

# Sandia National Laboratories Project Proposal Alzheimer's and FTD Classification from EEG Data

Brian Keith, January 2024

bkeith9@gatech.edu

**Abstract**—This paper will cover a high-level proposal for my project in conjunction with Sandia National Laboratory for the CSE 6748 course. The proposed project consists of taking data from electroencephalogram (EEG) readings to classify Alzheimer's disease in patients using machine learning techniques. The proposed dataset for the project was provided by Miltiadous et al. (2023)<sup>1</sup> under and open access Creative Common CC BY license<sup>i</sup>.

## 1 INTRODUCTION

The brain is one of the most crucial tools available to humans. However, there is a growing trend in Alzheimer's as well as other neurodegenerative diseases which is projected to continue increase further over the coming 30 years<sup>2</sup>. When it comes to treatment of these diseases, while there are no current treatments that can reverse Alzheimer's disease, accurate and early diagnosis can help prepare families of the affected and extend the good quality of life years for individuals with the disease.<sup>3</sup>

While any medical diagnosis should include a variety of factors such as consultations with medical professionals and clinical diagnostics, one proposed method to assist medical professionals in diagnosing neurodegenerative diseases is electroencephalography (EEG), which measures brain electrical activity. The signals from the EEG readings can then be used to automatically find patterns that might indicate the presence of neurodegenerative diseases using machine learning models<sup>4</sup>.

## 2 DATA SET

For this project, while at a later point I may determine that more data is needed, a recent publication by Miltiadous et al. (2023) provides a comprehensive dataset of EEG readings from 88 different patients (36 Alzheimer's patients, 23

---

<sup>i</sup> CC BY 4.0 Deed | Attribution 4.0 International | Creative Commons. (n.d.). <https://creativecommons.org/licenses/by/4.0/>

frontotemporal dementia (FTD) patients, and 29 healthy age-matched subjects). The data set contains data from 19 different sensors, Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, & O2 at various points on the scalp of the patients. Additional information on the specifics of the dataset can be found [here](#).

The main reason that I chose this data set for analysis is because it was published with the goal of being used for additional research so it's very well documented. In addition, the researchers also did some of the work required to preprocess the data and remove artifacts from the readings such as muscle activity, blinking, and swallowing which might be difficult to remove myself without significant medical knowledge.

### **3 PROPOSED GOAL**

The goal for the project will be to build various models to classify individuals with neurodegenerative diseases. This will likely require significant experimentation with different classification models such as random forests, logistic regression, support vector machines. Ideally, I would like to build out various models and compare the performance to get the best possible outcomes.

To generate the features for the model, I will use signal decomposition with fast Fourier transforms to decompose the EEG signal data into the relative band power for five different frequency ranges: Delta 0.5–4 Hz, Theta 4–8 Hz, Alpha 8–12 Hz, Beta 12–25 Hz, Gamma 25–48 Hz. The idea is to apply this transformation to all 19 of the sensors to use as features in the model for a total of 95 features.

One concern I have is that this may be too many features for some of the models I'm planning to develop so I may need to do some feature evaluation during the process to determine if there are some frequency ranges in specific portions of the brain which are more indicative of neurologic communication issues.

To validate the models, I plan to use leave one out cross validation for the training and testing datasets. As for measuring performance, I plan to use a variety of metrics related to classification models based on confusion matrices such as accuracy, sensitivity, and specificity.

## 4 CONCLUSION

While I do not have any experience in the medical field, I think it is a very interesting problem that fits within the scope of signal processing and modeling proposed by Sandia National Laboratories and would make a good project. I am open to hearing feedback or criticisms on the topic I've chosen and am looking forward to tackling this problem.

## 5 REFERENCES

- <sup>1</sup> Andreas Miltiadous and Katerina D. Tzimourta and Theodora Afrantou and Panagiotis Ioannidis and Nikolaos Grigoriadis and Dimitrios G. Tsalikakis and Pantelis Angelidis and Markos G. Tsipouras and Evripidis Glavas and Nikolaos Giannakeas and Alexandros T. Tzallas (2023). A dataset of EEG recordings from: Alzheimer's disease, Frontotemporal dementia and Healthy subjects. OpenNeuro. [Dataset] doi: doi:10.18112/openneuro.ds004504.v1.0.6
- <sup>2</sup> Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T., Hamad, H. A., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., . . . Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, 7(2), e105–e125. [https://doi.org/10.1016/s2468-2667\(21\)00249-8](https://doi.org/10.1016/s2468-2667(21)00249-8)
- <sup>3</sup> Rasmussen, J., & Langerman, H. (2019). Alzheimer's Disease – Why We Need Early Diagnosis Degenerative Neurological and Neuromuscular Disease, Volume 9, 123–130. <https://doi.org/10.2147/dnnd.s228939>
- <sup>4</sup> A. Miltiadous et al., "Machine Learning Algorithms for Epilepsy Detection Based on Published EEG Databases: A Systematic Review," in *IEEE Access*, vol. 11, pp. 564-594, 2023, doi: 10.1109/ACCESS.2022.3232563.