Written Traces of Alzheimer's Disease

Travis Allen
Cornell Tech
New York, NY 10044, USA
Twa24@cornell.edu

Renee Zacharowicz
Cornell Tech
New York, NY 10044, USA
Rz336@cornell.edu

Max Grusky
Cornell Tech
New York, NY 10044, USA
grusky@cs.cornell.edu

Deborah Estrin
Cornell Tech
New York, NY 10044
USA
destrin@cornell.edu

Abstract

An unsatisfied demand to diagnose and prognosticate Alzheimer's Disease (AD) for early intervention and disease monitoring has motivated work to build predictive models with high specificity and sensitivity. Given that language is a known early marker for AD and that an average individual is amassing robust sets of text-based digital traces, it seems right to leverage these datasets to diagnose and monitor AD onset and progression. These data suggest that while text composed "in the wild" sufficiently detects language decline, cognitive decline as indicative of Alzheimer's onset detection requires a combination of text and behavioral data streams.

1.INTRODUCTION

Alzheimer's Disease (AD) is a chronic neurodegenerative disorder that is characterized by a progressive loss of cognitive function. The diagnosis for AD typically occurs at a stage in the disease when the available treatments have limited-to-no efficacy. A commonly cited reason for this delay is patient reluctance to see a specialist until the symptoms become undeniable [27]. The shortcomings of the standard diagnostic process has created a need for developing tools that can detect Alzheimer's onset early and in the wild.

One of the earliest markers of Alzheimer's onset is language deficiency which slowly and progressively declines over time, making it a possible marker for early detection. In its current form, cognitive exams screening for AD assess patients' "connected speech", a form of language that is typically elicited as a verbal response to particular stimuli: picture description and word recognition tasks [2]. These exams are administered in an inorganic clinical setting at arbitrary time-points. By concept and context, connected-speech assessments tenuously reflect an individual's everyday language ability and may not detect trends.

Constant monitoring of natural language could address these shortcomings, detecting trends in everyday language ability. With the widespread adoption of email and social media, individuals are constantly communicating online and their everyday language is being recorded in ways unimaginable in decades past. The availability of this written data presents a new opportunity to understand how one's quotidian writing patterns may be affected by the early stages of Alzheimer's Disease. Passive analysis of written digital traces would both detect Alzheimer's onset by monitoring organic trends and promote patient awareness.

As a constant and protected mode of written communication, emails may be an ideal window through which to monitor cognitive trends through changes in written language patterns. Robust models would accurately and consistently ground emails to validated cognitive language metrics and flag cognitive decline as evident by natural language deterioration. As a step toward making this vision a reality, this work seeks to correlate written text features from records composed in the wild of unknowable content and context with validated cognitive language metrics. These findings influence machine learning of written text features to model Alzheimer's risk.

2. RELATED WORK

Language decline and AD onset

Snowdon's seminal work- colloquially referred to as the *Nun Study* -drew a correlation between writing patterns and Alzheimer's development later in life. This study sparked decades of clinical research in linguistic ability and AD [10], binding language decline to cognitive decline in Alzheimer's progression [14].

Rather than presenting as a symptom somewhere along the disease's progression, there is data to suggest that language decline serves as a marker, presenting an average of 12 months prior to clinical diagnosis with significant differences between controls and mild AD and between MCI and moderate AD: relentless deterioration in consonance with disease progression [14][11].

Current diagnostic standards which rely on clinical judgement and patient report/demonstration of significant interference feature word-finding difficulty as a flag of (non-amnestic) Probable-AD, regarding biomarkers as neither necessary nor sufficient [17].

Language-based neurocognitive exams are included in the battery of tests given to symptomatic patients. Some exams measure *connected speech* which anticipates speech output and score semantic units accordingly. Prominently featured among these tests is the *Boston Cookie Theft* which has been modeled to accurately discern between Alzheimer's and cognitive typical patients [3][4][12][15][16].

How language decline presents in early AD/MCI

Demonstrably, language decline is not only symptomatic of AD neurodegeneration but features of language decline manifest in MCI and prodromal Alzheimer's [20]. Specifically, preclinical patients experience lexico-semantic deficits in spontaneous-speech [20][22]: challenges with language fluency (immediate word repetitions, pauses etc.), word-retrieval and replacement, lower vocabulary, and circumlocution [21]. Furthermore, dementia significantly accelerates decline of grammatical complexity and proposition content as compared to typical aging [23].

NLP models classifying AD v CT

Efforts have been made to model these language features as a means of automatically detecting AD onset and has been met with encouraging accuracy rates [3][4][5][6][15][16]. Specifically, NLP SVM [3] and NN [3][5][6][15][16][18] models have proven successful in differentiating between AD and cognitive-typical (CT) patients in transcripts of clinic language and memory assessment exams (notably, the *Boston Cookie Theft* picture description task from the *DementiaBank* dataset). With more specific relevance to this project, Yancheva et al. (2015) used a Dynamic Bayes Network with continuous nodes to longitudinally model cognitive decline of *DementiaBank*[1] participants [16] and Weissenbacher et al. (2016) collected a corpus of native writing samples of an image description task.

Most of these aforementioned models were built from speech-to-text transcriptions of description-tasks and none fit a dataset of emails composed in the wild. Specifically, features which are enjoyed by transcriptions such as those which rely on utterances, revisions, and immediate repetitions are unextractable from written text composed and revised at one's leisure. Additionally, content of free-form text cannot be evaluated by content-expectations and descriptive semantic units.

This work builds off aforereferenced research in its objective to longitudinally model context- and content-independent writing samples composed in the wild.

3. MATERIALS

3.1 **Datasets**

Written responses to two distinct open-ended writing prompts (3.1.1) have been collected from two groups: Patients of the Alzheimer's Prevention Clinic (3.1.2) and registered members of the Alzheimer's Universe platform (3.1.3). Details on the prompts as well as the supplementary data collected from these two groups are elucidated below.

3.1.1 Writing Prompts

Memory Prompt: *Write a short sketch about a memory from your childhood and why it is memorable or important to you. Feel free to discuss an event with family or friends, a place you traveled, or a significant time in your life. Please limit your response to no more than 1-2 paragraphs*.

Opinion Prompt: *How does technology and social media impact the daily lives of you and your family? Please limit your response to no more than 1-2 paragraphs*.

These prompts were inspired by the writing prompt in the *Nun Study* [10] and informed by our conversations with neuropsychologist Dr. Bonnie Wong. Open-ended prompts are constructed to mimic emails of unknowable content and context and are restricted to conventional email-length. The distinct nature of these prompts with respect to each other promotes exploration independent of the subject matter.

3.1.2 Alzheimer's Prevention Clinic Patient Dataset

The Alzheimer's Prevention Clinic (APC) at Weill Cornell Medicine and NewYork-Presbyterian supports cognitively-typical patients who are at risk for developing AD. These patients' cognitive profiles are monitored and scored temporally. If relevant, patients are diagnosed and ultimately treated for the disease.

50 patients responded to either the memory or opinion prompt: 44 responded to both prompts, 6 responded to the opinion but not the memory prompt. 35 of these respondents are female, 15 male: ages ranging between 27 and 78 with a mean and median age of 58. 34 patients are cognitively-typical, 9 report Subjective Cognitive Impairment (SCI), 4 exhibit Mild Cognitive Impairment (MCI), 2 are diagnosed with Preclinical Alzheimer's and 1 is diagnosed with Preclinical Parkinson's/Lewy Body Dementia.

Though the clinic set is too small and imbalanced to be modeled for robust AD detection and prediction, each patient's cognitive profile is augmented by 13 validated cognitive scores, including MMSE composite and metrics indicative of global cognitive status, executive function, processing speed, language and verbal knowledge, learning

[1] https://dementia.talkbank.org/access/

and memory. Thus, the clinic set is useful to extract, engineer, and test features to influence feature selection for machine learning on the AlzU dataset.

3.1.3 Alzheimer's Universe (AlzU) Dataset

Alzheimer's Universe is a website created by neurologist Dr. Richard Isaacson that educates its ~70,000 users about Alzheimer's Disease through interactive lessons and gathers data on these users through periodic surveys and cognitive assessments, both of which are discussed below. The users are roughly ~80% female, and the median age of the population is 62. Its users are predominantly individuals concerned about developing Alzheimer's, including cognitively-typical individuals with family histories, people with MCI due to AD, as well as mild AD patients [7]. At the time of analysis, there were N=424 individuals who responded to at least one of the writing prompts, answered behavioral survey questions, and completed the Cognitive Function Test.

In addition to the aforementioned writing prompts, the data collected from these surveys falls into one of multiple categories: self-reported health information, general feelings and attitude, physical/leisurely activities, relationships, and dietary/sleep habits. Alongside general user demographic information, this data is represented as p=183 distinct features in this dataset.

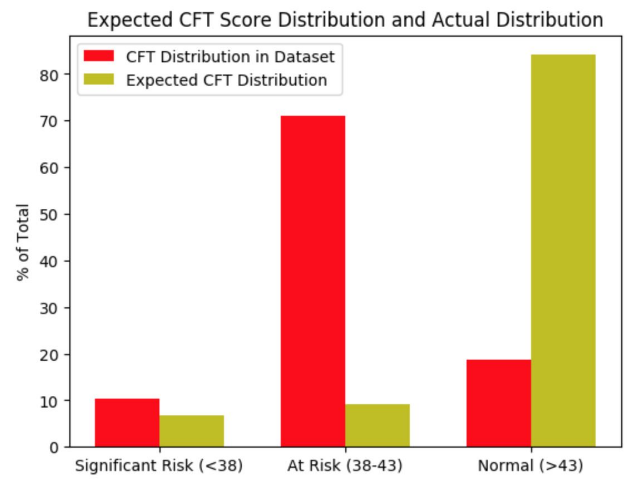
3.2 **Cognitive Scores**

3.2.1: *Cognitive Function Test (CFT)*

Validated by the *Interactive Journal of Geriatric Psychiatry* [9], the Cognitive Function Test provides an 'MCI risk' composite score comprised of episodic memory, executive function, and processing speed metrics [8].

Although this result is not a diagnosis, according to the CFT creators, a user's risk to develop MCI can be classified based on the range in which their score falls: scores above 43 indicates "No Risk", scores between 38 and 43 indicate "At Risk", and scores below 38 are categorized as "Significant Risk". As will be elucidated further in the "Methods" section below, this CFT score will therefore be used as a proxy for a user's MCI risk.

Notably, there is a drastically-different score distribution in the AlzU dataset as compared to the distribution of the general population. As can be seen in the figure below, AlzU has a much higher percentage of users that fall under the "At Risk" category than is expected, and a much lower percentage of "Normal" users. This asymmetry is to be expected, given the curated nature of a platform of users with AD-onset concerns.



4. METHODS

The overarching approach for the experiments was to leverage the Clinical dataset to identify the text features that correlate strongly with performance on well-established cognitive assessments and utilize those results to power robust machine learning prediction models on the AlzU dataset.

4.1: Identifying candidate text features

The APC clinic dataset was leveraged to identify candidate text features and inform a feature selection process. Previous modeling work of DementiaBank and related datasets of speech-to-text transcripts of picture description tasks leveraged both speech and text features [4][5][6][15][16][18]. Without the benefit of acoustic features such as mean length of utterance and filler-word rate or semantic content expectations, we explored features from natively-written text of unknowable content and considered how language deficits in AD onset may present through written communication. Thus, this work focused on lexico-semantic features as per expected linguistic presentation of emerging AD without neglecting exploratory efforts in syntactic and semantic-related features. Clinic substantiation, classifications of feature importance in related work, and expectations of attribute interactions influenced feature engineering efforts. Work was done in Python 3 using Stanford CoreNLP, NLTK Tree and POS tags, Numpy, Pandas, vaderSentiment, and TextBlob.

Lexical Features

Feature	**Description**
Response size	Word count
Sentence Count	
Honore's R statistic	Measure of vocabulary richness, insensitive to length, *R = 100log(N)/(1 - V1/V)* where N represents word count, V represents vocabulary size and V1 represents hapax legomena (proved unreliable on this dataset, specifically on short prompts comprised fully of unique words)
Brunet's W index	Measure of vocabulary richness, insensitive to length, W = N**(V −0.165) where N represents word count (proved unreliable on this dataset)
Vocabulary size	Unique word count (feature of Honore's and Brunet's)
Hapax legomena	Count of uniquely-featured words in a response (feature of Honore's and Brunet's)
Moving Average Type Token Ratio (MATTR)	Average rate of unique words across a sliding (10-word length) window

Semantic Features

Feature	Description
Noun rate	Ratio of common nouns to response size
Verb rate	Ratio of all verb forms to response size
Circumlocution	Ratio of pronouns and determiners to nouns (proper and common)
Paraphasia	Sum of words not found in dictionary i.e. non-word substitutions (conflated with misspellings)
Initiating with coordinating conjunctions	Number of sentences initiating with coordinating conjunctions
Function word rate	Ratio of stop-words

Sentiment and Syntactic Features

Feature	Description
Emotion recognition	Negativity, positivity, neutrality, compound 0-1 scores
Positivity sentiment	0-1 Polarity score
Subjectivity	0-1 Subjectivity score
Syntactic complexity	Height of constituency tree
Reduced sentences	Frequency of POS tags VBG VBN [3]

4.2.: Text Feature Correlation

Correlations are reported using Spearman's correlation to account for distribution skew and kurtosis. Experiments were performed to correlated features with clinical diagnosis, MMSE, and language and verbal knowledge. Given the constraints of a dataset from a prevention clinic, it became evident that clinical diagnosis of AD in a binary form and MMSE were too coarse and highly unbalanced (7/50 regarded as 'MCI' or 'Preclinical' and one participant with an MMSE score < 25, suggesting mild dementia) to report whether features significantly correlated with a binary AD classification. Instead, features were correlated with the *Picture Vocabulary Test* and *Oral Reading Recognition* scores. Administered via the NIH-TB [25], these exams are used to assess a patient's language and verbal knowledge.

4.3: ML on AlzU Dataset

Guided by the results of the preceding experiments, the logic was extended to the much larger AlzU dataset to generate more robust machine learning models.

Two subsets of the AlzU dataset were selected: "Memory Prompt Users" (those who had responded to the "Memory Prompt", N=364) and "Opinion Prompt Users" (those who had responded to the Opinion Prompt, N=424). Each of these subsets were further split into Training (64% of subset), Development (16%), and Held-out Test sets (20%). Models were initially tested and parameters were adjusted using the Training and Dev sets so as to avoid overfitting, and final results report on the Test data.

These models performed multiclass classification, using the three possible CFT categorizations ("Significant Risk", "At Risk", and "Normal") as the targets.

4.3.1: *Model Selection*

The models chosen for these experiments were k-Nearest Neighbors (k=3), Naive Bayes, Support Vector Machines, Decision Trees, and Random Forests.

kNN is a non-parametric classifier whose simplicity and ease of implementation make it an ideal choice as a baseline for many problems[28]. Similarly, NB often performs as a good baseline model for text classification [30]

Previous work found Support Vector Machines (SVM) as the most effective model in distinguishing AD patients from normal patients based on features extracted from verbal utterances [3].

Another study used Decision Trees and Random Forests to achieve >70% accuracy in determining whether or not a speech segment came from an individual with Alzheimer's or a healthy control [29].

4.3.2: *Performance on Non-Textual Data*

All models were evaluated for their performance in predicting the CFT score based on *non-textual* data (i.e. the data collected via surveys), to understand the important non-textual features (p=124) in predicting cognitive function.

4.3.2: *Performance on Textual Data*

All models were then tested on select text features (p=11), and various combinations of subsets of these features to identify accurate predictors of CFT scores.

4.3.2: *Combining Text and Non-Text data*

All text features were concatenated with all non-text features and the same models were tested for performance in predicting CFT scores. Additionally, various combinations of the text features were added to the non-text features to see if any further improvements on predictive performance could be obtained.

5. RESULTS

Significant correlations with fully-adjusted Oral Reading Recognition scores:

Feature	**Prompt**	**Spearman's Correlation**	**P value**
Response size	Memory	0.39	0.0054
	Opinion	0.39	0.0052
Sentence count	Memory	0.37	0.0073
Vocabulary size	Memory	0.42	0.0024
	Opinion	0.42	0.0024
Hapax legomena	Memory	0.42	0.0026
	Opinion	0.44	0.0012
Initiating with cc	Memory	0.32	0.024
Reduced sentences	Memory	0.30	0.031
Syntactic complexity	Opinion	0.30	0.036

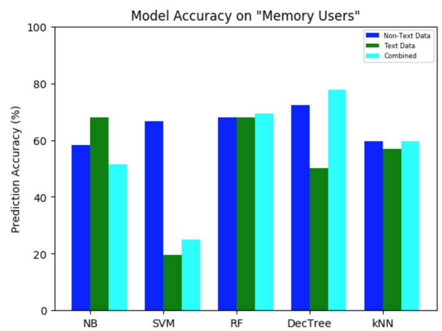
Considering a Bonferroni adjustment of 4 correlations on the same feature set, Picture Vocabulary and Oral Reading Recognition scores on both the opinion and memory writing prompts, significantly correlative features with a p-value of less than 0.0125 primarily highlight lexical features:

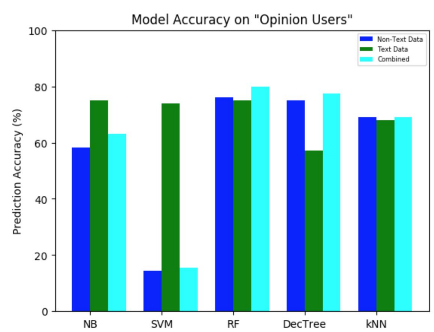
response size, sentence count, vocabulary size, and hapax legomena.

Alzheimer's Risk Prediction Results

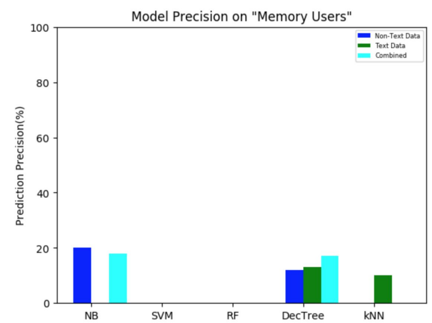
For all of the models tested, predictive performance metrics are shown for "Non-Text Data", "Text Data", and "Combined" (all non-text and text data). Specifically, the general predictive accuracy, as well as precision and recall of the "Significant Risk" category are all shown.

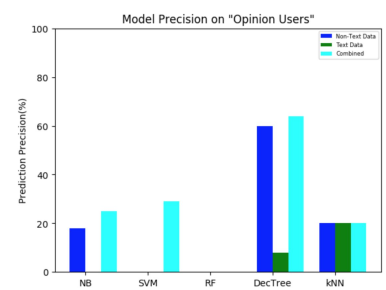
Accuracy



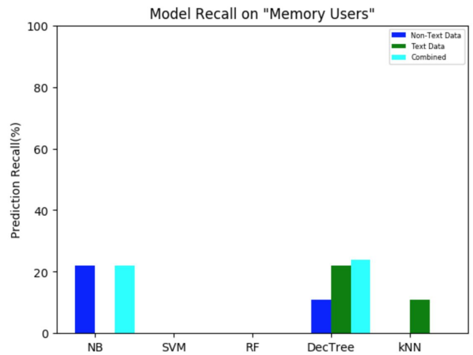


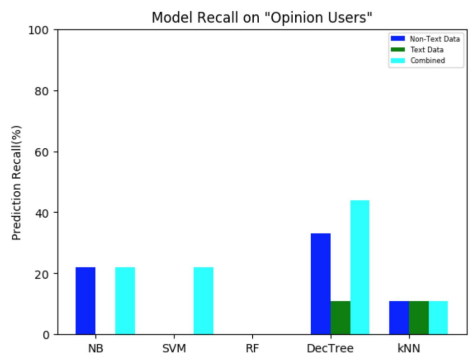
Precision





Recall





Although the models are able to predict the "Significant (MCI) Risk" class with reasonable accuracy, further inspection of the underlying distribution of the data and other performance metrics demonstrate that none of the models are adept at detecting users in the "Significant Risk" category with high sensitivity or specificity. The low precision observed indicates that users are erroneously being flagged as "Significant Risk", while the low recall indicates that those that actually do have significant risk are being missed. The high accuracy can be explained by the high proportion of users that fall into the "At Risk" category; a model that simply predicts "At Risk" for every query would perform reasonably well.

By doing a holistic assessment of all of the models tested, Decision Trees emerged as the best candidate for further exploration. Regardless of the performance metric being evaluated (Accuracy, Precision, or Recall) and the writing prompt (Memory or Opinion prompt), the top model was the Decision Tree using both textual and non-text data.

6. DISCUSSION

In exploring the relationship between free-form writing patterns and Alzheimer's progression, this work indicates that features of text composed in the wild are significantly correlated with cognitive language assessment scores. Thus, deficits or decline in significant lexical features can be associated with language deficiencies or decline. It is particularly encouraging for continued efforts in passive digital trace modeling that the significant features identified -response size, sentence count, vocabulary size, and hapax legomena - will persist in written traces despite adoption of writing-improvement plugins.

However, when moving beyond language assessments and into the broader task of predicting "Alzheimer's Risk", the features explored are not sufficient to predict "Significant Risk" of AD-MCI onset. The text features do a poor job of predicting one's "Risk" category, both failing to correctly classify Significant-Risk individuals and in some cases erroneously assigning Significant-Risk to those who are cognitive-typical. These results lend an early suggestion that text in the wild alone is not sufficient to detect an individual who may be on the verge of developing the disease.

Nonetheless, the results do offer some insight into the potential for analyzing free-form text in tandem with other behavioral data. Notably, this paper's highest performing model (Decision Tree) incorporated both text and demographic/behavioral features of clinical significance. This finding is consistent with the National Institute on Aging-Alzheimer's Association's [11] published set of criteria in diagnosis of AD, which underscores the inclusion of diverse datastreams in AD diagnosis. These results further demonstrate that language decline as evident by text features alone are informative but not sufficient to sensitively detect AD onset.

Further digital-trace features may be leveraged to account for learning and memory, processing speed, and executive functioning to detect AD onset with higher recall. In order to build off this paper's work on digital trace data, the authors recommend pivoting from modeling email exchanges to modeling social media profiles. Such a data source would incorporate significant text features together with demographic and behavioral data extracted through search history [25], location data, and public demographic information. With successful modeling of multiple cognitive metrics, the confluence of the data streams could detect and monitor AD onset and progression with higher sensitivity and specificity.

WORKS CITED

1) Wiles, Kayla. "Using Language Skills to Detect Alzheimer's." ScienceWriters (Www.NASW.org), 14 Mar. 2017, www.nasw.org/article/using-language-skills-detect-alzheimers.
2) Mueller, Kimberly Diggle et al. "Connected Language in Late Middle-Aged Adults at Risk for Alzheimer's Disease." Journal of Alzheimer's disease : JAD 54.4 (2016): 1539–1550. PMC. Web. 11 May 2018.
3) Orimaye, Sylvester O. et al. "Predicting Probable Alzheimer's Disease Using Linguistic Deficits and Biomarkers." BMC Bioinformatics 18 (2017): 34. PMC. Web. 14 May 2018.
4) Croisile B, Ska B, Brabant MJ, Duchene A, Lepage Y, Aimard G, Trillet M. Comparative study of oral and written picture description in patients with Alzheimer's disease. Brain Lang. 1996;53:1–19.
5) MacWhinney B. The CHILDES Project: The Database. New York: Psychology Press; 2000.
6) Covington MA, McFall JD (2010) Cutting the Gordian knot:
The moving-average type–token ratio (MATTR). J Quant Linguist
17, 94-100.
7) "Effectiveness of Alzheimer's Universe (Www.Alzu.Org) on Knowledge and Behavior." Full Text View - ClinicalTrials.gov, clinicaltrials.gov/ct2/show/NCT03149380.
8) "Frequently Asked Questions." About Schizophrenia and Psychosis, www.foodforthebrain.org/alzheimers-prevention/take-the-test/frequently-asked-questions.aspx.
9) Eve, Catharine Trustram, and Celeste A. De Jager. "Piloting and Validation of a Novel Self-Administered Online Cognitive Screening Tool in Normal Older Persons: the Cognitive Function Test." International Journal of Geriatric Psychiatry, vol. 29, no. 2, 2013, pp. 198–206., doi:10.1002/gps.3993.
10) Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, Markesbery WR. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. JAMA. 1996;275:528–532
11) Ahmed, Samrah et al. "Connected Speech as a Marker of Disease Progression in Autopsy-Proven Alzheimer's Disease." Brain 136.12 (2013): 3727–3737. PMC. Web. 13 May 2018.
12) Orimaye SO, Wong JS-M, Golden KJ. Learning predictive linguistic features for Alzheimer's disease and related dementias using verbal utterances. In: Proceedings of the 1st Workshop on Computational Linguistics and Clinical Psychology (CLPsych). Baltimore: Association for Computational Linguistics; 2014. p. 78–87.
13) Karlekar, Sweta, et al. Detecting Linguistic Characteristics of Alzheimer's Dementia by

Interpreting Neural Models. UNC Chapel Hill, 17 Apr. 2018, arxiv.org/pdf/1804.06440.pdf.

14) Verma M, Howard R. Semantic memory and language dysfunction in early Alzheimer's disease: a review. Int J Geriatr Psychiatr. 2012;27(12):1209–17
15) Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. Journal of Alzheimer's Disease,49(2), 407-422.
16) Yancheva, M., Fraser, K., & Rudzicz, F. (2015). Using Linguistic Features Longitudinally to Predict Clinical Scores for Alzheimer's Disease and Related Dementias.
17) McKhann, Guy M. et al. "The Diagnosis of Dementia due to Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease." Alzheimer's & dementia : the journal of the Alzheimer's Association 7.3 (2011): 263–269. PMC. Web. 13 May 2018.
18) Weissenbacher, Davy, et al. "Towards Automatic Detection of Abnormal Cognitive Decline and Dementia Through Linguistic Analysis of Writing Samples." Proceedings of NAACL-HLT 2016, 12 June 2016, pp. 1198–1207., www.aclweb.org/anthology/N16-1143.
19) "Interpreting Your Test Results." About Schizophrenia and Psychosis, www.foodforthebrain.org/alzheimers-prevention/take-the-test/interpreting-results.aspx.
20) Boschi, Veronica et al. "Connected Speech in Neurodegenerative Language Disorders: A Review" Frontiers in psychology vol. 8 269. 6 Mar. 2017, doi:10.3389/fpsyg.2017.00269
21) Szatloczki, Greta et al. "Speaking in Alzheimer's Disease, Is That an Early Sign? Importance of Changes in Language Abilities in Alzheimer's Disease." Frontiers in Aging Neuroscience 7 (2015): 195. PMC. Web. 16 Oct. 2018.
22) *Forbes-McKay KE, Venneri A Neurol Sci. 2005 Oct; 26(4):243-54.*
23) Kemper, S., Thompson, M., & Marquis, J. (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. Psychology and Aging, 16(4), 600-614. http://dx.doi.org/10.1037/0882-7974.16.4.600
24) The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author.*Garrard P, Maloney LM, Hodges JR, Patterson K. Brain. 2005 Feb; 128(Pt 2):250-60*
25) Weintraub, Sandra et al. "Cognition assessment using the NIH Toolbox"*Neurology* vol. 80,11 Suppl 3 (2013): S54-64.
26) Austin, Johanna, Kristy Hollingshead, and Jeffrey Kaye. "Internet Searches and Their Relationship to Cognitive Function in Older Adults: Cross-Sectional Analysis." Ed. Gunther Eysenbach. Journal of Medical Internet Research 19.9 (2017): e307. PMC. Web. 27 Aug. 2018.
27) Petersen RC. Early diagnosis of Alzheimer's disease: is MCI too late?. *Curr Alzheimer Res*. 2009;6(4):324-30.
28) Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell. 2000;22(1):4–37. doi: 10.1109/34.824819
29) Netscher, George**.** **(**2016**).** *Applications of Machine Learning to Support Dementia Care through Commercially Available Off-the-Shelf Sensing* (Unpublished master's thesis). Electrical Engineering and Computer Sciences University of California at Berkeley**,** Berkeley**,** CA**.**
30) Rennie, J.; Shih, L.; Teevan, J.; Karger, D. (2003). *Tackling the poor assumptions of Naive Bayes classifiers*(PDF). ICML