DETECTING ALZHEIMER DISEASE IN EEG DATA WITH MACHINE LEARNING AND THE GRAPH DISCRETE FOURIER TRANSFORM

Xavier S. Mootoo
Department of Electrical Engineering and Computer Science
York University
Toronto, ON M3J 1P3, Canada
xmootoo@gmail.com

Alice Fours
Faculty of Medicine
University of British Columbia
Vancouver, BC V6T 1Z3, Canada
afours@student.ubc.ca

Chinthaka Dinesh
School of Engineering Science
Simon Fraser University
Burnaby, BC V5A 1S6, Canada
hchintha@sfu.ca

Mohammad Ashkani
Vancouver School of Economics
University of British Columbia
Vancouver, BC V6T 1L4, Canada
mashkani@student.ubc.ca

Adam Kiss
School of Medicine
New York Medical College
Valhalla, NY 10595, United States
akiss@student.nymc.edu

Mateusz Faltyn
Department of Bioinformatics
University of British Columbia
Vancouver, BC V6T 1Z3, Canada
faltynm@student.ubc.ca

ABSTRACT

Alzheimer Disease (AD) poses a significant and growing public health challenge worldwide. Early and accurate diagnosis is crucial for effective intervention and care. In recent years, there has been a surge of interest in leveraging Electroencephalography (EEG) to improve the detection of AD. This paper focuses on the application of Graph Signal Processing (GSP) techniques using the Graph Discrete Fourier Transform (GDFT) to analyze EEG recordings for the detection of AD, by employing several machine learning (ML) and deep learning (DL) models. We evaluate our models on publicly available EEG data containing 88 patients categorized into three groups: AD, Frontotemporal Dementia (FTD), and Healthy Controls (HC). Binary classification of dementia versus HC reached a top accuracy of 85% (SVM), while multiclass classification of AD, FTD, and HC attained a top accuracy of 44% (Naive Bayes). We provide novel GSP methodology for detecting AD, and form a framework for further experimentation to investigate GSP in the context of other neurodegenerative diseases across multiple data modalities, such as neuroimaging data in Major Depressive Disorder, Epilepsy, and Parkinson disease.

1 INTRODUCTION

Alzheimer Disease (AD) is a progressive and irreversible neurodegenerative disorder characterized by memory loss, impaired problem-solving, reasoning, language difficulties, and physical decline (Scheltens et al., 2016); (Porsteinsson et al., 2021). Advancing age is the predominant risk factor for AD due to the accumulation of abnormal proteins like Aβ plaques and tau tangles, along with reduced neuronal repair capacity during the aging process (Lesné et al., 2006); (Binder et al., 2005); (Knopman et al., 2021). Genetic predisposition and co-morbidities further increase risk for this multi-factorial disease (Rabinovici, 2019); (Lindsay et al., 2002); (Silva et al., 2019).

AD and Frontotemporal Dementia (FTD) represent two distinct yet complex disorders with overlapping clinical features (Götz & Ittner, 2008); (Lindau et al., 2000); (Pachana et al., 1996). FTD is

another neurodegenerative disorder affecting cognitive functions. Unlike AD, which predominantly impacts the hippocampal and cerebral cortex regions, FTD is localized to the frontal and temporal lobes of the brain that govern behavior, personality, and language (Olney et al., 2017); (Levy et al., 1996). Subtypes of FTD manifest with specialized symptoms, including behavioral and language variants (Olney et al., 2017); (Pasquier et al., 2004). Given their symptomatic overlaps, distinguishing between AD and FTD in a clinical setting can be challenging (Warren et al., 2013); (Ritter et al., 2017); Rosen et al. (2002).

AD constitutes a significant majority of dementia cases, contributing to approximately 60-70% of diagnoses (Giridharan et al., 2022); (van der Flier & Scheltens, 2005). The World Health Organization (WHO) reported that, in 2020, a total of 50 million individuals lived with dementia globally, with projections suggesting a rise to 152 million by 2050. (Nichols et al., 2022); (Houmani et al., 2018); (Giridharan et al., 2022). Dementia-related disorders are diagnosed upon the emergence of symptoms through neurological assessments and managed through lifestyle modifications aimed at attenuating cognitive decline (Yiannopoulou & Papageorgiou, 2020); (Wahl et al., 2019); (Dominguez et al., 2021). With an aging demographic, there is a pressing need for the implementation of diagnostic tools capable of detecting the onset of these diseases (Knopman et al., 2021).

AD shows distinct alterations in the synchronicity and complexity of brain waves compared with normal brain and other types of dementia (Zande et al., 2018); (Houmani et al., 2018); (Dauwels et al., 2010); (Jeong, 2004). Electroencephalography (EEG), a non-invasive and cost-effective technique, stands as a promising modality to discern these changes (Thakor & Sherman, 2012); (Dauwels et al., 2010). EEG has proven to be invaluable, particularly in the detection and study of epilepsy, where it aids in detecting abnormal brain activity associated with seizures (Mansouri et al., 2012); (Smith, 2005). Previous studies have also explored the application of Graph Signal Processing (GSP) to model epileptic activity, showcasing the versatility and potential of utilizing EEG to detect clinically relevant conditions, for which we adapt to the context of dementia in this paper (Meena et al., 2022); (Mathur & Chakka, 2020).

2 Related Work

In recent years, the analysis of EEG signals has emerged as a promising avenue for the early detection of AD and several other neurodegenerative diseases (Cassani et al., 2018). Several studies have successfully harnessed the functional connectivity of electrode sites in EEG signals by feature engineering with GSP, utilizing standard machine learning (ML) models, or employing state-of-the-art Deep Learning (DL) methods such as Graph Neural Networks (GNNs) (Meena et al., 2022); (Mathur & Chakka, 2020); (Padole et al., 2018); (Song et al., 2021); (Vicchietti et al., 2023). This study contributes to this ongoing effort by benchmarking several ML models using GSP features created and the Graph Discrete Fourier Transform (GDFT) for AD detection in EEG data. We give an outline of the following sections below.

In (3.1) we provide a description of the dataset containing the EEG signals of AD, FTD, and Healthy Control (HC) patients. We describe our method for graph structure formation given from the raw EEG signal (3.2) and its derived GSP features including the GDFT (3.3). We benchmark these features on various ML and DL models (3.4), outlining model performance statistics in (4), before discussing our findings and conclusions in (5).

3 Methods

3.1 Data

We utilized the OpenNeuro ds004504 dataset, derived from recordings made at the 2nd Department of Neurology of AHEPA General Hospital, Thessaloniki (Miltiadous et al., 2023). The data was captured using a Nihon Kohden EEG 2100 clinical device fitted with 19 scalp and 2 reference electrodes, operating at a 500 Hz sampling rate and with a high-frequency filter at 70 Hz. The dataset contained three patient groups: AD, FTD, and HC, with average recording durations of 13.5 minutes, 12 minutes, and 13.8 minutes respectively. The cumulative durations for AD, FTD, and HC were 485.5 minutes, 276.5 minutes, and 402 minutes respectively.

During the preprocessing phase, raw data was transformed into a BIDS-compliant .set format, initially exporting it in .eeg format. The preprocessed and denoised recordings were stored in designated sub-0XX folders within a derivatives subfolder, segregating unprocessed and processed data. The preprocessing pipeline included a Butterworth band-pass filter at the 0.5-45 Hz range, re-referencing to A1-A2, and utilizing Artifact Subspace Reconstruction (ASR). Independent Component Analysis (ICA) was executed to shape the data into 19 ICA components, and automatic rejection of eye and jaw artifacts was implemented. The researchers abstained from including automatic annotations of artifacts to circumvent language compatibility issues. This preprocessed dataset can be found in the `derivatives` folder, containing the files used for this study. The following table below depicts additional information for each group including mean Mini-Mental State Examination (MMSE) and Median Disease Duration (MDD).

Table 1: Summary of patient characteristics for the OpenNeuro ds004504 dataset.

	Characteristics			
Group	n	Mean Age (years)	Mean MMSE	MDD (months)
Alzheimer Disease	36	66.4 ± 7.9	25	13.5
Frontotemporal Dementia	23	63.6 ± 8.2	N/A	12
Healthy Controls	29	67.9 ± 5.4	N/A	13.8

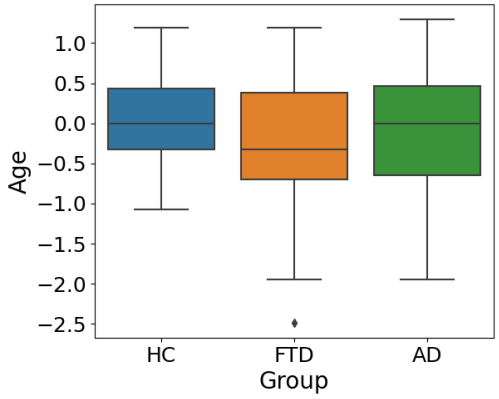

Figure 1: Box plot of age by group.

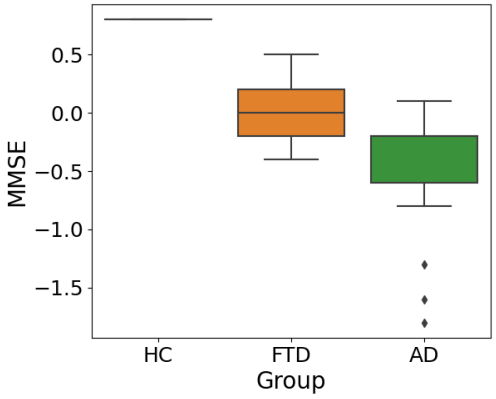

Figure 2: Box plot of MMSE by group.

3.2 Graph Structures

In graph theory, we define an undirected weighted graph by a triplet $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V}$ is a set of nodes (or vertices) representing individual objects, and $\mathcal{E}$ is an edge set representing relations between those objects with each element of the form $(i, j, \mathbf{W}_{i,j})$ for some $i, j \in \mathcal{V}$ and $\mathbf{W} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ being the weight matrix that represents the quantitative relationship between each node pair. An EEG signal is a matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$ where N is the number of electrodes and M is the number of discrete time points, written as:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(N)} \end{bmatrix}, \tag{1}$$

where each $\mathbf{x}^{(i)} \in \mathbb{R}^M$ is an individual electrode signal for $1 \leq i \leq N$. Given $\mathbf{X}$, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ is formed where the vertex set $\mathcal{V} = \{1, 2 \ldots, N\}$ are the nodes representing the set of electrodes. The edge set $\mathcal{E}$ and weight matrix $\mathbf{W}$ are created as follows. For parameters $\theta, k \in (0, \infty)$, we first define $\mathbf{W} \in \mathbb{R}^{N \times N}$ via a Gaussian kernel (ter Haar Romeny, 2003):

$$\mathbf{W}_{i,j} = \begin{cases} \exp\left(-\frac{d(\mathbf{x}^{(i)},\mathbf{x}^{(j)})^2}{2\theta^2}\right) & \text{if } d(\mathbf{x}^{(i)},\mathbf{x}^{(j)}) \leq k \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

for each $1 \leq i, j \leq N$, which represents the degree of connectivity between each electrode pair (i, j). For our experiments, we chose d to be the metric induced by the Euclidean norm $\| \cdot \|_2$, i.e. $d(\mathbf{x}, \mathbf{y}) = \| \mathbf{x} - \mathbf{y} \|_2$. The edge set is defined as the indices of the nonzero entries in $\mathbf{W}$:

$$\mathcal{E} = \{(i, j) \mid \mathbf{W}_{i,j} \neq 0,\ 1 \leq i, j \leq N\}. \tag{3}$$

3.3 Graph Signal Processing

We introduce the relevant GSP methods to create the input features from the EEG signal $\mathbf{X}$ and graph $\mathcal{G}$, before using these features to train our ML models in 3.4 (Ortega et al., 2018); (Dong et al., 2020);(Li et al., 2021). For standard methods in GSP see (Ortega, 2022). The first step is to obtain the **Combinatorial Graph Laplacian** matrix (Chung & Langlands, 1996):

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \tag{4}$$

where $\mathbf{D}$ is the diagonal degree matrix defined by:

$$\mathbf{D}_{i,j} = \begin{cases} \sum_{k=1}^{N} \mathbf{W}_{i,k} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

Since $\mathbf{L}$ is normal, by the Spectral Theorem we can write $\mathbf{L} = \mathbf{U}\,\mathbf{K}\,\mathbf{U}^*$ where the columns of $\mathbf{U} \in \mathbb{R}^{N \times N}$ are the eigenvectors of $\mathbf{L}$. Here, $\mathbf{L}$ is a real symmetric matrix. Therefore, all eigenvalues and eigenvectors are real. Hence, we can simply write $\mathbf{L} = \mathbf{U}\,\mathbf{K}\,\mathbf{U}^T$. The GDFT is given by (Mathur & Chakka, 2020):

$$\mathbf{X}_{\text{GDFT}} = \mathbf{U}^T\,\mathbf{X}\,. \tag{6}$$

Let $\mathbf{C}$ be the covariance matrix of $\mathbf{X}_{\text{GDFT}}$. We project $\mathbf{C}$ onto the space of the Laplacian's eigenvectors to obtain:

$$\mathbf{T} = \mathbf{U}^T\,\mathbf{C}\,\mathbf{U}, \tag{7}$$

for which we compute the **Stationary Ratio** defined as:

$$r = \frac{\|\text{diag}(\mathbf{T})\|_2}{\|\,\mathbf{T}\,\|_F}, \tag{8}$$

where $\|\cdot\|_F$ denotes the Frobenius norm (Meena et al., 2022). Let $\mathbf{P} = \mathbf{L}\,\mathbf{X}$. We define the **Tik-norm** by:

$$y = \left(\sum_{i,j} (\mathbf{X} \odot \mathbf{P})_{i,j}\right)^2, \tag{9}$$

where $\odot$ is the Hadamard product and the sum is taken over all entries of $\mathbf{X} \odot \mathbf{P}$ (Meena et al., 2022). Other input features include:

1. **Total Variation (TV):** Total Variation measures the overall variation or dissimilarity in the data represented by matrix $\mathbf{X}$ (Sandryhaila & Moura, 2014). It is defined as:

$$TV(\mathbf{X}) = \sqrt{\sum_{i,j} \mathbf{W}_{i,j} \,\|\, \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \,\|_2^2}.$$

   Higher Total Variation indicates greater dissimilarity among data points within the dataset.

2. **Graph Energy:** Graph Energy is a measure that captures the structural properties of a graph represented by the Laplacian matrix $\mathbf{L}$ (Balakrishnan, 2004). It is calculated by:

$$E(\mathcal{G}) = \sum_{i=1}^{N} |\lambda_i|,$$

   where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{L}$. A higher Graph Energy indicates a more complex and interconnected graph structure.

3. **Spectral Entropy:** Spectral Entropy quantifies the uncertainty or randomness in the spectral distribution of a matrix $\mathbf{U}$ Rényi (1961). It is defined as:

$$H(\mathbf{U}) = -\sum_{i,j} \mathbf{U}_{i,j}^2 \log\left(\mathbf{U}_{i,j}^2\right).$$

   Higher Spectral Entropy values indicate a more diverse and less predictable spectral distribution.

4. **Signal Energy:** Signal Energy measures the total energy or magnitude of the data represented by matrix $\mathbf{X}$ (Rihaczek, 1968). It is computed as:

$$\beta = \sum_{i,j} \mathbf{X}_{i,j}^2 \,.$$

   Higher Signal Energy values indicate data with higher overall magnitude of the signal.

5. **Signal Power:** Signal Power, denoted by σ^2, is equivalent to the variance of the data in matrix $\mathbf{X}$ defined by:

$$\sigma^2 = \text{Var}(\mathbf{X}) = \frac{1}{N}\sum_{i=1}^{N} \| \mathbf{x}^{(i)} - \boldsymbol{\mu} \|_2^2,$$

   where $\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}^{(i)}$ is the mean of the dataset $\mathbf{X}$ (Rao & Swamy, 2018). High Signal Power values indicate greater variability in the data.

6. **Unique Spectral Cluster Labels:** Unique Spectral Cluster Labels, denoted by γ, represents the number of distinct clusters or groups identified when applying spectral clustering to the weight matrix $\mathbf{W}$. Spectral clustering is a technique for grouping data points based on their connectivity in a graph. γ indicates the number of distinct clusters found. In our experiments we applied the `spectral_clutsering` method to $\mathbf{W}$ with the `scikit-learn` package to compute γ (Pedregosa et al., 2011).

7. **Average Degree:** Average Degree, denoted by $\mu_{\mathbf{D}}$, quantifies the average connectivity of nodes in a graph represented by the degree matrix $\mathbf{D}$ (Ortega, 2022). It is calculated as:

$$\mu_{\mathbf{D}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{D}_{ii} \,.$$

   A higher $\mu_{\mathbf{D}}$ indicates greater connectivity of the graph.

8. **Heat Trace:** Heat Trace, represented by h, is calculated by taking the trace (sum of diagonal elements) of a modified Laplacian matrix $\mathbf{L}'$, where each element is exponentiated by $\mathbf{L}'_{i,j} = \exp(-\mathbf{L}_{i,j})$ (Xiao et al., 2009). It is defined as:

$$h = \text{Tr}(\mathbf{L}').$$

   The Heat Trace is related to diffusion processes on graphs and can capture how information or "heat" spreads across the graph.

9. **Diffusion Distance:** Diffusion Distance, denoted by h', measures the overall spread or extent of diffusion in a graph (Hammond et al., 2013). It is calculated with:

$$h' = \sum_{i,j} \mathbf{L}'_{i,j} \,.$$

   The Diffusion Distance of a graph provides insight into how information or influence propagates across the graph.

For each patient EEG signal, we generate the following features using the methods above:

$$\text{Features} = (r, y, TV(\mathbf{X}), E(\mathcal{G}), H(\mathbf{U}), \beta, \sigma^2, \gamma, \mu_{\mathbf{D}}, h, h') \tag{10}$$

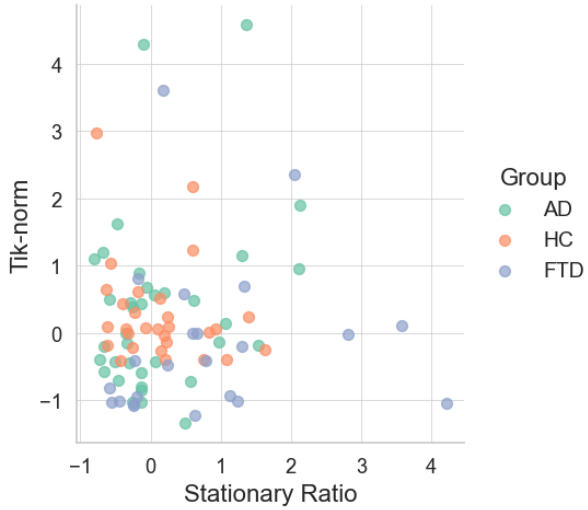

Figure 3: Plot of stationary ratio versus Tik-norm by group for AD, FTD, and HC.

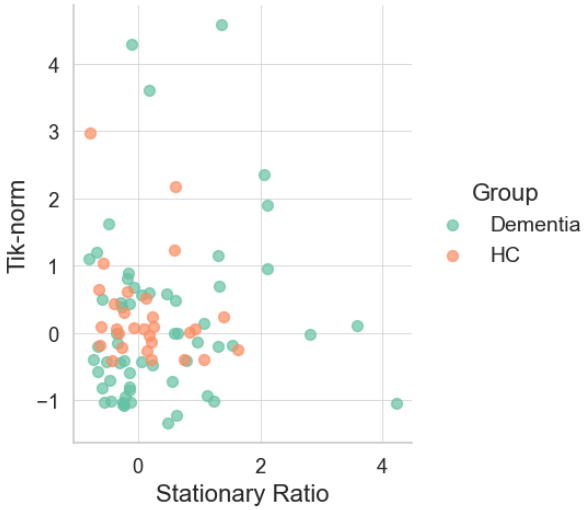

Figure 4: Plot of stationary ratio versus Tik-norm by group for Dementia and HC.

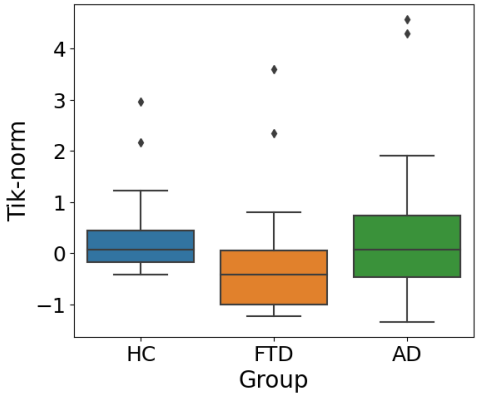

Figure 5: Box plot of Tik-norm by group.

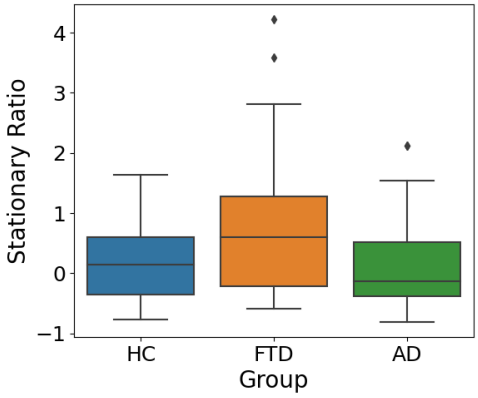

Figure 6: Box plot of stationary ratio by group.

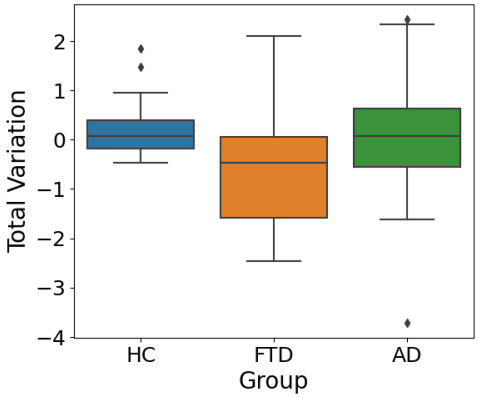

Figure 7: Box plot of total variation by group.

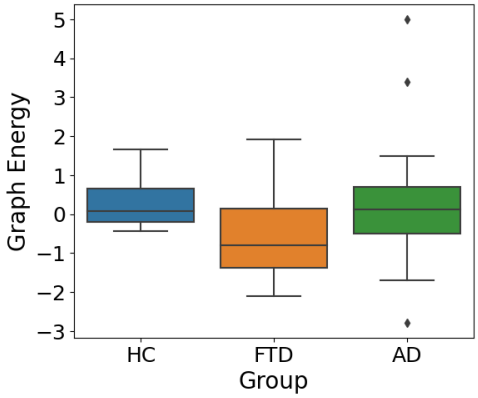

Figure 8: Box plot of graph energy by group.

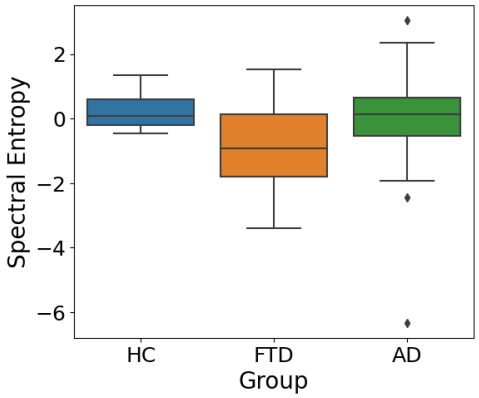

Figure 9: Box plot of spectral entropy by group.

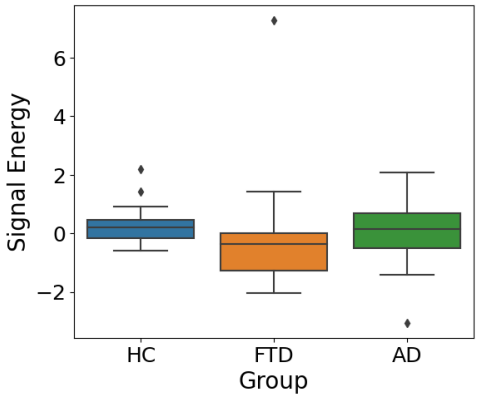

Figure 10: Box plot of signal energy by group.

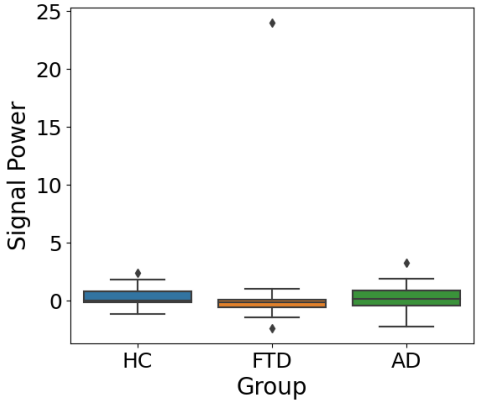

Figure 11: Box plot of signal power by group.

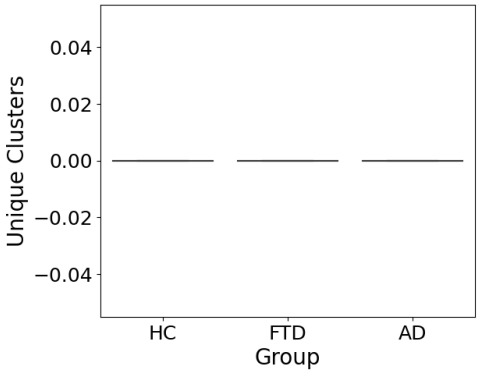

Figure 12: Box plot of unique spectral clusters by group.

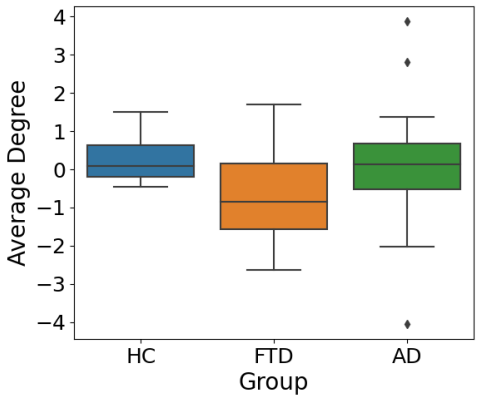

Figure 13: Box plot of average degree by group.

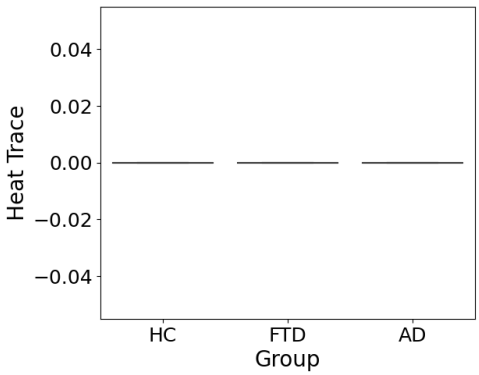

Figure 14: Box plot of heat trace by group.

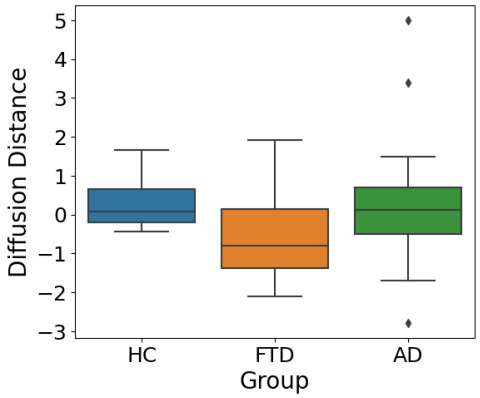

Figure 15: Box plot of diffusion distance by group.

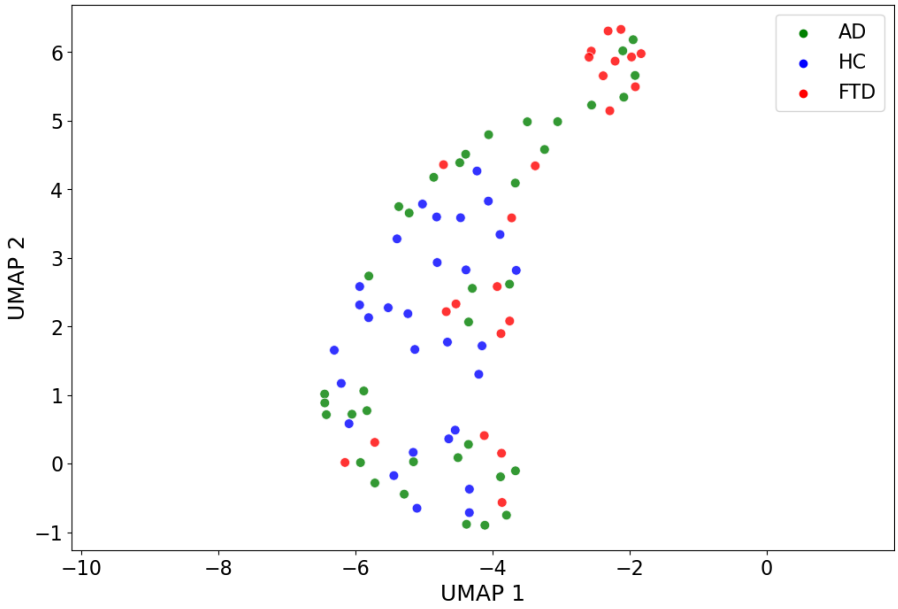

Figure 16: Plot of UMAP projected EEG features by group.

3.4 Models

We test our features from Equation 10 by training and evaluating several ML models below on the task of multiclass classification (AD, FTD, HC), and binary classification (Dementia, HC) where the Dementia group comprises of AD and FTD.

(a) **k-Nearest Neighbours (k-NN) Classification** : Given a new observation $\mathbf{x}_0$, the k-NN algorithm searches through the training dataset to find the k training examples that are closest to $\mathbf{x}_0$ based on a distance metric (e.g., Euclidean distance). The predicted label is then obtained by majority voting among the k nearest neighbours (Fix & Hodges, 1989).

(b) **Random Forest**: A Random Forest is a collection of decision trees $\{h(\mathbf{x}; \Theta_k), k = 1, \ldots\}$, where Θ_k are the parameters of each tree. A decision tree is a flowchart-like structure where each internal node represents a feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The root node is the feature that best splits the dataset into classes based on a decision criterion, and this process is recursively applied to the sub-sets of data reaching each internal node, constructing the tree until it reaches a specified depth or purity. The final prediction in a Random Forest is obtained by averaging the predictions of all the trees for regression, or by majority voting for classification (Ho, 1995).

(c) **XGBoost**: XGBoost builds an ensemble of decision trees. The prediction is given by $f(\mathbf{x}) = \sum_{k=1}^{K} h_k(\mathbf{x})$, where $h_k(\mathbf{x})$ is an individual tree, and K is the number of trees (Chen & Guestrin, 2016).

(d) **Logistic Regression**: Logistic Regression is a statistical model used for binary classification, where the probability of the positive class is modeled as $P(Y = 1|\mathbf{x}) = \frac{1}{1+\exp(-(\mathbf{w}^T\mathbf{x}+b))}$, where $\mathbf{w}$ is the weight vector, and b is the bias term (Menard, 2002). To convert this probability into a binary outcome, a threshold value, typically 0.5, is chosen. If $P(Y = 1|\mathbf{x}) \geq 0.5$, the prediction is 1; otherwise, the prediction is 0. The threshold can be adjusted based on the specific needs of a problem, such as minimizing false positives or negatives. In a multiclass setting, Logistic Regression can be extended using techniques like One-Versus-Rest (OvR) or Multinomial Logistic Regression, where the model estimates the probability of each class and picks the class with the highest estimated probability.

(e) **Support Vector Machine (SVM)**: SVM finds the hyperplane that maximizes the margin between the two classes. The decision function is given by $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}+b$, where $\mathbf{w}$ is the normal vector to the hyperplane (a subspace of 1 dimension lower) and b is the bias term (Cortes & Vapnik, 1995). For instance, in a 2-dimensional space (i.e. the Cartesian plane), a hyperplane is simply a line. If we have two classes of points scattered on the plane, SVM will find the line that best separates these two classes while keeping the maximum distance from the nearest points of each class to this line.

(f) **Naive Bayes**: Naive Bayes is a probabilistic classifier based on applying Bayes' Rule, which relates the conditional and marginal probabilities of random events, with the "naive" assumption of independence between features. Bayes' Rule is expressed as $P(c|\mathbf{x}) = \frac{P(\mathbf{x}|c)P(c)}{P(\mathbf{x})}$. However, computing $P(\mathbf{x}|c)$ directly can be challenging due to the high dimensionality of $\mathbf{x}$. The naive independence assumption simplifies this to $P(\mathbf{x}|c) = \prod_{i=1}^{n} P(x_i|c)$, leading to the formula $P(c|\mathbf{x}) \propto P(c)\prod_{i=1}^{n} P(x_i|c)$, where n is the number of features (Rish et al., 2001). This simplification allows for efficient computation and training of the model. The classifier then predicts the class c that maximizes $P(c|\mathbf{x})$, i.e., the class that is most probable given the feature vector $\mathbf{x}$.

(g) **TabTransformer**: TabTransformer employs self-attention mechanisms to capture interactions among features in tabular data (Huang et al., 2020). The self-attention mechanism is a key component of the transformer architecture, which allows the model to weigh the importance of different parts of the input differently when making predictions. Specifically, self-attention computes a weighted sum of all input features, where the weights are determined by the similarity between the query and key representations of the features, allowing each feature to attend over all others in a flexible manner. The self-attention formula is given by $\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}})\mathbf{V}$, where $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ are the query, key, and value matrices, respectively, and d_k is the dimension of the key vectors (Vaswani et al., 2017). This mechanism enables TabTransformer to effectively capture feature interactions in tabular data, even when relationships among features are complex and non-linear.

4 RESULTS

For each model, we utilized the full dataset, splitting it into a training set (35%), validation set (35%), and test set (30%). Tables 2 and 3 display accuracy for each classification task and model,

and Tables 4 and 5 display precision, recall, F1-score, and support for each class and each task.

$$\text{Accuracy} = \frac{TP}{TP + TN + FP + FN}, \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{13}$$

$$\text{F1-Score} = \frac{2(\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}}. \tag{14}$$

Here, TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives respectively. Precision measures the reliability of a model's positive classifications, while recall gauges its completeness in capturing positive samples. The F1-score offers a balanced evaluation of both, using their harmonic mean to provide a holistic view of model performance. Support refers to the number of occurrences of each class.

Table 2: Model performance on the test set for multiclass classification of AD, FTD, and HC. Highest accuracy is bolded.

Method	Accuracy
$k-$NN Classifier	0.37
Random Forest	0.33
XGBoost	0.41
Logistic Regression	0.41
SVM	0.37
Naive Bayes	**0.44**
TabTransformer	0.21

Table 2 displays Naive Bayes as the model with highest test accuracy at 44%, while the TabTransformer model had the lowest accuracy, achieving 21%. Aside from the TabTransformer, model performance is relatively close, implying that the TabTransformer may be overfitted to the training dataset, with poor generalization to the test set. With the strong assumption of independence of the features in Naive Bayes in 3.4, it may be the case that the features from 10 are sufficiently independent to harness the advantage of Naive Bayes. However, with a top accuracy that is just above random guessing rate (33%), this suggests that the choice of model is likely not to be substantially impacting accuracy, and points to an issue of separability of the classes and the inherent data distribution for AD, FTD, and HC; or the inability for the features in Equation (10) to incorporate enough information to distinguish between the three classes.

Table 3: Model performance on the test set for binary classification of dementia (AD, FTD) versus HC. Highest accuracy is bolded.

Method	Accuracy
$k-$NN Classifier	0.70
Random Forest	0.67
XGBoost	0.74
Logistic Regression	0.63
SVM	**0.85**
Naive Bayes	0.59
TabTransformer	0.71

Table 3 presents the performance of each model in the binary classification setting, trained to distinguishing between cases of dementia (AD, FTD) and HC. The SVM model, employing a linear kernel, displayed the highest accuracy of 85%, while the Naive Bayes model lagged with an accuracy of 59%. The SVM model's superior performance suggests that the dataset may be largely

linearly separable, although this should be further validated (Cortes & Vapnik, 1995). Conversely, the suboptimal performance of Naive Bayes is likely attributable to the violation of its feature independence assumption, signifying that the features are likely dependent in this binary classification context (Rish et al., 2001).

Table 4: Other model performance metrics on the test set for multiclass classification of AD, FTD, and HC.

Method	Precision	Recall	F1-Score	Support
AD Classification				
$k-$NN Classifier	0.50	**0.54**	**0.52**	13
Random Forest	0.40	0.46	0.43	13
XGBoost	0.44	0.31	0.36	13
Logistic Regression	**0.67**	0.31	0.42	13
SVM	0.60	0.23	0.33	13
Naive Bayes	0.50	0.15	0.24	13
TabTransformer	0.25	0.14	0.18	7
FTD Classification				
$k-$NN Classifier	**1.00**	0.20	0.33	10
Random Forest	0.00	0.00	0.00	10
XGBoost	0.67	0.40	0.50	10
Logistic Regression	0.75	0.30	0.43	10
SVM	**1.00**	0.30	0.46	10
Naive Bayes	0.75	**0.60**	**0.67**	10
TabTransformer	0.50	0.17	0.25	6
HC Classification				
$k-$NN Classifier	0.09	0.25	0.13	4
Random Forest	0.25	0.75	0.38	4
XGBoost	0.25	0.75	0.38	4
Logistic Regression	0.24	**1.00**	0.38	4
SVM	0.21	**1.00**	0.35	4
Naive Bayes	**0.27**	**1.00**	**0.42**	4
TabTransformer	0.12	**1.00**	0.22	1

Table 4 further evaluates each models' performance in the multiclass classification setting, displaying precision, recall, F1-Score, and support. For AD classification, Logistic Regression shows the highest precision at 67%, making it the most reliable in terms of positive predictive value, possibly due to its parametric nature which might be capturing the underlying data distribution effectively (Menard, 2002).

In FTD classification, $k-$NN and SVM both achieve perfect precision but differ substantially in recall and F1-Score; this could be indicative of k-NN's sensitivity to noise and outliers, as well as SVM's capacity for finding a more discriminative hyperplane in the feature space (Fix & Hodges, 1989); (Rish et al., 2001). Naive Bayes emerges as the most balanced model for FTD with the highest F1-Score of 67%, perhaps owing to its probabilistic framework that may better capture the conditional dependencies among features (Rish et al., 2001).

For HC, all models exhibit low precision, with Naive Bayes performing the best at 27%. This may indicate that the feature representation is not robust enough for effectively distinguishing this class, suggesting that the current features in Equation 10 are not sufficient for the task of multiclass classification. Additionally, given a low support of only $n = 4$ for HC, as a function of the relatively small dataset size of $n = 88$, values within the HC table are highly susceptible to the outcome of this random sample.

Table 5: Other model performance metrics on the test set for binary classification of Dementia and HC.

Method	Precision	Recall	F1-Score	Support
Dementia Classification				
$k-$NN Classifier	0.89	0.74	0.81	23
Random Forest	0.94	0.65	0.77	23
XGBoost	0.94	0.74	0.83	23
Logistic Regression	0.81	0.74	0.77	23
SVM	0.85	**1.00**	**0.92**	23
Naive Bayes	**1.00**	0.52	0.69	23
TabTransformer	**1.00**	0.69	0.82	13
HC Classification				
$k-$NN Classifier	0.25	0.50	0.33	4
Random Forest	0.27	0.75	0.40	4
XGBoost	**0.33**	0.75	**0.46**	4
Logistic Regression	0.00	0.00	0.00	4
SVM	0.00	0.00	0.00	4
Naive Bayes	0.27	**1.00**	0.42	4
TabTransformer	0.20	**1.00**	0.33	1

Table 4 extends the evaluation to the binary classification task, assessing the models' ability to distinguish between dementia cases and HC. In the dementia setting, all models scored relatively high for precision. Although, Naive Bayes achieved perfect precision, it had the lowest recall of 52% suggesting that its assumptions of feature independence may only hold in the positive class Rish et al. (2001). SVM achieved relatively high precision and perfect recall, suggesting the data distribution is likely linear separable in these two classes Cortes & Vapnik (1995). All models achieved relatively high F1-score with SVM being the highest.

The precision for all models in the HC classification was notably lower compared to the dementia classification, suggesting that while our features from Equation (10) are effective in identifying dementia traits, they may not be sufficient to distinguish attributes of healthy individuals. It is also worth noting that due the support value of 4, the values in HC Classification section are potentially subject to fluctuations and can vary by training run.

5 Conclusion

Our method introduces a novel approach utilizing GSP techniques with the GDFT for generating features to detect AD in EEG recordings. SVM was accurate in differentiating between dementia and HC in 85% of cases. However distinguishing AD from FTD was challenging in the multiclass classification setting across all models, yielding a top accuracy of only $\%44$. We conclude that our features from Equation (10) are useful for identifying healthy versus dementia patients, but they are not applicable for determining the type of dementia a patient might have.

The discrepancy between performance in binary classification and multiclass classification suggests that distinguishing AD and FTD is significantly more challenging and may require a more nuanced approach. We speculate that this challenge arises due to: (1) the similarities between AD and FTD features given by GSP; (2) the wide variability in FTD types (e.g., sporadic, behavioural); (3) the variability of disease stages in AD; (4) the lack of model fit (e.g., TabTransformer overfitting, traditional ML models underfitting).

In future work, we aim to address the variation seen within AD by incorporating multiclass classification of different stages of AD, including the pre-clinical, mild, moderate and severe stages, as well as implementing a similar approach for FTD types. This would not only provide more information for the machine learning model to use, but would also allow for potential earlier disease detection and informed diagnoses to prevent subsequent worsening of symptoms. We acknowledge the limitations of using mostly standard machine learning models, so we plan to implement state-of-the-art

deep learning (DL) methods such as graph neural networks (GNNs) that utilize node features, edge features, and graph-level features given by our GSP method. It is worth noting that these methods are useful to explore but require technical proficiency to integrate into a clinical setting. Therefore, future work can also include clinician education and improving ease of use.

6 Acknowledgements

We would like to express our gratitude to the researchers from the Department of Informatics and Telecommunications at the University of Ioannina, the Department of Electrical and Computer Engineering at the University of Western Macedonia, and the 2nd Department of Neurology at AHEPA University Hospital, Aristotle University of Thessaloniki, for generously sharing their dataset Miltiadous et al. (2023). Their open collaboration greatly contributed to the advancement of our research.

Xavier Mootoo also gratefully acknowledges the support received from the Vector Institute through the Vector Scholarship in AI and the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Canada Graduate Scholarship—Master's.

References

R Balakrishnan. The energy of a graph. *Linear Algebra and its Applications*, 387:287–295, 2004.

Lester I Binder, Angela L Guillozet-Bongaarts, Francisco Garcia-Sierra, and Robert W Berry. Tau, tangles, and alzheimer's disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1739(2-3):216–223, 2005.

Raymundo Cassani, Mar Estarellas, Rodrigo San-Martin, Francisco J Fraga, and Tiago H Falk. Systematic review on resting-state eeg for alzheimer's disease diagnosis and progression assessment. *Disease markers*, 2018, 2018.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

Fan RK Chung and Robert P Langlands. A combinatorial laplacian with vertex weights. *journal of combinatorial theory, Series A*, 75(2):316–327, 1996.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

Justin Dauwels, Francois Vialatte, and Andrzej Cichocki. Diagnosis of alzheimer's disease from eeg signals: where are we standing? *Current Alzheimer Research*, 7(6):487–505, 2010.

Ligia J Dominguez, Nicola Veronese, Laura Vernuccio, Giuseppina Catanese, Flora Inzerillo, Giuseppe Salemi, and Mario Barbagallo. Nutrition, physical activity, and other lifestyle factors in the prevention of cognitive decline and dementia. *Nutrients*, 13(11):4080, 2021.

Xiaowen Dong, Dorina Thanou, Laura Toni, Michael Bronstein, and Pascal Frossard. Graph signal processing for machine learning: A review and new perspectives. *IEEE Signal processing magazine*, 37(6):117–127, 2020.

Evelyn Fix and Joseph Lawson Hodges. Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*, 57 (3):238–247, 1989.

Vijayasree V Giridharan, Carlos E Barichello De Quevedo, and Fabricia Petronilho. Microbiota-gut-brain axis in the alzheimer's disease pathology-an overview. *Neuroscience Research*, 181: 17–21, 2022.

Jürgen Götz and Lars M Ittner. Animal models of alzheimer's disease and frontotemporal dementia. *Nature Reviews Neuroscience*, 9(7):532–544, 2008.

David K Hammond, Yaniv Gur, and Chris R Johnson. Graph diffusion distance: A difference measure for weighted graphs based on the graph laplacian exponential kernel. In *2013 IEEE global conference on signal and information processing*, pp. 419–422. IEEE, 2013.

Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pp. 278–282. IEEE, 1995.

Nesma Houmani, François Vialatte, Esteve Gallego-Jutglà, Gérard Dreyfus, Vi-Huong Nguyen-Michel, Jean Mariani, and Kiyoka Kinugawa. Diagnosis of alzheimer's disease with electroencephalography in a differential framework. *PloS one*, 13(3):e0193607, 2018.

Xin Huang, Ashish Khetan, Milan Cvitkovic, and Zohar Karnin. Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*, 2020.

Jaeseung Jeong. Eeg dynamics in patients with alzheimer's disease. *Clinical neurophysiology*, 115 (7):1490–1505, 2004.

David S Knopman, Helene Amieva, Ronald C Petersen, Gäel Chételat, David M Holtzman, Bradley T Hyman, Ralph A Nixon, and David T Jones. Alzheimer disease. *Nature reviews Disease primers*, 7(1):33, 2021.

Sylvain Lesné, Ming Teng Koh, Linda Kotilinek, Rakez Kayed, Charles G Glabe, Austin Yang, Michela Gallagher, and Karen H Ashe. A specific amyloid-β protein assembly in the brain impairs memory. *Nature*, 440(7082):352–357, 2006.

Morgan L Levy, Bruce L Miller, Jeffrey L Cummings, Lynn A Fairbanks, and Anne Craig. Alzheimer disease and frontotemporal dementias: behavioral distinctions. *Archives of neurology*, 53(7):687–690, 1996.

Rui Li, Xin Yuan, Mohsen Radfar, Peter Marendy, Wei Ni, Terence J O'Brien, and Pablo M Casillas-Espinosa. Graph signal processing, graph neural network and graph learning on biological data: a systematic review. *IEEE Reviews in Biomedical Engineering*, 2021.

M Lindau, O Almkvist, J Kushi, K Boone, SE Johansson, LO Wahlund, JL Cummings, and BL Miller. First symptoms–frontotemporal dementia versus alzheimer's disease. *Dementia and geriatric cognitive disorders*, 11(5):286–293, 2000.

Joan Lindsay, Danielle Laurin, René Verreault, Réjean Hébert, Barbara Helliwell, Gerry B Hill, and Ian McDowell. Risk factors for alzheimer's disease: a prospective analysis from the canadian study of health and aging. *American journal of epidemiology*, 156(5):445–453, 2002.

Alireza Mansouri, Aria Fallah, and Taufik Valiante. Determining surgical candidacy in temporal lobe epilepsy. *Epilepsy research and treatment*, 2012:706917, 02 2012. doi: 10.1155/2012/706917.

Priyanka Mathur and Vijay Kumar Chakka. Graph signal processing of eeg signals for detection of epilepsy. In *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 839–843. IEEE, 2020.

Hemant Kumar Meena, Ramnivas Sharma, Abhinav Tailor, Harshil Verma, and Rajveer Saini. Detection of epilepsy using graph signal processing of eeg signals with three features. In *Pattern Recognition and Data Analysis with Applications*, pp. 569–578. Springer, 2022.

Scott Menard. *Applied logistic regression analysis*. Number 106. Sage, 2002.

Andreas Miltiadous, Katerina D. Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G. Tsalikakis, Pantelis Angelidis, Markos G. Tsipouras, Evripidis Glavas, Nikolaos Giannakeas, and Alexandros T. Tzallas. "a dataset of eeg recordings from: Alzheimer's disease, frontotemporal dementia and healthy subjects", 2023.

Emma Nichols, Jaimie D Steinmetz, Stein Emil Vollset, Kai Fukutaki, Julian Chalek, Foad Abd-Allah, Amir Abdoli, Ahmed Abualhasan, Eman Abu-Gharbieh, Tayyaba Tayyaba Akram, et al. 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, 2022.

Nicholas T Olney, Salvatore Spina, and Bruce L Miller. Frontotemporal dementia. *Neurologic clinics*, 35(2):339–374, 2017.

Antonio Ortega. *Introduction to graph signal processing*. Cambridge University Press, 2022.

Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst. Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106 (5):808–828, 2018.

Nancy Ann Pachana, Kyle Brauer Boone, Bruce L Miller, Jeffrey L Cummings, and Nancy Berman. Comparison of neuropsychological functioning in alzheimer's disease and frontotemporal dementia. *Journal of the International Neuropsychological Society*, 2(6):505–510, 1996.

Himanshu Padole, S.D. Joshi, and Tapan K. Gandhi. Early detection of alzheimer's disease using graph signal processing on neuroimaging data. In *2018 2nd European Conference on Electrical Engineering and Computer Science (EECS)*, pp. 302–306, 2018. doi: 10.1109/EECS.2018.00062.

Florence Pasquier, Florence Richard, and Florence Lebert. Natural history of frontotemporal dementia: comparison with alzheimer's disease. *Dementia and geriatric cognitive disorders*, 17(4): 253–257, 2004.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

AP Porsteinsson, RS Isaacson, Sean Knox, MN Sabbagh, and I Rubino. Diagnosis of early alzheimer's disease: clinical practice in 2021. *The journal of prevention of Alzheimer's disease*, 8:371–386, 2021.

Gil D Rabinovici. Late-onset alzheimer disease. *Continuum: Lifelong Learning in Neurology*, 25 (1):14, 2019.

K Deergha Rao and Madisetti NS Swamy. *Digital signal processing: Theory and practice*. Springer, 2018.

Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, volume 4, pp. 547–562. University of California Press, 1961.

A Rihaczek. Signal energy distribution in time and frequency. *IEEE Transactions on information Theory*, 14(3):369–374, 1968.

Irina Rish et al. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pp. 41–46, 2001.

Aaron R Ritter, Gabriel C Leger, Justin B Miller, and Sarah J Banks. Neuropsychological testing in pathologically verified alzheimer's disease and frontotemporal dementia: how well do the uniform data set measures differentiate between diseases? *Alzheimer disease and associated disorders*, 31 (3):187, 2017.

Howard J Rosen, KM Hartikainen, W Jagust, JH Kramer, BR Reed, JL Cummings, K Boone, W Ellis, C Miller, and BL Miller. Utility of clinical criteria in differentiating frontotemporal lobar degeneration (ftld) from ad. *Neurology*, 58(11):1608–1615, 2002.

Aliaksei Sandryhaila and Jose MF Moura. Discrete signal processing on graphs: Frequency analysis. *IEEE Transactions on Signal Processing*, 62(12):3042–3054, 2014.

Philip Scheltens, Kaj Blennow, Monique MB Breteler, Bart De Strooper, Giovanni B Frisoni, Stephen Salloway, and Wiesje Maria Van der Flier. Alzheimer's disease. *The Lancet*, 388(10043): 505–517, 2016.

Marcos Vinícius Ferreira Silva, Cristina de Mello Gomide Loures, Luan Carlos Vieira Alves, Leonardo Cruz de Souza, Karina Braga Gomes Borges, and Maria das Graças Carvalho. Alzheimer's disease: risk factors and potentially protective measures. *Journal of biomedical science*, 26:1–11, 2019.

Shelagh JM Smith. Eeg in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(suppl 2):ii2–ii7, 2005.

Xiaofan Song, Mingyi Mao, and Xiaohua Qian. Auto-metric graph neural network based on a meta-learning strategy for the diagnosis of alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, 25(8):3141–3152, 2021. doi: 10.1109/JBHI.2021.3053568.

Bart M ter Haar Romeny. The gaussian kernel. *Front-End Vision and Multi-Scale Image Analysis: Multi-Scale Computer Vision Theory and Applications, written in Mathematics*, pp. 37–51, 2003.

Nitish V Thakor and David L Sherman. Eeg signal processing: Theory and applications. In *Neural Engineering*, pp. 259–303. Springer, 2012.

Wiesje M van der Flier and Philip Scheltens. Epidemiology and risk factors of dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(suppl 5):v2–v7, 2005.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Mário L Vicchietti, Fernando M Ramos, Luiz E Betting, and Andriana SLO Campanharo. Computational methods of eeg signals analysis for alzheimer's disease classification. *Scientific Reports*, 13(1):8184, 2023.

Devin Wahl, Samantha M Solon-Biet, Victoria C Cogger, Luigi Fontana, Stephen J Simpson, David G Le Couteur, and Rosilene V Ribeiro. Aging, lifestyle and dementia. *Neurobiology of disease*, 130:104481, 2019.

Jason D Warren, Jonathan D Rohrer, and Martin N Rossor. Frontotemporal dementia. *Bmj*, 347, 2013.

Bai Xiao, Edwin R Hancock, and Richard C Wilson. Graph characteristics from the heat kernel trace. *Pattern Recognition*, 42(11):2589–2606, 2009.

Konstantina G Yiannopoulou and Sokratis G Papageorgiou. Current and future treatments in alzheimer disease: an update. *Journal of central nervous system disease*, 12:1179573520907397, 2020.

Jessica Zande, Alida Gouw, Inger van Steenoven, Philip Scheltens, Cornelis Stam, and Afina Lemstra. Eeg characteristics of dementia with lewy bodies, alzheimer's disease and mixed pathology. *Frontiers in Aging Neuroscience*, 10:190, 07 2018. doi: 10.3389/fnagi.2018.00190.