DDT: Dementia Detection Tool

Erdos Institute Fall 2023 Bootcamp

Team DDTs: Himanshu Khanchandani, Clark Butler, Cisil Karaguzel, Selman Ipek, Shreya Shukla

Motivation

Alzheimer's disease (AD) is one of the most common types of dementia and frequently affects the elderly. **Electroencephalography** (EEG) is a non-invasive technique to measure the brain activity using external electrodes and may help provide improved diagnosis of AD.

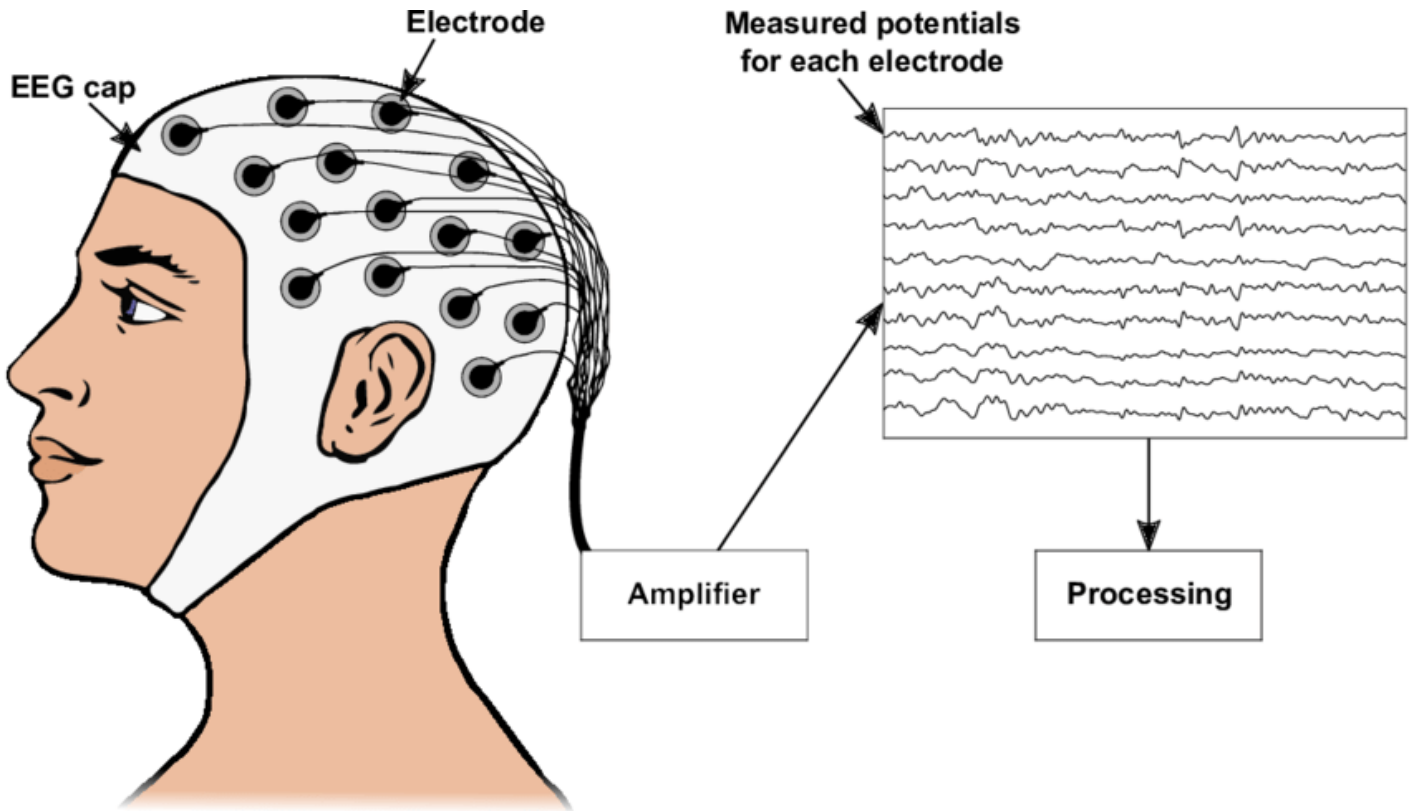
Goals: (1) Develop an accurate classifier based on EEG data for predicting if a patient has AD or is healthy. (2) Infer which features in EEG data can be used to distinguish AD versus healthy.

Use Cases: Useful in hospital and lab settings to screen patients before subjecting them to more invasive treatments or tests.

KPI: Prediction accuracy in distinguishing patients with AD from healthy subjects.

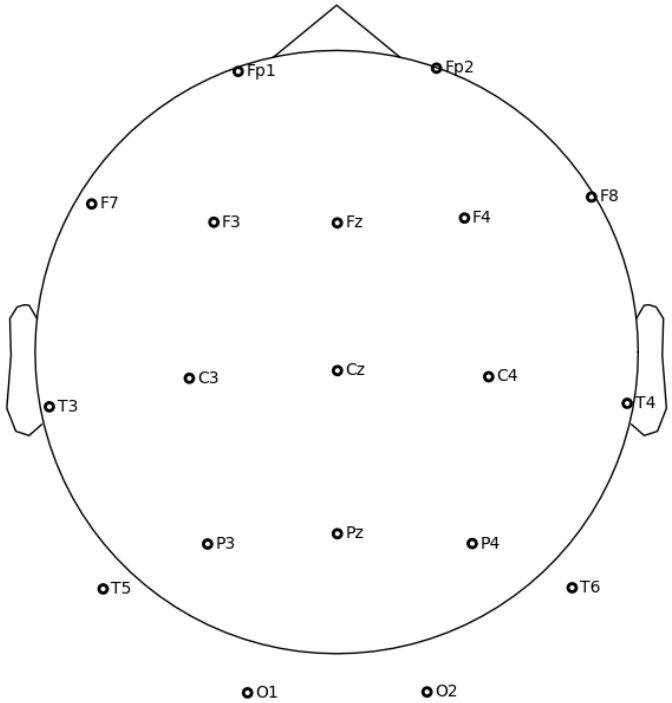
Data Visualization

EEG data collection process:



Data Visualization

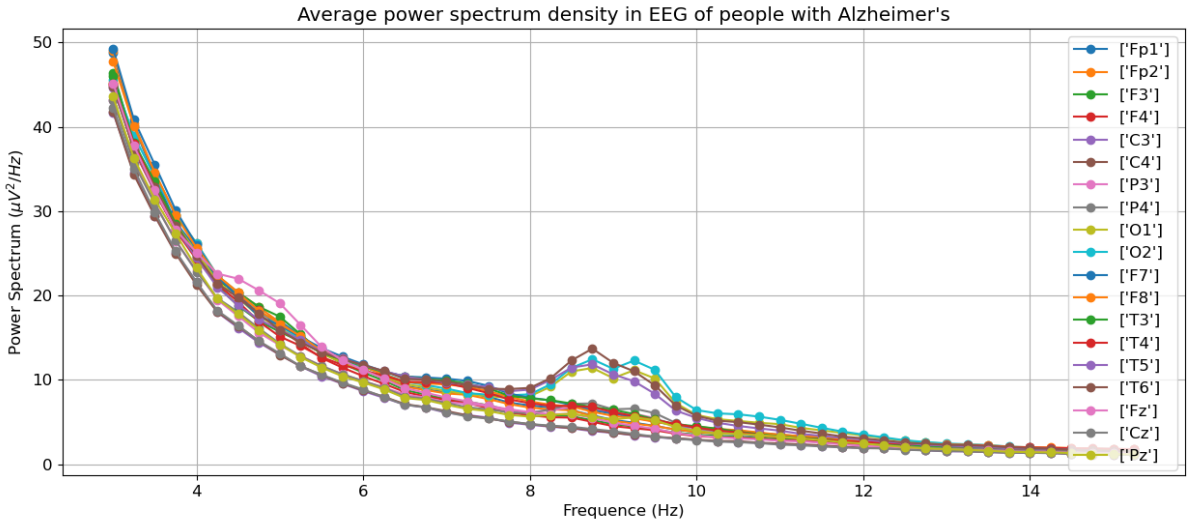
Our data[1] consists of EEG signals from 19 channels (or electrodes) on the scalp, arranged as follows:

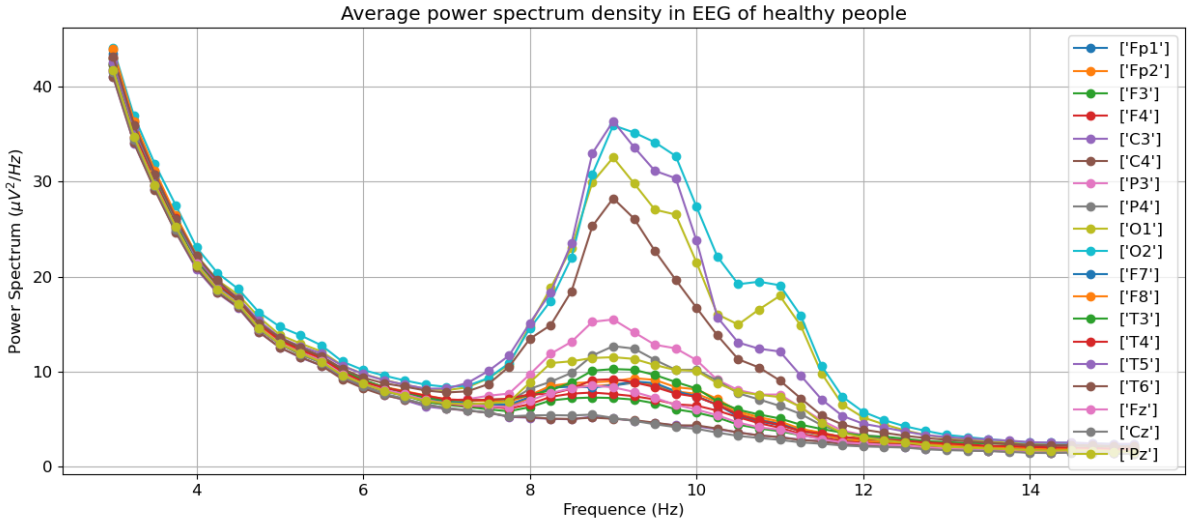


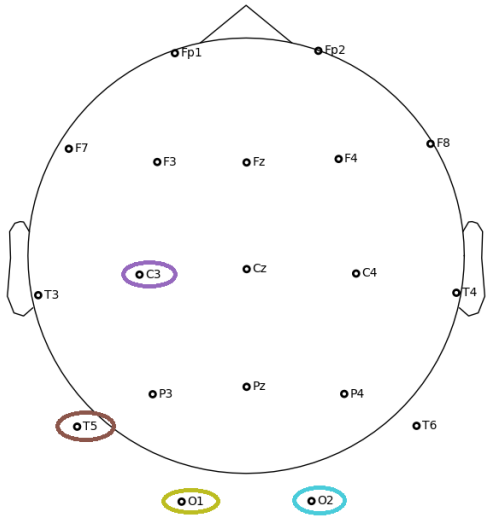
Analyzing the power spectrum of the EEG data in the frequency domain is more distinctive than studying the time series directly.

[1] EEG Dataset: https://openneuro.org/datasets/ds004504

Data Visualization







Data Source: openneuro.org/datasets/ds004504

Feature and Model Selection

A common practice is to divide the full length EEG into smaller time windows, or epochs, with some overlap.

In each epoch we compute the power spectrum and then the relative band power (RBP) for various frequency bands.

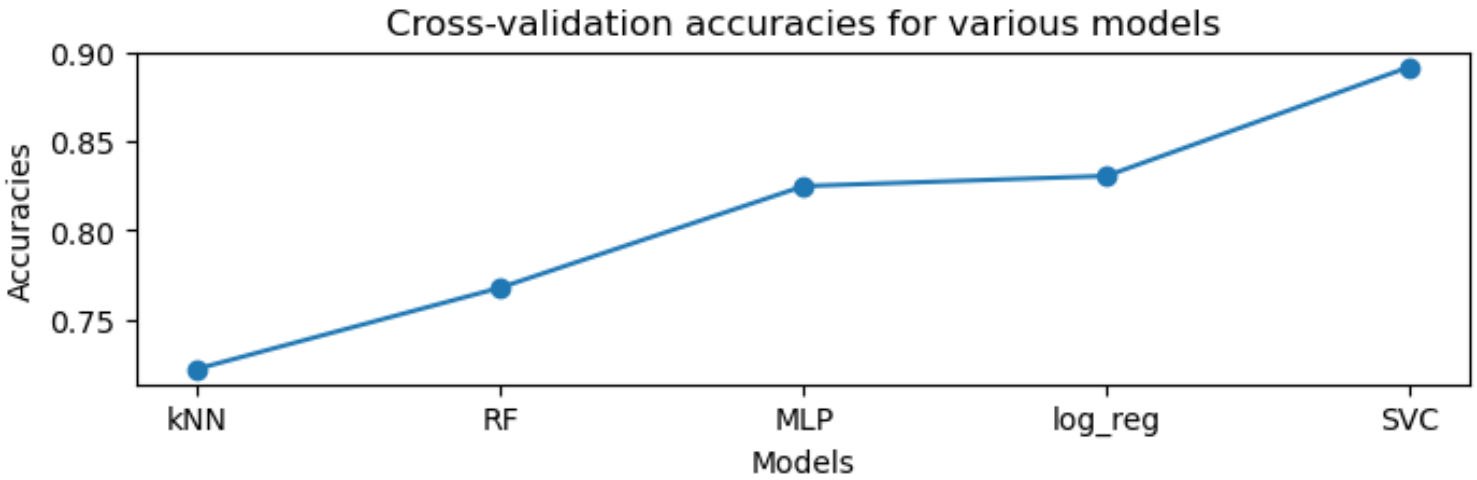
Based on our exploratory data analysis and cross-validation, we chose optimal frequency bands, epoch lengths, and channels:

	Literature	Erdos DDT
Bands	(0.5- 4 Hz), (4- 8 Hz), (8 - 13 Hz), (13- 25 Hz), (25 - 45 Hz)	(0.5- 4 Hz), (4 - 7 Hz) (7 - 9 Hz), (9 - 11 Hz), (11- 13 Hz), (13 - 25 Hz)
Epochs	4s - 30s	2 minutes
Channels	All channels	Left out F3, F4, C4 F8

Model performance

Leave-one-subject-out (LOSO) cross-validation was used for model selection and tuning hyperparameters.

	acc.	sens.	spec.	F1
kNN	72%	76%	67%	74%
RF	76%	79%	74%	78%
MLP	82%	84%	80%	83%
log. reg.	83%	85%	79%	84%
SVC	89%	91%	86%	90%



Model performances

SVC model with a linear kernel gives the best performance in cross-validation, with a validation accuracy of 89% and a test set accuracy of 81%.

Performance comparison with existing literature models				
	acc.	sens.	spec.	F1
Our SVC model	89%	91%	86%	90%
Best literature RBP based method[2]	77%	78%	81%	75%
DICE-NET: A CNN based method [3]	83%	79%	87%	84%

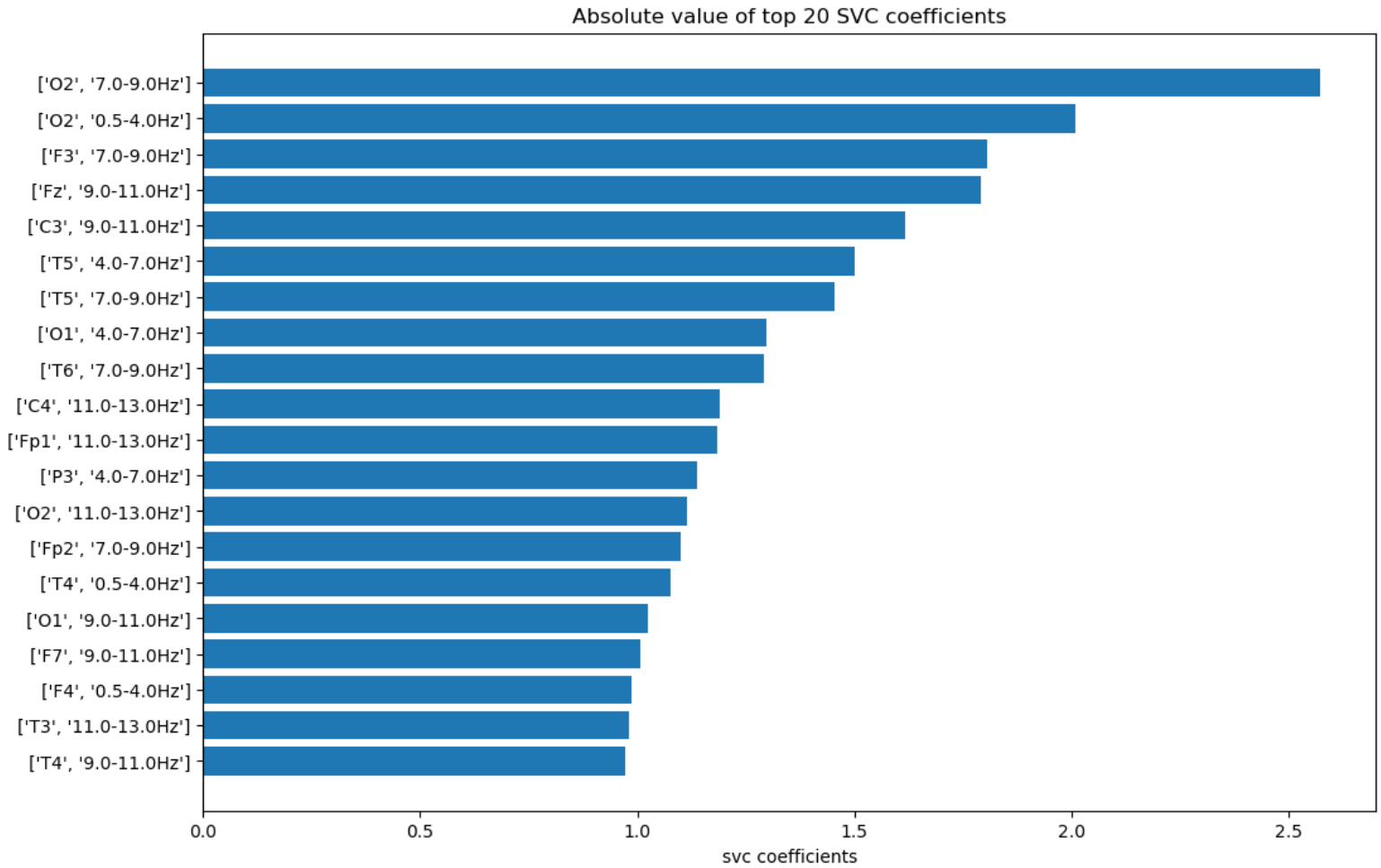
We achieved a better LOSO validation accuracy than all the existing models in literature.

[2] This is RF model in OpenNeuro, doi: 10.18112/openneuro.ds004504.v1.0.1

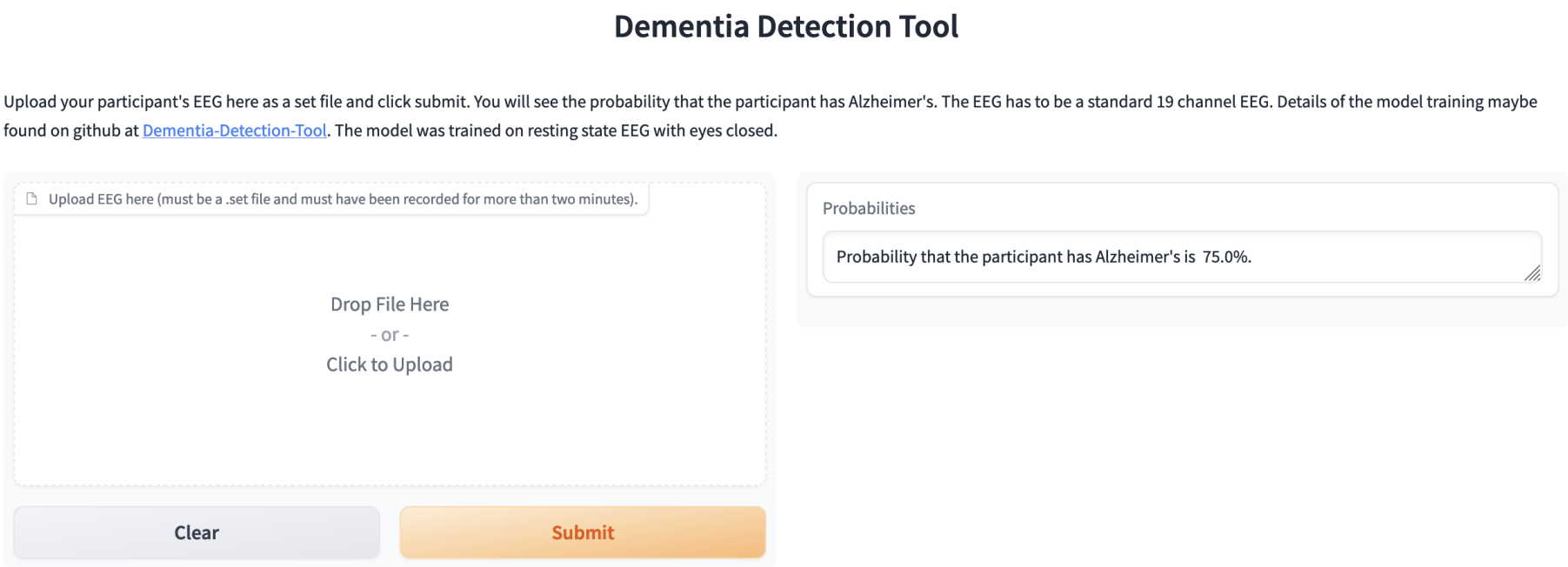
[3] This is a convolutional neural network based model in IEEE, doi: 10.1109/ACCESS.2023.3294618

Feature importances

To gauge which of the 90 features (15 channels and 6 freq. bands) are more important, we plot the absolute value of top twenty coefficients from SVC classifier.



Product: App



We have hosted the trained model on a web-based app on Hugging Face. It takes in a EEG file and uses our model to predict the probability that the EEG belongs to a patient with Alzheimer's.

Conclusions and Future Directions

We showed that a support vector classifier trained on power spectrum of EEG identifies Alzheimer's in patients of age above 50 yrs with an accuracy of over 89%.

We observed that modifying the frequency bands compared to what is used in the literature and in particular: splitting beta band (8-13 Hz) into subbands greatly improves the accuracy.

We also identified that 'O2' channel of the EEG (located at the rear of the scalp) in the frequency range 7 - 9 Hz is the most useful distinguishing feature for Alzheimer's.

This will help in building an early screening tool for Alzheimer's.

For future: it will be helpful to use this technique for also distinguishing Alzheimer's from other types of dementia: such as Frontotemporal dementia. We are currently exploring usefulness of other features such as spectral connectivity, in addition to power spectrum, for this problem.