BioData Mining

RESEARCH

Open Access

Multivariate longitudinal clustering reveals neuropsychological factors as dementia predictors in an Alzheimer's disease progression study

Patrizia Ribino[1*], Claudia Di Napoli[2], Giovanni Paragliola[2], Davide Chicco[3,4,5] and Francesca Gasparini[3,4]

*Correspondence:
patrizia.ribino@icar.cnr.it

[1] Consiglio Nazionale delle Ricerche (CNR), Istituto di Calcolo e Reti ad Alte Prestazioni, Palermo, Italy
[2] Consiglio Nazionale delle Ricerche (CNR), Istituto di Calcolo e Reti ad Alte Prestazioni, Naples, Italy
[3] Dipartimento di Informatica Sistemistica e Comunicazione, Università di Milano-Bicocca, Milan, Italy
[4] Neuromi, Milan Center for Neuroscience, Università di Milano-Bicocca, Milan, Italy
[5] Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

Abstract

Dementia due to Alzheimer's disease (AD) is a multifaceted neurodegenerative disorder characterized by various cognitive and behavioral decline factors. In this work, we propose an extension of the traditional *k*-means clustering for multivariate time series data to cluster joint trajectories of different features describing progression over time. The algorithm we propose here enables the joint analysis of various longitudinal features to explore co-occurring trajectory factors among markers indicative of cognitive decline in individuals participating in an AD progression study. By examining how multiple variables co-vary and evolve together, we identify distinct subgroups within the cohort based on their longitudinal trajectories. Our clustering method enhances the understanding of individual development across multiple dimensions and provides deeper medical insights into the trajectories of cognitive decline. In addition, the proposed algorithm is also able to make a selection of the most significant features in separating clusters by considering trajectories over time. This process, together with a preliminary pre-processing on the OASIS-3 dataset, reveals an important role of some neuropsychological factors. In particular, the proposed method has identified a significant profile compatible with a syndrome known as Mild Behavioral Impairment (MBI), displaying behavioral manifestations of individuals that may precede the cognitive symptoms typically observed in AD patients. The findings underscore the importance of considering multiple longitudinal features in clinical modeling, ultimately supporting more effective and individualized patient management strategies.

Keywords: Alzheimer's disease, Dementia, Neuropsychological symptoms, Clustering, Multivariate longitudinal study

Background

Alzheimer's disease (AD) is the most common neurological disorder causing dementia in the elderly population aged 65 years and older, affecting an estimated 6.5 million people worldwide [1]. Alzheimer's disease typically begins with mild memory problems that gradually worsen, leading to loss of brain function. In 2011, the National Institute of Ageing and Alzheimer's Association presented revised guidelines for diagnosing AD [2], which identifies three AD stages: (1) preclinical AD, in which measurable changes in biological and pathological markers occur but no outward changes are observed; (2) Mild Cognitive Impairment (MCI), during which slight memory and cognitive complaints are observed and measured; and (3) dementia due to AD, when the patient cannot perform daily life activities due to memory and cognition problems.

Since the number of people with dementia worldwide is expected to increase from 57.4 million in 2019 to nearly 153 million in 2050 [3], there is an urgent need to improve the early detection of Alzheimer's disease to optimize interventions in preclinical phases and potentially improve the success rates of trials of disease-modifying therapies. The enormous impact of Alzheimer's disease has stimulated the collection of huge datasets over the past two decades to facilitate the sharing and advancement of knowledge. These datasets provide not only a significant amount of data but also protocols that enable uniform data collection and comparison. Most of these are longitudinal data collections, where participant's outcomes, biological markers, and clinical markers are gathered at multiple follow-up assessments. This type of study typically results in numerous measurements for each subject. Longitudinal data collection is particularly important for developing predictive models and identifying the correlation among different variables that may be related to disease progression. Identifying distinct developmental factors using longitudinal data is a frequent research objective to gain a comprehensive understanding of patients' disease profiles. This method is notably beneficial for investigating heterogeneous diseases, such as Alzheimer's, which exhibits varied and complex characteristics. A great contribution to data collection comes from the Open Access Series of Imaging Studies (OASIS) project [4]. This project provides open access to four significant databases: OASIS-1 described in work by Marcus et al. [5], OASIS-2 described in [6], OASIS-3, outlined by LaMontagne and colleagues [7], and the OASIS-4 clinical cohort [8].

The longitudinal nature of repeated measures necessitates the use of specialized statistical techniques to account for the correlation within individuals, ensuring the validity of the analysis and inference of the data. The identification of clinically relevant subgroups through the analysis of longitudinal data is crucial for facilitating targeted interventions, improved disease management, and the efficient allocation of healthcare resources for both patients and their caregivers.

In the past, biomedical researchers and medical doctors utilized several methods based on traditional statistics to analyze longitudinal trajectories of patients with mental health diseases [9–11]. The recent development of artificial intelligence (AI) and machine learning (ML) models, both traditional or based on deep learning, has given rise to an enormous amount of research in the field of AD, in particular analyzing longitudinal data, evidenced by the hundreds of review or survey papers on these topics in 2024 alone. Some of these works are mainly focused on the application of deep learning

for Alzheimer's disease prediction [12] or on challenges, research gaps, and future directions in AD diagnosis [13]. Others provide useful references to understand how to adopt machine learning in dementia research and clinical practice [14, 41] or to classify different stages of the disease progression [15]. Other works highlight the growing role of AI for the prediction of early stages and diagnosis of AD [16–18], and the potential of explainable AI [19, 20], to name a few among many. When gold standard and ground truth labels are unavailable, unsupervised machine learning methods, including clustering algorithms, can be employed to detect significant groups of patients with clinical relevance. Several longitudinal methods for clustering exists: KML [21, 22], clustra [23], and traj [24], for example, are based on *k*-means, while clusterMLD is based on hierarchical clustering [25]. Trajectories can also be analyzed through mixed models, which are particular statistical models incorporating parameters representing both fixed and random effects of a data trend [26–28]. A comparison of longitudinal clustering methods in the case of data with slowly changing trends has been presented in [29]. The authors studied the performance of five methods using Monte Carlo simulations on synthetic data sets, with the aim of understanding under which conditions one method is preferable to another. They considered KML [21], growth mixture modeling (GMM) [30], a growth curve K-means model (GCKM) [31], a group-based trajectory model (GBTM) [32] and a mixed time-varying effect model MixTVEM [33]. However, a deeper description of these methods goes beyond the scope of this article.

In particular, longitudinal clustering methods have been employed in a number of studies on mental health. Liesbeth Bruckers and colleagues [34], for example, introduced a clustering method that can be applied to trajectories of patients, considering multiple variables, and showed its application to real electroencephalography (EEG) data. Niek den Teuling et al. [29], instead, used simulated datasets to demonstrate the effectiveness of their proposed longitudinal method based on growth mixture modeling. Shih-Tsung Huang and coauthors [35], on the other hand, used multi-visit longitudinal data of patients with mental health to develop a temporal disease network. They used a combination of traditional static machine learning models, such as support vector machines. This solution stands for its originality among the studies listed here. Regarding more traditional approaches, instead, Carolyn Zhu and her collaborators [36] applied the KML method [21] to data of patients with Alzheimer's to group them into significant clinical groups. Among them, Bayesian cluster analysis offers substantial benefits over algorithmic approaches by providing not only point estimates but also uncertainty in the clustering structure and patterns within each cluster [37–40].

Finally, many works differ from the point of view of the features considered for the analysis of mental health patients. To cite a few, the meta-analysis of Modupe Odusami and colleagues [15] listed some features more relevant than others useful to distinguish mental health patients from healthy controls: grey matter volume in the medial temporal lobe and the entorhinal cortex, cortical thickness in the entorhinal cortex and the inferior temporal gyrus, grey matter volume in the hippocampus, amygdala, and temporal lobe, and cortical thickness in the medial temporal lobe. The studies of Giulia Lorenzon and colleagues [38] and Konstantinos Poulakis et al. [40] indicated that some variables had a major role: age, education, sex, American

National Adult Reading Test (ANART) test results, carrier of the APOE gene, mini-mental state examination (MMSE), geriatric depression scale (GDS).

To summarize, most of the studies found in the literature were supervised and had the goal of predicting dementia diagnosis [12–14, 17–20, 41], while a few of them were designed to predict different stages of AD [15, 16]. Regarding longitudinal clustering, four studies were aimed at identifying clinically relevant clusters of patients with dementia [35, 39, 40] by mainly looking at the longitudinal changes of brain atrophy, while another one was designed to detect clusters of patients with dementia for medical expenses [36] and one to identify clusters of patients with different cerebrovascular and cognitive profiles [38].

In this study, we propose a multivariate clustering approach for multiple longitudinal features to investigate the existence of different marker trajectories as specific indicators of cognitive decline over time in a cohort of individuals participating in an Alzheimer progression study. Joint trajectories are used to understand how multiple variables co-vary or evolve about each other. Clustering these joint trajectories allows the identification of distinct subgroups within a population based on their longitudinal profiles, providing a more detailed understanding of how individuals develop over time across multiple dimensions. Unlike approaches that examine anatomic brain changes, this study's primary objective is to identify feature patterns that are not directly linked to the effects of Alzheimer's disease on brain structure. By doing so, we aim to explore less invasive and more cost-effective methods for the early identification of risk factors associated with the progression of Alzheimer's disease.

Hence, we developed a multivariate longitudinal *k*-means, which is based on the *k*-means technique [42] specifically designed to cluster joint trajectories, that also integrates a feature selection strategy. We applied the proposed approach to a public dataset derived from Electronic Mental Health Records (EMHRs) called Open Access Series of Imaging Studies-3 (OASIS-3) [7], which contains data on patients suffering from dementia. EMHRs are a specialized Electronic Health Records (EHR) subset designed to digitally manage and store patients' mental health information. EMHRs contain detailed records of a patient's mental health history, diagnoses, treatments, medication records, and progress over time [43].

Among all features in the OASIS-3 dataset, the proposed approach detected different trajectories of neuropsychiatric symptoms (NPS) related to cognitive functioning. Thus, it highlights that NPS symptoms are prevalent in patients with dementia and evolve differently according to the dementia stage.

As a noteworthy result, the proposed approach has identified an interesting cluster of individuals with slight behavioral impairments that may occur before the cognitive decline of AD patients emerges, i.e., before MCI can be diagnosed. According to recent literature, multiple longitudinal and cross-sectional studies have supported an association between NPS and AD progression, identifying a validated syndrome, Mild behavioral impairment (MBI), as an early manifestation of dementia [44]. MBI is characterized by the emergence of neuropsychiatric symptoms in elderly persons that

serve as a sensitive transitional state marker for dementia syndromes. The identification of individuals with a profile compatible with an MBI syndrome is an encouraging result concerning the prognostic value of MBI for dementia development and for predicting different dementia subtypes.

The rest of the paper is organized as follows. Methods and data section presents the proposed approach by introducing all materials and methods, together with the validation of our longitudinal clustering algorithm on synthetic datasets, with details reported in the Appendix. Longitudinal clustering results section shows the result of the proposed longitudinal clustering on the case study dataset. Finally, Discussion on the case-study dataset section discusses the obtained results, and in Conclusions section conclusions are drawn.

Methods and data

Multivariate *k*-Means for longitudinal clustering

To investigate the existence of different trajectories of specific indicators of cognitive decline over time in a cohort of individuals participating in an Alzheimer's progression study, we developed a version of *k*-Means clustering for multivariate time series. The multivariate *k*-Means longitudinal clustering algorithm has been implemented to cluster joint trajectories of different features describing the individual progression over time. Such a method is based on time series *k*-Means clustering [45], a relatively novel method commonly used to identify univariate time series patterns. The following is a formal presentation of our approach.

Let us consider a set P of p patients. For each patient i, a set of n outcome variables (i.e., features) are measured at each visit t, defining the patient's health state at time t, as follows:

$$F_i(t) = \{f_{i1}(t), f_{i2}(t), \ldots f_{in}(t)\} \tag{1}$$

The joint trajectory for patient i for the time window $t \in [0, m]$ is defined as follows:

$$JointTraj_i = (F_i(0), F_i(1), \ldots F_i(m)) \tag{2}$$

where $F_i(0), F_i(1), \ldots F_i(m)$ are the health states of patient i at $t = 0$ (i.e., at the first visit that represents the baseline) and at the follow-up assessments.

According to this notation, the algorithm aims to find k homogeneous sub-groups within the set of P patients with respect to the joint trajectory trend by specifically adopting *k*-Means clustering. In the Algorithm 1 box, the pseudocode of the algorithm is presented.

Algorithm 1 Multivariate k-Means for longitudinal clustering

Data: $Data = \{JointTraj_i | i \in [1, p]\}$
Result: $Centroids : \{C_k\} = \{c_{kj}(t) | j \in [1, w], w \leq n, t \in [1, m]\}, \forall k \in [1, Nclust]$
$Initial\_Features \leftarrow \{f_1, f_2, \dots f_n\}$;
$Data_{scaled} \leftarrow scale(Data)$;
$k \leftarrow 2$;
while $k < Nclust$ **do**
  $iteration \leftarrow 0$;
  $Centroids \leftarrow \varnothing$;
  $Clusters \leftarrow \varnothing$;
  $Data_{current} \leftarrow get(Data_{scaled}, Initial\_Features)$;
  /*Cluster Initialization*/
  for $c = 1, ..., k$ **do**
    $C_c \leftarrow RandomChoose(Data_{current})$;
    $Clust_c \leftarrow RandomAssign(Data_{current}, C_c)$;
  $Centroids \leftarrow \{C_1, ..., C_k\}$;
  $Clusters \leftarrow \{Clust_1, ..., Clust_k\}$;
  /*Cluster Formation*/
  repeat
    $iteration \leftarrow iteration + 1$;
    for $i = 1, ..., p$ **do**
      $List\_Dist \leftarrow \varnothing$
      for $c = 1, ..., k$ **do**
        $dist(c) \leftarrow SoftDTWDist(JointTraj_i, C_c)$;
      $List\_Dist \leftarrow \{dist(1), ..., dist(k)\}$;
      $closestC \leftarrow indexOf(min(List\_Dist))$;
      $Clusters \leftarrow assign(JointTraj_i, closestC)$;
    $Centroids \leftarrow SoftDTWBarycenter(Clusters)$;
    /*Clustering Stop Condition Evaluation */
    for $c = 1, ..., k$ **do**
      $\delta(c) \leftarrow SoftDTWDist(C_{c_{iter-1}}, C_{c_{iter}})$;
    $\Delta \leftarrow \{\delta(1), ..., \delta(k)\}$
    if $\Delta < Threshold$ **then**
      stop=True;
  until $iteration \neq maxIteration \ OR \ stop = True$;
  $EvaluateMetrics(Clusters)$;
  /*Feature Selection */
  $Features\_Selected \leftarrow \varnothing$;
  for $j = 1, ..., n$ **do**
    $selected \leftarrow False$;
    for $h = 1, ..., (k-1)$ **do**
      for $z = (h+1), ..., k$ **do**
        $dist_j(hz) \leftarrow SoftDTWDist(c_h, c_z)$;
        if $dist_j(hz) > threshold$ **then**
          $selected \leftarrow True$;
          $Append(Features\_Selected, f_j)$;
  if $Features\_Selected = Initial\_Features$ **then**
    $k \leftarrow k + 1$
  else
    $Initial\_Features \leftarrow Features\_Selected$

As it is well known, *k*-Means [42] is a popular clustering algorithm that aims to partition p elements into k clusters, in which each observation belongs to the cluster with the nearest center. *k*-Means can also be seen as a particular case of a Expectation-Maximization (EM) algorithm for an iterative convergence [46]. The idea behind *k*-Means is to simply separate a dataset with p observations into k homogeneous groups. For the initialization of this method, the number of clusters k is usually needed beforehand as an input. Thus, it starts by randomly assigning the clusters centroid in the space. Then, each data point is assigned to one of the clusters based on its distance from the cluster's centroid.

To reach the optimal partition, the algorithm alternates between two steps: Expectation (E) and Maximization (M). In the E step, the distance between the observations and the centroids of each cluster is calculated. The M step then consists of assigning each observation to its nearest cluster. These two phases are done repeatedly and iteratively until a stabilization in the cluster assignment is reached. In the context of longitudinal data analysis, "*cluster centers*" are defined as the average trajectory of each group, representing the mean of all individual trajectories within the respective clusters. For an individual i, the "*nearest cluster*" denoted as C can be defined as the cluster that minimizes the distance between i and the mean trajectory of cluster C.

Normally, *k*-Means use Euclidean distance. However, in the case of time series, it generally performs poorly. This paper implements *k*-Means for multivariate time series by employing soft-DTW distance. Soft-DTW [47] is a differentiable loss function suitable for Dynamic Time Warping. This allows for the application of gradient-based algorithms in the context of time series analysis. The barycenter is defined as the time series that minimizes the aggregate distance between itself and the other time series within a given dataset. Finally, a further centroid-based feature selection method was implemented in the K-mean-based longitudinal clustering, where the features with the closest similarity between cluster centroids (that means overlapping) were discarded (since such features decrease cluster separation), and the algorithm performs a new execution with the new set of features. It is worth noting that a propaedeutic step to clustering is to scale data opportunity. In the case under study, we adopt z-score normalization.

Validation

To systematically assess the efficacy of the proposed longitudinal clustering algorithms in accurately grouping subjects based on their similar joint trajectories of pertinent features, we conducted a series of experiments utilizing synthetic datasets. These experiments were designed to evaluate the algorithm's performance across varying degrees of data variability, achieved by manipulating the levels of within-group and between-group correlations in the synthetic datasets. These datasets were designed to simulate realistic scenarios where subjects exhibit longitudinal data across multiple time points, with temporal dependencies influencing the clustering structure. Hence, to validate the proposed approach, five scenarios have been simulated by increasing the level of variability. We started with well-separated groups characterized by very low within-group variability (i.e., high temporal consistency within groups) and high between-group separation to

total overlapping groups, characterized by high within-group variability (i.e., low temporal consistency within groups) and low between-group correlation leading to substantial overlap and noise in the data. The intermediate cases concern both well-separated groups with moderate intra-group variability and high between-group separation and moderate overlapping groups characterized by an increased within-group and between-group variability with respect to the previous case, resulting in a moderate overlapping between groups.

The proposed longitudinal clustering algorithm proved to be efficient in detecting clusters even when moderate overlapping among groups is present. However, in the presence of high within-group variability and low between-group correlation, the algorithm's ability to effectively identify clusters diminished significantly. Further details of the validation analyses and full results are reported in the Appendix.

The case-study dataset

Data used in this paper were obtained from the Open Access Series of Imaging Studies-3 (OASIS-3) database [7]. OASIS-3, collected by Washington University Knight Alzheimer Disease Research Center provided Magnetic Resonance (MR) imaging and related clinical data of 1098 participants, consisting of 605 cognitively normal adults and 493 individuals at various stages of cognitive decline ranging in age from 42 to 95 years. OASIS-3 is unique in that the initial enrollment focused on the pre-clinical cohort and followed longitudinal progression. For each participant, OASIS-3 documents the corresponding entries in a time series.

Participants were assessed through clinical protocols following the National Alzheimer's Coordinating Center Uniform Data Set (UDS) [48]. UDS assessments included participant demographics, family history of AD, medical history, physical examination, and neurological evaluation, each recorded in a dedicated dataset.

Dementia status was assessed for the UDS using the Clinical Dementia Rating (CDR) Scale [49]. In addition, also Clinical Dementia Rating Scale Sum of Boxes (CDR-SB) [50] and Mini-Mental State Examination (MMSE) [51] are included in the dataset as indicators of dementia, with the CDR-SB providing more stratified categories. Table 1 reports dementia stages according to these metrics.

All participants were required to have a $CDR \leq 1$ at the baseline visit. Participants who reached $CDR = 2$ were no longer eligible for in-person assessments.

Table 1 Cognitive metrics for dementia stages

	Normal	Very Mild Impairment		Mild Impairment	Moderate Dementia	Severe Dementia
CDR	0	0.5		1	2	3
CDR SB		**Questionable Dementia**	**Very Mild Dementia**			
	0	[0.5, 2]	[2.5, 4]	[4.5, 9]	[9.5, 15.5]	[16, 18]
MMSE	30	[26, 29]		[21, 25]		[0, 10]

The Neuropsychiatric Inventory (NPI) was used to assess neuropsychiatric symptoms (NPS). This 12-item informant-based interview is a widely accepted measure of NPS in dementia. NPS consists of 12 domains: delusions, hallucinations, agitation/aggression, depression, apathy, irritability, anxiety, euphoria, aberrant motor behavior, disinhibition, and sleep and appetite disturbance. Each NPS domain is rated according to its severity (0–3). The behavioral assessment was performed using the Geriatric Depression Scale (GDS). The GDS taps the affective and behavioral symptoms of depression and excludes most symptoms that may be confused with somatic disease (for example, slowness, insomnia, etc.). It is a yes/no self-report questionnaire completed by the patient.

Finally, we take the final diagnosis of a patient given by a clinician from the feature diagnosis (DX) of the dataset. The clinician diagnosis classifies each individual as either cognitive normal (CN), with dementia (AD), or with an uncertain diagnosis (U).

Dataset pre-processing and feature selection

Before performing cluster analysis, data was pre-processed to collect patients with at least five visits. Moreover, a feature selection process has been applied to take into account only relevant features. Such a pre-processing phase on the OASIS-3 dataset was carried out in several steps.

Firstly, all the UDS datasets were merged, and a new dataset was obtained from which only the participants with 5 consecutive visits occurring with an annual frequency, with a tolerance of 2 months, were selected. This step resulted in a complete dataset of 166 subjects.

Subsequently, a feature selection process was conducted to select the most relevant features to reduce the cardinality of the input features, thus improving the computational cost of modeling and the model's performance. This process is articulated as follows.

First, since our aim is to uncover feature patterns that are not directly connected to the effects of Alzheimer's disease on brain structure, we removed the features related to the MRI and PET images that account for the brain structural changes.

Then, the UDS features related to the final clinical assessment are discharged, as these indicators are widely utilized in assessing Alzheimer's disease. Therefore, CDR, CDR-SB, and MMSE scores are excluded from the clustering analysis. For the same reason, the diagnosis feature DX is also excluded from the analysis. After that, the features with at least 20% of undefined values for all 5 visits were eliminated because they did not significantly contribute to the study and could wrongly affect the clustering as well as features containing values labeled as *unknown*. Then, filtering using the Pearson correlation coefficient was performed to remove high-correlated features. After these steps, thirty-four features have been collected. The list of these features and the range of their values are reported in the supplementary material. Then, this set of features is given as input of our longitudinal clustering as described in Algorithm 1 that performs a further feature selection discharging twenty-seven further features. The seven features that were finally selected refer to neuropsychological factors: GDS, agitation, dysphoria, anxiety, disinhibition, irritability, and apathy.

Longitudinal clustering results

As previously said, a total of 166 subjects were identified in the OASIS-3 dataset after data pre-processing to be included in the proposed longitudinal analysis, whose diagnosis at the baseline visit ($t = 0$), reported by the variable DX in the original dataset, are so distributed: 123 (i.e., 74.1%) have a diagnosis of cognitive normal (CN) and a CDR score equal to 0, 21 (i.e., 12.7%) have dementia diagnosis (AD) and a CDR score in the range of [0.5, 1], and 22 (i.e., 13.2%) have an uncertain diagnosis (U) and a CDR score equal to 0.5. All participants had four follow-up assessments, each occurring at one-year intervals, on average.

At the final visit ($t = 4$), 118 (i.e., 71.1%) individuals were diagnosed with normal cognition (CN), 32 (i.e., 19.3%) with dementia (AD), and 16 (i.e., 9.6%) remained with an uncertain diagnosis (U). Among AD individuals, the majority had mild dementia (i.e., 87.5%, CDR score $\in [0.5, 1]$), while the remaining (i.e., 12.5%) had moderate dementia (CDR=2). No severe diagnoses are presented. All individuals with uncertain diagnoses at the final visit show a CDR score of 0.5.

Given the unsupervised nature of the problem, meaning that subtypes do not have known labels, there is no standard way to use prediction ability to drive model selection. To determine the number of clusters, we evaluated the performance of the longitudinal clustering by varying the number of clusters from 2 to 6 using the three most common clustering metrics: the *Silhouette score* (SS) [52] utilized to assess the cohesion and separation of clusters in the $[-1; +1]$ interval (the higher, the better); the *Davies-Bouldin index* (DBI) [53]) that measures the ratio of within-cluster distances to between-cluster distances in the $[0; +\infty)$ interval (the lower the better); and the *Calinski-Harabasz index* (CHI) [54] that evaluates the ratio of between-cluster dispersion to within-cluster dispersion in the $[0; +\infty)$ interval (the higher the better). According to the results reported in Table 2, the best values of the metrics are obtained for $k = 2$.

Figure 1 shows the results of the longitudinal clustering with k=2. Mainly, here we can observe the trend of the selected NPS features (DISN - Disinhibition; IRR - Irritability; APA -Apathy; AGIT-Agitation; DYSPH - Dysphoria; ANX- Anxiety; GDS- Geriatric Depression Scale) at the baseline (referred to as 0 on the x-axis) and during the follow-up assessments (points 1, 2, 3 and 4 on the x-axis). As we can note, the two clusters are clearly separated, making it easy to identify subjects without neuropsychiatric symptoms in *Cluster 2* and those with neuropsychiatric symptoms in *Cluster 1*.

To deepen the analysis, we also examine the characteristics of these clusters from the perspective of the cognitive profile determined by the CDR and MMSE that were not included in the clustering process. As shown in Fig. 2, there is a clear separation during the whole follow-up assessments between healthy subjects, with an average MMSE ≈ 29 and an average CDR ≈ 0, and cognitively impaired individuals that show a slight decrement of the MMSE and a slight increment of the CDR-SB. This trend is very likely because, with $k = 2$, the obtained separation is simply between healthy subjects and all the other subjects having some kind of cognitive problems.

It is evident that although the clustering metrics are higher with $k = 2$, there is no stratification of individuals with different profiles in terms of neuropsychological symptoms and cognitive decline. Usually, in deciding the number of clusters, two types of errors may occur. The first error occurs when k groups are identified in the data, but in

reality, the number of groups is less than k, resulting in a solution with too many clusters. The second type of error occurs when a number of groups lower than those actually present in the data is set, thus obtaining a partition with too few clusters. Although the severity of the two error types varies depending on the application context, erroneously considering too few groups leads to a loss of information stemming from the merging of distinct groups.

Thus, since $k = 2$ and $k = 3$ show comparable performance metrics (see Table 2), we decided to examine the cluster internal composition with respect to the diagnosis feature that was also not included in the clustering process, both for $k = 2$ and $k = 3$ to reveal further stratifications in the population of individuals with cognitive decline, considering their neuropsychological symptoms.

Tables 3 and 4 report the composition of clusters obtained with $k = 2$ and $k = 3$, respectively, accordingly to the clinical diagnosis.

For both $k = 2$ and $k = 3$, longitudinal clustering determines a cluster of cognitively normal subjects (respectively *Cluster 2* for $k = 2$ and *Cluster 1* for $k = 3$), but with a better distribution obtained for $k = 3$ with the 86.3% of healthy individuals w.r.t. the 84.2 % for $k = 2$. Moreover, for $k = 3$, the algorithm identifies a subset of individuals with dementia (i.e., *Cluster 3*) that, for $k = 2$, were entirely contained in *Cluster 1*. It is also worth noting that with $k = 3$, in *Cluster 3*, there are no CN (healthy) individuals at the final visit, denoting the greater discriminative capacity achieved by increasing the number of clusters.

Further analysis concerning the individuals belonging to clusters obtained with different k values clearly reveals the increased individual stratification obtained with $k = 3$ as reported in the diagram of Fig. 3.

The longitudinal clustering with $k = 3$ was carried out on the same features, and the results of their trends for each cluster are reported in Fig. 4. As shown, *Cluster 1* groups individuals with a very slight level of depression (GDS) and with a stable trend of the relevant NPI domains, without problems in each domain over time. Regarding *Cluster 2* and *Cluster 3*, individuals show similar problems concerning dysphoria and disinhibition, but they differ in what concerns agitation, anxiety, irritability, and apathy. In particular, individuals of both *Cluster 2* and *Cluster 3* present a relatively stable trend of dysphoria, with a little increment of disinhibition for individuals of *Cluster 3* that instead remains stable for the ones of *Cluster 2*. Differences between the two clusters arise from the third assessment since individuals in *Cluster 3* experienced a further decline in the NPI domains related to anxiety, agitation, irritability, and apathy.

Table 2 Clustering performance results

# Clusters	SS	CHI	DBI
2	0.688	241.05	0.659
3	0.610	220.42	0.835
4	0.584	205.56	0.866
5	0.348	148.87	1.164
6	0.344	140.09	1.273

Silhouette score (*SS*) interval: $[-1; +1]$, the higher the better. Calinski-Harabasz index (*CHI*) interval: $[0; +\infty)$, the higher the better. Davies-Bouldin index (*DBI*) interval: $[0; +\infty)$, the lower the better

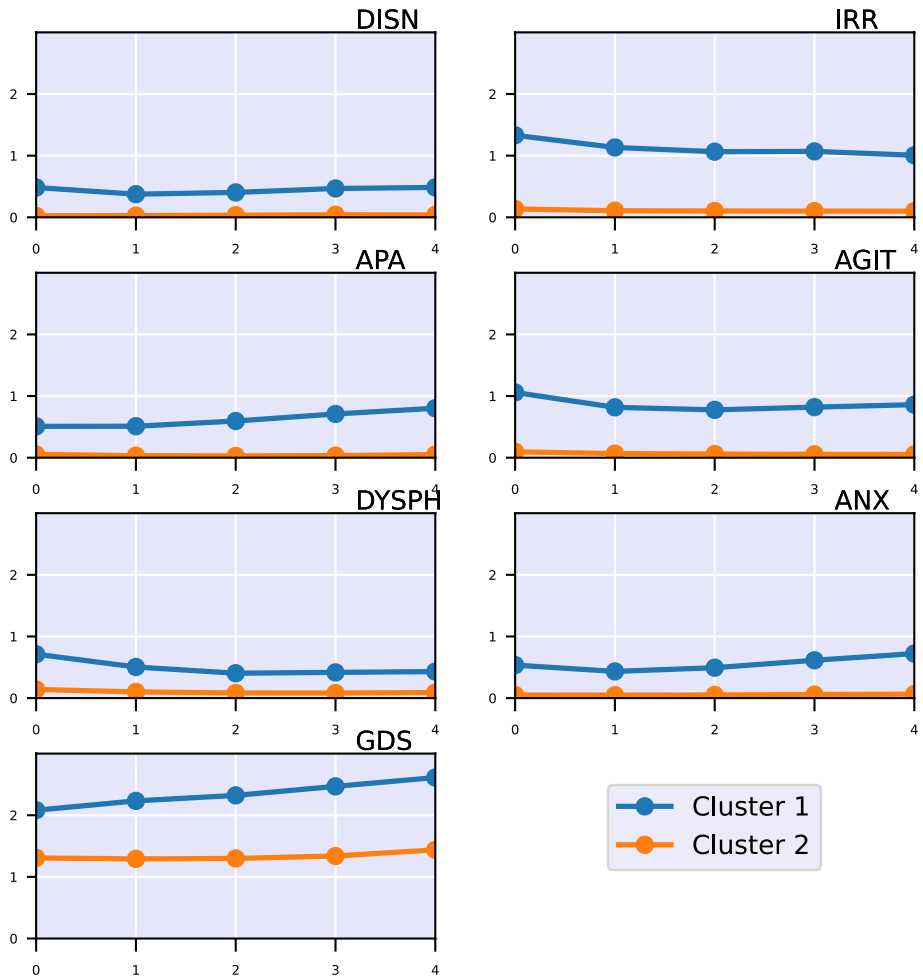

Fig. 1 Trends of the clusters' centroids for the analyzed variables of the OASIS-3 dataset with $k = 2$, starting from the baseline 0, and considering the four subsequent visits (1–4). Each cluster centroid trend is represented with a different color. AGIT (Agitation): 0 no, 1 mild, 2 moderate, 3 severe. ANX (Anxiety): 0 no, 1 mild, 2 moderate, 3 severe. APA (Apathy): 0 no, 1 mild, 2 moderate, 3 severe. DISN (Disinhibition): 0 no, 1 mild, 2 moderate, 3 severe. DYSPH (Dysphoria): 0 no, 1 mild, 2 moderate, 3 severe. GDS (Geriatric Depression Scale): 0–4 no significant depression, 5–9 mild depression, 10–15 severe depression. IRR (Irritability): 0 no, 1 mild, 2 moderate, 3 severe. *x* axis: time in years, starting from the baseline at year 0, and considering the four subsequent visits (1–4). *y* axis: feature value

For what concerns the cognitive aspects, as reported in Fig. 5, *Cluster 1* groups individuals whose cognitive profile is stable over the five visits with a value of $CDR \approx 0$, $CDR - SB \approx 0$ and $MMSE \approx 29$ on average, thus identifying subjects that do not show a cognitive decline. *Cluster 2* also groups individuals whose cognitive profile is stable over the five visits, but with different values of the diagnostic measures, $CDR \approx 0.5$, $1 < CDR - SB < 2$ and $MMSE \approx 28$ on average, thus identifying subjects that show a little cognitive impairment. *Cluster 3* groups individuals who show a cognitive decline during the follow-up assessments, as highlighted by the increasing CDR toward a value of 1, the increasing CDR-SB toward a value of 5, and the decreasing MMSE toward a value below 22 (Table 1). Thus, also, the cognitive profiles reported in Fig. 5 confirm the increased stratification with $k = 3$.

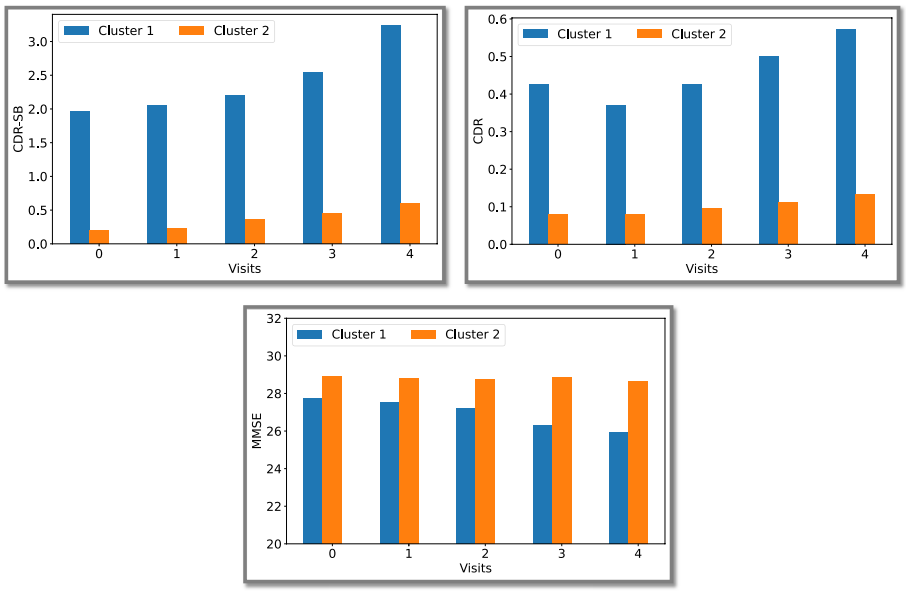

Fig. 2 Average values of CDR, CDR-SB and MMSE within the clusters for $k = 2$

Table 3 Cluster composition with $k = 2$ versus diagnosis

DX Baseline	Cluster 1 *N*=27	Cluster 2 *N*=139	DX Final	Cluster 1 *N*=27	Cluster 2 *N*=139
CN	22.2%	84.2%	**CN**	25.9%	79.9%
AD	51.9%	5.0%%	**AD**	66.7%	10.1%
U	25.9%	10.8%	**U**	7.4%	10.1%

DX diagnosis, *CN* cognitive normal, *AD* Alzheimer's disease, *U* uncertain diagnosis

Table 4 Cluster composition with $k = 3$ versus diagnosis

DX Baseline	Cluster 1 *N*=131	Cluster 2 *N*=26	Cluster 3 *N*=9	DX Final	Cluster 1 *N*=131	Cluster 2 *N*=26	Cluster 3 *N*=9
CN	86.3%	34.6%	11.1%	**CN**	80.9%	46.2%	0%
AD	4.6%	38.5%	55.6%	**AD**	9.9%	42.3%	88.9%
U	9.1%	26.9%	33.3%	**U**	9.1%	11.5%	11.1%

DX diagnosis, *CN* cognitive normal, *AD* Alzheimer's disease, *U* uncertain diagnosis

Hence, there is evidence that the results with $k = 3$ stratify more than $k = 2$, separating individuals with a higher cognitive decline from those in borderline situations. Thus, despite the slightly better clustering metrics with $k = 2$, we decided to deeply analyze the results with $k = 3$.

Statistical analysis for three clusters

A statistical analysis was conducted to evaluate and compare the neuropsychiatric symptoms and cognitive performance among the identified clusters for each visit. We utilized the 0.05 p-value threshold for statistical significance.

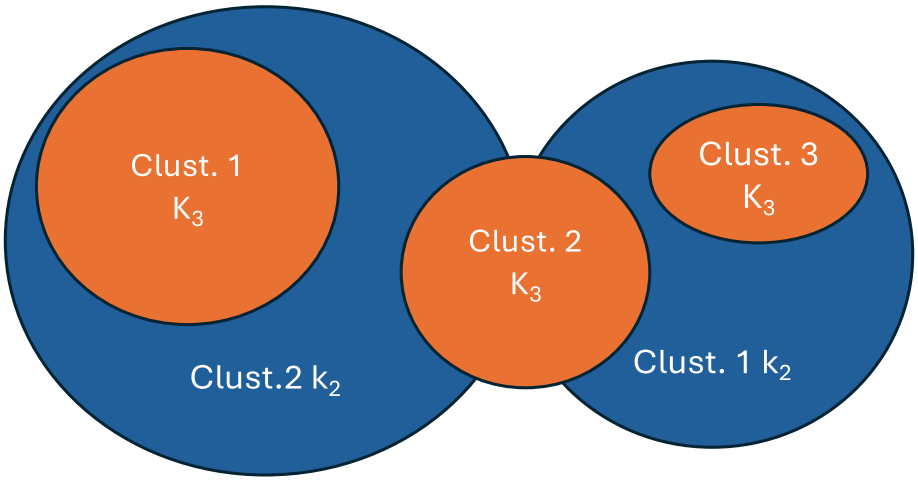

Fig. 3 Comparing individuals distributions for $k = 2$ and $k = 3$

The categorical variables were examined using a χ^2 test [55] to determine statistically significant differences among clusters. The quantitative variables were examined using the ANOVA or the Kruskal-Wall tests [56] based on the normality of their distribution. A pairwise comparison with post-hoc Tukey's test [57] was performed on variables showing statistical differences between clusters. All statistical analyses were performed using Python libraries.

Table 5 reports the characteristics of individuals in each cluster along with the related p-value at the baseline. Qualitative variables were represented in terms of frequency and percentage, while quantitative variables were represented in terms of mean and standard deviation ($mean \pm SD$).

As we can see in Table 5, there are no statistical differences between clusters in terms of age and education at the baseline. Individuals in each cluster show an average age of 73 years and an education level of 15 years. There is a statistical difference for gender ($p = 3.2 \cdot 10^{-3}$). A χ^2 post-hoc reveals differences between *Cluster 1* and *Cluster 2* and *Cluster 1* and Cluster 3, while no gender difference between *Cluster 2* and *Cluster 3* are revealed. Since *Cluster 2* and *Cluster 3* group subjects with major psychological concerns, the prevalence of male subjects within these clusters suggests that males are more affected than females by psychological disorders in elderly age. On the contrary, as shown in Table 5, there are statistically significant differences among clusters at the baseline for the MMSE ($p = 4 \cdot 10^{-2}$), the CDR ($p = 4.7 \cdot 10^{-11}$) and CDR-SB ($p = 7.4 \cdot 10^{-13}$).

A statistical analysis of the five visits was performed, leading to the following considerations regarding the features characterizing the severity of dementia, i.e., MMSE, CDR, and CDR-SB, and those related to the NPS domains.

Cluster1 vs Cluster 2 - For the MMSE, the post-hoc Tukey's test shown in Fig. 6 reveals that *Cluster 1* and *Cluster 2* are significantly different at visits 2 ($p < 0.05$), 3, and 4 ($p < 0.01$), while no statistical difference is present at visits 1 and 5. For the CDR, a statistically significant difference among clusters is highlighted at baseline ($p = 4.7 \cdot 10^{-11}$ in Table 5). The post-hoc χ^2 test on CDR shown in Fig. 6 reveals that the two clusters are significantly different at all visits ($p < 0.001$). The same occurs for the CDR-SB. These

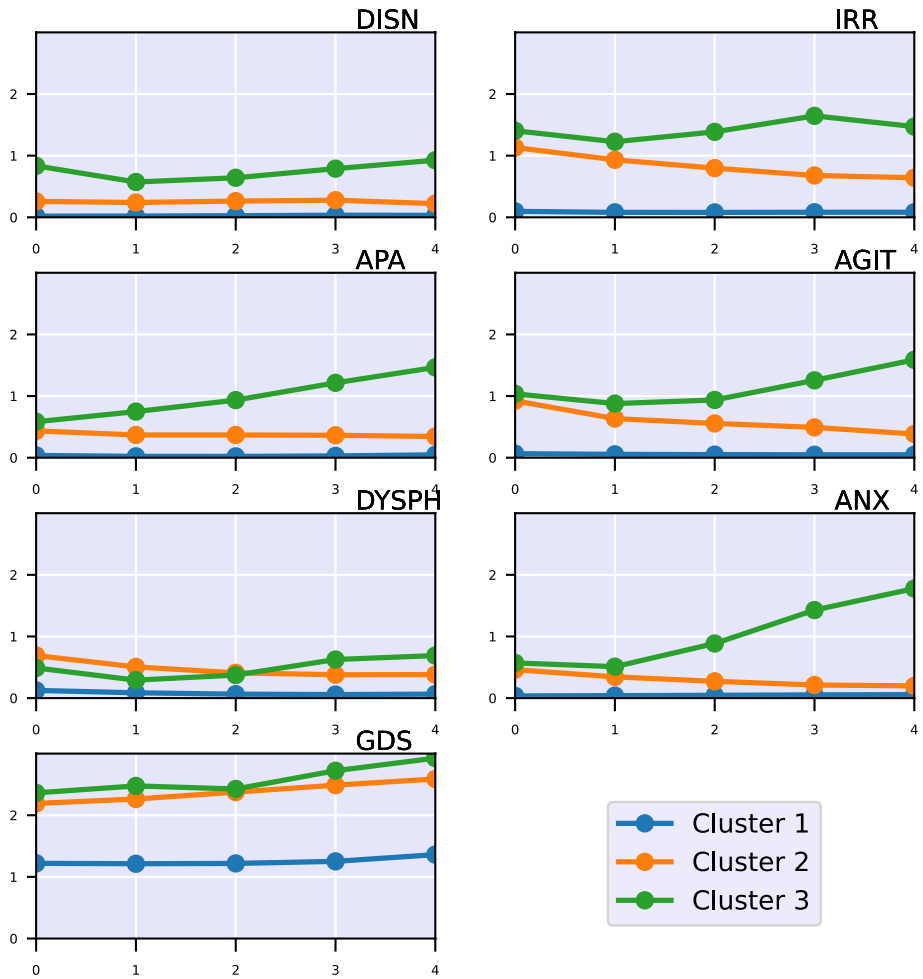

Fig. 4 Trends of the clusters' centroids for the analyzed variables of the OASIS-3 dataset with $k = 3$. Each cluster's centroid trend is represented with a different color. AGIT (Agitation): 0 no, 1 mild, 2 moderate, 3 severe. ANX (Anxiety): 0 no, 1 mild, 2 moderate, 3 severe. APA (Apathy): 0 no, 1 mild, 2 moderate, 3 severe. DISN (Disinhibition): 0 no, 1 mild, 2 moderate, 3 severe. DYSPH (Dysphoria): 0 no, 1 mild, 2 moderate, 3 severe. GDS (Geriatric Depression Scale): 0–4 no significant depression, 5–9 mild depression, 10–15 severe depression. IRR (Irritability): 0 no, 1 mild, 2 moderate, 3 severe. *x* axis: time in years, starting from the baseline at year 0, and considering the four subsequent visits (1–4). *y* axis: feature value

results confirm that *Cluster 1* and *Cluster 2* group individuals with different cognitive profiles, with healthy subjects in *Cluster 1*, as derived from the cognitive indicators' values. Regarding the NPS domains (Fig. 7), *Cluster 1* and *Cluster 2* present a statistically significant difference for agitation, apathy, dysphoria, irritability, and disinhibition for all visits, while for anxiety, there is a difference at all visits except the last visit. For geriatric depression, *Cluster 1* and *Cluster 2* present a statistically significant difference for all visits, with $p < 0.05$ for visits 1, 2, and 5, and $p < 0.01$ for visits 3 and 4.

Cluster 1 vs Cluster 3 - For the MMSE, the post-hoc Tukey's test shown in Fig. 6 reveals that the two clusters are significantly different at all visits ($p < 0.01$). The same occurs for CDR and CDR-SB with $p < 0.001$. Regarding the NPS domains (Fig. 7), *Cluster 1*

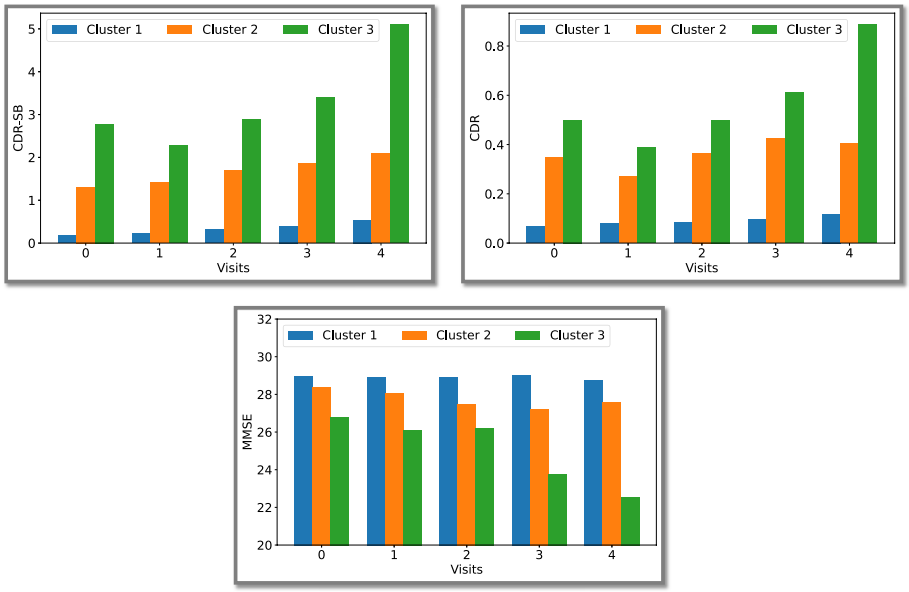

Fig. 5 Average values of CDR, CDR-SB and MMSE for $k = 3$

and *Cluster 3* present a statistically significant difference for agitation, anxiety, apathy, irritability, and disinhibition for all visits, while for dysphoria, there is a difference at all visits except visit 2. For geriatric depression, *Cluster 1* and *Cluster 3* present a statistically significant difference only at visit 4, with $p < 0.05$ for visits 1, 2, and 5, and $p < 0.01$ for visits 3 and 4.

Cluster 2 vs Cluster 3 - For the MMSE, the post-hoc Tukey's test shown in Fig. 6 reveals that the two clusters are significantly different at all visits ($p < 0.01$ at visit 1 and $p < 0.001$ for the others), except at visit 3. Nevertheless, there are no statistically significant differences for CDR and CDR-SB at all visits except for visit 5 (with respectively $p < 0.05$ and $p < 0.01$). For what concerns the NPS domains (Fig. 7), *Cluster 2* and *Cluster 3* present a statistical difference for agitation, anxiety, and apathy only at visits 4 and 5 (with $p < 0.05$, $p < 0.01$, $p < 0.001$). Instead, both for dysphoria and geriatric depression, there is not a statistically significant difference at all visits. For irritability, there is a statistically significant difference at visits 3 and 4 (with $p < 0.05$, $p < 0.01$), and for disinhibition, the difference occurs only at the last visit ($p < 0.01$).

Discussion on the case-study dataset

From the reported clustering results, it emerges that individuals with dementia develop neuropsychiatric symptoms at some stage of the disease or even well before the disease onset. This result confirms what is reported in the literature [58] that although Alzheimer's disease is a cognitive disorder, it is also linked to the development of neuropsychiatric symptoms. The frequency of NPS is much higher in people with AD than in the general elderly population. As reported in the literature, the main NPS domains that are affected by the possible disease onset are affective dysregulation, characterized by symptoms of anxiety, dysphoria, changeability, or euphoria, impulse dyscontrol, characterized

Table 5 Participants baseline characteristics

Features	Cluster 1 (*N*=132)	Cluster 2 (*N*=26)	Cluster 3 (*N*=9)	*p*-value
Age (years)				0.89[a]
mean (SD)	73 ± 6.4	73 ± 8.1	73 ± 6.7	
[min, max]	[64, 89]	[50,87]	[62, 83]	
Education (years)				0.97[a]
mean (SD)	15.7 ± 2.5	15.8 ± 2.3	15 ± 3.2	
[min, max]	[10, 20]	[12,20]	[10, 18]	
MMSE				$4.0 \cdot 10^{-02a}$
mean (SD)	29.0 ± 1.3	28.4 ± 2.4	26.8 ± 2.8	
[min, max]	[23, 30]	[19, 30]	[22, 30]	
GDS				$1.3 \cdot 10^{-02b}$
mean (SD)	1.24 ± 1.6	2.1 ± 1.8	2.5 ± 3	
[min, max]	[0, 9]	[0, 5]	[1, 10]	
CDR-SB				$7.4 \cdot 10^{-13a}$
mean (SD)	0.17 ± 1.47	1.3 ± 1.3	2.8 ± 1.7	
[min, max]	[0, 3]	[0, 4.5]	[0, 5]	
CDR				$4.7 \cdot 10^{-11b}$
=0	113 (86%)	9 (35%)	1 (11%)	
=0.5	18 (14%)	16(61%)	7 (78%)	
≥ 1	0 (0%)	1(4%)	1 (11%)	
Gender				$3.2 \cdot 10^{-03b}$
Female	74 (56%)	8 (31%)	1 (11%)	
Male	57 (44%)	18(69%)	8 (89%)	
AGIT				$1.3 \cdot 10^{-15b}$
=0	123 (94%)	8 (30%)	3 (33%)	
=1	7 (5%)	9(35%)	2 (22%)	
≥ 2	1 (1%)	9(35%)	4 (44%)	
DEPD				$2.3 \cdot 10^{-05b}$
=0	119 (90%)	14 (54%)	5 (56%)	
=1	6 (5%)	7(27%)	2 (22%)	
≥ 2	6 (5%)	5(19%)	2 (22%)	
ANX				$2.5 \cdot 10^{-07b}$
=0	128 (97%)	17 (65%)	6 (66%)	
=1	2 (2%)	5(19%)	1 (11%)	
≥ 2	1 (1%)	4(16%)	2 (22%)	
DISN				$3.5 \cdot 10^{-13b}$
=0	130 (99%)	20 (77%)	4 (45%)	
=1	0 (0%)	5(19%)	2 (22%)	
≥ 2	1 (1%)	1(4%)	3 (33%)	
IRR				$4.6 \cdot 10^{-16b}$
=0	118 (90%)	8 (31%)	1 (11%)	
=1	12 (9%)	8(31%)	3 (33%)	
≥ 2	1 (1%)	10(38%)	5 (56%)	
APATHY				$4.9 \cdot 10^{-08b}$
=0	127 (96%)	16 (62%)	6 (67%)	
=1	2 (2%)	7(27%)	1 (11%)	
≥ 2	2 (2%)	3(11%)	2 (22%)	

[a] Kruskal-Wallis H test

[b] χ^2 test

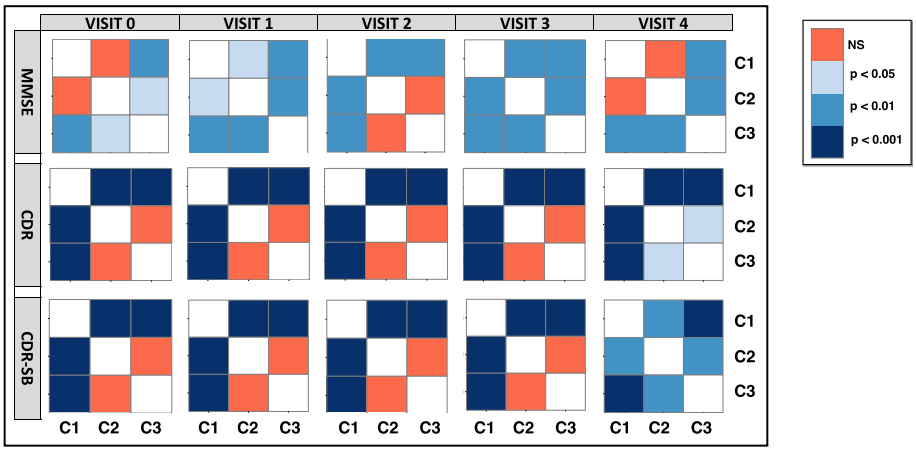

Fig. 6 Statistical analysis results for cognitive metrics. MMSE (Mini Mental State Examination). CDR (Clinical Dementia Rating). CDR-SB (Clinical Dementia Rating Sum of Box). C1 (Cluster 1). C2 (Cluster 2). C3 (Cluster 3)

by agitation, irritability, disinhibition, and decreased motivation, characterized by apathy. In line with these outcomes, from our results, it emerged that in individuals with AD, there are moderately increasing values for dysphoria and disinhibition and a more evident increment for irritability, apathy, agitation, and anxiety, with the last two symptoms worsening significantly during the last visit.

However, a deeper analysis is necessary to interpret the results concerning individuals with an uncertain diagnosis.

The trend of their neuropsychiatric symptoms is less clear, showing increased irritability and agitation but not increased anxiety, dysphoria, apathy, and disinhibition w.r.t. the general elderly population. As reported in recent Literature [44], new-onset persistent personality and behavior changes after age 50, rather than signaling a new-onset psychiatric disorder, could be an early manifestation of neurodegenerative disease. Indeed, Mild cognitive impairment (MCI) is often a transitional state between normal cognition and dementia, but it can preceded or accompanied by behavioral changes. Mild Behavioral Impairment (MBI) is a neurobehavioral syndrome that can reveal a high risk for incident dementia by leveraging risk associated with new-onset and persistent neuropsychiatric symptoms (NPS).

As we have seen from our results, NPS occurs in almost all individuals with an uncertain diagnosis, including agitation, depression, anxiety, apathy, disinhibition, and irritability; these symptoms fluctuate during the follow-up assessments, and they could be ascribable to an MBI condition. The clinical features of MBI and criteria were developed and validated by an international panel led by the International Society to Advance Alzheimer's Research and Treatment (ISTAART) NPS Professional Interest Area [59]. The ISTAART-MBI diagnostic criteria require demonstration that a change has occurred in at least 1 of the 5 core domains of NPS: decreased drive or motivation (apathy), affective dysregulation (mood or anxiety symptoms), impulse dyscontrol (for example, agitation, impulsivity), social inappropriateness (for example, impaired social cognition), and abnormal perception or thought (i.e., delusions or hallucinations). However, to meet the criteria for MBI, NPS should emerge after age 50, persist for longer than 6 months, represent a change in personality or behavior, and are not explained by another psychiatric disorder.

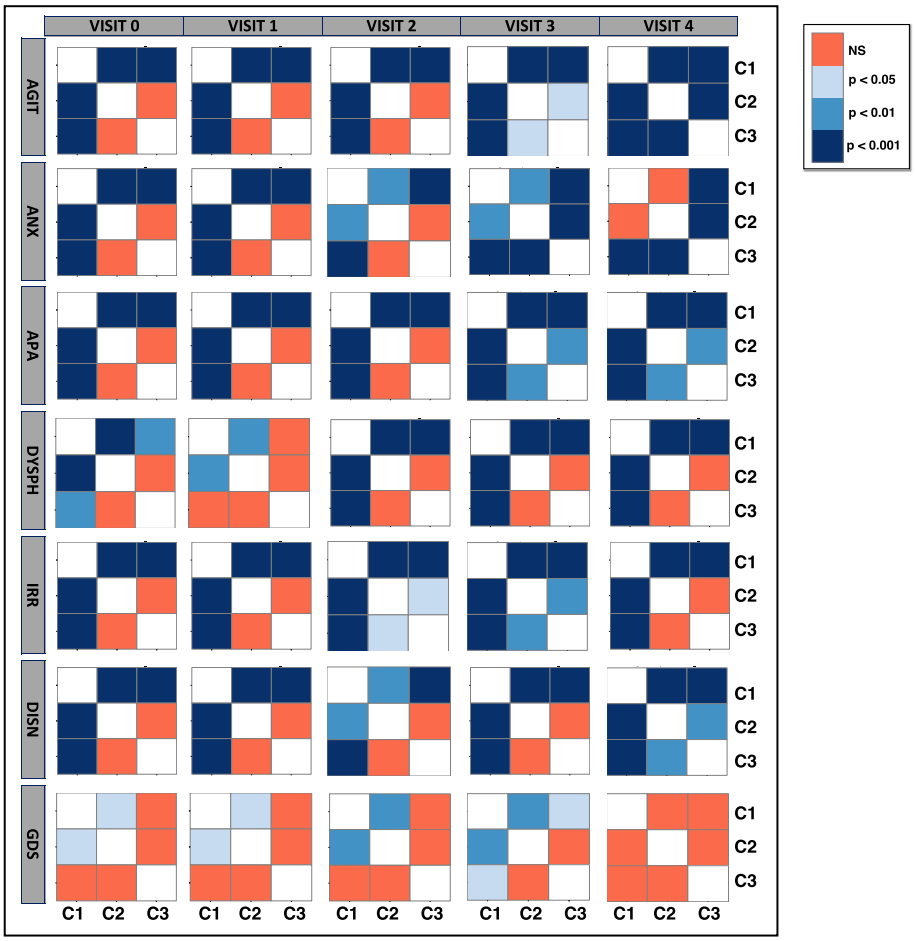

Fig. 7 Statistical analysis results for neuropsychological symptoms. AGIT (Agitation). ANX (Anxiety). APA (Apathy). DISN (Disinhibition). DYSPH (Dysphoria). GDS (Geriatric Depression Scale). IRR (Irritability). C1 (Cluster 1). C2 (Cluster 2). C3 (Cluster 3)

The characterization of individuals with an uncertain diagnosis, i.e., belonging to *Cluster 2*, that emerges from the trend of their neuropsychological traits seems to be compatible with the MBI diagnostic criteria, so providing additional evidence that MBI may be an early sign and predictor of Alzheimer's disease dementia and that more attention should be paid on the number and duration of MBI symptoms [60].

Hence, neuropsychological measures can provide reliable indicators of the presence of cognitive decline while being cost-effective and minimally invasive. Clinical and experimental neuropsychological research is beginning to uncover the earliest preclinical cognitive changes that might predict the subsequent development of dementia. Moreover, it has delineated different cognitive profiles that distinguish AD from other age-associated neurodegenerative disorders, which enhances an accurate differential diagnosis of dementia subtypes.

Conclusions

In conclusion, the adopted longitudinal clustering with three clusters was able to differentiate not only healthy subjects from the ones affected by dementia, as occurred with two clusters, but also to differentiate individuals that present risk factors to progress

toward dementia. The obtained results are encouraging in adopting unsupervised techniques and longitudinal studies to identify patterns of behaviors in the classes of individuals that are more critical in providing useful insights for early predictors of dementia.

We acknowledge that our study, based on a single dataset, needs further investigation to generalize the obtained results. To confirm the obtained results and prove the generalizability of our longitudinal clustering algorithm, we plan to apply our approach to other multi-visit datasets of EHRs, including NPS. Nevertheless, the results regarding the role of NPS in AD progression highlight the strength of our algorithm's intrinsic feature selection.

The validation results on synthetic datasets confirm the effectiveness of the proposed algorithm in discriminating groups, increasing the levels of within-group variability and between-group overlapping, also changing the group cardinality. In addition, the proposed algorithm can be applied to categorical data since a trajectory may be represented as a sequence of states rather than a continuum.

In future work, we plan to apply the algorithm to larger datasets to evaluate its computational performance.

Appendix

Several experiments were performed on synthetic datasets created by manipulating the levels of within-group and between-group correlations to evaluate the algorithm's performance considering varying degrees of data variability. Each experiment under the same variability condition consisted of 50 tests to calculate average performance values. Each test was executed on a different synthetic dataset randomly generated with the same within-group and between-group correlations. For the experimental design, we considered a synthetic dataset composed of 150 subjects. To give the datasets a longitudinal structure, we considered that each subject was observed at 10 distinct time points, with two features recorded per subject at each time point. This design aimed to capture the temporal evolution of each subject's features, reflecting typical longitudinal study settings where subjects are repeatedly measured over time. The subjects in the datasets were labeled as belonging to three different groups. The number of subjects per group was varied randomly in each test of each experiment, thereby introducing variability in the group sizes to better simulate real-world datasets. The features for each subject were generated with a degree of within-group correlation, meaning that the feature values for each subject followed consistent temporal patterns over time, with varying levels of correlation depending on the experimental condition. The within-group correlation was manipulated to represent low, moderate, or high temporal consistency. The datasets were generated to exhibit also varying degrees of between-group separation with different experimental conditions designed to create well-separated, overlapping, or highly variable group structures. These experiments allowed us to assess the algorithm's sensitivity to different levels of group separation and temporal variability. Hence, to validate the proposed approach, five scenarios have been simulated by increasing the level of variability:

- **Case 1: Well-separated Groups** - In this scenario, the groups were characterized by very low within-group variability (i.e., high temporal consistency within groups) and high between-group separation. In this case, the data for each group exhibit distinct and well-defined temporal patterns. Figure 8 illustrates this case, demonstrating clear separation between groups. Figure 9 shows the temporal trends of the two features across ten time points for the three groups, showing distinct trajectories and low variability within groups.
- **Case 2: Well-separated Groups with a moderate intra-group variability** - In this scenario, the groups were characterized by moderate temporal consistency within groups and high between-group separation. Figures 10 and 11 illustrate this case, demonstrating clear separation between the groups and highlighting trajectories with moderate variability within groups.
- **Case 3: Partial Overlapping Groups** - In this case, we introduced higher within-group and between-group variability with respect to the previous case, resulting in partial overlap between groups, as shown in Figs. 12 and 13.
- **Case 4: Moderate Overlapping Groups** - In this case, within-group and between-group variability is increased with respect to the previous case, resulting in a moderate overlapping between groups (see Figs. 14 and 15).
- **Case 5: Total Overlapping Groups** - The last case was designed to test the algorithm in an extreme situation, namely in the presence of high within-group variability (i.e., low temporal consistency within groups) and low between-group correlation leading to substantial overlap and noise in the data. This setup represented a more difficult scenario for clustering, where the temporal patterns across groups were highly variable and inter-group distinctions were weak. Figure 16 shows Feature 1 vs Feature 2 with significant overlap and less discernible separation between the groups, reflecting the high variability and weak between-group correlation. Figure 17 for this case depicts temporal trends with high within-group variability, showing total overlapping trajectories and substantial noise in the feature values over time.

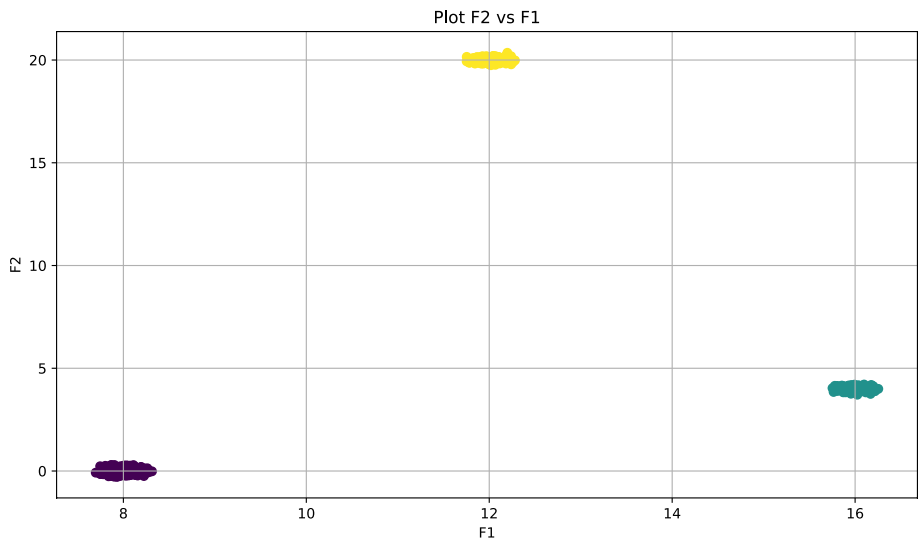

Fig. 8 Scatter plot of Feature 1 vs Feature 2 - CASE 1

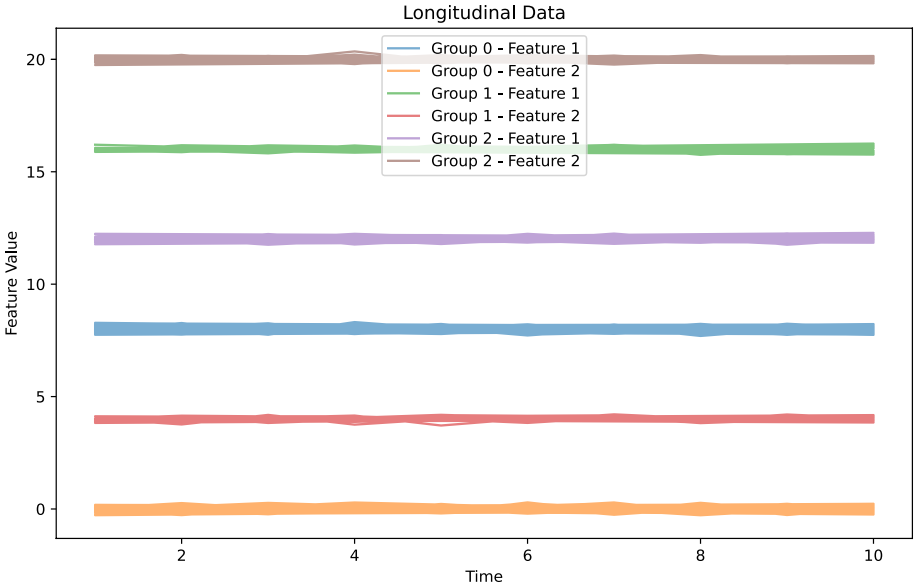

Fig. 9 Temporal trends for features across time for each group in CASE 1

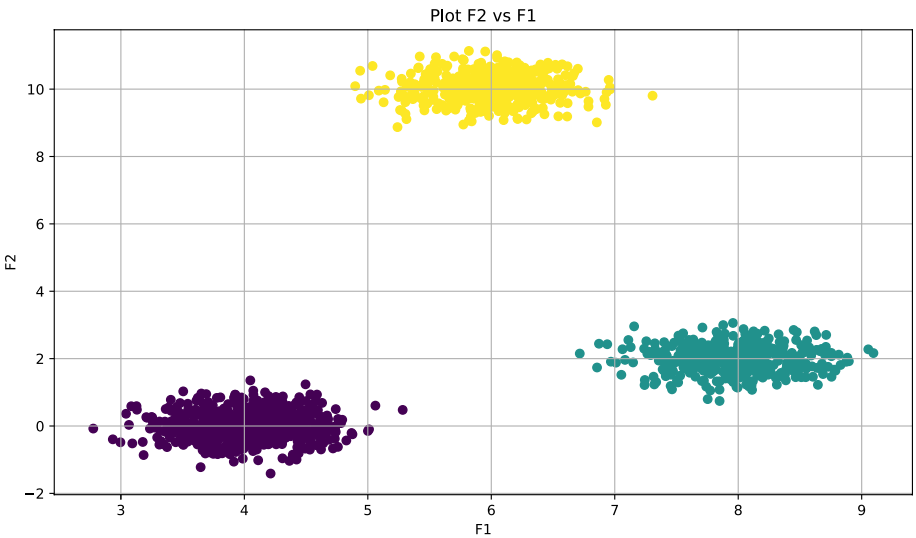

Fig. 10 Scatter plots of Feature 1 vs Feature 2 - CASE 2

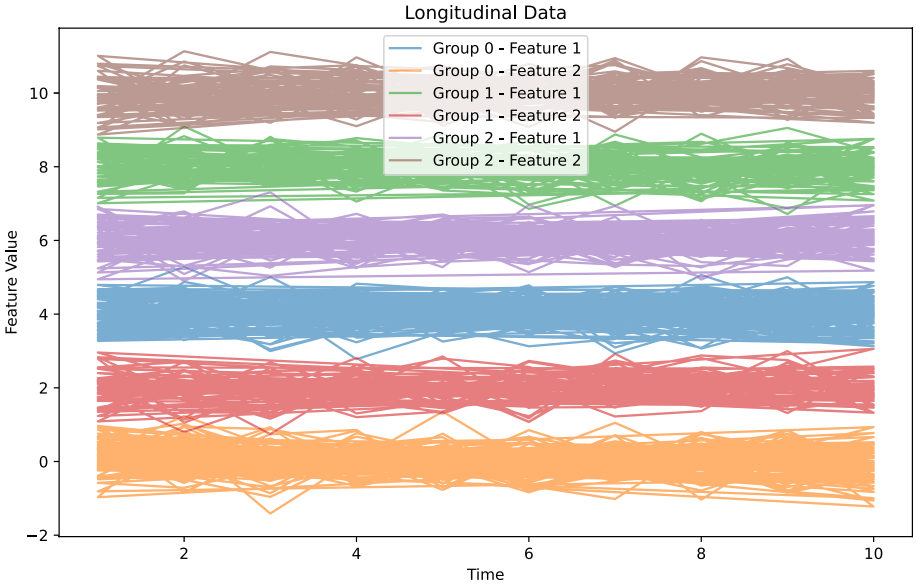

Fig. 11 Temporal trends for features across time for each group in CASE 2

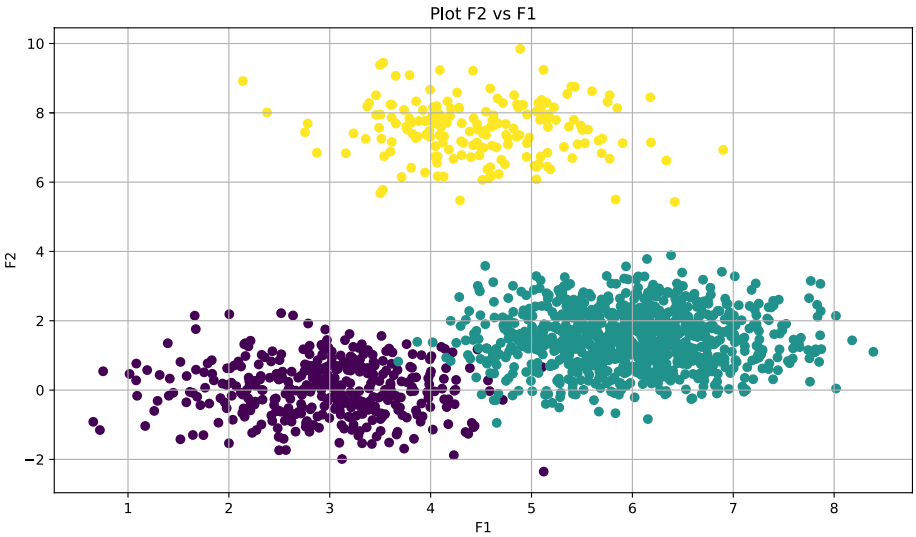

Fig. 12 Scatter plots of Feature 1 vs Feature 2 - CASE 3

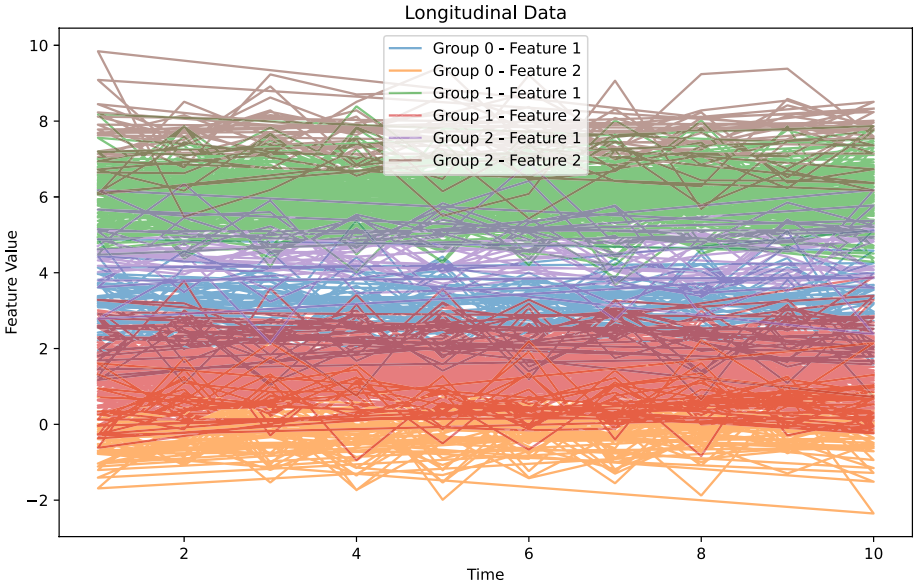

Fig. 13 Temporal trends for features across time for each group in CASE 3

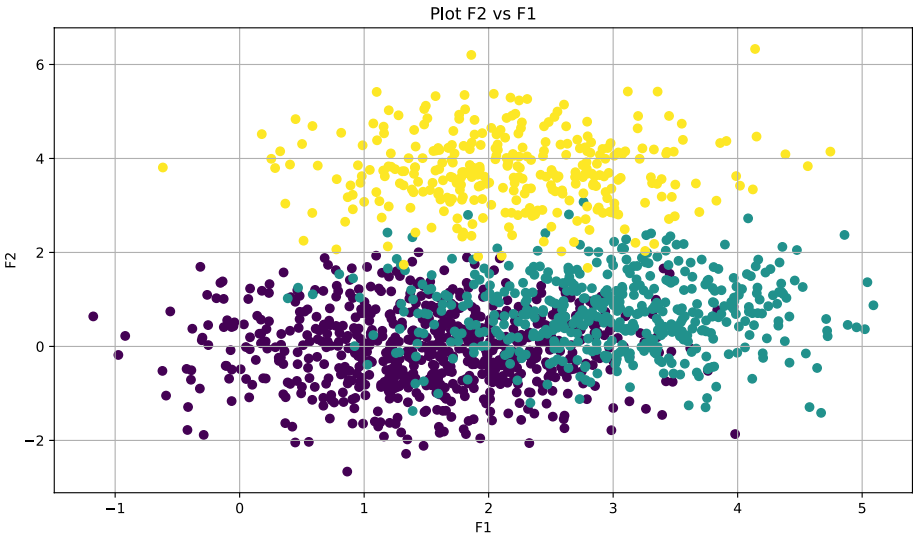

Fig. 14 Scatter plots of Feature 1 vs Feature 2 - CASE 4

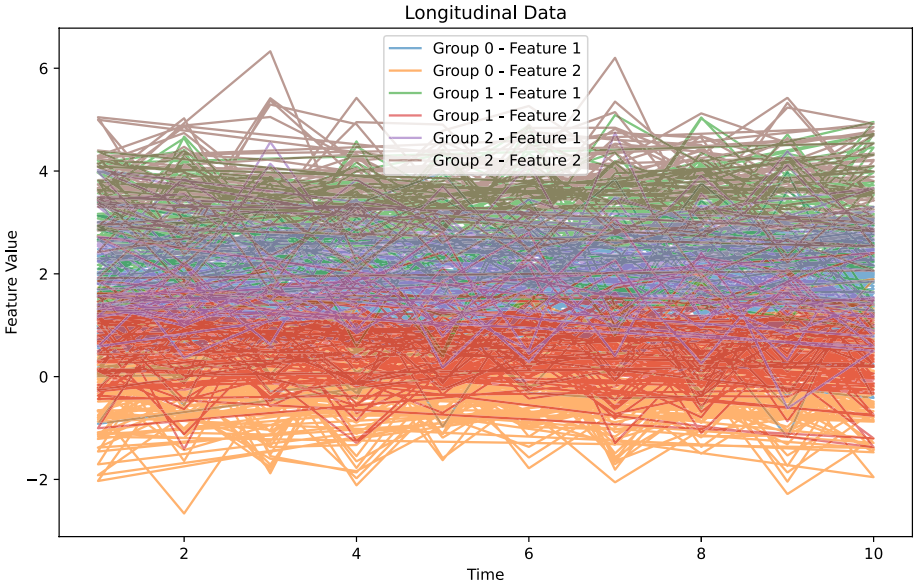

Fig. 15 Temporal trends for features across time for each group in CASE 4

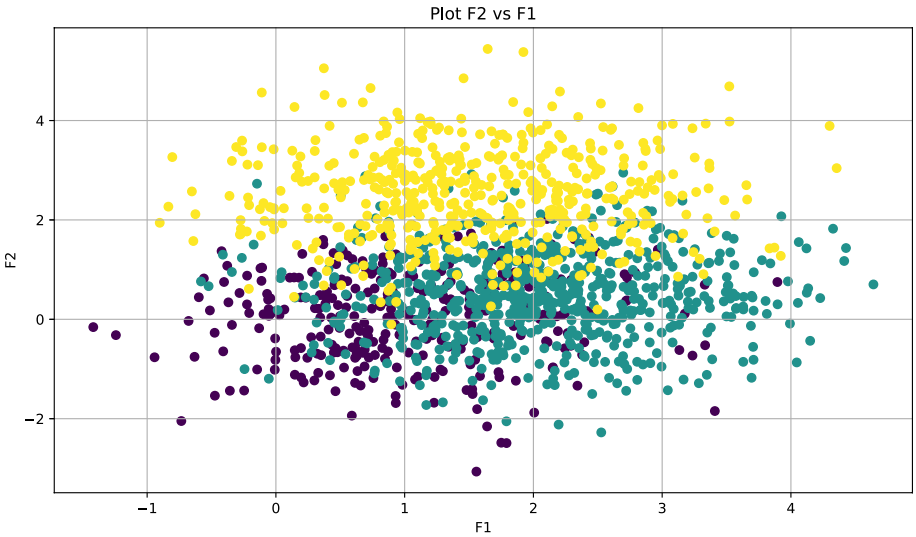

Fig. 16 Scatter plots of Feature 1 vs Feature 2 - CASE 5

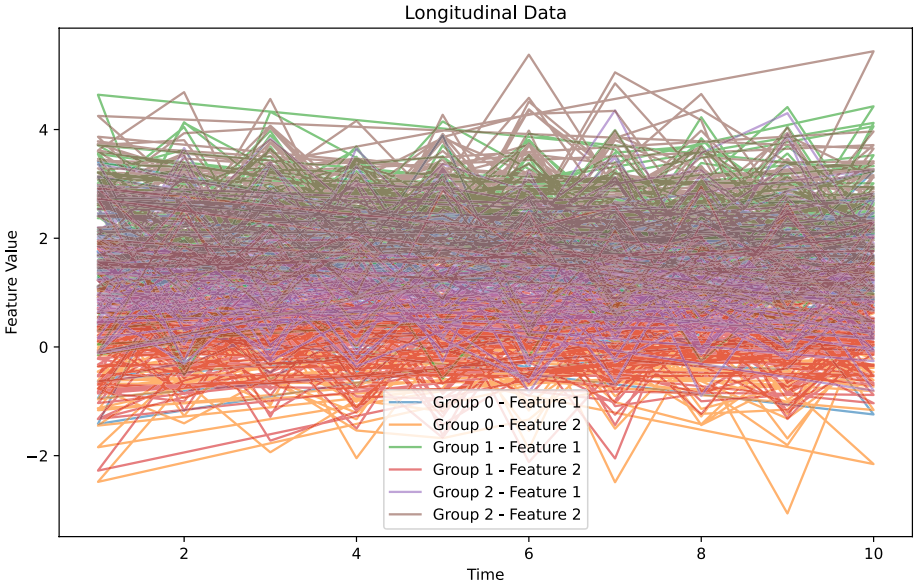

Fig. 17 Temporal trends for features across time for each group in CASE 5

Evaluation Results - The performance of the clustering algorithm was assessed using both internal and external clustering performance metrics. Internal clustering evaluation metrics are utilized to assess the quality of clustering results based exclusively on the intrinsic properties of the data, independent of any external information or predefined labels. Here we adopted the *Silhouette score* [52] utilized to assess the cohesion and separation of clusters in the $[-1; +1]$ interval (the higher, the better); the *Davies-Bouldin index* (DBI) [53]) that measures the ratio of within-cluster distances to between-cluster distances in the $[0; +\infty)$ interval (the lower the better); and the *Calinski-Harabasz index* (CHI) [54] that evaluates the ratio of between-cluster dispersion to within-cluster dispersion in the $[0; +\infty)$ interval (the higher the better).

On the contrary, external clustering evaluation metrics measure the quality of clustering results by comparing them to external labels, like the data's actual class labels. They assess how well the clustering results match a pre-defined structure [61]. Here, we adopted the *Homogeneity score* (HS) that measures how many data points within a cluster belong to the same class. In other words, a clustering result is homogeneous if most of the data points in a cluster are from the same true class. The homogeneity score ranges from [0; 1], where 1 means perfect homogeneity (i.e., all points in each cluster belong to the same true class). *Completeness score* (CS) evaluates whether all data points that belong to the same class are assigned to the same cluster. A clustering result is complete if all data points from a single class appear in the same cluster. The *Completeness score* ranges from [0; 1], where 1 indicates perfect completeness, meaning that all members of a class are grouped together in one cluster. Finally, *Normalized Mutual Information* (NMI) quantifies the amount of shared information between the clustering and the true class labels, normalized to account for the number of clusters

and the number of classes. It provides a balance between homogeneity and completeness by measuring how much information the clustering reveals about the true class distribution, while also taking into account the number of clusters and the inherent class distribution. NMI ranges from [0; 1], where 1 indicates perfect correlation (i.e., the clustering perfectly reflects the true class labels), and 0 indicates no mutual information (i.e., the clustering provides no insight into the true labels).

Table 6 reports the mean performance values for all metrics across the 50 tests for each considered scenario.

Table 6 Performance comparison under different variability conditions

Performance Comparison						
	SS	**CHI**	**DBI**	**HS**	**CS**	**NMI**
CASE 1	0.9998	1196749	0.02	1	1	1
CASE 2	0.9887	19868.1	0.08	1	1	1
CASE 3	0.9177	2551.7	0.23	1	1	1
CASE 4	0.6342	655.0	0.61	0.97	0.98	0.97
CASE 5	0.3830	235.5	1.14	0.81	0.86	0.83

Silhouette score (*SS*) interval: $[-1; +1]$, the higher the better. Calinski-Harabasz index (*CHI*) interval: $[0; +\infty)$, the higher the better. Davies-Bouldin index (*DBI*) interval: $[0; +\infty)$, the lower the better. Homogeneity score (*HS*) interval: [0; 1] the higher the better. Completeness score (*CS*) interval: [0; 1] the higher, the better. Normalized Mutual Information (*NMI*) interval: [0; 1] the higher, the better

As we can see, in CASE 1, the algorithm exhibited strong performance across all evaluation metrics, with the highest value of silhouette score, a very high value of Davis-Bouldin Index, and a value of Calinski-Harabasz Index ≈ 0. These results are consistent with the clear separation of groups and low within-group variability. The values of the external metrics of Completeness, Homogeneity, and Normalized Mutual Information highlight that all groups are clusterized correctly.

Regarding CASE 2, also in this case, the algorithm obtained very high performance with respect to all metrics. The values of external metrics show a very slight decrease due to the low variability of temporal trends for each subject.

In the partial overlapping case (CASE 3), the algorithm's performance remained good, though a slight decrease in performance was observed across the first three metrics compared to the previous cases. Silhouette scores were somewhat lower, indicating some overlap between the clusters, while DBI and CHI showed moderate degradation. However, CS, HS, and NMI highlight that all groups are clusterized correctly.

In the moderate overlapping case (CASE 4), the algorithm's performance starts to decrease. A decrease in performance was observed across the first three internal metrics compared to the previous cases, reflecting more overlapping groups. However, CS, HS, and NMI remained high, suggesting that the algorithm was still able to assign most subjects to their correct group despite some misclassifications at the boundaries of the overlapping regions.

Finally, in extreme CASE 5, the algorithm's performance significantly declined in the presence of high within-group variability and low between-group correlation. The Silhouette scores were notably lower, reflecting poor separation between clusters. Both the DBI and CHI notably worsened, indicating reduced inter-cluster separation

and increased intra-cluster variability. Completeness and Homogeneity scores also dropped, suggesting that the algorithm struggled to correctly assign subjects to their respective groups. The algorithm performance in this case is markedly worse due to the high noise and weak group separation in the data.

It is important to underline that Soft Dynamic Time Warping has been chosen as a distance metric for its ability to better work with time series, especially in situations where there is overlap or variability in features. This is why the performance of CASE 3 and CASE 4 benefit from these metric characteristics. Indeed, Soft DTW [47] is a smoother version of Dynamic Time Warping (DTW). Unlike traditional DTW, which gives a binary "match" or "mismatch" between points in two time-series sequences, Soft DTW assigns a soft cost for the warping path. This means that it allows for more flexibility in aligning time-series data that may have noise, outliers, or slight misalignments. Its ability to focus on trends rather than precise values helps when there is some overlap in the features, as the algorithm can still find common patterns in the data even when the exact feature values are not the same. Soft-DTW provides a principled way to differentiate between overlapping trajectories by leveraging global alignment costs rather than local point-wise differences. Even in cases of significant overlap, the clustering results remain valid due to Soft-DTW's ability to capture temporal and structural variations in trajectory patterns [62]. Hence, because of its soft matching approach, soft DTW is less sensitive to noise and outliers than traditional methods. Thus, if there is some overlap or noise in the features of the longitudinal data, Soft DTW's smoother approach to warping helps it tolerate these inconsistencies and still identify the underlying patterns, improving clustering quality. This flexibility helps in cases where there is some overlap or variation in the features being measured, leading to more accurate clustering results despite those variations. Soft DTW focuses on the temporal alignment and overall trends of the time series rather than on exact feature values at specific time points.

In conclusion, the proposed longitudinal clustering algorithm demonstrated a high degree of efficacy in identifying well-separated clusters when the data exhibited clear temporal patterns and low within-group variability. Under conditions where groups exhibited moderate overlap or varying levels of temporal correlation, the algorithm maintained reasonable performance, although some degradation in clustering quality was observed. However, in the presence of high within-group variability and low between-group correlation, the algorithm's ability to effectively identify clusters diminished significantly. The results, as detailed in Table 6, highlight the algorithm's strengths in scenarios with well-defined temporal patterns but also reveal its limitations in handling very noisy, highly variable longitudinal data.

Abbreviations

AD	Alzheimer's disease
AGIT	Agitation
ANOVA	Analysis of variance
ANX	Anxiety
APA	Apathy
CDR	Clinical Dementia Rating
CHI	Calinski-Harabasz index
CN	Cognitive normal
CS	Completeness score
DBI	Davies-Bouldin index
DISN	Disinhibition

DTW	Dynamic Time Warping
DYSPH	Dysphoria
DX	Diagnosis
EHR	Electronic health record
EMHRs	Electronic mental health records
GDS	Geriatric Depression Scale
HS	Homogeneity score
IRR	Irritability
ISTAART	International Society to Advance Alzheimer's Research and Treatment
MBI	Mild Behavioral Impairment
MCI	mild cognitive impairment
MMSE	Mini-Mental State Examination
NMI	Normalized Mutual Information
NPS	Neuropsychiatric symptoms
OASIS-3	Open Access Series of Imaging Studies 3
SD	Standard deviation
U	Uncertain diagnosis
UDS	National Alzheimer's Coordinating Center Uniform Data Set

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13040-025-00441-0.

Supplementary Material 1.

Acknowledgements
This study work was funded by the European Union - Next Generation EU programme, in the context of The National Recovery and Resilience Plan,Investment Partenariato Esteso PE8 "Conseguenze e sfide dell'invecchiamento", Project Age-It (Ageing Well in an Ageing Society). This work was also partially supported by Ministero dell'Università e della Ricerca of Italy under the "Dipartimenti di Eccellenza 2023-2027" ReGAInS grant assigned to Dipartimentodi Informatica Sistemistica e Comunicazione at Università di Milano-Bicocca. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Authors' contributions
All authors conceived the proposed study and reviewed the manuscript. P.R. designed the algorithm, preprocessed data, performed the tests, supervised the study, and contributed to the article's writing. C.D.N. supervised the study, preprocessed data, and contributed to the article's writing. G.P. worked on the algorithm, preprocessed data, and contributed to the article's writing. D.C. collected the dataset, supervised the study, and contributed to the article's writing. F.C. supervised the study, performed the literature review, and contributed to the article's writing.

Funding
This study work was funded by the European Union - Next Generation EU programme, in the context of The National Recovery and Resilience Plan, Investment Partenariato Esteso PE8 "Conseguenze e sfide dell'invecchiamento", Project Age-It (Ageing Well in an Ageing Society) (PE0000015). National Recovery and Resilience Plan (NRRP) -- PE8 -- Mission 4, C2, Intervention 1.3, CUP B83C22004880006. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Data availability
Due to OASIS-3's data policy, unfortunately we are not authorized to release the OASIS-3 dataset we analyze in this study. Access to this dataset can be requested at: https://sites.wustl.edu/oasisbrains/home/access/.

Declarations

Competing interests
The authors declare no competing interests.

Received: 14 December 2024 Accepted: 17 March 2025
Published online: 28 March 2025

References

1. Association Alzheimer's. 2015 Alzheimer's disease facts and figures. Alzheimers Dement. 2015;11(3):332–84. https://doi.org/10.1016/j.jalz.2015.02.003.
2. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):280–92.
3. Collaborators GDF. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health. 2022;7(2):e105–25.

4. OASIS. Open Access Series of Imaging Studies (OASIS). 2024. https://sites.wustl.edu/oasisbrains/. Accessed 5 Dec 2024.
5. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci. 2007;19(9):1498–507. https://doi.org/10.1162/jocn.2007.19.9.1498.
6. Marcus DS, Fotenos AF, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. J Cogn Neurosci. 2010;22(12):2677–84.
7. LaMontagne PJ, Benzinger TL, Morris JC, Keefe S, Hornbeck R, Xiong C, et al. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medRxiv. 2019;(2019-12):1–37.
8. Koenig LN, Day GS, Salter A, Keefe S, Marple LM, Long J, et al. Select Atrophied Regions in Alzheimer disease (SARA): an improved volumetric model for identifying Alzheimer disease dementia. Neuroimage Clin. 2020;26:102248. https://doi.org/10.1016/j.nicl.2020.102248.
9. Cheng G, Huang C, Deng H, Wang H. Diabetes as a risk factor for dementia and mild cognitive impairment: a meta-analysis of longitudinal studies. Intern Med J. 2012;42(5):484–91. https://doi.org/10.1111/j.1445-5994.2012.02758.x.
10. Biggins CA, Boyd JL, Harrop FM, Madeley P, Mindham R, Randall JI, et al. A controlled, longitudinal study of dementia in Parkinson's disease. J Neurol Neurosurg Psychiatry. 1992;55(7):566–71. https://doi.org/10.1136/jnnp.55.7.566.
11. Asher S, Stephen R, Mäntylä P, Suominen AL, Solomon A. Periodontal health, cognitive decline, and dementia: a systematic review and meta-analysis of longitudinal studies. J Am Geriatr Soc. 2022;70(9):2695–709. https://doi.org/10.1111/jgs.17978.
12. Malik I, Iqbal A, Gu YH, Al-Antari MA. Deep Learning for Alzheimer's Disease Prediction: a Comprehensive Review. Diagnostics. 2024;14(12):1281. https://doi.org/10.3390/diagnostics14121281.
13. Nazir A, Assad A, Hussain A, Singh M. Alzheimer's disease diagnosis using deep learning techniques: datasets, challenges, research gaps and future directions. Int J Syst Assur Eng Manag. 2024;1–35. https://doi.org/10.1007/s13198-024-02441-5.
14. Wang Y, Liu S, Spiteri AG, Huynh ALH, Chu C, Masters CL, et al. Understanding machine learning applications in dementia research and clinical practice: a review for biomedical scientists and clinicians. Alzheimers Res Ther. 2024;16(1):175. https://doi.org/10.1186/s13195-024-01540-6.
15. Odusami M, Maskeliūnas R, Damaševičius R, Misra S. Machine learning with multimodal neuroimaging data to classify stages of Alzheimer's disease: a systematic review and meta-analysis. Cogn Neurodyn. 2024;18(3):775–94. https://doi.org/10.1007/s11571-023-09993-5.
16. Rudroff T, Rainio O, Klén R. AI for the prediction of early stages of Alzheimer's disease from neuroimaging biomarkers–a narrative review of a growing field. Neurol Sci. 2024;1–11. https://doi.org/10.1007/s10072-024-07649-8.
17. Veneziani I, Marra A, Formica C, Grimaldi A, Marino S, Quartarone A, et al. Applications of artificial intelligence in the neuropsychological assessment of dementia: A systematic review. J Personalized Med. 2024;14(1):113. https://doi.org/10.3390/jpm14010113.
18. Bazarbekov I, Razaque A, Ipalakova M, Yoo J, Assipova Z, Almisreb A. A review of artificial intelligence methods for Alzheimer's disease diagnosis: Insights from neuroimaging to sensor data analysis. Biomed Signal Process Control. 2024;92:106023. https://doi.org/10.1016/j.bspc.2024.106023.
19. Viswan V, Shaffi N, Mahmud M, Subramanian K, Hajamohideen F. Explainable artificial intelligence in Alzheimer's disease classification: a systematic review. Cogn Comput. 2024;16(1):1–44. https://doi.org/10.1007/s12559-023-10192-x.
20. Vimbi V, Shaffi N, Mahmud M. Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. Brain Inform. 2024;11(1):10.
21. Genolini C, Falissard B. KmL: k-means for longitudinal data. Comput Stat. 2010;25(2):317–28.
22. Genolini C, Alacoque X, Sentenac M, Arnaud C. kml and kml3d: R packages to cluster longitudinal data. J Stat Softw. 2015;65:1–34. https://doi.org/10.18637/jss.v065.i04.
23. Ostrouchov G, Gagnon D, Gerlovin H, Wei-Chen C, Drew S. clustra: Clustering Longitudinal Trajectories. 2023. https://doi.org/10.32614/CRAN.package.clustra. Accessed 20 Sep 2023.
24. Leffondré K, Abrahamowicz M, Regeasse A, Hawker GA, Badley EM, McCusker J, et al. Statistical measures were proposed for identifying longitudinal patterns of change in quantitative health indicators. J Clin Epidemiol. 2004;57(10):1049–62. https://doi.org/10.1016/j.jclinepi.2004.02.012.
25. Zhou J, Zhang Y, Tu W. clustermld: an efficient hierarchical clustering method for multivariate longitudinal data. J Comput Graph Stat. 2023;32(3):1131–44. https://doi.org/10.1080/10618600.2022.2149540.
26. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. J Stat Softw. 2017;78(2):1–56. https://doi.org/10.18637/jss.v078.i02.
27. Komárek A, Komárková L. Capabilities of R package mixAK for clustering based on multivariate continuous and discrete longitudinal data. J Stat Softw. 2014;59:1–38. https://doi.org/10.18637/jss.v059.i12.
28. Nagin DS, Jones BL, Passos VL, Tremblay RE. Group-based multi-trajectory modeling. Stat Methods Med Res. 2018;27(7):2015–23. https://doi.org/10.1177/0962280216673085.
29. Den Teuling NG, Pauws SC, van den Heuvel ER. A comparison of methods for clustering longitudinal data with slowly changing trends. Commun Stat Simul Comput. 2023;52(3):621–48. https://doi.org/10.1080/03610918.2020.1861464.
30. Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. J Am Stat Assoc. 1996;91(433):217–21.
31. Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: a comparison between methods. J Clin Epidemiol. 2012;65(10):1078–87.
32. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Ann Rev Clin Psychol. 2010;6(1):109–38.
33. Dziak JJ, Li R, Tan X, Shiffman S, Shiyko MP. Modeling intensive longitudinal data with mixtures of nonparametric trajectories and time-varying effects. Psychol Methods. 2015;20(4):444.
34. Bruckers L, Molenberghs G, Drinkenburg P, Geys H. A clustering algorithm for multivariate longitudinal data. J Biopharm Stat. 2016;26(4):725–41.

35. Huang ST, Hsiao FY, Tsai TH, Chen PJ, Peng LN, Chen LK. Using hypothesis-led machine learning and hierarchical cluster analysis to identify disease pathways prior to dementia: longitudinal cohort study. J Med Internet Res. 2023;25:e41858. https://doi.org/10.2196/41858.
36. Zhu CW, Lee S, Ornstein KA, Cosentino S, Gu Y, Andrews H, et al. Dependence clusters in Alzheimer disease and medicare expenditures. Alzheimer Dis Assoc Disord. 2020;34(4):293–8. https://doi.org/10.1097/WAD.0000000000000402.
37. Corneli M, Erosheva E, Qian X, Lorenzi M, Alzheimer's Disease Neuroimaging Initiative. A Bayesian approach for clustering and exact finite-sample model selection in longitudinal data mixtures. Comput Stat. 2024;1–37. https://doi.org/10.1007/s00180-024-01501-5.
38. Lorenzon G, Poulakis K, Mohanty R, Kivipelto M, Eriksdotter M, Ferreira D, et al. Frontoparietal atrophy trajectories in cognitively unimpaired elderly individuals using longitudinal Bayesian clustering. Comput Biol Med. 2024;182:109190. https://doi.org/10.1016/j.compbiomed.2024.109190.
39. Poulakis K, Ferreira D, Pereira JB, Smedby O, Vemuri P, Westman E. Fully Bayesian longitudinal unsupervised learning for the assessment and visualization of AD heterogeneity and progression. Aging. 2020;12(13):12622–47.
40. Poulakis K, Pereira JB, Muehlboeck JS, Wahlund LO, Smedby O, Volpe G, et al. Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease. Nat Commun. 2022;13(1):4566. https://doi.org/10.1038/s41467-022-32202-6.
41. Elazab A, Wang C, Abdelaziz M, Zhang J, Gu J, Gorriz JM, et al. Alzheimer's disease diagnosis from single and multimodal data using machine and deep learning models: achievements and future directions. Expert Syst Appl. 2024;124780. https://doi.org/10.1016/j.eswa.2024.124780.
42. Likas A, Vlassis N, Verbeek JJ. The global *k*-means clustering algorithm. Pattern Recogn. 2003;36(2):451–61.
43. Miller RH, West C, Brown TM, Sim I, Ganchoff C. The Value Of Electronic Health Records In Solo Or Small Group Practices. Health Aff. 2005;24(5):1127–37. https://doi.org/10.1377/hlthaff.24.5.1127.
44. Jin P, Xu J, Liao Z, Zhang Y, Wang Y, Sun W, et al. A review of current evidence for mild behavioral impairment as an early potential novel marker of Alzheimer's disease. Front Psychiatr. 2023;14:1099333.
45. Tavenard R, Faouzi J, Vandewiele G, Divo F, Androz G, Holtz C, et al. Tslearn, a machine learning toolkit for time series data. J Mach Learn Res. 2020;21(118):1–6.
46. McLachlan, Geoffrey J, Thriyambakam Krishnan. The EM algorithm and extensions. John Wiley & Sons, 2008.
47. Cuturi M, Blondel M. Soft-dtw: a differentiable loss function for time-series. In: International Conference on Machine Learning. PMLR; 2017. pp. 894–903.
48. Morris JC, Weintraub S, Chui HC, Cummings J, DeCarli C, Ferris S, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. Alzheimer Dis Assoc Disord. 2006;20(4):210–6.
49. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int Psychogeriatr. 1997;9(S1):173–6.
50. O'Bryant SE, Lacritz LH, Hall J, Waring SC, Chan W, Khodr ZG, et al. Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. Arch Neurol. 2010;67(6):746–9.
51. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189–98. https://doi.org/10.1016/0022-3956(75)90026-6.
52. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65.
53. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. 1979;PAMI-1(2):224–7.
54. Calinski T, Harabasz J. A dendrite method for cluster analysis. Commun Stat Theory Methods. 1974;3(1):1–27.
55. Chi-Square Test. New York: Springer New York; 2008. pp. 77–79. https://doi.org/10.1007/978-0-387-32833-1_57.
56. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 1952;47(260):583–621.
57. Tukey JW. Comparing individual means in the analysis of variance. Biometrics. 1949;5(2):99–114.
58. Lyketsos CG, Carrillo MC, Ryan JM, Khachaturian AS, Trzepacz P, Amatniek J, et al. Neuropsychiatric symptoms in Alzheimer's disease. Alzheimers Dement. 2011;7(5):532–9.
59. Ismail Z, Smith EE, Geda Y, Sultzer D, Brodaty H, Smith G, et al. Neuropsychiatric symptoms as early manifestations of emergent dementia: provisional diagnostic criteria for mild behavioral impairment. Alzheimers Dement. 2016;12(2):195–202. https://doi.org/10.1016/j.jalz.2015.05.017.
60. Jiang F, Cheng C, Huang J, Chen Q, Le W. Mild behavioral impairment: an early sign and predictor of Alzheimer's disease dementia. Curr Alzheimer Res. 2022;19(6):407–19.
61. Han J, Kamber M, Pei J. 10 - Cluster Analysis: Basic Concepts and Methods. In: Han J, Kamber M, Pei J, editors. Data Mining (Third Edition). 3rd edn. The Morgan Kaufmann Series in Data Management Systems. Boston: Morgan Kaufmann; 2012. pp. 443–95. https://doi.org/10.1016/B978-0-12-381479-1.00010-1.
62. Tavenard R. Differentiability of DTW and the case of soft-DTW. 2021. https://rtavenar.github.io/blog/softdtw.html. Last access 21 Mar 2025.

Publisher's Note