IDEA: Integrative Detection of Early-stage Alzheimer's disease

Wei Ye* Bianca Wackersreuther* Christian Böhm* Michael Ewers† Claudia Plant‡

Abstract

Data integration and selecting only the relevant information for solving biological and environmental problems is one of the most important challenges in today's data mining. One urgent problem in the medical community is to support the classification of dementia caused by Alzheimer's disease and even its detection in the predementia phase to optimize the medical treatment of a disease that accounts for 60 to 80 percent of dementia cases and affects more than 35 million people world-wide. In this paper we present IDEA, a fully automated, easy-to-use and clinically interpretable diagnostic software for early-stage Alzheimer's. The main contribution of our framework is that it allows for a combined analysis of various feature types such as neuroimaging data sourcing from different modalities, and non-image data that consist of numerical and categorical values, resulting in high classification accuracy results. Using advanced information theory, we select only subsets out of the rich pool of information that build high-predictive feature combinations. In an extensive medical case-study on a large real-world data set, we show that already small feature subsets are adequate to derive significant classification accuracies. And, as IDEA usually determines more than one suitable feature set, it even can be used for an optimized analysis process by selecting the assessment tools that produce minimal cost (in terms of money or stress for the patients) without loosing accuracy.

1 Introduction

Analyzing mixed-type attributes or also known as integrative data mining is among the top 10 challenging problems in data mining research identified in panel discussions [19] and position papers [25]. Moreover, it is essential for solving many of the other top 10 challenges, including data mining in social networks and data mining for biological and environmental problems. In this paper, we address the application of integrative data mining for the detection of early-stage patterns for Alzheimers's disease (AD) dementia, by a combined analysis of different medical imaging modalities together with multiple numerical and categorical attributes, resulting from neuropsychological tests or genetic and biochemical screenings.

AD is the most common form of dementia, that usually develops slowly and includes gradual onset of cognitive impairment in episodic memory and at least one other domain [16]. Although, there is currently no cure for Alzheimer's that stops the disease from progressing, medical treatment can temporarily slow down the worsening of dementia symptoms. However, the benefit of this treatment strongly correlates with a reliable early detection of AD in predementia stages such as mild cognitive impairment (MCI). But, cerebral or cognitive changes are only of subtle degree at MCI stages, and therefore much harder to detect.

Usually AD is diagnosed on the basis of a patient's medical history and a variety of cognitive tests. Most of these tests produce sets of continuous numerical values or categorize a certain screening result in predefined bins. In order to exclude other cerebral pathology or subtypes of dementia, advanced medical imaging techniques, like initially computed tomography (CT) and then magnetic resonance imaging (MRI), are often used. Structural MRI detects tissue changes in the grey and white matter of the human brain. Cognitive task-related changes in brain activity and basal brain activity during resting state are assessed by functional MRI (fMRI). The positron emission tomography (PET) visualizes and quantifies abnormal structures called plaques caused by the protein amyloid-beta (Aβ) in the brains of patients with AD, even in stages of MCI or complete presymptomatic states. Figure 1 shows a hypothetical model of the predicted utility during the progression of AD for different biomarkers, following the studies of Jack et al. [14].

Consequently, we do not rely on single test modes in this project, but rather combine different sources to determine individual risk profiles. We develop IDEA, a new Integrative Detection framework for Early-stage AD patterns. We exploit an unprecedented amount of heterogeneous knowledge sources, including multimodal neuroimaging, biochemical markers and neuropsychological tests. However, the essential effort (in terms of money, time and stress factor for the patients) for collecting the data strongly depends on the different data acquisition tools. Consequently, we select a set of relevant key features yielding best possible classification results concerning both accuracy and cost-effectiveness based on an information-theoretic driven feature selection, and pro-

*Department of Computer Science, LMU, Munich, Germany
{ye, wackersreuther, boehm}@dbs.ifi.lmu.de
†Institute for Stroke and Dementia Research, LMU, Munich, Germany
{michael.ewers@med.uni-muenchen.de}
‡Faculty of Computer Science, University of Vienna, Austria
{claudia.plant@univie.ac.at}

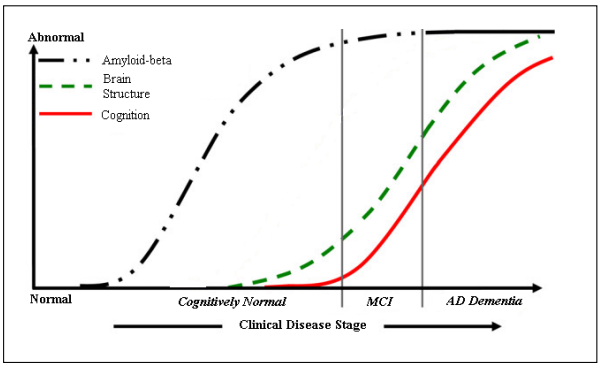


Figure 1: Predicted utility of various biomarkers during the progression of Alzheimer's.

vide a suggestion for the most promising association of different assessment tools. Therefore, IDEA provides two main contributions.

1. A combined analysis of image and non-image data achieves more accurate prediction results.
2. Unavailable measures (due to any reason) can be replaced by equivalent sets of feature combinations.

The rest of this paper is organized as follows: Section 2 gives a brief survey of the large previous work on integrative data mining and related research for early-stage detection of Alzheimer's disease. Section 3 presents our new diagnosis framework which performs heterogeneous data mining for image, numerical and categorical data to achieve high accurate risk profiles for Alzheimer's disease. Section 4 documents a medical case-study, where we present each processing step on a real-world data set provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (`http://adni.loni.usc.edu/`). Finally, Section 5 summarizes the paper.

2 Related Work

In this section, we survey relevant research in the field of data integration and describe related classification approaches for neuroscience application.

Integrative Data Mining. Several papers address the problem of finding dependencies among heterogeneous data. Most integrative clustering approaches, as for instance the algorithms K-Prototypes [13], CFIKP [26], CAVE [12], and K-means-mixed [1] rely on the basic algorithmic paradigm of K-means [11]. While K-means focuses on clustering numerical data, the aforementioned approaches typically use several different optimization goals, one for each data type. Whenever these goals disagree, a manually chosen weighting factor has to decide how to resolve this tie situation. But, it is not trivial to select a suitable weighting factor that is valid for different clusters or for a complete clustering process (while the clusters evolve). Moreover, such approaches implicitly assume the independence between attributes of different types. More advanced solutions, like INTEGRATE [2] or INCONCO [20], consider the task of learning weighting factors and even the number of expected clusters K to detect dependencies between attributes (of the same or different type) as part of the overall clustering process.

The proposed ideas for integrative clustering can be easily applied for a classification scenario. But none of these approaches are suitable for the combination of numerical, categorical and *imaging* data. Rather, we present a solution for this clinically relevant task without the need of challenging parameter settings by using advanced information-theoretic techniques.

Classification of Neuroimaging Data for Early Stage AD Detection. Pattern classifiction methods on the basis of high-dimensional neuroimaging data sets are promising tools to aid the clinical diagnosis of Alzheimer's fully automatically. Support vector machines (SVM) have been applied in a number of studies to exploit structual or functional MRI and PET images for the early diagnosis of AD in MCI and healthy controls [7, 15] and also have been applied to multicenter MRI data sets [5]. However, the cross-validation results of SVM derived patterns show only limited robustness for the prediction of clinical progression in MCI. Other classification algorithms such as Bayes statistics and voting feature intervals show clinically acceptable accuracy ($> 85\%$) for the detection of AD dementia, but insufficient accuracy for the prediction of AD dementia at the MCI stage [21, 3]. A major reason for the limited clinical applicability for the early detection is the inherent heterogeneity of brain changes that are characteristic of AD. In keeping with the diagnostic guidelines, we propose here to source different types of measures including neuroimaging, biochemical markers, genetic features and neuropsychological tests.

The most related approach is the work by Shuo Xiang et al. [24], that examines AD prediction on the basis of heterogeneous data with the focus on missing values. However, besides balancing missing attributes, IDEA tries to find an optimal set of independent features by identifying redundant information sources.

3 Integrative Detection of Early-stage AD Patterns

The first step of the integrative diagnosis framework IDEA is selecting the most informative features of each data modality (neuroimaging, numerical or categorical). This step deserves high diligence, because selecting subsets of strong discriminating features is indispensable for reliable classification results.

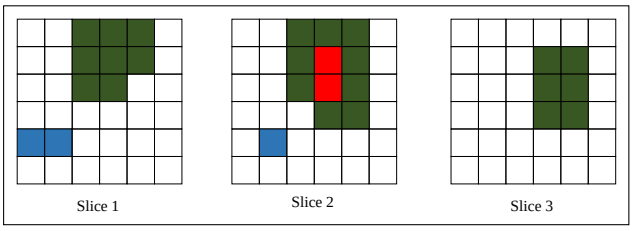


Figure 2: Example for a 3-dimensional DBSCAN used for density-based clustering of neuroimaging data to find brain regions with high-discriminatory power.

3.1 Feature Selection

With the Information Gain (IG) [22, 10], we perform class separation based on the concept of entropy, as IG rates the interestingness of a specific attribute (e.g. one voxel of neuroimage scan) for the separation. To formalize the IG, first the definition of the entropy of the class distribution is needed.

DEFINITION 3.1. (ENTROPY OF CLASS DISTRIBUTION) *Given a class c_i (e.g. AD patients) and its corresponding class probability $p(c_i)$, the entropy of the class distribution is defined as follows:*

$$H(C) = -\sum_{c_i \in C} p(c_i) \cdot log_2(p(c_i)).$$

$H(C)$ corresponds to the required amount of bits to predict the class of an unknown subject and scales between 0 and 1.

The entropy of the class distribution before and after observing an attribute a refers to the information gain (IG) of a and is formally defined as follows:

DEFINITION 3.2. (INFORMATION GAIN) *Given an attribute a (e.g. a voxel), the information gain of a is:*

$$IG(a) = H(C) - H(C|a).$$

In the case of $k = 2$ (e.g. if we consider the classes MCI and AD), IG scales between 0 and 1, where 0 means that the attribute a provides no information on class label of the subject. An IG of 1 means that the class label of all subjects can be derived from the corresponding attribute a without any error.

We can compute the IG for each attribute type, regardless of being an image, numerical or categorical attribute. For features with continuous values (e.g. voxel intensities), we apply the discretization algorithm by Fayyad and Irani [8], which divides the attribute range into class pure intervals, where the IG of the split defines the cut points. To avoid a disproportional high number of cut points, the MDL principle is used to determine the optimal number and location of the cut points. For all attributes, regardless of arising in an image data or not, we hereby calculate class-separation information without the need for data format transformations, which means that we combine the different data types without loss. Only features that have an IG value above a specified threshold IG_{opt} are kept for further processing.

However, the huge amount of information present especially in the neuroimaging data (each image consists of more than two million voxels) poses a major problem for the automated analysis including noisy data and replicability, irrelevant information, and costs in terms of data acquisition and processing time. For this purpose, we apply a density-based clustering approach on the spatially complex imaging data. Thereby, we receive connected brain *regions* which are much more informative for further processing than single voxels.

3.2 Clustering of Neuroimaging Data

In general, clustering algorithms aim at deriving a partitioning of the data into groups (clusters) such that similar objects are grouped together. To identify groups of adjacent voxels that commonly share high IG values, and to remove noise in the imaging data, we use a variant of the well-established density-based clustering approach DBSCAN [6] as recommended in the paper of Plant et al. [21]. Density-based clustering algorithms are designed to find clusters of arbitrary shape in noisy data.

The notion of the original DBSCAN algorithm, which was designed for clustering data objects represented by feature vectors, is defined as follows. An object O is called **core object** if it has at least $MinPts$ objects in its ϵ-range, i.e. $|N_\epsilon(O) \geq MinPts|$, where $N_\epsilon(O) = \{O' | dist(O, O') \leq \epsilon\}$. An object O is **directly density reachable** from another object P w.r.t. ϵ and $MinPts$ if P is a core object and $O \in N_\epsilon(P)$. An object O is **density-reachable** from an object P w.r.t. ϵ and $MinPts$ if there exists a sequence of objects $O_1, \cdots, O_n$ such that $O_1 = P$ and $O_n = O$ and O_{i+1} is directly density-reachable w.r.t. ϵ and $MinPts$ from O_i for $1 \leq i \leq n$. Two objects O and P are **density-connected** w.r.t. ϵ and $MinPts$ if there exists an object Q such that both O and P are density-reachable from Q. A **density-based cluster** is the maximum set of density-connected objects, i.e. the transitive closure of the density reachability relation.

To adapt this algorithm to the setting of neuroimage data, where each object is represented by 3-dimensional voxels, a **core voxel** is a voxel, which is surrounded by at least six voxels that commonly share an IG value higher than IG_{opt}. Figure 2 illustrates an example. It shows three sequent slices in the brain, each of which contains 6×6 voxels. Colored voxels (red, blue or green) indicate voxels with high IG-values. The red voxels are core voxels w.r.t. $\epsilon = 1$ and $MinPts = 6$. The blue voxels are noise and the green voxels are density-reachable w.r.t. to the given values of ϵ and $MinPts$.

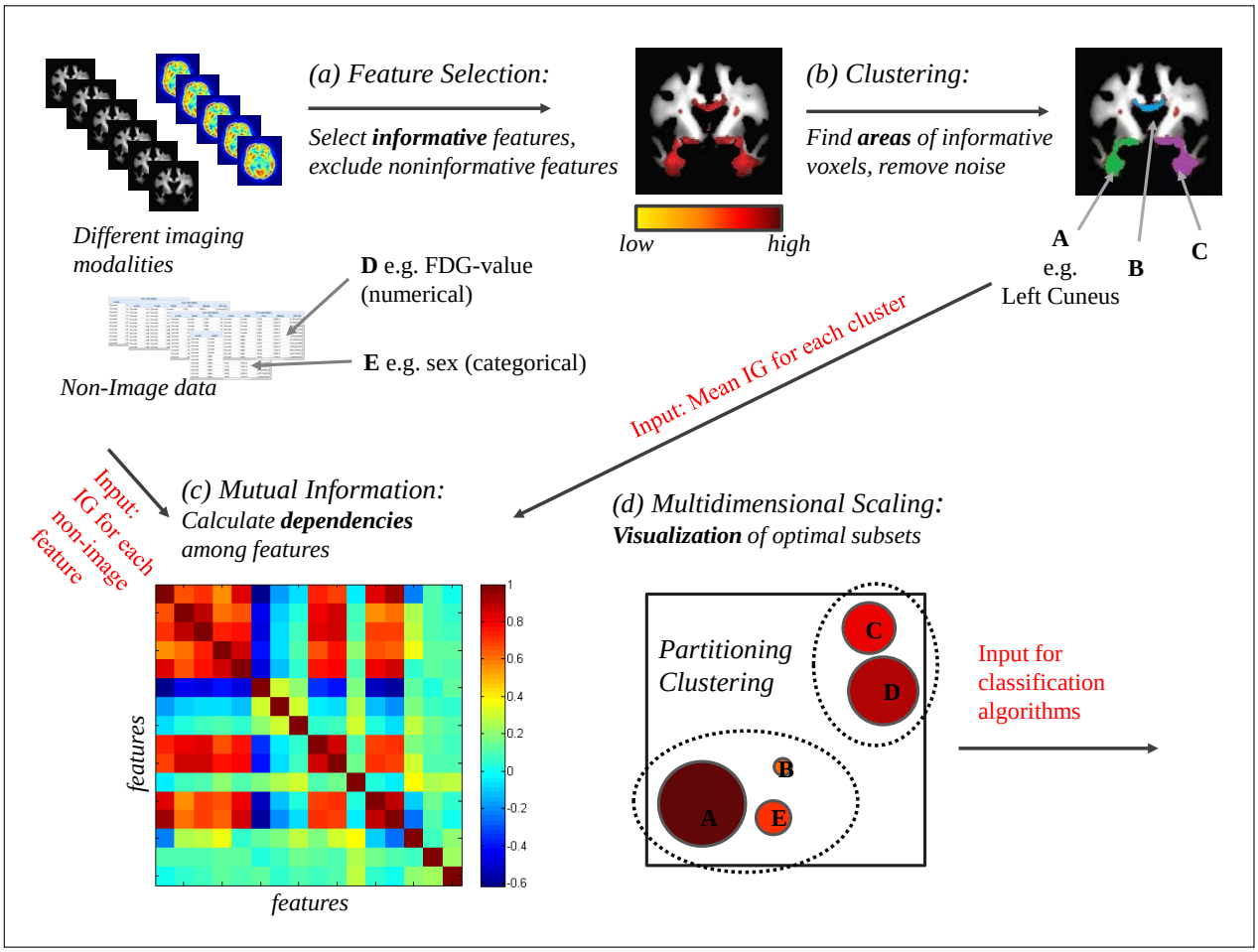


Figure 3: Data analysis stream from heterogeneous data sources to the visualization of optimal feature sets. Input of the combined analysis are the IG values of numerical and categorical non-image data and a representative IG-value per informative voxel cluster for each neuroimage modality. The calculation of the pairwise mutual information leads to feature subsets that provide maximum information for the classification process, visualized by multidimensional scaling. The feature sets, determined by partitioning clustering, serve as input for the classification algorithms.

After selecting single informative features, IDEA computes dependencies among all possible pairs of attributes regardless of being a clustered neuroimaging feature or being a numerical or categorical non-image assessment value in the next step.

3.3 Calculating Dependencies Among Features

To build sets of informative features, we use the concept of mutual information (MI) as suggested by Peng et al. [17]. Hereby, IDEA rates the information dependencies among the different attributes. Informative brain regions that result from the aforementioned feature selection step are represented by the mean values of the corresponding voxels. MI is not limited to real-valued random variables like the correlation coefficient, but rather MI is more general and determines how similar the joint distribution of two random variables x and y is.

DEFINITION 3.3. (MUTUAL INFORMATION) *Given two random variables x and y, their mutual information is defined in terms of their probabilistic density functions $p(x)$, $p(y)$ and $p(x,y)$:*

$$MI(x;y) = \int\int p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy.$$

The resulting MI-matrix forms a metric space, which enables us to determine irrelevant or redundant information sourcing from the various analysis methods for the clinical diagnosis of AD. This means, that the clinician might chose only one assessment modality out of multiple redundant features to

reduce cost, or receives a recommendation for further tests that maximize the accuracy of the classification.

3.4 Visualization of Feature Subsets

Finally, the results of the MI-matrix are representable in 2-dimensional space to facilitate the application of our heterogeneous data mining approach in the clinical environment. For this purpose, we use a standard technique in statistics called multidimensional scaling (MDS) [4]. For a measure of the global similarity among a set of features (in our case the MI matrix), MDS provides a spatial configuration, in which the distances between the objects match their proximities as closely as possible. Each object in the spatial configuration (each point in the visual mapping) is one assessable attribute, its radius visualizes its IG, which is an additional criterion for an optimal subset configuration. The smaller the distance between two objects is, the higher is the amount of redundant information. Therefore, an optimal subset of measures consists of attributes with large radius and high distance to each other.

In order to build sets of informative features, IDEA performs partitioning clustering (e.g. K-means [11]), where each cluster represents one source of independent information. As each cluster usually contains several attributes, we select the features of one cluster according their IG-values. If one attribute can not be assessed (due to expensive costs or accessability) a feature in its direct neigborhood is chosen instead.

3.5 Summary and Implementation Details

Figure 3 summarizes the overall workflow for our integrative diagnosis tool IDEA. After identifying the most informative voxels in all imaging modalities in step (a), a clustering algorithm groups these voxels into areas of interest in step (b) that can be mapped to real anatomical brain regions, e.g. 'Left Cuneus'. Together with the non-image data (e.g. the FDG-value and the sex of the subject), the pairwise mutual information is calculated in step (c). By use of multidimensional scaling, the pairwise dependencies are visualized. This can be used to decide which measure should be assessed to achieve best accuracy with minimal number of tests. In our example, 'Left Cuneus' (feature A) and the FDG-value (feature D) provide the highest IG (radius sizes of the circles correspond to IG values) and therefore should be favored. However, if A is not an option for any reason, feature E is closest to A and thus the best alternative, as E and A share a lot of common information, while C (higher IG) is redundant to D. The detected feature sets are the input data for the classification algorithms.

The implementation of IDEA roughly consists of three parts. Part (1) determines the best IG threshold value for each fold of image data. Part (2) is dedicated to masking the training data and test data in each fold, and part (3) integrates data from different sources. For the first step, we store the candidate IG threshold values in a vector $\boldsymbol{t}$, and select the optimal value by 10-fold cross validation. For part (2), we perform an IG-based feature selection on the training data and mask the test data in each fold, i.e. we keep voxels in the test data which have the same positions as those kept in the training data. Part (3) is the core part of IDEA. Here, each image cluster is represented by its mean image intensity value. We combine the mean value matrix with non-image data and compute pairwise MI. After applying partitioning clustering in the space returned by MDS on MI-matrix, each cluster is represented by the feature with highest IG value. Finally, IDEA performs Support Vector Machine (SVM) classification with polynomial kernel [23] on selected features.

4 Experimental Evaluation

In this section, we present our medical case-study for early-stage AD pattern detection on an open-source data set.

4.1 The Data

We evaluate IDEA on a study that was conducted in the years 2005 to 2007 and attended by 395 participants. The corresponding data set is obtained from the Alzheimer's Disease Neuroimaging Initiative (`http://adni.loni.usc.edu/`). It includes scans for 98 healthy control subjects (HC), 201 patients with MCI (amnestic w.r.t. the study by *Petersen et al.* [18]) and 96 patients with clinically probable AD dementia (referring to *McKhann et al.* [16]). All subjects underwent volumetric MRI (performed on a T1 MRI scanner) and PET, resulting in $121 \times 145 \times 121$ voxels per scan. In addition, the data set provides information for multiple clinical examinations. Table 1 summarizes eight non-image attributes we used for further processing, including demographic variables (e.g. age and sex), biochemical measures (e.g. FDG), genetics (e.g. ApoE genotype) and neuropsychological test scores (e.g. MMSE). The epsilon 4 allele of APOE is the strongest known genetic risk factor for AD with a two- to three-fold increased risk for AD in people with one allele of this kind, rising up to approximately 12-fold in those with two alleles.

4.2 IG-based Feature Selection

Our medical case-study includes three different settings, namely AD vs. HC, AD vs. MCI and HC vs. MCI. To process the neuroimaging data, all scans were randomly divided and stratified w.r.t. the diagnosis into ten folds using 10-fold cross-validation. For each experiment, we also used 10-fold cross-validation on the training data to select a suitable information gain threshold IG_{opt} in a range of $0.02, 0.04, \cdots, 0.5$. To determine relevant brain regions, IDEA performs density-based clustering (cf. Section 3.2)

Table 1: Demographic, biochemical, genetic and neuropsychological variables for the different groups. For each numerical attribute, we report mean and standard deviation of the underlying values. For each categorical variable, we specify the number of subjects in each category.

Attribute	Type	HC	MCI	AD
Age	numerical	$\mu_{Age} = 74.75$ $\sigma_{Age} = 6.90$	$\mu_{Age} = 75.50$ $\sigma_{Age} = 6.60$	$\mu_{Age} = 75.30$ $\sigma_{Age} = 6.61$
Sex	categorical	female: 37 (37.76 %) male: 61 (62.24 %)	female: 64 (31.84 %) male: 137 (68.16 %)	female: 38 (39.58 %) male: 58 (60.42 %)
Years of education	numerical	$\mu_{Education} = 15.95$ $\sigma_{Education} = 3.02$	$\mu_{Education} = 15.76$ $\sigma_{Education} = 2.87$	$\mu_{Education} = 14.61$ $\sigma_{Education} = 3.20$
Race	categorical	white: 90 (91.84 %) black: 7 (7.14 %) asian: 1 (1.02 %)	white: 187 (93.03 %) black: 10 (4.98 %) asian: 4 (1.99 %)	white: 89 (92.71 %) black: 5 (5.21 %) asian: 2 (2.08 %)
Marital status	categorical	never married: 6 (6.12 %) married: 71 (72.45 %) divorced: 8 (8.16 %) widowed: 13 (13.27 %)	never married: 3 (1.49 %) married: 151 (75.12 %) divorced: 18 (8.96 %) widowed: (14.43 %)	never married: 3 (3.13 %) married: 83 (86.46 %) divorced: 4 (4.17 %) widowed: 6 (6.25 %)
Number of ApoE4 alleles	categorical	0: 73 (74.49 %) 1: 23 (23.47 %) 2: 2 (2.04 %)	0: 94 (46.77 %) 1: 81 (40.30 %) 2: 26 (12.94 %)	0: 33 (34.38 %) 1: 48 (50.00 %) 2: 15 (15.63 %)
FDG value	numerical	$\mu_{FDG} = 6.09$ $\sigma_{FGD} = 0.76$	$\mu_{FDG} = 5.85$ $\sigma_{FDG} = 0.76$	$\mu_{FDG} = 6.06$ $\sigma_{FDG} = 0.64$
MMSE-Score	categorical	none: 90 (91.84 %) ($28 \leq MMSE \leq 30$) mild: 8 (8.16 %) ($25 \leq MMSE \leq 27$) moderate: 0 (0.00 %) ($20 \leq MMSE \leq 24$) severe: 0(0.00 %) ($MMSE < 20$)	none: 92 (45.77 %) mild: 93 (46.27 %) moderate: 16 (9.96 %) severe: 0 (0.00 %)	none: 0 (0.00 %) mild: 39 (40.63 %) moderate: 56 (58.33 %) severe: 1 (1.04 %)

MMSE: The Mini Mental State Examination (also known as Folstein test is a 30-point neuropsychological questionnaire, used in clinical and research settings to measure general cognitive impairment [9].

with a parametrization of $MinPts = 4$ voxels and $\epsilon = 1$ voxel. We only keep robust clusters that are detected across all folds. Figure 4a shows ten robust clusters detected in MRI data for the setting AD vs. HC. The two identified clusters of the PET data are illustrated in Figure 4b, respectively. Single informative voxels, which distinguish AD patients from HC are spread all over the brain (162,532 voxels in MRI and 110,117 voxels in PET). To interpret the detected clusters, we map them to real brain regions according their anatomical location information using the Talairach Daemon software available at `http://www.talairach.org`. This mapping is presented in Table 2.

Only a few features (49 voxels in MRI and 675 voxels in PET) classify HC from MCI. For AD vs. MCI, 64,265 voxels in MRI and 37 voxels in PET have an IG value above IG_{opt}. Consequently, IDEA did not detect any informative neuroimaging clusters for AD vs. MCI and HC vs. MCI.

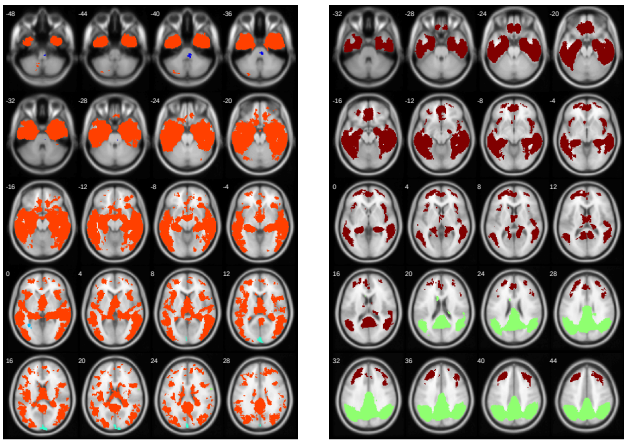
Finally, Table 3 summarizes the IG values of each attribute for the non-image data (cf. Table 1). For further processing, IDEA selects all attributes with an IG value higher than zero.

Table 3: IG values for each attribute of the non-image data for the settings AD vs. HC, AD vs. MCI and HC vs. MCI, respectively.

	AD vs. HC	AD vs. MCI	HC vs. MCI
Age	0.06	0.00	0.05
Sex	0.00	0.00	0.00
Years of education	0.00	0.00	0.00
Race	0.00	0.00	0.00
Marital status	0.00	0.00	0.00
Number of ApoE4 alleles	0.12	0.00	0.15
FDG value	0.41	0.15	0.07
MMSE-Score	0.83	0.49	0.20

4.3 Dependencies among Features

For all features identified in the aforementioned section and each experimental setting, we calculate the pairwise MI and visualize it using MDS, as described in Sections 3.3 and 3.4. Figure 5a shows the MI-matrix of informative attributes



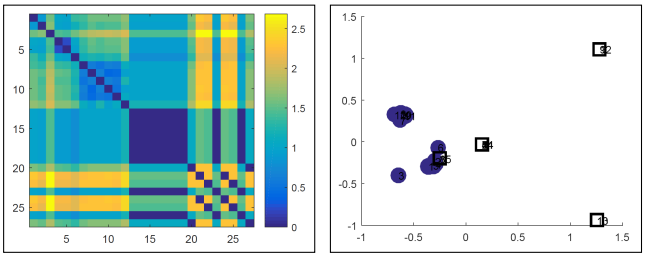
(a) MRI: AD vs. HC **(b)** PET: AD vs. HC

Figure 4: Selected informative clusters appearing in all folds of MRI and PET data for the AD vs. HC study.

Cluster ID	Cluster size	Brain region
MRI data		
1	4	Left Cerebellar Tonsil
2	4	Left Cingulate Gyrus
3	4	Right Precuneus
4	36	Right Medial Frontal Gyrus
5	51	Right Precentral Gyrus
6	66	Left Parahippocampal Gyrus
7	236	Right Cerebellar Tonsil
8	355	Right Superior Parietal Lobule
9	576	Left Cuneus
10	161,200	Right Middle Temporal Gyrus
PET data		
1	37,197	Left Precuneus
2	72,920	Left Middle Temporal Gyrus

Table 2: Mapping of detected clusters in the neuroimaging data to real brain regions using the Talairach Daemon software for the study AD vs. HC.

sourcing from neuroimaging scans and non-image data for the class of MCI patients in one fold. Figure 5b illustrates the corresponding dependencies by MDS. The depicted distance of two objects in this plot, directly correlates with their joined degree of information. Hence, it is obvious that some features provide redundant information. By partitioning clustering, IDEA determines *different* kind of information. To represent the discriminatory attributes, shown in Figure 5, only five features (one of each cluster) are adequate to achieve strong classification results (cf. Section 4.4).



(a) MI-matrix of merged attributes sourcing from neuroimaging and non-imaging data.

(b) 2D-representation by MDS. Circles indicate neuroimaging attributes, squares formalize non-image features.

Figure 5: Calculation and illustration of dependencies among features for the group of MCI patients.

4.4 Classification Results

For classification, we use the WEKA implementation (available at `http://www.cs.waikato.ac.nz/ml/weka`) of the Support Vector Machine (SVM) with polynomial kernel. For each classification result, we report accuracy (acc), sensitivity (sen) and specificity (spec). Table 4 presents the results on neuroimaging data w.r.t. using all voxels of the detected clusters versus the mean value of the underlying voxels of each cluster.

The next experiments document the benefit of an integrative classification procedure as performed by IDEA. Again, we distinguish between image cluster representations by all voxels or mean values. The classification results described by accuracy, sensitivity and specificity are represented in Table 5. The accuracy of AD vs. HC of MRI and PET image data combined with the non-image attributes is approximately the same due to the number of features of image data dominate the number of non-image attributes. However, when combining mean value of clusters with the informative non-image features, the classification results are improved above 90%.

As stated in the aforementioned section, IDEA automatically provides small feature sets that achieve accurate classification results. For this experiment, we evaluate the classification results on a set of features that was built by partitioning clustering with $k = 5$ and an IG-driven feature selection for each cluster on the data illustrated in Figure 5b. The corresponding results are presented in Table 6. Compared with Table 5, where we were using all available attributes, selecting the right set of (few) features yield to similar classification accuracies.

5 Conclusion

With IDEA, we presented a data mining framework for Integrative Detection of Early-stage Alzheimer's disease based on multimodal neuroimaging and heterogeneous non-image data types. The combination of information gain, mutual information, multidimensional scaling and clustering enables us to find feature combinations that have a high potential to predict Alzheimer's at an early stage. In near future, we per-

Table 4: Classification results on neuroimaging data using all voxels of a cluster vs. using the mean value of the voxels to represent a cluster.

	AD vs. HC	AD vs. MCI	HC vs. MCI
MRI data			
acc (all)	0.8029	0.6959	0.6556
acc (mean)	0.7458	0.7095	0.6723
sen (all)	0.8067	0.3700	0.1433
sen (mean)	0.69221	0.1656	0
spec (all)	0.7978	0.8510	0. 9055
spec (mean)	0.7978	0.9650	1
PET data			
acc (all)	0.8763	0.7128	0.6956
acc (mean)	0.7513	0.7024	0.6723
sen (all)	0.8422	0.3478	0.2900
sen (mean)	0.7378	0.2089	0
spec (all)	0.9100	0.8857	0.8900
spec (mean)	0.7600	0.9355	1

Table 5: Classification results on neuroimaging data in combination with non-image data using all voxels of a cluster vs. using the mean value of the voxels to represent a cluster.

	AD vs. HC	AD vs. MCI	HC vs. MCI
MRI data			
acc (all)	0.8789	0.9187	0.9433
acc (mean)	1	0.9292	0.9667
sen (all)	0.8667	0.8289	0.9600
sen (mean)	1	0.8189	0.9900
spec (all)	0.8900	0.9600	0.9352
spec (mean)	1	0.9800	0.9555
PET data			
acc (all)	0.8866	0.9359	0.9199
acc (mean)	0.9950	0.9392	0.9532
sen (all)	0.8633	0.8189	0.9089
sen (mean)	0.9900	0.8300	0.9589
spec (all)	0.91	0.9900	0.9255
spec (mean)	1	0.9900	0.9505

form a big data study on a second data set contributed by partners of the Institute for Stroke and Dementia Research (ISD), University of Munich based on a compact data representation. Here, we again expect new insight to the development and diagnosis of a disease that causes problems with memory, thinking and behavior for a multitude of elderly people. Furthermore, we currently work on a user-optimized graphical presentation based on scatter-plots that enable the medical scientists to rate the individual risk profile of a particular subject.

Table 6: Classification results after feature selection on neuroimaging data using average voxels of a cluster in combination with non-image data

	AD vs. HC	AD vs. MCI	HC vs. MCI
MRI data			
acc	0.8497	0.9121	0.9067
sen	0.8622	0.7433	0.8800
spec	0.8389	0.9900	0.9207
PET data			
acc	0.7971	0.9160	0.8700
sen	0.7867	0.7778	0.8400
spec	0.8067	0.9800	0.8800

References

[1] Ahmad, A., Dey, L.: A k-mean clustering algorithm for mixed numeric and categorical data. Data & Knowledge Engineering 63(2), 503–527 (2007)

[2] Böhm, C., Goebl, S., Oswald, A., Plant, C., Plavinski, M., Wackersreuther, B.: Integrative Parameter-Free Clustering of Data with Mixed Type Attributes. In: Advances in Knowledge Discovery and Data Mining, 14th Pacific-Asia Conference, PAKDD 2010, Hyderabad, India, June 21-24, 2010. Proceedings. Part I. pp. 38–47 (2010)

[3] Böhm, C., Oswald, A., Plant, C., Wackersreuther, B.: A Data Mining Framework for Classification of High-resolution Magnetic Resonance Images. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Disvovery and Data Mining (KDD 2008), Workshop on Mining Medical Data, Las Vegas, USA (2008)

[4] Borg, I., Groenen, P.J.: Modern Multidimensional Scaling: Theory and Applications. Springer (2005)

[5] Dyrba, M., Ewers, M., Wegrzyn, M., Kilimann, I., Plant, C., Oswald, A., Meindl, T., Pievani, M., Bokde, A., Fellgiebel, A., Filippi, M., Hampel, H., Kloppel, S., Hauenstein, K., Kirste, T., Teipel, S., the EDSD study group: Robust automated detection of microstructural white matter degeneration in Alzheimer's disease using machine learning classification of multicenter DTI data. PLOS ONE 8(5), 1–14 (2013)

[6] Ester, M., Kriegel, H., Sander, J., Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA. pp. 226–231 (1996)

[7] Fan, Y., Batmanghelich, N., Clark, C., Davatzikos, C., ADNI: Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. NeuroImage 39(4), 1731–1743 (2008)

[8] Fayyad, U.M., Irani, K.B.: Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence. Chambéry, France, August 28 - September 3, 1993. pp. 1022–1029 (1993)

[9] Folstein, M.F., Folstein, S.E., McHugh, P.R.: Mini-Mental State (a practical method for grading the state of patients for the clinician). Journal of Psychiatric Research 12(3), 189–198

(1975)
[10] Hall, M.A., Holmes, G.: Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. IEEE Transactions on Knowledge and Data Engineering 15(6), 1437–1447 (2003)
[11] Hartigan, J.A.: Clustering Algorithms. John Wiley & Sons, Inc., New York, NY, USA, 99th edn. (1975)
[12] Hsu, C., Chen, Y.: Mining of mixed data with application to catalog marketing. Expert Systems with Applications 32(1), 12–23 (2007)
[13] Huang, Z.: Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Mining and Knowledge Discovery 2(3), 283–304 (1998)
[14] Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q.: Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. The Lancet Neurology 9(1), 119–128 (2010)
[15] Kloppel, S., Stonnington, C., Chu, C., Draganski, B., Scahill, R., Rohrer, J., Fox, N., Jack, C., Ashburner, J., Frackowiak, R.: Automatic classification of MR scans in Alzheimer's disease. Brain 131(3), 681–689 (2008)
[16] McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M.: Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34(7), 939–944 (1984)
[17] Peng, H., Long, F., Ding, C.H.Q.: Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8), 1226–1238 (2005)
[18] Petersen, R.C., Doody, R., Kurz, A., Mohs, R.C., Morris, J.C., Rabins, P.V., Ritchie, K., Rossor, M., Thal, L., Winblad, B.: Current Concepts in Mild Cognitive Impairment. Archives of neurology 85(12), 1985–1992 (2001)
[19] Piatetsky-Shapiro, G., Djeraba, C., Getoor, L., Grossman, R., Feldman, R., Zaki, M.J.: What are the grand challenges for data mining?: KDD-2006 panel report. SIGKDD Explorations 8(2), 70–77 (2006)
[20] Plant, C., Böhm, C.: INCONCO: INterpretable Clustering Of Numerical and Categorical Objects. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011. pp. 1127–1135 (2011)
[21] Plant, C., Teipel, S.J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M.: Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. NeuroImage 50(1), 162–174 (2010)
[22] Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann (1993)
[23] Vapnik, V.N., Chervonenkis, A.Y.: Theory of Pattern Recognition [in Russian]. Nauka (1974)
[24] Xiang, S., Yuan, L., Fan, W., Wang, Y., Thompson, P.M., Ye, J.: Multi-source learning with block-wise missing data for Alzheimer's disease prediction. In: The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013. pp. 185–193 (2013)
[25] Yang, Q., Wu, X.: 10 Challenging Problems in Data Mining Research. International Journal of Information Technology & Decision Making 5(4), 597–604 (2006)
[26] Yin, J., Tan, Z.: Clustering Mixed Type Attributes in Large Dataset. In: Parallel and Distributed Processing and Applications, Third International Symposium, ISPA 2005, Nanjing, China, November 2-5, 2005, Proceedings. pp. 655–661 (2005)