

# Multilayer clustering: Biomarker driven segmentation of Alzheimer's disease patient population

**Dragan Gamberger**

Rudjer Bošković Institute, Zagreb, Croatia

**Bernard Ženko**

Jožef Stefan Institute, Ljubljana, Slovenia

**Alexis Mitelpunkt**

Tel Aviv University, Israel

**Nada Lavrač**

Jožef Stefan Institute, Ljubljana, Slovenia

University of Nova Gorica, Slovenia

## Abstract

Identification of biomarkers for the Alzheimer's disease is a challenge and a very difficult task both for medical research and data analysis. In this work we present results obtained by application of a novel clustering tool. The goal is to identify subpopulations of the Alzheimer's disease (AD) patients that are homogeneous in respect of available clinical and biological descriptors. The result presents a segmentation of the Alzheimer's disease patient population and it may be expected that within each subpopulation separately it will be easier to identify connections between clinical and biological descriptors. Through the evaluation of the obtained clusters with AD subpopulations it has been noticed that for two of them relevant biological measurements (whole brain volume and intracerebral volume) change in opposite directions. If this observation is actually true it would mean that the diagnosed severe dementia problems are results of different physiological processes. The observation may have substantial consequences for medical research and clinical trial design. The used clustering methodology may be interesting also for other medical and biological domains.

## 1 Introduction

Identification of connections between biological and clinical characteristics of Alzheimer's disease patients is a long term goal that could significantly improve the understanding of the Alzheimer's disease (AD) pathophysiology, improve

clinical trial design, and help in predicting outcomes of mild cognitive impairment [1]. The difficulty of the task is in the fact that AD is clinically described as a set of signs and symptoms that can be only indirectly measured and that have been integrated into various scoring systems like Clinical Dementia Rating Sum of Boxes, Alzheimer’s Disease Assessment Scale, and Montreal Cognitive Assessment [2]. All of these scales as well as everyday cognition problems have proved their usefulness in the diagnostic process but a unique reliable measure does not exist. On the other side, although relations between some biological descriptors and AD diagnosis have been undoubtedly demonstrated [3, 4], currently available biological descriptors are non-specific (e.g., whole brain or hippocampal volume) and their changes may be a consequence of various physiological processes. It means that potentially useful information related to biological causes of the cognitive status of a patient is hidden in the large “noise” of interfering biological processes.

Technically speaking, we are looking for relevant relations in a very noisy data domain (biological descriptors) in which the target function is defined by a large set of imprecise values (clinical descriptors). A simplified approach in which medical AD diagnosis is used as the target function has enabled the detection of some relations, like importance of decreased FDG-PET values for the AD diagnosis, but all the detected relations including those obtained by complex supervised approaches [5] have low predictive quality and did not help significantly in expert understanding of the disease. In line with the approach proposed in [6], our work aims at finding homogeneous subpopulations of AD patients in which it will be easier to identify statistically and logically relevant relations between clinical and biological descriptors. The approach is based on finding subpopulations with clustering algorithms.

Clustering is a well-established machine learning methodology but it still suffers from problems such as definition of the distance measure and optimal selection of the number of resulting clusters. Typically the obtained clustering results are unstable because they significantly depend on user selectable parameters. This is especially true for noisy domains and domains with statistically related descriptors (attributes). Recently we have developed a novel clustering approach called *multilayer clustering* that successfully solves some of the basic problems of data clustering [7]. In this methodology, the quality of the resulting clusters is ensured by the constraint that clusters must be homogeneous at the same time in two or more data layers, i.e., two or more sets of distinct data descriptors. By defining the clinical descriptors as one data layer and biological descriptors as the other layer, we can expect not only more reliable clusters but clusters which will be potentially good candidates for the detection of relevant relations between the clinical and biological descriptors. The AD domain fulfills all the requirements of the ideal setting in which multilayer clustering may demonstrate its advantages.

The rest of the paper is structured as follows. Section 2 presents the multilayer clustering methodology, and Section 3 presents the concrete results for the AD domain. Medical meaning and the results significance are analysed in Section 4.

## 2 Multilayer clustering

Clustering is an optimisation task which tries to construct subpopulations of instances so that distances between instances within each subpopulation are small while distances between instances in different subpopulations are as large as possible [8]. The most commonly used distance measure is the Euclidean distance that is well defined for numerical attributes.

Redescription mining is a novel clustering approach in which the quality of the results is ensured by the constraint that the resulting clusters must have meaningful interpretations in at least two independent attribute layers [9]. It is possible, and it occurs often, that some of the training instances remain outside the identified clusters but the detected clusters are more likely really relevant. A very important property of this approach is that the constructed clusters have human interpretable descriptions in all attribute layers.

However, redescription mining has some issues as well. For the approach to be applied, both numerical and nominal attributes have to be transformed into a transactional form [10] or some on-the-fly approaches have to be implemented [11]. Also, selection of the appropriate minimal necessary support level is not a trivial task. Low values may result in unacceptably long execution times of the algorithms and unreliability of the results, while too high values may prevent detection of any useful clusters [12]. An even more serious problem is that in all real life domains some level of attribute noise can be expected. In such cases an error in a single attribute value may prevent the identification of correct descriptions. Such an error does not only cause that the erroneous example is not detected as a member of a cluster, but it causes the descriptions in different attribute layers not to cover all the subsets of examples it would have covered otherwise. As a result, some of the actual clusters may not be detected.

In this work we use an approach to more reliable clustering that reuses the basic idea of redescription mining in a novel setting, proposed in [7]. The first step is to determine the similarity of instances in each attribute layer independently and then to search for clusters that satisfy similarity conditions for *all* layers. The main characteristic of the approach is that the resulting clusters are small but very coherent.

### 2.1 Single layer algorithm

Let us assume a basic clustering task in which we have only one layer of attributes. The clustering approach consists of two steps. In the first step we estimate pair-wise similarity between all examples in the training data set. In the second step we use this similarity estimation in order to construct clusters.

#### **Similarity estimation.**

In the first step we compute the so called *example similarity table (EST)*. It is an  $N \times N$  symmetric matrix, where  $N$  is the number of examples. All its

values are in the  $[0, 1]$  range, where large values denote large similarity between examples.

We start from the original set of  $N$  examples represented by nominal and numerical attributes that may contain unknown values. We define an “artificial” binary classification problem on a data set constructed as follows. The first part consists of examples from the original set, these examples are labelled as positive. The second part consists of the same number of examples which are generated by randomly shuffling attribute values of original examples (within each attribute separately), these examples are labelled as negative.

Next, we use supervised machine learning to build classifiers for discrimination between positive cases (original examples) and negative cases (examples with shuffled attribute values). The goal of learning are not the predictive models themselves, but the resulting information on the similarity of the original (positive) examples. Machine learning approaches in which we can determine if some examples are classified “in the same way” (meaning they are somehow similar) are appropriate for this task. For example, in decision tree learning this means that examples fall in the same leaf node, while in covering rule set induction this means that examples are covered by the same rule. In order to statistically estimate the similarity of the examples, it is necessary to use a sufficiently large number of classifiers. Additionally, a necessary condition for a good result is that the classifiers are as diverse as possible and that each of them is better than random. All these conditions are satisfied by Random Forest [13] and Random Rules algorithms [14]. Here we use the later one with which we construct a large number of rules (100,000) for each EST computation.

Finally, the similarity of examples is estimated so that for each pair of examples we count how many rules cover both examples. The example similarity table presents the statistics for positive examples only. A pair of similar examples will be covered by many rules, while no rules or a very small number of rules will cover pairs that are very different in respect of their attribute values. The final EST values are normalised with the largest detected count value.

### Clustering with the CRV score.

In the second step we use the EST values to perform a bottom-up clustering. The agglomeration of clusters is guided by the *Clustering Related Variability (CRV)* score [7]. The score measures the variability of the EST similarity values in a cluster of examples with respect to all other clusters. It is defined as follows.

Let  $x_{ij}$  be the similarity between examples  $i$  and  $j$  from the EST matrix. The CRV score of a single example  $i$  from cluster  $C$  is the sum of within cluster and outside of cluster components:  $CRV(i) = CRV_{wc}(i) + CRV_{oc}(i)$ ,  $i \in C$ . The two components are sums of squared deviations from the mean value within (or outside of) cluster  $C$ :  $CRV_{wc}(i) = \sum_{j \in C} (x_{ij} - \bar{x}_{i,wc})^2$  and  $CRV_{oc}(i) = \sum_{j \notin C} (x_{ij} - \bar{x}_{i,oc})^2$ . Finally, the *CRV* score of cluster  $C$  is the mean value of  $CRV(i)$  values of all examples in the cluster:  $CRV(C) = \sum_{i \in C} CRV(i) / |C|$ .

The clustering algorithm starts with each example being in a separate cluster and then iteratively tries to merge clusters together. In each iteration for each

possible pair of clusters we compute the potential variability reduction that can be obtained by merging the clusters. The variability reduction of joining clusters  $C_1$  and  $C_2$  is computed as:  $\text{DIFF}(C_1, C_2) = (\text{CRV}(C_1) + \text{CRV}(C_2))/2 - \text{CRV}(C_2 \cup C_1)$ . The pair of clusters with the largest variability reduction is then merged into a single cluster. The iterative process repeats until no pair of clusters exists for which the variability reduction is positive. A more detailed description of the algorithm including some examples can be found in [7].

The algorithm produces a hierarchy of clusters and, in contrast to most other clustering algorithms, it has a very well defined stopping criterion. The algorithm stops when additional merging of clusters cannot further reduce the variability, measured by the CRV score. This means that the algorithm automatically determines the optimal number of clusters and that some examples may stay unclustered, i.e., some clusters may only include a single example.

## 2.2 Multilayer algorithm

The single layer approach for clustering presented in the previous section can be easily extended to clustering in multi-layer domains. For each attribute layer we compute the example similarity table independently. Regardless of the number and type of attributes in different layers, the EST tables will always be  $N \times N$  matrices, because the number of examples in all layers is the same.

After having all the EST tables, we proceed with the clustering. The clustering procedure for multiple layers is basically the same as for a single layer, except that we merge two clusters only if a variability reduction exists in all layers. For each possible pair of clusters we compute potential variability reduction for all attribute layers. Then we find the minimal variability reduction in all layers, and merge the pair of clusters for which this value is largest. As previously, we only merge two clusters if the (minimal) variability reduction is positive.

## 3 Data and results

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database. ADNI is a long term project aimed at the identification of biomarkers of the disease and understanding of the related pathophysiology processes. The project collects a broad range of clinical and biological data about patients with different cognitive impairment.<sup>1</sup> In our work we started from a set of numerical descriptors extracted from the AD-NIMERGE table, the joined dataset from several ADNI data tables. We have used baseline evaluation data for 916 patients in total with 5 different medical diagnoses: cognitive normal CN (187 patients), significant memory concern

---

<sup>1</sup>The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations (<http://www.adni-info.org> and <http://adni.loni.usc.edu>).

Table 1: Properties of five largest clusters. The clusters are ordered by the decreasing median value of the CDRSB score for patients included in the clusters.

Cluster	Number of patients	Distribution of diagnoses					CDRSB
		AD	LMCI	EMCI	SMC	CN	
A	42	41	1	-	-	-	5.5
B	19	12	4	1	-	2	4.0
C	20	14	2	-	-	4	3.75
D	34	10	6	5	5	8	2.5
E	27	6	3	1	6	11	0

SMC (106), early mild cognitive impairment EMCI (311), late mild cognitive impairment LMCI (164), and Alzheimer’s disease AD (148). The patients are described by a total of 10 biological and 23 clinical descriptors. Biological descriptors are genetic variations of APOE4 related gene, PET imaging results FDG-PET and AV45, and MRI volumetric data of: Ventricles, Hippocampus, WholeBrain, Entorhinal, Fusiform gyrus, Middle temporal gyrus (MidTemp), and intracerebral volume (ICV). Clinical descriptors are: Clinical Dementia Rating Sum of Boxes (CDRSB), Alzheimer’s Disease Assessment Scale (ADAS13), Mini Mental State Examination (MMSE), Rey Auditory Verbal Learning Test (RAVLT immediate, learning, forgetting, percentage of forgetting), Functional Assessment Questionnaire (FAQ), Montreal Cognitive Assessment (MOCA) and Everyday Cognition which are cognitive functions questionnaire filled by the patient (ECogPt) and the patient study partner (ECogSP) (Memory, Language, Visuospatial Abilities, Planning, Organization, Divided Attention, and Total score).

The clustering process started from one table with biological data consisting of 916 rows and 10 columns and one table with clinical data consisting of 916 rows and 23 columns. Each of the two tables represented one attribute layer. The information about medical diagnoses of the patients have not been included into the tables with the intention to use it only for the evaluation of the clustering results. The goal of the clustering process has been to identify as large as possible groups of patients that are similar according to both layers, i.e., biological and clinical characteristics.

The result is a set of five clusters. The largest among them includes 42 patients and the smallest only 19 patients. Table 1 presents the distribution of medical diagnoses for patients included into the clusters. We notice that the largest cluster A is very homogeneous. It has 42 patients and 41 of them have diagnosis AD while only one has diagnosis LMCI. Least homogeneous clusters are D and E.

We have identified the CDRSB score as the clinical characteristic that best discriminates between the constructed clusters. The rightmost column in Table 1 presents median CDRSB values for patients included in each cluster. Val-

Table 2: Median values and standard deviations of biological descriptors discriminating clusters A and C. Distinguishing values are typeset in bold. Actual values for ICV, Whole brain, Fusiform, and MidTemp values are 1,000 times larger than presented in the table.

Cluster	FDG-PET	AV45	ICV	Whole Brain	Fusiform	MidTemp
A	<b>4.37 / 0.52</b>	<b>1.45 / 0.13</b>	1404 / 193	918 / 121	<b>15.4 / 2.7</b>	<b>15.6 / 3.2</b>
B	5.67 / 0.67	1.43 / 0.23	1372 / 149	935 / 84	16.0 / 2.7	17.8 / 2.5
C	5.55 / 1.06	1.35 / 0.25	<b>1634 / 126</b>	<b>1107 / 72</b>	18.2 / <b>2.0</b>	19.4 / 2.7
D	6.28 / 0.85	1.20 / 0.26	1453 / 230	1005 / 144	17.2 / 2.4	19.6 / 2.4
E	6.45 / 0.70	1.19 / 0.25	1445 / 174	1022 / 93	18.0 / 3.	19.6 / 2.5
AD	5.36 / 0.75	1.42 / 0.21	1490 / 175	986 / 115	16.4 / 2.5	17.6 / 3.2
CN	6.57 / 0.54	1.05 / 0.17	1483 / 155	1051 / 101	18.4 / 2.3	20.5 / 2.4

ues above 3.0 demonstrate that patients in clusters A–C have problems with severe dementia. Patients in cluster D typically have moderate to severe dementia, while the majority of patients in cluster E do not have problems with dementia.

### 3.1 Distinguishing properties of patients in clusters A and C

For each cluster we have computed median values and standard deviations of all biological and clinical descriptors. The intention is to identify distinguishing properties of patients included in the clusters. It is worth focusing on clusters that have either extremely high or extremely low median values or very low standard deviation of some descriptor. In the former case the patients in the cluster have this property (significantly) increased or decreased, while in the later case most of the patients have very similar values of the descriptor. We can also find a distinguishing descriptor that has both low standard deviation and extreme median value. In all such situations we interpret the descriptor as a distinguishing property.

Clusters A, B, and C are especially interesting because most of the included patients have diagnosis AD and these clusters may be regarded as relatively homogeneous subsets of the AD patient population. The difference between clusters A and C turned out to be especially intriguing. Table 2 presents the values of biological descriptors for which patients in clusters A and C have distinguishing values. Presented are median values and standard deviations for all five clusters as well as for the complete AD population consisting of 148 patients and the complete cognitive normal (CN) population of 187 patients.

These results demonstrate that the patients in cluster A have extremely low values of the FDG-PET descriptor. The median for this cluster is 4.37 while median for the complete AD population is 5.36 and the median for cognitive normal patients is 6.57. Surprisingly, we can notice that the difference between the median of AD patients in cluster A and the median of all AD patients is

almost as large as the difference between the median of all AD patients and the median of CN patients. Additionally, it must be noted that the patients in cluster A have very small standard deviation for FDG-PET meaning that the consistency of these small values is high. A similar pattern can be observed for the AV45 descriptor. When compared with other clusters and with the whole AD and CN populations, patients in cluster A have the largest median value 1.45 and the smallest standard deviation of 0.13 . We can conclude that cluster A is characterized by outstandingly low FDG-PET values and outstandingly high AV45 values.

In contrast to cluster A, cluster C is characterized by outstandingly high values for ICV and Whole brain descriptors, which are typeset in bold in the fourth and fifth column of Table 2, respectively. There is a substantial difference between biological descriptors that characterize clusters A and C. For cluster A the extreme values of FDG-PET and AV45 descriptors are relevant, while for cluster C the median values of these descriptors change (compared to CN patients) in the same direction but with less intensity. The situation with ICV and Whole brain descriptors is very different. Median values of ICV for CN and all AD patients are very similar: 1483 and 1490, respectively. Patients in cluster C have a high median value of 1634, while patients in cluster A have a low median value of 1404 that is lower than the CN median value. A similar situation is with the Whole brain descriptor. CN patients have a median value of 1051, patients in cluster C have an increased value of 1107, while patients in cluster A have a very low median value equal to 918. Additionally, patients in cluster C have a low standard deviation of ICV and Whole brain descriptors.

Differences between clusters A and C can also be seen in some other biological descriptors. The rightmost two columns in Table 2 demonstrate that patients in cluster A have very low Fusiform and MidTemp volumes. For patients in cluster C these values are almost normal or only slightly decreased.

A good property of the multilayer clustering is that similarity of patients in both attribute layers is needed if they are to be included in the same cluster. Table 3 presents clinical descriptors with distinguishing values for patients in cluster A. CDRSB is a distinguishing descriptor but it is not included in Table 3 because we have already demonstrated in Table 1 that it has very high values for cluster A. The median value for the complete AD population for CDRSB descriptor is 4.5 while the median value for cluster A is 5.5 with a low standard deviation. Clinical descriptors with distinguishing values for patients in cluster C are presented in Table 4.

## 4 Discussion

Very elaborate medical studies have recently shown that mild and severe cognitive impairment as well as AD diagnosis are correlated with some measurable changes in the human brain. Our results are completely in agreement with these results. Firstly, it has been shown that the progressive reduction in fluorodeoxyglucose positron emission tomography (FDG-PET) measurement of the

Table 3: Median values and standard deviations of clinical descriptors characteristic for patients in cluster A. Distinguishing values are typeset in bold.

Cluster	ADAS13	MOCA	Functional Assessment	RAVLT Perc. Forgetting	Ecog SP Organization
A	<b>35</b> / 8.5	<b>14.5</b> / 4.9	<b>16</b> / 6.8	100 / <b>10</b>	<b>3.17</b> / <b>0.76</b>
B	28 / 11.1	17 / 5.2	11 / 8.3	100 / 26	2.83 / 1.07
C	28 / 10.7	19 / 5.0	12 / 9.0	100 / 35	3.08 / 1.07
D	19.5 / 11.8	21.5 / 5.3	6.5 / 7.0	63 / 37	1.83 / 1.03
E	10 / 8.8	25 / 3.6	0 / 4.8	42 / 35	1.00 / 0.94
AD	31 / 8.4	18 / 4.5	13 / 7.1	100 / 20	2.83 / 0.86
CN	9 / 4.5	26 / 2.4	0 / 0.6	31 / 27	1.00 / 0.42

Table 4: Median values and standard deviations of clinical descriptors characteristic for patients in cluster C. Distinguishing values are typeset in bold.

Cluster	Ecog Pt Memory	Ecog Pt Organization	Ecog SP Divided Attention
A	2.12 / 0.80	1.45 / 0.72	3.00 / 0.74
B	2.38 / 0.89	1.67 / 0.55	3.25 / 0.96
C	<b>2.62</b> / 0.85	<b>1.73</b> / 0.78	<b>3.50</b> / 1.06
D	1.88 / 0.61	1.33 / 0.57	2.00 / 1.08
E	1.75 / 0.65	1.00 / 0.63	1.00 / 0.95
AD	2.38 / 0.75	1.50 / 0.74	3.25 / 0.93
CN	1.50 / 0.44	1.00 / 0.38	1.00 / 0.48

regional cerebral metabolic rate for glucose is related with cognitive impairment [3]. This follows from a series of increasing values in the second column of Table 2 and how it nicely correlates with the decreasing values of the CDRSB score (last column of Table 1) practically for all clusters A-E. Also, there is a recent result presented in [4] which demonstrates that the increased values of Florbetapir F18-AV-45 PET (AV45) are positively correlated with dementia. This effect can be noticed in the third column of Table 2 where we have constantly increasing values for clusters A-E in the order of their cognitive impairment severity.

There is a statistical analysis of ADNI data which demonstrated that whole brain atrophy and ventricular enlargement differ between cognitive impaired patients and healthy controls [15]. The result is very interesting because it is based on the same data source as our analysis. Our analysis detected the same relation for clusters A, B, D, and E. So it is not surprising that in [15] this property has been detected for the complete AD population. In our results the

lowest median value for the whole brain volume is for cluster B (1372) but it is also very low for cluster A (1404). The real novelty of our result is cluster C in which the corresponding median value of 1107 is *higher than in all other clusters and also higher than in the complete cognitive normal population* with median value of 1051.

This result is important because it potentially suggests that we have two different AD related physiological processes. One, that results in significantly decreased IC and whole brain volumes and is characteristic for the majority of AD patients, and second, that results in increased IC and whole brain volumes and appears in only 10-15% of cases. If this hypothesis is actually true, it will have substantial consequences for medical research and treatment of Alzheimer’s disease. Namely, it is very likely that each of the two physiological processes will require different research and treatment procedures.

An additional result of our analysis is an observation that decreased FDG-PET values, increased AV45 values and decreased whole brain volume are especially characteristic for cluster A that is the largest and most homogeneous AD subpopulation. So it is not a surprise that these properties have been previously recognized as relevant for the complete AD population [3, 4, 15]. But if these studies would have concentrated on the AD subpopulation in cluster A only, the statistical significance and usefulness of their results would most probably be even higher.

Especially interesting is the question whether different physiological development of the AD disease may affect the clinical status of the patient and the patient’s clinical prognosis. Figure 1 illustrates the differences between clusters A and C for ICV and ADAS13 values. Black circles denote patients in cluster A while empty squares are patients in cluster C. Signs + and – denote the position of median values for the whole AD and CN populations, respectively. The big black circle is the median value for cluster A patients and the big square is the median value for cluster C patients. Results presented in Table 3 suggest that patients characterized by *decreased ICV, fusiform, whole brain, and MidTemp results* (type A patients) have clinically very intensive general picture of AD identified by high values of CDRSB, ADAS13, MOCA, and Functional Assessment Questioner scores. In contrast, patients characterized by *increased ICV, almost normal fusiform, slightly increased whole brain, and slightly decreased MidTemp* (type C patients) have main problems with Everyday cognition, especially with Memory, Organization, and Divided Attention (see Table 4). This is potentially an interesting issue for further medical expert evaluation. A preliminary observation is that type C patients have self-awareness and insight to the condition. This pattern is not the usual course of classic AD where the level of insight decreases as the disease progresses.

Four squares at the bottom of Figure 1 denote four patients with cognitive normal diagnosis included into cluster C. This is a potentially interesting fact suggesting that the physiological process of type C (increased IC and whole brain volumes) may in a small number of patients, in spite of significant physiological changes, result in only slightly changed clinical status, which is diagnosed as cognitive normal. In contrast, the significant physiological changes result-

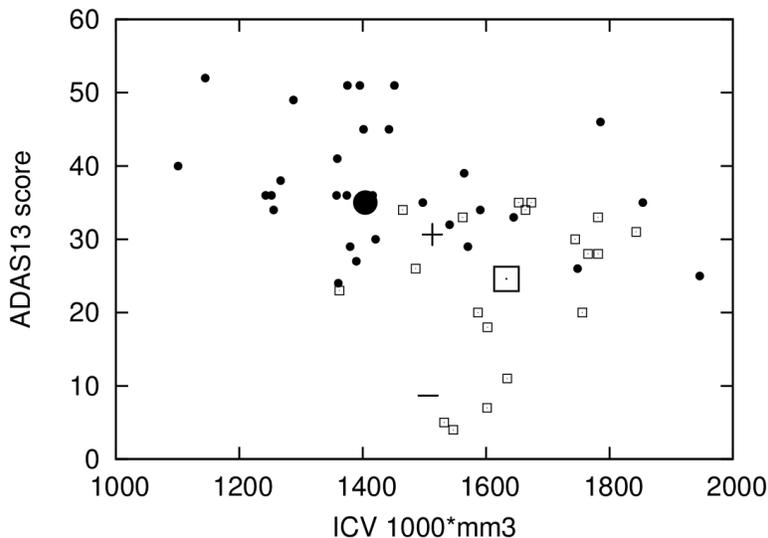


Figure 1: Alzheimer’s disease patients in cluster A (circles) and those in cluster C (squares) presented in the space defined by the ICV (Intracerebral Volume) and ADAS13 (Alzheimer’s Disease Assessment Scale).

ing from the process of type A (low FDG-PET, high AV45), according to the available data, practically always results in severe dementia diagnosed as AD.

## 5 Conclusions

This work presents the application of a novel clustering methodology to discover relations between two distinct sets of descriptors. The result is a set of small but homogeneous clusters of examples (patients). The main advantages of the methodology are that it may be successfully used on instances described by both numeric and nominal attributes and that it has a well-defined stopping criterion. Additionally, its unique property is that the quality of clusters is ensured by the requirement that examples must be similar in at least two different attribute layers. As a consequence the resulting clusters present the segmentation of the complete population where each subpopulation has some homogeneous properties in both layers. AD domain is a good example of a task in which relations between layers can be different or even contradictory for various subpopulations. In such cases it is very difficult to identify relevant relations on a complete population by classical statistical analysis and supervised machine learning approaches. But after successful segmentation even a simple analysis of median values and standard deviations may enable identification of relevant relations.

The most relevant problem of the methodology is that constructed clusters are small and that they tend to be even smaller if more than two layers are used. In the concrete AD domain we got some useful insight about less than a half of AD patients and practically no insight about cognitive normal patients and patients with mild impairment. Additionally, the methodology has high time complexity, which is growing quadratically with the number of examples. Because of this the methodology is currently applicable only to domains with an order of 1,000 examples.

The results of the analysis are potentially relevant medical hypotheses. It will be necessary to extend the clinical measurements and to further study the identified relations in order to confirm correctness of these hypotheses.

## Acknowledgement

We would like to acknowledge the support of the European Commission through the Human Brain Project (Grant number 604102) and MAESTRA project (Grant number 612944) as well as the support of the Croatian Science Foundation (under the project number 9623 Machine Learning Algorithms for Insightful Analysis of Complex Data Structures) and Slovenian Research Agency supported grants: program (Knowledge Technologies) and project (Development and Applications of New Semantic Data Mining Methods in Life Sciences).

## References

- [1] Weiner M. W., et al. (2013) The Alzheimer’s Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer’s & Dementia*, 9(5):e111-e194.
- [2] Smith G. E. and Bondi M. W. (2013) *Mild Cognitive Impairment and Dementia*. Oxford University Press.
- [3] Langbaum J. B., et al. (2009) Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the Alzheimer’s Disease Neuroimaging Initiative (ADNI). *Neuroimage*, 45(4):1107-1116.
- [4] Doraiswarny P. M., et al. (2014) Florbetapir F 18 amyloid PET and 36-month cognitive decline: a prospective multicenter study. *Molecular Psychiatry*, 19(9):1044-1051.
- [5] Hinrichs C., et al. (2011) Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *Neuroimage*, 55(2):574-589.
- [6] Galili T., et al. (2014) Categorize, Cluster, and Classify: a 3-C Strategy for Scientific Discovery in the Medical Informatics Platform of the Human

- Brain Project. In *Proc. of the 17th International Discovery Science Conference, DS 2014*, pp.73–86.
- [7] Gamberger, D., Mihelčič, M., Lavrač, N. (2014) Multilayer clustering: A discovery experiment on country level trading data. In *Proc. of the 17th International Discovery Science Conference, DS 2014*, pp. 87–98.
- [8] Gan, G., Ma, C., Wu, J. (2007) Data Clustering: Theory, Algorithms, and Applications. *Society for Industrial and Applied Mathematics*.
- [9] Parida L. and Ramakrishnan N. (2005) Redescription mining: Structure theory and algorithms. In *Proc. of the Association for the Advancement of Artificial Intelligence, AAAI '05*, pp. 837-844.
- [10] Ramakrishnan N., Kumar D., Mishra B., Potts M., Helm R. F. (2004) Turning cartwheels: an alternating algorithm for mining redescription. In *Proc. of the 10th ACM Intern. Conf. on Knowledge Discovery and Data Mining*, pp. 266-275.
- [11] Galbrun, E. and Miettinen, P. (2012) From black and white to full color: extending redescription mining outside the boolean world. *Statistical Analysis and Data Mining*, pp. 284-303.
- [12] Zaki M. J. and Ramakrishnan N. (2005) Reasoning about sets using redescription mining. In *Proc. of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, KDD '05*, pp. 364-373.
- [13] Breiman, L. (2001) Random forests. *Machine Learning* 45(1):5-32.
- [14] Pfahringer, B., Holmes, G., Wang, C. (2004) Millions of random rules. In *Proc. of the Workshop on Advances in Inductive Rule Learning, 15th European Conference on Machine Learning, ECML 2004*.
- [15] Evans M. C., etal. (2010) Volume changes in Alzheimer’s disease and mild cognitive impairment: cognitive associations. *European Radiology*, 20(3):674-680.