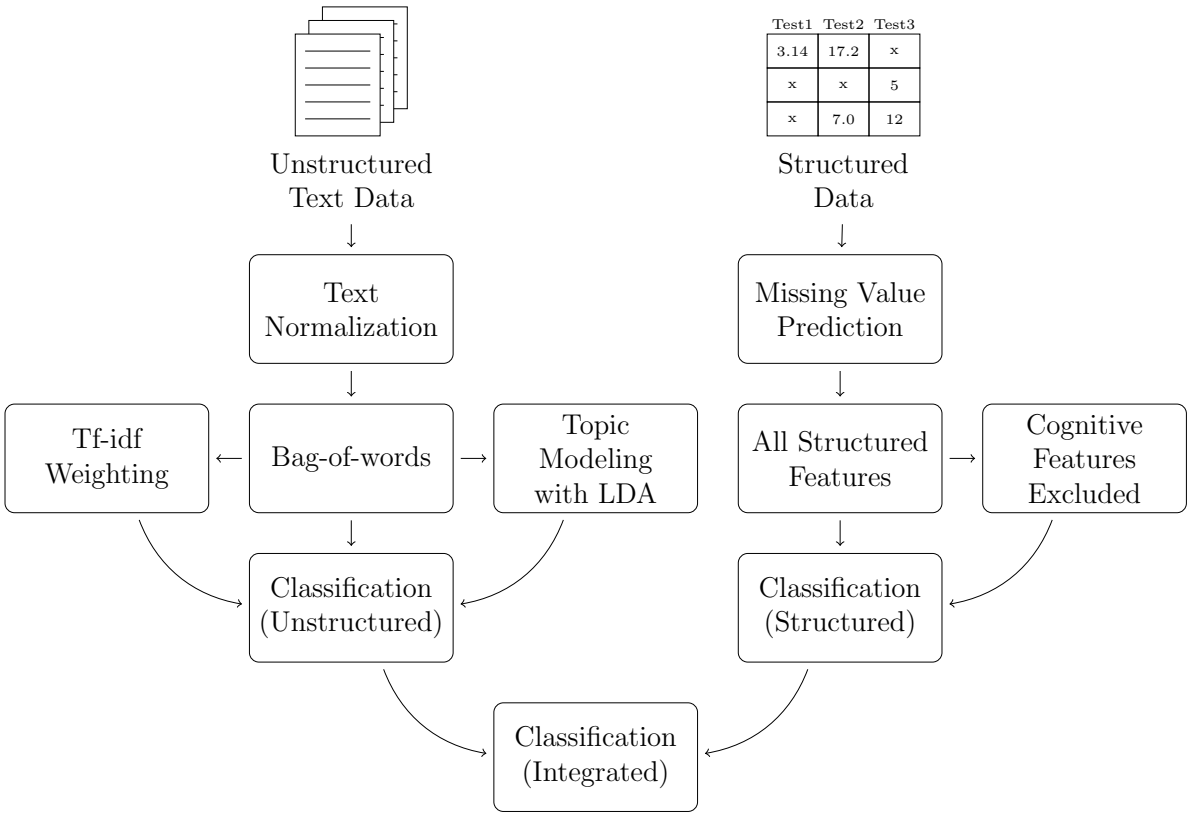

Figure 1: Modeling stages. Arrows indicate where the output of a stage feeds into the next.