Diagnosing Alzheimer's Disease using NLP in Icelandic

1st Elena Callegari
Language & Technology
University of Iceland
Reykjavík, Iceland
ecallegari@hi.is

2nd Anton Karl Ingason
Language & Technology
University of Iceland
Reykjavík, Iceland
antoni@hi.is

3rd Omer Korat
Language & Technology
University of Iceland
Reykjavík, Iceland
korat@hi.is

4th Agnes Sólmundsóttir
Language & Technology
University of Iceland
Reykjavík, Iceland
ags46@hi.is

***Abstract*—We present the purposes and logic behind the ACoDe project ("Assessing Cognitive Decline using automatic language analysis"). Our goal is to develop speech analytics technology for Icelandic that can be used to diagnose Alzheimer's disease through automatic language analysis.**

***Index Terms*—Alzheimer's disease, Icelandic, EEG, Automatic Language Analysis**

I. Introduction

Alzheimer's disease (henceforth, *AD*) is a particularly pressing concern in Iceland: according to a 2016 study [1], people of Icelandic heritage are considerably more likely than other European populations to carry a genetic mutation that results in a greater risk to develop AD. A timely Alzheimer's diagnosis provides patients with a better chance of benefiting from existing treatments. Currently available procedures to diagnose this pathology (neurocognitive tests in combination with PET or MRI, and/or the sampling of cerebrospinal fluid), however, are costly and can only be administered at specialized clinics which are often overbooked; waiting times at Reykjavík's Memory Clinic, for instance, average 2 to 4 months. This results in delayed diagnoses but also in greater difficulties in monitoring the evolution of the pathology over time. AD affects what we say and how we say it [2], [3]; spoken language can thus offer a universal and accessible means for assessing neurological health [4]. Our goal is then to collect speech samples from individuals with different degrees of AD, extract key linguistic features from these samples and train a number of classifiers on them to determine whether these can distinguish between healthy and AD participants, and among different degrees of AD.

A. Participants

We will be testing a total of 120 participants: 30 patients with Mild Alzheimer's Dementia (MD), 30 patients with Mild Cognitive Decline (MCD), 30 patients with Subjective Cognitive Decline (SCD), and 30 healthy controls. All participants will be aged between 60 and 80 and will be monolingual speakers of Icelandic. At the time of writing (March 2023), we have tested all 30 controls and have started testing patients.

This research study is funded by Rannís, the Icelandic Research Center.

B. Experimental protocol

ACoDe sits at the intersection between NLP, machine learning and neuroscience. The project consists of 3 components:

1) **Speech analysis**: We will elicit speech samples from each participant by asking them to describe: (i) the "picnic scene" by The Arizona Alzheimer's Disease Center; (ii) how they would plan a trip to Akureyri, a city in the north of Iceland; (iii) their childhood. The planning-a-trip kind of narrative was chosen as it is effective in eliciting complex-event descriptions while engaging episodic memory. The childhood scenario was chosen as it naturally elicits long responses from participants.
2) **Neuroimaging data**: We will record resting-state, eyes-closed EEG for each participant. EEG data will be used to formulate a *dementia index* [5] for each of our participants; this is a clinically approved method used for diagnosing AD based on one's EEG results. This method outputs a score ranging from 0 to 100 that categorizes individuals on a scale of severity for AD, with 0 being equal to healthy, 100 to severely impaired.
3) **Machine Learning**: We are developing an open-source library that can automatically extract linguistic features with clinical relevance to AD from the transcriptions of speech samples. Once we have tested all participants, we will train different classifiers on different combinations of said features. We will determine whether these classifiers can distinguish between patients and controls, between different stages of AD, with which accuracy and which combinations of features work best.

C. Novelty

To our knowledge, ACoDe is the first study using NLP to diagnose AD that is directly relating linguistic data to EEG results. EEG has excellent clinical value in diagnosing AD [6]: by directly relating language data to dementia index scores, we will be able to estimate with considerably more precision how advanced a patient's cognitive decline is. Moreover, this is the very first time NLP techniques are used to formulate clinical diagnoses for Icelandic. Icelandic has complex inflectional morphology and rare verbal syntax (with V2 in embedded clauses), and hence represents a good addition to the research effort studying how AD affects language production.

D. Feature Selection

The features we will use for our model are based on those shown to be effective at predicting cognitive decline in previous studies on diagnosing AD from speech. However, our features will be adapted to account for the specific landscape of Icelandic as well as its existing datasets. The types of features are divided into lexical, syntactic, and acoustic.

1) Lexical: We will extract various lexical features that have been found to be useful predictors of cognitive decline, including Part of Speech (PoS), lexical variability, perplexity, type-token ratio, proportion of cohesion words, among others [7]–[9]. Higher perplexity (lower predictability) of texts is a measure of AD, since it sometimes leads to speech with lower coherence, and is therefore very unpredictable. Icelandic's rich inflectional morphology includes case marking on nouns and person markers on verbs, and as such PoS features are of particular interest for our study. We will use the Greynirseq package[1]'s Icelandic POS tagger for these features, and the IceBERT Icelandic language model [2] to determine perplexity of texts. Lastly for this category, we will count the ratios of cohesion words, such as pronouns and determiners, in the text, as they have been demonstrated to interact with AD [10].

2) Syntactic: We will use the Greynirseq dependency parser (yet to be released at the time of writing, see Footnote 1) to extract syntactic features from our dataset. We expect overall simpler structure, lower tree depth, shorter phrases, and fewer relative clauses for individuals with AD [11]

3) Acoustic: We will apply a battery of state-of-the-art acoustic feature extractors to the data, modeled after [12], [13], in which they have been demonstrated to successfully predict levels of autism spectrum disorder in children. We will use the UPenn Speech Activity Detector (SAD) [14] to detect pauses, and the openSMILE [15] and covarep [16] packages to detect a variety of acoustic features such as F0 range, pitch, jitter, shimmer, harmonic-to-noise ratio (HNR), and spectral moments (1st order: centroid, 2nd order: standard deviation, 3rd order: skewness, 4th order: kurtosis).

E. Model Selection

To decide on our final model, we will use standard cross-validation and parameter-tuning on a battery of classifiers, choosing those that perform the best. The models we test will range from simple to complex, starting from linear regression, going through non-linear models such as Support Vector Machines and decision trees, and then finally state-of-the-art artificial neural networks, such as deep AdaBoost and transformer models. We will also include techniques that have been shown to be effective at evaluating the performance of clinical diagnostic/predictive models, such as Area Under the receiver operating characteristics (ROC) Curve (AUC) [17] and Generalized Additive Models (GAMs) [18].

This is a more general learning approach than what has been attempted in existing studies (such as the ones described above), in which the statistical models have been chosen manually, for the most part. The potential advantage in our approach is the ability to learn high-order non-linear relationships between the feature space and the classes of cognitive decline of arbitrary complexity, which might be necessary for accurate predictions.

[1]https://github.com/mideind/GreynirSeq

[2]https://huggingface.co/jonfd/convbert-base-igc-is

References

[1] J. Jakobsdottir, S. J. van der Lee, J. C. Bis, V. Chouraki, D. Li-Kroeger, S. Yamamoto, M. L. Grove, A. Naj, M. Vronskaya, J. L. Salazar *et al.*, "Rare functional variant in tm2d3 is associated with late-onset alzheimer's disease," *PLoS genetics*, vol. 12, no. 10, p. e1006327, 2016.

[2] G. Kavé and A. Dassa, "Severity of alzheimer's disease and language features in picture descriptions," *Aphasiology*, vol. 32, no. 1, pp. 27–40, 2018.

[3] S. Ahmed, A.-M. F. Haigh, C. A. de Jager, and P. Garrard, "Connected speech as a marker of disease progression in autopsy-proven alzheimer's disease," *Brain*, vol. 136, no. 12, pp. 3727–3737, 2013.

[4] K. C. Fraser, J. A. Meltzer, N. L. Graham, C. Leonard, G. Hirst, S. E. Black, and E. Rochon, "Automated classification of primary progressive aphasia subtypes from narrative speech transcripts," *cortex*, vol. 55, pp. 43–60, 2014.

[5] K. Johnsen, R. Cromarty, G. H. Jóhannesson, M. Johannsson, and J.-P. Taylor, "P1-140: Validation of the sigla indices in an independent cohort: A new way forward for the differential diagnosis of dementias?" *Alzheimer's & Dementia*, vol. 10, pp. P350–P351, 2014.

[6] K. Engedal, M. L. Barca, P. Høgh, B. B. Andersen, N. W. Dombernowsky, M. Naik, T. E. Gudmundsson, A.-R. Øksengaard, L.-O. Wahlund, and J. Snaedal, "The power of eeg to predict conversion from mild cognitive impairment and subjective cognitive decline to dementia," *Dementia and geriatric cognitive disorders*, vol. 49, no. 1, pp. 38–47, 2020.

[7] B. S. Kim, Y. B. Kim, and H. Kim, "Discourse measures to differentiate between mild cognitive impairment and healthy aging," *Frontiers in aging neuroscience*, vol. 11, p. 221, 2019.

[8] S. Cho, N. Nevler, S. Ash, S. Shellikeri, D. J. Irwin, L. Massimo, K. Rascovsky, C. Olm, M. Grossman, and M. Liberman, "Automated analysis of lexical features in frontotemporal degeneration," *Cortex*, vol. 137, pp. 215–231, 2021.

[9] S. Cho, N. Nevler, S. Shellikeri, N. Parjane, D. J. Irwin, N. Ryant, S. Ash, C. Cieri, M. Liberman, and M. Grossman, "Lexical and acoustic characteristics of young and older healthy adults," *Journal of Speech, Language, and Hearing Research*, vol. 64, no. 2, pp. 302–314, 2021.

[10] S. X. Tang, R. Kriz, S. Cho, S. J. Park, J. Harowitz, R. E. Gur, M. T. Bhati, D. H. Wolf, J. Sedoc, and M. Y. Liberman, "Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders," *npj Schizophrenia*, vol. 7, no. 1, p. 25, 2021.

[11] V. Boschi, E. Catricala, M. Consonni, C. Chesi, A. Moro, and S. F. Cappa, "Connected speech in neurodegenerative language disorders: a review," *Frontiers in psychology*, vol. 8, p. 269, 2017.

[12] S. Cho, R. Fusaroli, M. R. Pelella, K. Tena, A. Knox, A. Hauptmann, M. Covello, A. Russell, J. Miller, A. Hulink, J. Uzokwe, K. Walker, J. Fiumara, J. Pandey, C. Chatham, C. Cieri, R. Schultz, M. Liberman, and J. Parish-morris, "Identifying stable speech-language markers of autism in children: Preliminary evidence from a longitudinal telephony-based study," in *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*. Seattle, USA: Association for Computational Linguistics, Jul. 2022, pp. 40–46. [Online]. Available: https://aclanthology.org/2022.clpsych-1.4

[13] S. Cho, M. Liberman, N. Ryant, M. Cola, R. T. Schultz, and J. Parish-Morris, "Automatic Detection of Autism Spectrum Disorder in Children Using Acoustic and Text Features from Brief Natural Conversations," in *Proceedings of Interspeech 2019*, 2019, pp. 2513–2517.

[14] N. Ryant, "Linguistic Data Consortium Broad Phonetic Class Speech Activity Detector (ldc-bpcsad)," 2023. [Online]. Available: https://github.com/Linguistic-Data-Consortium/ldc-bpcsad

[15] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462.

[16] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "Covarep—a collaborative voice analysis repository for speech technologies," in *2014 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE, 2014, pp. 960–964.

[17] L. Mickes, "Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory," *Journal of Applied Research in Memory and Cognition*, vol. 4, no. 2, pp. 93–102, 2015.

[18] P. C. Austin, "A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting ami mortality," *Statistics in medicine*, vol. 26, no. 15, pp. 2937–2957, 2007.