IADIS International Journal on Computer Science and Information Systems
Vol. 21, No. 2, pp. 161-173
ISSN: 1646-3692

MODELING BRAIN CONNECTIVITY TO DISTINGUISH COGNITIVE STATES IN ALZHEIMER'S DISEASE

Dobromir Boyadzhiev[1], Olga Georgieva[1,2] and George Petkov[2]
[1]*FMI, Sofia University "St. Kliment Ohridski", Sofia 1164, Bulgaria*
[2]*GATE Institute, Sofia University "St. Kliment Ohridski", 1164 Sofia, Bulgaria*

ABSTRACT

There is currently no effective treatment for Alzheimer's disease, and medications only help to relieve symptoms. Therefore, early diagnosis is essential to control its progression. Electroencephalography (EEG) allows detection of abnormalities in brain waves associated with certain neurological disorders and offers greater ease and convenience. This paper aims to identify and develop an effective methodology for distinguishing groups of people in distinct cognitive states of Alzheimer's disease by processing their EEG data. The preprocessed data are analyzed to extract informative features based on brain functional connectivity estimation. The selected features form a data space that enables distinction among the three investigated cognitive groups: healthy controls, persons with mild cognitive impairment, and Alzheimer's patients. Group identification is based on clustering. For this purpose, different algorithms are applied, and the obtained results are rigorously compared according to their ability to recognize the known data labels. The proposed methodology could provide complementary decision support for patients' diagnosis and subsequent monitoring and, thereby, help reduce possible delays in disease development.

KEYWORDS

EEG Data Analysis, Data-Mining, Machine Learning, Clustering Analysis, Cognitive State, Alzheimer's Disease

1. INTRODUCTION

Dementia is a group of neurodegenerative diseases characterized by cognitive impairment, memory loss, and difficulty in daily activities due to the progressive degeneration of brain cells. The social and economic consequences of dementia are enormous, as it places enormous pressure on individuals, their families, society, and health systems. According to the World Health Organization (WHO), more than 55 million people are currently living with dementia, with this number projected to increase to 78 million by 2030 (WHO, 2026; Prince, 2015).

Alzheimer's disease is the most common type of dementia (Cassani et al., 2018). There is currently no effective treatment for Alzheimer's disease, and medications only help to relieve symptoms. Therefore, early diagnosis is essential to control its progression. Mild cognitive impairment represents an intermediate stage between normal aging and dementia (Michaud et al., 2017), and its diagnosis is more critical than diagnosing Alzheimer's disease.

Various approaches for distinguishing disease stages have been widely used in recent decades, such as enquiry by magnetic resonance imaging, positron emission tomography, and computed tomography. However, they are all costly and require more time for examination. Electroencephalography (EEG), an alternative approach, allows the detection of abnormalities in brain waves associated with certain neurological disorders (Wang et al., 2014; Escobar-López & Salazar-Varas, 2025). The findings of authors in (Rossini et al., 2020) suggest EEG rhythms analysis as a valid screening tool to predict AD conversion.

In particular, the EEG data obtained and used in specific machine learning (ML) algorithms could contribute to the identification of an appropriate methodology offering accurate diagnosis of patients in stages of Alzheimer's disease. ML algorithms play an increasingly important role in EEG data analysis, enabling automated extraction and interpretation of complex neurophysiological patterns that traditional visual or statistical methods often miss. Combined with EEG's noninvasive, inexpensive, and widely available nature, ML significantly increases the clinical value of electroencephalography for defining, tracking, and understanding cognitive states and their neural mechanisms (Papaliagkas, 2025). Supervised machine learning methods such as SVM (Wang et al., 2014; McBride et al., 2013; Sheng et al., 2019) and Random Forest are used as classifiers that separate different cognitive states as normal cognition, mild cognitive impairment, and Alzheimer's through optimally defined decision boundaries in the feature space. Other algorithms for distinguishing patients with Alzheimer's disease, frontotemporal dementia, and healthy controls, are: KNN, which classifies new observations based on their proximity to the closest examples; XGBoost - an advanced version of the gradient boosting, which sequentially builds models that correct the errors of the previous ones; Extra Trees is also an ensemble algorithm of decision trees, which introduces additional randomness in the selection of the dividing points, which helps reduce variation and improves the generalization of the model (Lal at al., 2024). The KNN algorithm reported the best results. In recent years, algorithms such as deep learning have also emerged, opening possibilities for automatic feature extraction directly from raw or spectrally transformed EEG data (Bashivan et al., 2015). Unlike the supervised learning approach, unsupervised machine learning methods are not frequently applied to the recognition of dementia states, yet they hold potential for addressing various issues related to grouping patients at different stages of the disease. Thus, a promising methodology employing the *k*-means clustering algorithm demonstrates the ability to group patients according to their cognitive state (Akrofi et al., 2008).

The aim of the paper is to present an effective methodology for distinguishing groups of people in different cognitive states of Alzheimer's disease by processing their EEG recordings. The proposed solution follows a data-mining step procedure by applying statistical analysis for feature selection and machine learning approach for data grouping. The presented and detailed methodology has been proved to data of individuals with varying levels of cognitive decline. The proposed methodology could provide complementary decision support for patients' diagnosis and subsequent monitoring and, thereby, help reduce possible delays in disease development.

2. METHODOLOGY FOR COHERENCE ANALYSIS OF EEG RECORDINGS

Here, we implement a step-wise data-mining methodology based on functional connectivity estimation of brain channels. To address the research objective, the methodology detects the participants' groups. The techniques used are motivated by the nature of EEG signals and the goal of detecting unknown internal dependencies to separate cognitive groups. EEG signals represent a high-dimensional, nonlinear, and highly time-varying system in which boundaries between cognitive states are not always clearly defined. Thus, algorithms that do not impose prior assumptions about class labels are appropriate for identifying cognitive groups. Such are clustering algorithms as they could find latent structures in the data. It should be underlined that the present methodology uses class labels in the feature selection step in order to determine the data space where the clusters are to be found. The labels are also used in the final assessment of the resulting clusters to validate the methodology.

The methodology consists of successive steps for data acquisition, data preprocessing, feature extraction and selection of significant features, clustering analysis for data grouping and for assessment of the results (Figure 1). Each step will be discussed in detail below.

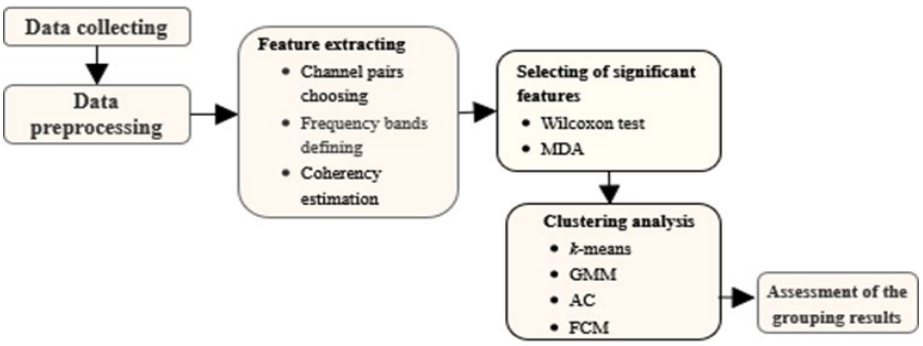


Figure 1. Data-mining methodology of cognitive state assessment by brain functional connectivity

2.1 Data Collecting and Preprocessing

The EEG recordings transcript of the electrical activity of the brain collected by electrodes of a special equipment placed on the scalp. The raw recordings should be preprocessed to ensure the subsequent data analysis. The typical tasks in this process are recalculating the signals relative to the reference signals and detection and removal of artifacts such as muscle activity, body movements, electrical interference. A bandpass filter is applied to the data to pass only signal components in a specific frequency range from 0.5 Hz to 45/50 Hz. This removes very low-frequency artifacts (< 0.5 Hz), caused, for example, by signal delay, changes in electrode movement, or potential artifacts from electrode contact with the skin. It also removes high-frequency noise (> 45/50 Hz).

2.2 Feature Extraction

After the EEG signals have been collected and preprocessed, the significant features for group discrimination must be selected. The concept of distinguishing the participants' groups by assessing the functional connectivity of the brain channels requires defining the important for this estimation channel pairs. The channel pairs are determined by a neuroanatomically reasoned approach, which builds on the conceptual and methodologically developed framework for existing connections of the brain signals (Akrofi et al., 2008; Leuchter et al., 1992). This originally introduced channel pairing is for the Alzheimer and Mild Cognitive Impairment discrimination realized for 16 channels assembly. We adopt the approach, but due to the 19-channel setup in our case, we also consider the recordings from the central EEG electrodes Fz, Cz, and Pz. Thus, the two central pairs Fz- Cz and Cz-Pz have been also investigated in the feature selection analysis. Following the representation of the anatomical interpretability of the functional connectivity accepted for 19-channel Standard 10-20 System, we determine channel pairs shown in Figure 2 and their respective pairing given in Table 1.

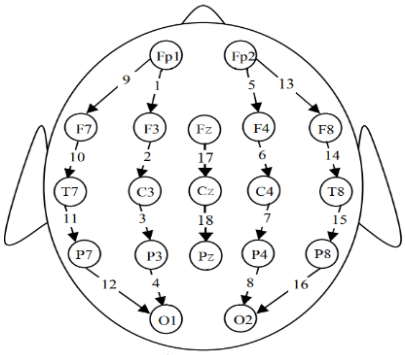


Figure 2. Connections in bipolar assembly with 19 EEG electrodes

Functional connectivity assessments intend to describe the statistical dependence and synchronization between EEG pairing signals recorded from the electrodes. It captures interactions between spatially distributed neuronal populations, measures their degree of synchronization, and assesses how they communicate. One of the most commonly used measures for such functional dependence is the coherence. It measures the linear relationship between two signals in a specific frequency domain. This search is based on the knowledge that Alzheimer patients demonstrate lower coherence compared to MCI and HC people (Leuchter et al., 1992; Jeong, 2004).

Table 1. Channel pairing for bipolar mounting

Channel Pair Electrodes	Pair Number	Channel Pair Electrodes	Pair Number	Channel Pairs	Received Pair Number
P3-O1	4	F3-C3	2	P3-O1 - F3-C3	42
P3-O1	4	Fp1-F3	1	P3-O1 - Fp1-F3	41
P3-O1	4	Fp1-F7	9	P3-O1 - Fp1-F7	49
P4-O2	8	Fp2-F4	5	P4-O2 - Fp2-F4	85
P4-O2	8	Fp2-F8	13	P4-O2 - Fp2-F8	813
P4-O2	8	F4-C4	6	P4-O2 - F4-C4	86
Fp1-F3	1	C3-P3	3	Fp1-F3 - C3-P3	13
Fp1-F3	1	T5-O1	12	Fp1-F3 - T5-O1	112
Fp2-F4	5	C4-P4	7	Fp2-F4 - C4-P4	57
Fp2-F4	5	T6-O2	16	Fp2-F4 - T6-O2	516
P3-O1	4	C3-P3	3	P3-O1 - C3-P3	43
P3-O1	4	T3-T5	11	P3-O1 - T3-T5	411
P4-O2	8	C4-P4	7	P4-O2 - C4-P4	87
P4-O2	8	T4-T6	15	P4-O2 - T4-T6	815
Fp1-F3	1	F3-C3	2	Fp1-F3 - F3-C3	12
Fp1-F3	1	Fp1-F7	9	Fp1-F3 - Fp1-F7	19
Fp2-F4	5	F4-C4	6	Fp2-F4 - F4-C4	56
Fp2-F4	5	Fp2-F8	13	Fp2-F4 - Fp2-F8	513
Fz-Cz	17	Cz-Pz	18	Fz-Cz - Cz-Pz	1718

The chosen bipolar pairs (Table 1) were used to calculate the Coherence $C_{xy}(f)$ between signals of two EEG channels as a normalized value that indicates the degree of dependence between the two signals at a certain frequency:

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)}, \quad (1)$$

where $S_{xy}(f)$ is the cross-spectral density of the signals $x(t)$ and $y(t)$, t denotes time, and $S_{xx}(f)$ and $S_{yy}(f)$ are the auto-spectral densities of the signals $x(t)$ and $y(t)$, respectively, at the frequency f estimated by the Welch method. The value of $C_{xy}(f)$ close to 1 indicates a high degree of linear synchronization between the two signals at frequency f, whereas a value close to 0 indicates a lack of linear dependence. The Coherence is averaged over a predefined frequency band. This gives an average coherence over a given frequency range:

$$\bar{C}_{xy} = \frac{1}{N_f} \sum_{f \in \text{band}} C_{xy}(f), \quad (2)$$

where N_f is the number of frequency points in the predefined frequency band.

As an additional method to estimate the functional connectivity, Imaginary coherence (IC) is also considered in the present study. It is a modification of the classical coherence that is introduced to reduce the influence of bulk conductivity. While coherence uses the entire modulus of the cross-spectral density, imaginary coherence uses only the imaginary part of the normalized cross-spectrum:

$$ICoh_{xy}(f) = \text{Im}\left\{\frac{S_{xy}(f)}{\sqrt{S_{xx}(f)S_{yy}(f)}}\right\}. \quad (3)$$

IC isolates only the component that reflects a phase shift other than 0 or π (180°) between the signals. Further, it is also averaged over a predefined frequency band.

The selection of the frequency ranges for the average Coherence and IC calculations is based on both the physical and mathematical properties of spectral analysis and the known neurophysiological significance of brain rhythms. The standard range for separating in five frequency bands, namely Delta (δ) 0.5-4 Hz, Theta (θ) 4-8 Hz, Alpha (α) 8-13 Hz, Beta (β) 13-30 Hz and Gamma (γ) > 30 Hz, was examined. However, additional investigation of coherence at the boundary between the Alpha and Beta frequency bands is reasonable, given the known differences in brain activity in these bands in cognitive diseases (Cassani et al., 2018; Jeong, 2004). Their finer separation allows investigation of subranges in which groups' differences are more significant.

2.3 Selecting Significant Features

The selection of statistically significant features is based on the Wilcoxon test applied to the averaged coherence values in predefined frequency bands for each pair of channels. The choice of the Wilcoxon rank-sum test was justified by the fact that it does not require normality of the distribution as it is not proved in our case. It is also robust to outliers and works well for small sets of data. The statistical test is applied to a pairwise comparison of the groups, namely AD vs. CH, AD vs. MCI, MCI vs. CH. Only the channel pairs for which the test threshold value is kept are selected as significant features, and their values are included for the subsequent analysis. To ensure greater confidence in the significance of the selected pairs, a lower threshold of $p \leq 0.03$ than the default one of 0.05 has been adopted.

The Wilcoxon test can identify a varying number of significant features for a particular data set under investigation. This implies that the obtained data space for identifying cognitive groups can be potentially high-dimensional, especially relative to the low size of the dataset. This leads to a distortion of the cluster analysis results. To avoid this problem, Multiple Discriminant Analysis (MDA) is applied on the selected statistically significant EEG features. It linearly transforms the feature space and yields into a two-dimensional data space with increased informativity grouping by boosting the intra-class separability in comparison to the unreduced data space.

2.4 Clustering Analysis for Grouping Assessment

In the final step of the methodology, several different unsupervised machine learning algorithms are applied to categorize the individuals into distinct groups. Clustering to identify groups based on cognitive status across stages of Alzheimer's disease is performed on the EEG data processed in the preceding steps. This enables the discovery of the data structure without relying on

existing data labels. Due to the latent nature of EEG data, selecting a specific clustering algorithm in advance is challenging. A comparative analysis involving the application of algorithms that differ in type and clustering approach aims to achieve more accurate grouping results. The use of various algorithms is motivated by the fact that each employs a unique clustering approach and methodology, making it suitable for specific data types and distributions.

Four clustering algorithms are used to differentiate groups by cognitive decline: *k*-means, Gaussian Mixture Model (GMM), Agglomerative Clustering (AC), and Fuzzy-C-Means (FCM). *k*-means clustering finds compact, spherical groups around the cluster prototype, whereas GMM separates the data according to the normal distribution within the groups. FCM finds spherical but overlapping clusters, meaning that each data point belongs to every cluster with a different membership degree. Agglomerative clustering finds clusters by merging the closest data points.

Data fitting is avoided by 10-fold cross-validation and presenting average values of each of the metrics used after the last iteration. It offers a good trade-off between bias and variance and suits small to medium-sized data sets. Cross-validation is used solely to improve stability and the generalization ability of clustering algorithms, not to train them.

For a reliable assessment of the effectiveness of the algorithms, several indices were used to evaluate the clustering results, both for the overall data structuring and for particular groups:

1) Adjusted Rand Index (ARI) is a measure of clustering quality that measures the statistical degree of agreement between two different groupings of the same set of objects.

2) Normalized Mutual Information (NMI) is an information-theoretic metric for assessing the similarity between two groupings – clusters and grouping by the participants' diagnosis.

3) Classification accuracy (CR) estimates the degree to which the clustering results correspond to the existing groups. For each cluster, the group with the greatest match is determined. CR measures the proportion, a value in the interval from 0 (for no match) to 1 (for full match).

3. METHODOLOGY APPLICATION

3.1 Data Used

The EEG data analyzed in this study were originally collected and introduced in (Stothart et al., 2016). Brain electrical activity was recorded in a resting state with eyes open. While the original dataset was acquired using a high-density 64-channel EEG configuration, the current analysis utilizes a subset of 32 channels. These 32 channels correspond to the standardized electrode placement system recommended by the American Electroencephalographic Society (Malmivuo & Plonsey, 1995).

The decision to reduce the dataset from 64 to 32 channels was guided by spatial and signal characteristics. The 64 electrodes are effectively divided into two interleaved sets of 32 channels each. Because scalp electrodes in such a high-density montage are positioned in close spatial proximity, the two subsets recorded virtually identical physiological signals, with differences confined to microscopic measurement noise. Consequently, analyzing the primary 32-channel subset provides a mathematically robust and representative description of the underlying cortical activity without data redundancy.

For each participant, continuous long-term EEG recordings were performed. To ensure data quality and uniformity, a trained clinician visually inspected the recordings and selected a single, continuous 20-second artifact-free segment per subject. Each chosen segment strictly reflects a stable resting state condition with eyes open, establishing a reliable baseline for cross-group comparisons.

Participants were not undergoing any drug therapy that could alter EEG dynamics at the time of testing. The cognitive status of the participants was characterized by the Mini-Mental State Examination (MMSE). Qualified specialists conducted a thorough clinical and medical examination for each participant to confirm their diagnosis. The cohort was then divided, according to the given diagnoses, into three distinct groups: healthy elderly controls (HC), individuals with mild cognitive impairment (MCI), and patients diagnosed with Alzheimer's disease (AD). Table 2 summarizes the age, MMSE, and clinical profiles of these groups.

Table 2. Description of the three cognitive groups

Group	HC	MCI	AD
Average age	76 (± 7)	78 (± 7)	79 (± 9)
Average MMSE score	29 (± 1)	26 (± 2)	23 (± 3)
Number of participants	26	25	21

3.2 Grouping According the Cognitive State

The average coherence (2) is calculated for each defined channel pairs (Table 1) distinctly for the three defined diapasons of frequency bands. Thus, three data spaces are organized, in which every person is represented with coherence values for the accepted frequency bands of the respective space.

Further, the three data spaces are explored to recognize the cognitive groups. For this aim the significant features are selected by Wilcoxon test for each data space. Following this the standard division into the five basic frequency bands is investigated and selected significant channel pairs are presented at Table 3. By dividing the Beta band into three parts (Table 4) and the Alpha band into two parts (Table 5), we are looking for more informative features to achieve better group separation. It is seen that the three spaces identify the same signal connectivity paths. Figure 3 shows them as bold colored arrows. It is concluded that the most important brain path identified according to coherence as a functional connectivity marker is path Fp2-F4-C4-P4-O2. In addition, the connections of Fp1, O1, and O2 should also be underlined as significant for group distinguishing.

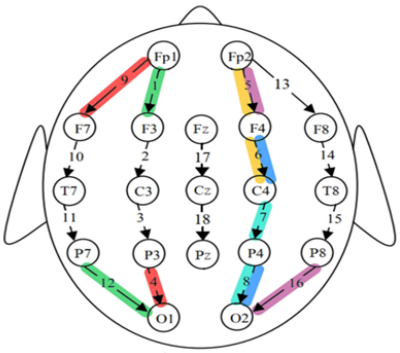


Figure 3. The established signal connectivity pathways under bipolar montage proved significant for distinguishing between the three cognitive groups

Table 3. Significant features found for the five frequency bands

Band (Hz) / Groups' pair	0.5 – 4 Hz	4 – 8 Hz	8 – 13 Hz	13 – 30 Hz	> 30 Hz
AD vs. HC	87		87	86, 516	86, 112
AD vs. MCI	87	112			516
MCI vs. HC			56	86	

Table 4. Significant features found for dividing the Beta frequency into three bands

Band (Hz) / Groups' pair	0.5–4 Hz	4 – 8 Hz	8 – 12 Hz	12–16 Hz	16 – 20 Hz	20–24 Hz
AD vs. HC	87		87	87	86	
AD vs. MCI	87	112		516		
MCI vs. HC			56			86

Table 5. Significant features found for dividing the Alpha frequency into two bands

Band (Hz) / Groups' pair	0.5–4 Hz	4 – 6 Hz	6 – 9 Hz	9 – 13 Hz	13 – 30 Hz	>30 Hz
AD vs. HC	87	49, 56		87	86, 516	112, 86
AD vs. MCI	87					516
MCI vs. HC		49	56	516, 56	86	

After applying the MDA algorithm, three reduced data space are formed of each of the considered data spaces. The Elbow method is applied to each, in order to identify the proper number of clusters, as this is an initialization parameter for the clustering algorithms *k*-means, GMM, and FCM. The results for the inertia reduction of Within-cluster sum of squares (WCSS) in respect to the number of clusters *k* show that best data space division is in 3 clusters for the three data spaces (Figure 4).

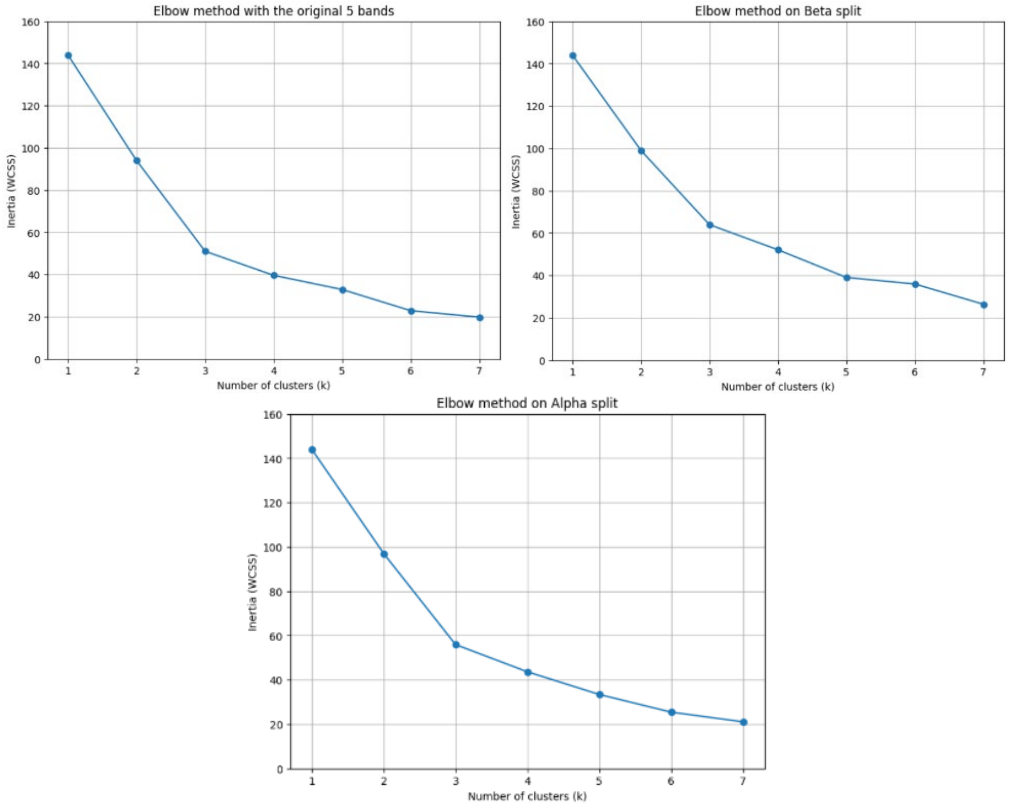


Figure 4. Results of Elbow method of the three formed data spaces

3.3 Grouping Results and Analysis

The clustering algorithm results by coherence are given in Table 6, and for Imaginary coherency in Table 7, where the best index values for each data space are given in bold. In the case of Coherence used for channel connectivity estimation, the best result, CR = 0.6554 for 5 basic frequency bands, is obtained by k-means clustering, whereas for division of the Beta frequency band, the best value (CR = 0.5964) is achieved for Agglomerative clustering with a single link as the connection criterion and cosine distance as the distance measure. Another important observation concerns the Alpha band division, where GMM is best presented among all methods, followed by FCM algorithm. The grouping accuracy considered for the MCI group achieves high accuracy (CR-MCI = 0.80) for FCM clustering. The observation is that the space with Alpha band division outperforms the other two.

Table 6. Clustering results by features determined by Coherency

	Five band frequencies				Beta in three frequency bands				Alpha in two frequency bands			
Index	***k*-means**	**GMM**	**AC**	**FCM**	***k*-means**	**GMM**	**AC**	**FCM**	***k*-means**	**GMM**	**AC**	**FCM**
ARI	**0.1464**	0.0089	0.1248	0.0820	-0.1128	0.0024	-0.007	-0.029	0.2225	**0.2607**	0.1849	0.2104
NMI	**0.4502**	0.3372	0.4400	0.4048	0.2895	0.2815	**0.3768**	0.3391	0.5178	**0.5290**	0.5069	0.5050
CR	**0.6554**	0.5875	0.6411	0.6125	0.5429	0.5250	**0.5964**	0.5446	0.6696	**0.6964**	0.6679	0.6821
CR-AD	**0.68**	0.48	0.63	0.68	0.53	0.50	0.53	0.53	0.63	0.48	0.58	0.63
CR-HC	0.78	**0.83**	0.81	0.75	0.63	0.62	0.66	**0.67**	0.71	**0.85**	0.72	0.62
CR-MCI	0.52	0.43	0.47	0.42	0.51	0.38	**0.63**	0.45	0.68	0.72	0.70	**0.80**

The results of Imaginary coherency for the functional connectivity estimation (Table 7) are not better than those obtained by Coherence estimation (Table 6). For the five frequency bands, the best value is obtained by the AC method (CR = 0.6089). In the case of Beta band division, *k*-means performs best for NMI, CR, and CR-AD, but particularly the index of CR-HC=0.80 is maximized for GMM clustering. In the case of Alpha band division, FCM clustering is best according to all indices. Recognition of healthy people by GMM is best among all data spaces and all clustering algorithms, CR-HC=0.83.

Table 7. Clustering results by data space with features determined by Imaginary coherency

	Five band frequencies				Beta in three frequency bands				Alpha in two frequency bands			
Index	***k*-means**	**GMM**	**AC**	**FCM**	***k*-means**	**GMM**	**AC**	**FCM**	***k*-means**	**GMM**	**AC**	**FCM**
ARI	-0.072	-0.0301	**0.1015**	-0.0038	0.0845	-0.006	0.0812	0.0344	0.0199	0.0407	0.0312	**0.0394**
NMI	0.2439	0.2050	**0.4458**	0.3643	**0.4246**	0.2906	0.4168	0.3875	0.3882	0.2963	0.4207	**0.4270**
CR	0.5411	0.4875	**0.6089**	0.5821	**0.6375**	0.5696	0.6250	0.5946	0.5732	0.5714	0.5839	**0.5857**
CR-AD	0.51	**0.61**	0.51	0.51	**0.61**	0.51	0.58	0.56	**0.56**	0.31	0.53	0.51
CR-HC	0.68	0.58	**0.78**	0.63	0.63	**0.80**	0.60	0.62	0.75	**0.83**	0.75	0.75
CR-MC	0.38	0.20	0.48	**0.58**	0.65	0.31	**0.70**	0.55	0.40	0.48	0.50	**0.50**

Pairwise participants grouping for the Coherency feature is interesting to follow, as it helps observe the distinguishing ability of two particular cognitive states. Thus, AD group is best distinguished from healthy people by Agglomerative clustering in the data space formed by alpha-band division (Table 8). Results of the AD group versus MCI group (Table 9) show that the best distinction is accomplished by the GMM method for beta-band division. AD group is well identified by all methods with accuracy close to or over 0.80, whereas the MCI group is less well recognized. It can be seen that the ARI index is very low, even negative for some groupings. This result could be explained partly by the statistical character of the index, which, in the case of a small dataset, is not able to provide a reliable metric.

Table 8. Results of the AD vs. HC pair of groups at the different frequency ranges

	Five band frequencies				Beta in three frequency bands				Alpha in two frequency bands			
Index	***k*-means**	**GMM**	**AC**	**FCM**	***k*-means**	**GMM**	**AC**	**FCM**	***k*-means**	**GMM**	**AC**	**FCM**
ARI	-0.031	0.0173	0.112	-0.0308	-0.1019	-0.0365	-0.100	-0.0904	0.2526	0.1276	**0.3679**	0.2526
NMI	0.2774	0.3134	0.384	0.2774	0.2111	0.2253	0.2095	0.2277	0.4654	0.3674	**0.5452**	0.4654
CR	0.7050	0.7250	0.725	0.7050	0.6550	0.6800	0.6600	0.6600	0.7750	0.7350	**0.8150**	0.7750
CR-AD	0.87	0.87	0.78	0.86	0.67	0.60	0.66	0.68	0.78	0.76	0.78	0.78
CR-HC	0.58	0.61	0.71	0.58	0.65	0.73	0.66	0.67	0.80	0.73	0.86	0.80

Table 9. Results of the AD vs. MCI pair of groups at the different frequency ranges

	Five band frequencies				Beta in three frequency bands				Alpha in two frequency bands			
Index	***k*-means**	**GMM**	**AC**	**FCM**	***k*-means**	**GMM**	**AC**	**FCM**	***k*-means**	**GMM**	**AC**	**FCM**
ARI	-0.0321	0.0026	0.068	-0.0321	-0.0654	0.083	0.035	-0.0654	-0.0654	-0.0404	-0.006	-0.0558
NMI	0.1796	0.1956	0.28	0.2139	0.2572	**0.359**	0.323	0.2572	0.1884	0.1864	0.2753	0.2066
CR	0.6500	0.6500	0.700	0.6750	0.6950	**0.740**	0.720	0.6950	0.6450	0.6450	0.6950	0.6450
CR-AD	0.73	**0.90**	0.81	0.77	0.77	0.82	0.76	0.77	0.87	0.73	0.83	0.83
CR-MCI	0.53	0.40	0.58	0.58	0.65	0.68	0.70	0.65	0.45	0.57	0.60	0.50

Other importance for distinguishing MCI versus HC (Table 10) is best performed with agglomerative clustering of the data space with alpha-band division. Close to these results are the two objective function clustering algorithms.

Table 10. Results of the MCI vs. HC pair of groups at the different frequency ranges

	Five band frequencies				Beta in three frequency bands				Alpha in two frequency bands			
Index	***k*-means**	**GMM**	**AC**	**FCM**	***k*-means**	**GMM**	**AC**	**CM**	***k*-means**	**GMM**	**AC**	**FCM**
ARI	0.0389	0.0904	0.1558	0.0571	-0.047	-0.033	0.0468	-0.047	0.3015	0.2598	0.3495	0.3015
NMI	0.2489	0.2185	0.3256	0.2890	0.1641	0.0805	0.2925	0.1641	0.4719	0.4307	**0.5079**	0.4719
CR	0.6867	0.6700	0.7267	0.7067	0.6300	0.6100	0.6867	0.6300	0.7900	0.7700	**0.8100**	0.7900
CR-HC	0.68	0.58	0.77	0.62	0.62	0.48	0.60	0.62	0.85	0.80	0.88	0.85
CR-MCI	0.68	0.70	0.65	0.80	0.65	0.65	0.82	0.65	0.73	0.73	0.73	0.73

Summarized Coherency as a biomarker provides better division than Imaginary Coherency. The data space with division of the Alpha band improves the grouping abilities. Different clustering algorithms are well presented, but mostly the preference is given to the AC method, and next are FCM and GMM. It is seen that spaces tend to find groups with no strong boundaries (FCM, GMM) or without a well-determined structure (AC).

4. CONCLUSION

This study presents a methodology for distinguishing groups in different cognitive states of Alzheimer's disease by processing EEG data. Significant features of coherence for estimating functional connectivity in different frequency bands were extracted. Four different clustering algorithms were investigated to find three cognitive groups.

Conclusions about coherence as a disease biomarker, the informativity of the different frequency bands as well as the effectiveness of the applied clustering algorithms are summarized. Using alpha subdivision bands is preferred by the fact that it improves the estimates of the results in most of the experiments conducted. GMM, FCM and AC achieved good results in separating the data into three groups, and AC also performed well in the pairwise clustering experiments. The groups found in the explored data spaces present no well distinguished boundaries and structure.

The methodology is still in the exploratory stage and requires validation on larger and independent datasets. However, its achievements confirms that it could serve as complementary decision support for AD diagnosis, patient monitoring, and care planning.

ACKNOWLEDGEMENT

This research is part of the GATE project funded by the Horizon 2020 WIDESPREAD-2018-2020 TEAMING Phase 2 programme under grant agreement no. 857155, the programme "Research, Innovation and Digitalization for Smart Transformation" 2021-2027 (PRIDST) under grant agreement no. BG16RFPR002-1.014-0010-C01 and funded by the Science Fund of Sofia University under project no. 80-10-191/2026.

REFERENCES

Akrofi, K. et al. (2008). Clustering and modeling of EEG coherence features of Alzheimer's and mild cognitive impairment patients.*30th Annual Int. Conference of the IEEE Engineering in Medicine and Biology Society*, v.2008, pp.1092–1095.

Bashivan, P. et al. (2015). Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks. *Cornell University*, *1*, 1-15.

Cassani, R. et al. (2018). Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment, *Disease Markers, 2018,* 5174815.

Escobar-López, M. & Salazar-Varas, R. (2025). Analysis of EEG coherence according to the onset of Alzheimer's disease. *Neuroscience, 588*, 40-48.

Jeong, J. (2004). EEG Dynamics in Patients with Alzheimer's Disease. *Clinical Neurophysiology, 115*(7), 1490-1505.

Lal, U. et al. (2024). A Comparative Study on Feature Extraction Techniques for the Discrimination of Frontotemporal Dementia and Alzheimer's Disease with Electroencephalography in Resting-State Adults. *MDPI Brain Science*, *14*(4), 335.

Leuchter, A. F. et al. (1992). Changes in brain functional connectivity in Alzheimer-type and multi-infarct dementia. *Brain, 115*(5), 1543–1561.

Malmivuo, J. & Plonsey R. (1995). *Electroencephalography. Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetis Fields* (Chapter 13, pp. 247-264).

McBride, J. et al. (2013). Resting EEG discrimination of early stage Alzheimer's disease from normal aging using inter-channel coherence network graphs. *Ann Biomed Eng, 41*(6), 1233-1242.

Michaud, T. L. et al. (2017). The risk of incident mild cognitive impairment and progression to dementia considering mild cognitive impairment subtypes, *Demen. Geriatr. Cognit. Disorders Extra*, *7*(1), 15–29.

Papaliagkas, V. (2025). The Role of Quantitative EEG in the Diagnosis of Alzheimer's Disease. *Diagnostics*, *15*, 1965.

Prince, M. et al. (2015). World Alzheimer Report 2015, The Global Impact of Dementia: an Analysis of Prevalence, Incidence, Cost and Trends. *Alzheimer's Disease International.*

Rossini, P. M. et al. (2020), Early diagnosis of Alzheimer's disease: the role of biomarkers including advanced EEG signal analysis. Report from the IFCN-sponsored panel of experts. *Clinical Neurophysiology, 131*(6), 1287-1310.

Sheng, J. et al. (2019), A novel joint HCPMMP method for automatically classifying Alzheimer's and different stage MCI patients. *Behavioural Brain Research, 365*, 210-221.

Stothart, G. et al. (2016). Graph-theoretical measures provide translational markers of large-scale brain network disruption in human dementia patients and animal models of dementia. *International Journal of Psychophysiology*, *108*, 71.

Wang, R. et al. (2014). Power spectral density and coherence analysis of Alzheimer's EEG. *Cognitive Neurodynamics, 9*(3), 291-304.

World Health Organization [WHO] (2026). *Dementia. 2026*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/dementia