

# **Analysis of semantics and early linguistic symptoms to develop machine learning predictive modeling of Alzheimer's Disease**

## **Abstract:**

Over 6 million Americans are living with Alzheimer's disease, and dementia deaths have seen an increase of 16% during the COVID-19 pandemic. Dementia was predicted to cost Americans \$355 billion by the end of 2021. With an increasing percentage of cases and rising healthcare costs, the need for accessible prediction of dementia utilizing the analysis of properties of digital biomarkers is necessary. The majority of individuals who experience the symptoms of Alzheimer's are 65 years or older. Starting from the age of 65 years, the risk of succumbing to Alzheimer's significantly increases - doubling every 5 years. There is an evident overlap between

the high-risk age groups of COVID-19 and Alzheimer's Disease. Thus, the demand for digital prediction tools has unquestionably increased. The purpose of conducting this research is to employ digital and physically contact-free technology to construct an artificial intelligence model to identify early symptoms of Alzheimer's Disease. Data was collected from fluency data from UCSD. A machine learning Random Forest Model was the most effective at predicting the advancement of Alzheimer's Disease in patients who are able to function cognitively normally with an accuracy percentage of 93%. Further research into this area of building digital Alzheimer's Disease prediction tools will help aid the early

diagnosis of dementia, lower the burden on medical professionals, and assist in meeting the rising healthcare needs of countries with underdeveloped healthcare systems.

## **I. Introduction**

Alzheimer's Disease is a type of cognitive disorder that results in patients developing neurological issues within the brain that increasingly negatively impact daily tasks and quality of life. Such tasks include inhibition of memory, behavior, and decision making. The progression of Alzheimer's worsens overtime, causing the negative effects of this cognitive brain disorder to impact one's quality of life exponentially, over the course of a few years. Additionally, Alzheimer's Disease has no known cure, but a recently discovered treatment of aducanumab, which has

proven helpful in reducing the dangers of Alzheimer's Disease. However, to maximize the benefits of the treatment available, it is integral to advocate for the early detection of Alzheimer's Disease.

## **II. Methods and Procedures**

Semantic fluency data was acquired from the University of California San Diego Shiley-Marcos Alzheimer's Disease Research Center (ADRC). The data from the participants were collected between the years of 1985 to 2016. Participants were meant to visit the lab one time per year. The average number of years of participation of each participant was about 9 years. The health of the participants ranged from those who are not diagnosed with Alzheimer's Disease to those who are diagnosed with cognitive disorders such as, but not limited to, Alzheimer's Disease. In the duration of

the annual visit, participants listed as many animals as they could in one minute. Once the data was received, the data was cleaned and sorted through to form groups to differentiate between participants who did not have Alzheimer's Disease, which was the normal control group, participants who were at risk of developing Alzheimer's Disease, and participants that had already developed Alzheimer's Disease. Next, different metrics were analyzed to see which provided the most accurate metric and the most easily calculable metric if needed to be reimplemented. Finally, it was determined that calculating the means number of animals per participant and measuring them with the metrics of sex, education, and number of animals listed per visit gave the most accurate and

reimplementable metrics to use in the machine learning models.

### **III. Results**

A random forest machine learning model resulted in the highest accuracy score of 93%. Other machine learning models were also used, such as Logistic Regression and Support Vector Machine, which had approximately 80% and 89% accuracy respectively.

### **IV. Discussion**

The implications of this research are as follows: using a metric such as mean number of animals listed provides an easy-to-use resource that encourages patients to monitor the progress of Alzheimer's Disease from the comfort of one's home. This model can be further implemented in noninvasive and

inexpensive tools to benefit populations that have been affected by the lack of access due to or enhanced by the COVID-19 pandemic.

#### Bibliography:

- Zemla, J. C., & Austerweil, J. L. (2019, May 2). Analyzing knowledge retrieval impairments associated with alzheimer's disease using network analyses. *Complexity*. Retrieved February 23, 2022, from <https://www.hindawi.com/journals/complexity/2019/4203158/>
- Gerkin, R. C., Ohla, K., Veldhuizen, M. G., Joseph, P. V., Kelly, C. E., Bakke, A. J., Steele, K. E., Farruggia, M. C., Pellegrino, R., Pepino, M. Y., Bouysset, C., Soler, G. M.,

Pereda-Loth, V., Dibattista, M., Cooper, K. W., Croijmans, I., Pizio, A. D., Ozdener, M. H., Fjaeldstad, A. W., ... Parma, V. (2020, January 1). The best covid-19 predictor is recent smell loss: A cross-sectional study. *medRxiv*. Retrieved February 23, 2022, from <https://www.medrxiv.org/content/10.1101/2020.07.22.20157263v1>