Using EEG to make Biomarkers for Alzheimer's Research Project

Siddharth Nori [*], Ranganathan Srinivasan [†]

Abstract

Alzheimer's disease (AD) is associated with disrupted neural synchrony and altered network connectivity, which can be assessed using electroencephalography (EEG). In this paper, we employed a generalized Kuramoto model to quantify directed functional coupling between EEG channels in the alpha frequency range (8–12 Hz) for both AD patients and cognitively normal (CN) controls. The model extends the classic Kuramoto framework by estimating interaction strengths and preferred phase lags for each directed channel pair via regularized linear regression, yielding coupling matrices K_{ij} and phase offsets Φ_{ij}. From these matrices, we computed node-level graph measures of out-strength and in-strength to characterize the influence each channel exerts on and receives from others. Comparative analysis revealed distinct coupling patterns between AD and CN groups, with AD networks exhibiting altered connectivity profiles. To evaluate the diagnostic utility of these network features, we conducted a classification experiment using Support Vector Machine (SVM) with RBF kernel, XGBoost, and Multi-Layer Perceptron (MLP) models on 14 electrodes after data preprocessing and channel selection. An 80/20 train-test split was applied. MLP achieved the highest average test accuracy (0.8824), outperforming XGBoost (0.8235) and SVM (0.7647). Electrode-specific performance indicated that occipital (O1, O2) and frontal (F3, F7) regions contributed most to classification accuracy, with O2 reaching 0.9412 accuracy using MLP. These findings suggest that generalized Kuramoto-based coupling measures combined with machine learning classification can effectively discriminate AD from CN EEG recordings, and highlight specific brain regions as potential EEG biomarkers for early detection.

Introduction

Alzheimer's disease is a progressive neurodegenerative disorder that disrupts neural communication and gradually erodes memory, reasoning, and other cognitive abilities[5]. Its underlying science that causes these effects includes beta-amyloid plaque deposition and tau protein tangles, both of which impair synaptic signaling and contribute to widespread dysfunction throughout the brain network. While current diagnostic tools rely heavily on cognitive testing and neuroimaging, these methods often detect the disease only after significant neuronal loss has occurred. In recent years, electroencephalography (EEG) has gained attention as a non-invasive, cost-effective technique for capturing subtle alterations in brain activity that may emerge much earlier in the disease process[2, 3]. Along with mathematical models of neural synchronization, EEG analysis provides insight into the functional connectivity of the brain, allowing researchers to identify potential biomarkers for earlier diagnosis. Building on this, classification models can be applied to EEG data to distinguish between Alzheimer's patients and healthy controls, while also highlighting which channels contribute most strongly to predictive accuracy. This study applies the Generalized Kuramoto Model to EEG datasets from Alzheimer's patients and controls to quantify network connectivity patterns and integrates classification analysis to pinpoint the electrodes and brain regions most affected in the early stages of the disease.

[*]South Brunswick High School
[†]Lynbrook High School

Methods

Standard Kuramoto Model

The classical Kuramoto model[1, 4] describes the dynamics of N coupled phase oscillators according to

$$\frac{d\theta_i}{dt} = \omega_i + \frac{1}{N}\sum_{j=1}^{N} W_{ij}\sin(\theta_j - \theta_i), \tag{1}$$

where:

- $i = 1, \ldots, N$ indexes the oscillators.
- ω_i are the natural frequencies drawn from a distribution $p(\omega)$. A Lorentzian distribution is often chosen for analytical tractability while a Gaussian distribution is commonly used in simulations.
- $W_{ij} > 0$ indicates attractive interaction that promotes phase synchrony, and $W_{ij} < 0$ indicates repulsive interaction that promotes desynchronization.
- $\theta_i(t)$ is the instantaneous phase of oscillator i.

In the case of a fully connected network with uniform coupling $W_{ij} = K$, the model simplifies to

$$\frac{d\theta_i}{dt} = \omega_i + \frac{K}{N}\sum_{j=1}^{N} \sin(\theta_j - \theta_i). \tag{2}$$

The collective synchronization of the system can be characterized by the Kuramoto order parameter[6]

$$r(t)e^{i\psi(t)} = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_j(t)}, \tag{3}$$

where $r(t) \in [0, 1]$ measures the degree of phase coherence across oscillators and $\psi(t)$ is the average phase.

Generalized Kuramoto Model

To estimate directed coupling from EEG data, we generalized the Kuramoto model by allowing the interaction strength and phase offset to vary for each directed pair of oscillators. The dynamics for oscillator i are modeled as

$$\frac{d\theta_i}{dt} = \omega_i + \sum_{j\neq i} \left[a_{ij}\sin(\theta_j - \theta_i) + b_{ij}\cos(\theta_j - \theta_i)\right] + \varepsilon_i(t), \tag{4}$$

where a_{ij} and b_{ij} are coefficients estimated from the data using regularized linear regression, and $\varepsilon_i(t)$ is residual noise. The coupling magnitude and preferred phase lag are recovered as

$$K_{ij} = \sqrt{a_{ij}^2 + b_{ij}^2}, \quad \Phi_{ij} = \arctan 2\left(b_{ij}, a_{ij}\right). \tag{5}$$

Here K_{ij} quantifies the strength of influence from oscillator j to oscillator i, and Φ_{ij} quantifies the phase offset at which this influence is maximal.

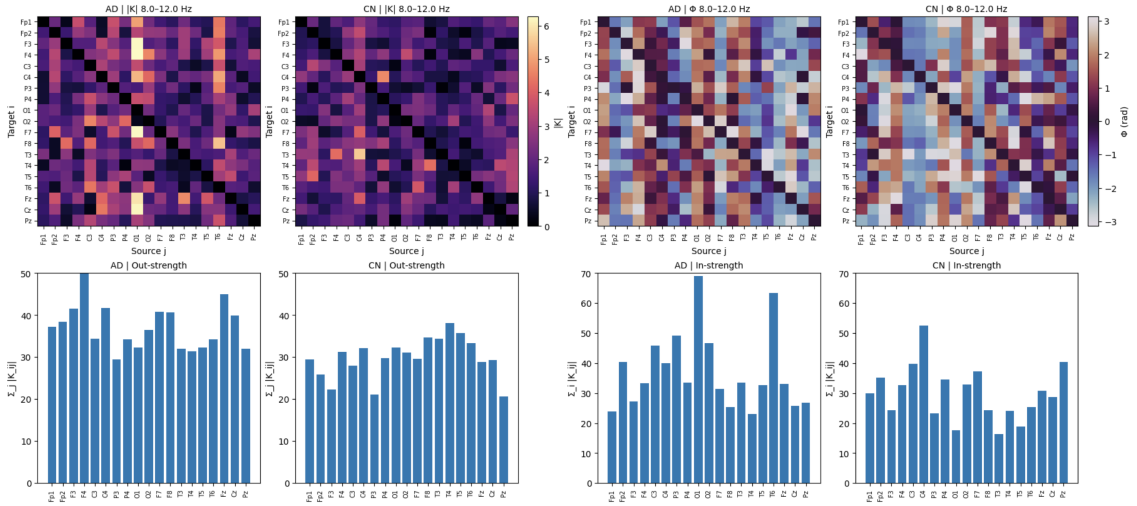


Figure 1: Directed coupling analysis and network strength measures for Alzheimer's disease (AD) and control (CN) groups in the alpha frequency band (8–12 Hz). Top row: Directed coupling matrices K_{ij} (left) and preferred phase lag matrices Φ_{ij} (right) estimated using the generalized Kuramoto model for representative subjects from AD and CN groups. Bottom row: Node-level graph measures showing out-strength $\sum_j |K_{ij}|$ (left) and in-strength $\sum_j |K_{ji}|$ (right) for each EEG channel, quantifying the total influence exerted on and received from other channels. Differences in coupling patterns and strength distributions between groups highlight increased network strength in AD.

Connection to Alzheimer Disease EEG Data

EEG recordings were preprocessed by selecting channels of interest, resampling to reduce computational cost, and bandpass filtering to isolate the alpha frequency range (8–12 Hz). The analytic signal was computed for each channel using the Hilbert transform, and the instantaneous phase was extracted as the argument of the analytic signal. The phase time series were unwrapped in time and numerically differentiated to estimate $d\theta_i/dt$.

For each channel i, the predictors consisted of the sine and cosine of the pairwise phase differences with every other channel j. A linear regression model with either Lasso regularization was used to fit these predictors to the measured $d\theta_i/dt$. The intercept provided an estimate of the intrinsic frequency ω_i, and the fitted coefficients yielded a_{ij} and b_{ij}, which were then converted to K_{ij} and Φ_{ij} as above.

From the directed coupling matrix K, we computed node-level graph measures. The out strength of node i was defined as $\sum_j |K_{ij}|$ and quantifies the total influence of i on other nodes. The in strength of node i was defined as $\sum_j |K_{ji}|$ and quantifies the total influence that i receives from other nodes. These measures were compared between the Alzheimer disease group and the control group.

Results

By analyzing the EEG data and observing the cohen's d value between the control and alzheimers population, we can deduce which electrodes hold the most discriminatory power. In this case, Cohen's d is a measure of difference between averages. The magnitude of this value tells us the most discrimatory electrode shown by differences in the data. The results showed the O2 electrode as the biggest differentiator. This result remains consistent with the classification experiment produced below.

The EEG-based classification experiment evaluated the performance of Support Vector Ma-

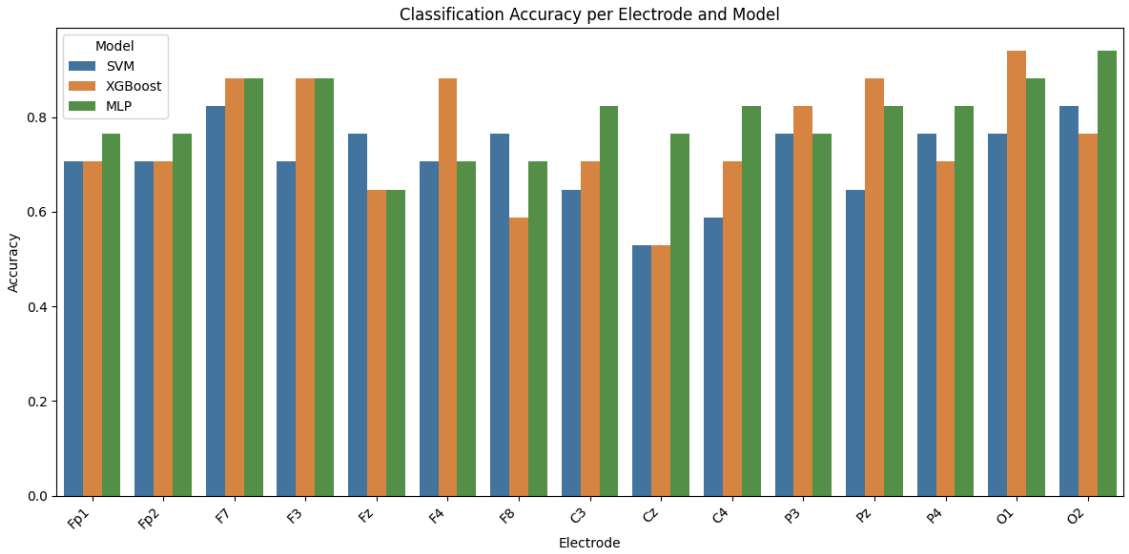


Figure 2: Bar chart showing per-electrode performance across the 15 EEG channels, with occipital electrodes (O1, O2) achieving the highest accuracies. The MLP model at O2 reached peak performance (94.12%), demonstrating the superior discriminative power of posterior brain regions for distinguishing Alzheimer's disease from control subjects.

chine (SVM) with RBF kernel, XGBoost, and Multi-Layer Perceptron (MLP) models across 19 electrodes, using an 80% training and 20% testing dataset split. Due to missing feature data, electrodes T7, T8, P7, P8, and Oz were excluded from analysis, resulting in 14 electrodes being evaluated. Accuracy varied across electrodes, with MLP frequently outperforming the other models. For instance, electrode O2 achieved the highest recorded accuracy of 0.9412 with MLP, followed by O1 with 0.8824. XGBoost demonstrated strong performance on certain electrodes, such as F7 and F3, both achieving 0.8824 accuracy, while SVM showed competitive results on electrodes like F7 (0.8235) and Fz (0.7647). Overall, MLP yielded the highest average test accuracy across all electrodes at 0.8824, outperforming XGBoost (0.8235) and SVM (0.7647). These results indicate that certain brain regions, particularly the occipital and frontal areas, may provide more discriminative EEG biomarkers for Alzheimer's detection.

Conclusion

In this paper, we have effectively addressed the late diagnosis of Alzheimer's by pinpointing which regions of the brain exhibit the most differences between AD and Control patients. Namely, we see that, when testing with MLP and SVM, there was a prevalence for electrodes such as O2 and O1, as well as electrodes such as F7, revealing more definitive biomarkers to differentiate patients. This will aid medical facilities in diagnosing Alzheimer's in patients before severe damage has occurred in the brain.

The next steps towards more accurate detection of early onset Alzheimer's would be to work with more data and more powerful devices which can handle more computationally expensive loads. Considering the results provided from taking Cohen's d to find the greatest differentiator of Alzheimer's, we can utilize MLP and SVM and target EEG scans for certain electrodes such as 02. Primarily, we can now hone our search results to the frontal and occipital regions of the brain, which can help aid in an earlier diagnosis of the disease.

Despite these promising results, we must account for the relatively small dataset considering the millions of people with Alzheimer's. In addition, we can consider using more advanced neuroimaging techniques such as those MEG or fMRI, which will provide more accurate spacial

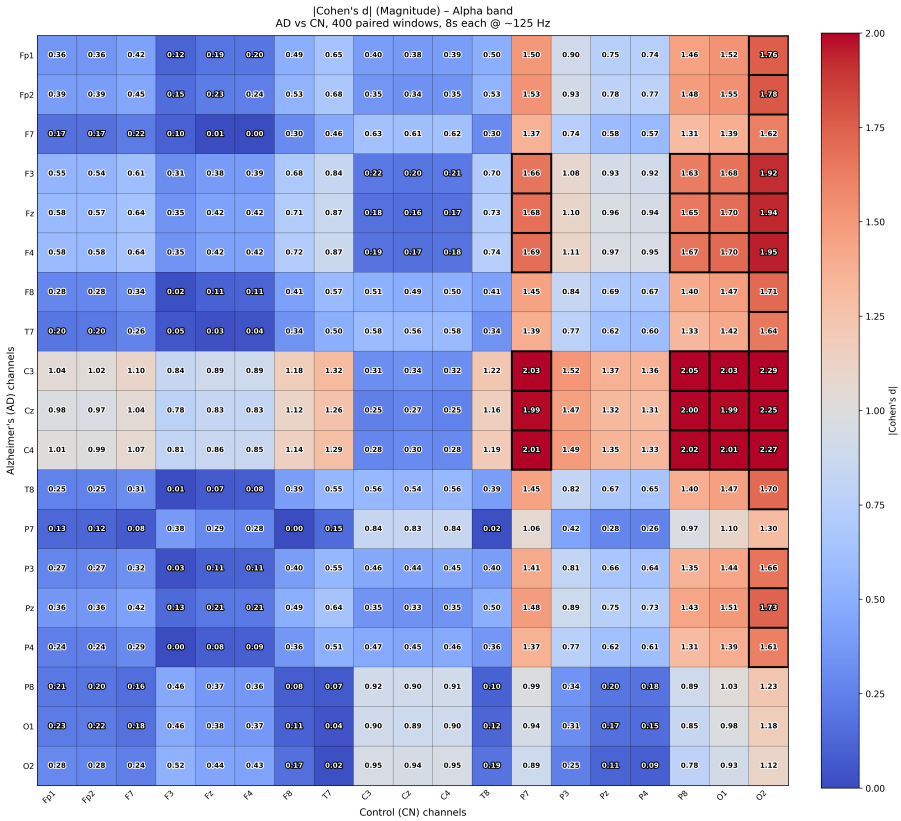
Overall Model Test Accuracy

Model	Test Accuracy
MLP	**0.8824**
XGBoost	0.8235
SVM (RBF)	0.7647

Per-Electrode Accuracy by Model

Electrode	SVM	XGBoost	MLP
Fp1	0.7059	0.7059	0.7647
Fp2	0.7059	0.7059	0.7647
F7	**0.8235**	0.8824	0.8824
F3	0.7059	0.8824	0.8824
Fz	0.7647	0.6471	0.6471
F4	0.7059	0.8824	0.7059
F8	0.7647	0.5882	0.7059
C3	0.6471	0.7059	0.8235
Cz	0.5294	0.5294	0.7647
C4	0.5882	0.7059	0.8235
P3	0.7647	0.8235	0.7647
Pz	0.6471	0.8824	0.8235
P4	0.7647	0.7059	0.8235
O1	0.7647	**0.9412**	0.8824
O2	**0.8235**	0.7647	**0.9412**

Figure 3: Overall and per-electrode classification accuracy for SVM (RBF), XGBoost, and MLP models trained on EEG data to distinguish Alzheimer's disease from control subjects. The dataset was split into 80% training and 20% testing. The top table shows overall test accuracies, with MLP achieving the highest (88.24%). The bottom table reports per-electrode accuracy, with bold values indicating the highest for each electrode. The O2 electrode with the MLP model achieved the best individual accuracy (94.12%).



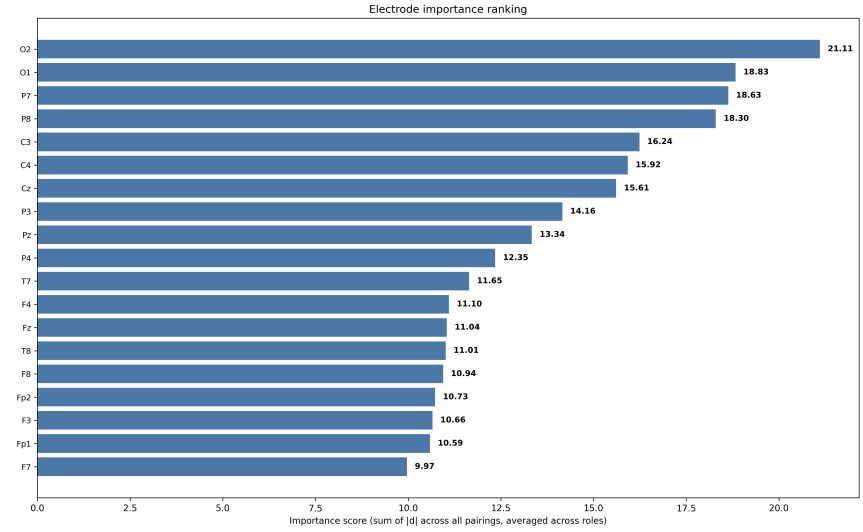


Figure 4: AD vs Control differences in EEG for the alpha band. The top is a 19 x 19 heatmap for all electrodes from an AD and a CN patient. The magnitude of Cohen's d difference is used to assign values of difference. Significant boxes have a bolded border. The bottom takes the sum of the d values to find the electrode with the greatest differentiable potential

resolution (which would help us further pin the region in the head where Alzheimers is present. However, these methods are extremely expensive and would be possible in the future in a controlled environment. We must also account for the time taken during data collection using these advanced methods. Thus, our results will begin to help us in the mitigation of Alzheimer's disease, however, there is far more to be discovered in the future if given the access and time to expand our dataset and/or use more accurate forms of neuroimaging.

References

[1] Juan A Acebrón et al. "The Kuramoto model: A simple paradigm for synchronization phenomena". In: *Reviews of modern physics* 77.1 (2005), pp. 137–185.

[2] András Attila Horváth et al. "EEG and ERP biomarkers of Alzheimer's disease: a critical review". In: (2018).

[3] Fernando Maestú et al. "The importance of the validation of M/EEG with current biomarkers in Alzheimer's disease". In: *Frontiers in human neuroscience* 13 (2019), p. 17.

[4] Francisco A Rodrigues et al. "The Kuramoto model in complex networks". In: *Physics Reports* 610 (2016), pp. 1–98.

[5] Philip Scheltens et al. "Alzheimer's disease". In: *The Lancet* 397.10284 (2021), pp. 1577–1590.

[6] Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Chapman and Hall/CRC, 2024.