**Journal of Medical Systems manuscript No.**
(will be inserted by the editor)

# Machine learning for predicting cognitive diseases: methods, data sources and risk factors

**Brankica Bratić · Vladimir Kurbalija · Mirjana Ivanović · Iztok Oder · Zoran Bosnić**

Received: date / Accepted: date

**Abstract** Machine learning and data mining approaches are being successfully applied to different fields of life sciences for the past 20 years. Medicine is one of the most suitable application domains for these techniques since they help model diagnostic information based on causal and/or statistical data and therefore reveal hidden dependencies between symptoms and illnesses. In this paper we give a detailed overview of the recent machine learning research and its applications for predicting cognitive diseases, especially the Alzheimer's disease, mild cognitive impairment and the Parkinson's disease. We survey different state-of-the-art methodological approaches, data sources and public data, and provide their comparative analysis. We conclude by identifying the open problems within the field that include an early detection of the cognitive diseases and inclusion of machine learning tools into diagnostic practice and therapy planning.

B. Bratić
University of Novi Sad, Faculty of Sciences, Trg Dositeja Obradovića 4, Novi Sad
E-mail: brankica.bratic@dmi.uns.ac.rs

V. Kurbalija
University of Novi Sad, Faculty of Sciences, Trg Dositeja Obradovića 4, Novi Sad
E-mail: kurba@dmi.uns.ac.rs

M. Ivanović
University of Novi Sad, Faculty of Sciences, Trg Dositeja Obradovića 4, Novi Sad
E-mail: mira@dmi.uns.ac.rs

I. Oder
University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, Ljubljana
E-mail: iztok.oder@gmail.com

Z. Bosnić
University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, Ljubljana
E-mail: zoran.bosnic@fri.uni-lj.si

**Keywords** Cognitive diseases · Machine learning · Data mining · Alzheimer's disease · Parkinson's disease

## 1 Introduction

A wide range of different artificial intelligence techniques have been shown to be promising in helping medical experts with understanding causes and risks of illnesses, providing appropriate diagnoses and courses of therapy for validating their suggestions and decisions. Data commonly produced during medical practice falls into the category of complex, i.e., big data. Prominent examples of such big data vary from multimedia objects (images, sound, etc.), temporal data, time-series (i.e., sequences of data points measured at successive times) towards semi-structured and unstructured text documents [76]. Due to their usefulness, complex data are prevalent in many contemporary medical systems. For a wide range of serious diseases it is necessary to obtain numerous kinds of tests that have a purpose to examine different aspects of patients' health. Those tests usually provide very complex and interrelated data. Human cognition and behavior disorders are among diseases which influence older adults, causing many serious neurological diseases (brain ischemia, brain haemorrhagia, brain tumors, neurodegenerative disorders, demyelination disorders, dementia). Different kinds of human cognition and behavior disorders represent a whole spectrum of conditions resulting in a deteriorated mentality. For example, dementia is an acquired loss of intellectual functions, associated with personality disturbances that interfere with the patient's professional and social relations. Dementias represent a whole spectrum of different conditions resulting in deteriorated mentality, among which the most common is Alzheimer's disease. There are numerous studies on cognition disorders, dealing with the risk factors for its development [43], causes [69], diagnostic procedures relying on data from different kind of questionnaires [22,68] or magenetic resonance imaging (MRI) scans [11,47,15,57], predictors [46], etc.

Although there exist numerous techniques for analysis of complex big data, they have been seldom rigorously tested and put into practical use within a centralized medical repository. Developing a decision support system for that task would bring together, under a single roof, numerous artificial intelligence and machine learning techniques.

In this paper we provide an overview of machine learning development in the field of cognitive diseases. Numerous papers that we collected and analyzed cover a wide range of algorithms and techniques of machine learning applied on data collected from patients that suffer from various kind of cognitive diseases - Alzheimer's disease (AD), mild cognitive impairment (MCI), vascular dementia, Parkinson's disease (PD), etc.

The paper is organized as follows. Section 2 gives an overview of development and use of machine learning techniques in the field of cognitive diseases presenting current state-of-the-art and applications of different contemporary machine learning techniques. Datasets that are mostly used in relevant works

are presented in Section 3. Section 4 summarizes open problems in the field and Section 5 summarizes some conclusions.

## 2 Machine learning techniques in cognitive diseases

Machine learning grew up from the idea of making machines (i.e. computers) learn from a wide range of complex, unstructured and semi-structured data. The field started to flourish in the nineties and has proven to be very useful tool for solving many real-life problems in different domains and good replacement of the traditional statistical methods.

Medicine is an example field that gained a lot from machine learning. Examination of cognitive diseases was relatively early encouraged by machine learning algorithms. First papers started with a really basic task: binary and ternary classification of patients. The essential task of binary classification was to determine if the patient is demented or not. Ternary classifications were more difficult - their task was to determine if the patient is healthy, cognitively impaired or demented. The mentioned two types of classifications are still useful for making automated diagnosis, which helps doctors optimize their work. Namely, professional staff needs to devote a fair amount of time to score different types of questionnaires in order to determine dementia diagnoses. Besides the fact that those scoring tasks are time consuming, scores are also not ideally objective and consistent. It is clear that proper classification algorithms would be beneficial.

In this section we provide an overview of different machine learning approaches that were developed in the field of cognitive diseases, disorders and impairment. The section is divided into 8 subsections. We first present research based on questionnaires data, followed by research based on Magnetic Resonance Imaging data and on other less conventional data sources for the field of cognitive diseases. We proceed by surveying research papers that are trying to fuse different datasets in order to achieve higher classification accuracy, and provide comparative analyses of classification algorithms and unconventional machine learning tasks. Finally, we provide a tabular summary of all methods and data that were used in the presented papers.

### 2.1 Classification tasks on questionnaire-based data

When machine learning algorithms began to be applied in the field of cognitive diseases, the first choice of the available learning data were questionnaires. The most used questionnaire types were: the six-item Blessed Orientation, Memory and Concentration test (BOMC [29]), Functional Activities Questionnaire (FAQ [59]), Dementia Rating Scale (DRS [44]) and Mini-Mental Status Exam (MMSE [30]). A choice of classification algorithms depended on authors, but still, a very frequent choice were rule-based classifiers (e.g. C4.5 [63]), Naive Bayes (NB) [25] and IB1 [6] (a local nearest-neighbor classifier).

Several papers that applied machine learning techniques within the field of cognitive diseases stems from the research team of University of California, Irvine [22,68–70]. They used data provided by UC Irvine Alzheimer's Disease Research Center (ADRC). The experiments on these data showed that the common classifiers (C4.5, C4.5 Rules, Naive Bayes, IB1 and PL) can obtain reasonably good results in the aforementioned classification tasks. As reported in [22], authors achieved accuracy above 80% for every classifier they used, which is interpreted as a valuable result. Although the overall accuracy is high, there was still one major problem – the accuracy for the class that represents cognitively impaired patients (i.e., the sensitivity) was around 60% [68], which means that cognitively impaired patients were not discriminative enough from normal controls (NC). This also means that the symptoms of patients with cognitive diseases were not frequent enough in the examined questionnaires, which caused the machine learning algorithms to perform poorly in classification of these patients.

Interesting part of mentioned papers was experimentation with the decision rules that were obtained from decision trees. Rule based classifiers are particularly interesting in medical classification tasks. By analyzing generated rules, one can obtain information about which features are important and what are the features' marginal values. Professional staff can afterwards evaluate and confirm these findings in practice.

As an example, Figure 1 shows a set of rules obtained by Shankle et al. in [69]. After obtaining the rules, the authors and the the neurologists jointly corrected them in order to eliminate the so-called "nonsense rules". As a result they obtained a higher classification accuracy. The authors used datasets from the following tests: FAQ [59], MMSEPLUS [29] and Ishihara Color Plates [41]. With the FAQ test data they obtained the accuracy of 88% for C4.5, and 85% for C4.5 rules; with the MMSEPLUS dataset they achieved 79% for C4.5, and 77% for C4.5 rules; and for the Ishihara Color Plates 66% for C4.5, and 63% for C4.5 rules.

**Rule 1:** age $> 56$ *and* job $> 2 \implies$ class **impaired**
**Rule 2:** money $> 0$ *and* forget $> 0 \implies$ class **impaired**
**Rule 3:** gender $= 0$ *and* *age* $> 56$ *and* forget $> 0 \implies$ class **impaired**
**Rule 4:** age $> 56$ *and* *age* $\leq 64$ *and* forget $> 0 \implies$ class **impaired**
**Rule 5:** age $> 73$ *and* forget $> 0 \implies$ class **impaired**
**Rule 6:** forget $\leq 0 \implies$ class **normal**
**Rule 7:** Default $\implies$ class **impaired**

**Fig. 1:** A C4.5 rule set taken from [69]. In order to classify a new instance, the given rules were applied in the presented order until reaching the rule that is satisfied. For example, a patient with data [age=60, job=1, money=0, forget=1, gender=1] would be classified with the rules as follows. The rule 1 would not be satisfied since the job is not grater than 2. The rule 2 would also not be satisfied since money is not greater than 0. The rule 3 would not be satisfied since the gender is not equal to 0. The rule 4 would finally be satisfied since all conditions are met. Therefore, the class of a given example would be **impaired**.

Tierney et al. [73] researched the possibility of using a battery of neuropsychological tests for predicting onset of AD. They analysed AD and NC group differences on test results with ANCOVA and discovered that both MMSE and DRS scores were significantly different. Prediction power of test scores was additionally examined with logistic regression. First all features were used to calculate intercorrelation matrix. Those with a high correlation were removed to reduce multicollinearity. Remaining features were used for prediction. The accuracy of this model was 89%.

Finally, Table 1 summarizes approaches that are presented in this section, giving their short overviews, class values and classification accuracies.

| Authors | Method | Classes | Accuracy |
|---|---|---|---|
| Tierney et al. [73] | Predicting AD using neuropsychological test results and logistic regression. | NC<br>AD | 89% |
| Shankle et al. [69] | Detecting very early stages of dementia from normal aging with machine learning methods. | normal<br>impaired | 88% |
| Datta et al. [22] | Applying machine learning to an Alzheimer's database. | healthy<br>MCI<br>demented | $> 80\%$<br>(60% for MCI) |

**Table 1:** An overview of classification methods and accuracies that were obtained from questionnaire data.

## 2.2 Classification tasks on medical imaging data

After the initial exploration of machine learning algorithms in different areas and domains, medical imaging started to be the main source of data. This includes magnetic resonance imaging (MRI) scans and positron emission tomography (PET) scans. In the following section we present a few papers in which MRI or PET scans were used as the main data source.

The idea of using the MRI or PET scans was to extract features from the images and then run classifiers in order to detect unhealthy subjects. To reduce the feature count, features were usually grouped and averaged in accordance to regions of interest (ROI), which present certain portions of images that are relevant for a given medical task. As an example of the described approach, Burge et al. [11] extracted causal relationships from the functional magnetic resonance imaging (fMRI) data by using the Dynamic Bayesian Network (DBN) [56]. DBN is a Bayesian Network that is extended with the concept of time – at any point in time T, the value of a network's variable can be calculated. Based on extracted causal relationships, the authors built neural-anatomical networks that classified patients into healthy or demented. Classification accuracy of this approach was around 73%.

Li et al. [47] built a framework for group analysis of fMRI data using Dynamic Bayesian Networks. The authors noticed that there were certain

downsides when using plain Dynamic Bayesian Network on the whole group data, so they decided to build a DBN for each subject separately.

Challis et al. [13] compared performance of Bayesian Gaussian process logistic regression models using linear and non-linear covariance functions with Support Vector Machine model. The input data was fMRI as well. Built models were comparable and achieved accuracy around 80% for prediction of MCI vs. NC.

Patil et al. [57] used Artificial Neural Networks (ANN) upon Open Access Series of Imaging Studies (OASIS) database [4]. That database includes individuals with early-stage Alzheimer's disease, each individual having 3 or 4 MRI scans. The authors achieved overall accuracy that is above 95%, which is not surprising since they applied only binary classification.

Convolutional neural networks provide another way for classifying MRI scans. Meszlényi et al. [51] introduced a convolutional neural network architecture for functional connectome classification called connectome-convolutional neural network (CCNN). They used their model for MCI classification of fMRI scans and managed to outperform SVM, LASSO, simple neural network, and deep neural network. The highest obtained classification accuracy of CCNN was reported to be 71.9%.

Payan et al. [58] also used 3D-CNN for AD prediction using MRI data. The authors initially used a sparse autoencoder to learn filters for convolution operations, and integrated the learned filters into the first layer of a convolutional neural network. They managed to reach following accuracies: 95.39% for AD vs. NC classification, 86.84% for AD vs. MCI classification and 92.11% for NC vs. MCI classification.

Moradi et al. [53] proposed an approach to categorize subjects into three classes: subjects with Alzheimer's disease, subjects with mild cognitive impairment (MCI) and cognitively normal subjects. First, the authors selected features from a larger pool of MRI scan features by using sparse logistic regression. After that, they built hierarchical classification framework that consists of multiple binary classifiers. They used Alzheimer's Disease Neuroimaging Initiative (ADNI) database [2] to train classifiers, and for testing purposes they used data provided by [3]. Classification accuracy of this setup was around 77%.

Diffusion Tensor Imaging (DTI) [42] is a MRI-based neuroimaging technique that is used for characterizing the microstructural organization of tissue in vivo. It is shown that measures of structural disconnection captured by DTI are a promising marker of AD [65]. Authors of [26] applied machine learning techniques upon the DTI dataset that contains images of 137 AD patients and 143 healthy controls, all of which were captured with nine different MRI scanners. Authors applied binary classification by using Support Vector Machine (SVM) and Naive Bayes (NB) classifiers. Results of these experiments suggest that SVM is more suitable for this purpose by achieving accuracy of 83%, while the NB achieved significantly lower accuracy of 75%.

Chu et al. [17] explored 4 different feature selection methods for retrieval of informative voxels from raw MRI. They compared preselected brain atlas

based ROI selection, t-test filtering, t-test filtering in ROI and recursive feature elimination. Selected features were fed into SVM for classification of AD vs. NC and MCI vs. NC. Using feature selection methods based on ROI produced significantly better classification accuracies than no feature selection in both classification tasks. The accuracies were 85% and 67% respectively.

Szenkovits et al. [72] presented a new feature selection method called Minimalist Genetic Algorithm (mGA), that is developed specifically for fMRI data. The method is used as a technique for improving fMRI scans classification accuracy. The authors evaluated the algorithm on a MCI binary classification task, and managed to outperform LASSO feature selection method with accuracy of 62.7%.

Liu et al. [48] suggested using ensemble technique using Sparse representation-based classifier (SRC) as a weak classifier on subsamples of patches from raw MRI. SRC constructs a dictionary using all training data across all classes and seeks for the sparse representation of a test sample in it in prediction step. SRC codes the test sample for each possible class as a sparse linear combination of all training samples belonging to that class by L1-norm minimization. Minimum reconstruction error over all classifiers is then used for predicting class. Classification accuracy of the proposed method was 90.8% and 87.85% when classifying AD vs. NC and MCI vs. NC respectively. Additional analysis of weak classifiers with the highest accuracies showed that patches used for their training came from brain regions often associated with AD (hippocampus, parahippocampal gyrus, entorhinal cortex and amygdala).

Authors of [37] proposed the use of Linear Programming Boosting for predicting AD using MRI. A combination of L1 sparsity and spatial constraints for voxels led to increase of classification accuracy. Additional analysis of voxels selected during training phase revealed that they were mostly concentrated in hippocampus and parahippocampal gyrus which have been previously associated with AD. Some samples had more than 65% of voxels incorrectly classified, which suggests that the samples were not categorized correctly by experts in the first place.

Lopez et al. [50] proposed a method for selecting principal components for subspace representation calculated with Principal Component Analysis (PCA). Fisher score was used to determine best components instead of percentage of total variance contained in principal components. Projected features were then fed into ANN and SVM. Using PET data the accuracies of this approach were 91% and 90% for each method respectively.

Kippenhan et al. [45] split PET voxels into 67 brain regions and extracted metabolism activity features. These features were used to compare predictive power of Linear Discriminant Analysis (LDA) and ANN. ANN performed better than LDA. Its performance was comparable to performance of an independent expert who classified samples based on examining only PET images.

Cho et al. [16] used LDA as well. However their input data was PCA reduced cortical thickness data extracted from MRI. This approach had accuracy of 87% for AD vs. NC classification.

Use of longitudinal data can provide valuable information on how the disease is progressing. Gray et al. [36] showed that using data extracted from baseline and 12-month follow-up MRI. They extracted average voxel intensities for 83 brain regions for each image and used them to train the SVM. The classification accuracies of 2 clinically relevant groups AD vs. NC and MCI vs. NC were 88% and 81% respectively. Both were significantly better than the accuracies of predictions using only baseline data.

Huang et al. [39] experimented with longitudinal MRI data as well. Instead of building a single classifier with an optimal subset of features they proposed a hierarchical classification method that builds multiple multilevel classifiers that are able to address the issues of high feature dimensionality and incorporate spatial information. Their model consisted of 3 levels. The first layer was built on raw MRI voxel intensities and its output was fed into the second layer of classifiers. Each second-level classifier used a patch of predictions from the previous layer. The final layer used all outputs to build the last classifier. The proposed method outperformed base classifier with accuracy of 79%.

Gomar et al. [34] and Ewers et al. [28] investigated the ability of biomarkers and cognitive markers to predict a change from MCI to AD. Gomar et al. [34] used ANOVA followed by t-tests to determine which features were significant among the two groups. The output was then used to train Logistic regression model. Its accuracy was 72%. Ewers et al. [28] first established input features for the best Logistic regression model for discrimination between AD and NC. They tested different models for single and combined modalities. The best model was then used for predicting the conversion from MCI to AD. Using Cox regression they modeled time to conversion and determined its best predictors. Among others entorhinal cortex and hippocampus proved to have huge impact.

To conclude this section, we provide Table 2 that summarizes the most relevant approaches based on MRI scans, along with their classification accuracies.

### 2.3 Classification tasks on other data

Researches also used other data sources that might contain valuable information for distinguishing between different disease diagnoses. These comprise blood sample data and Electroencephalography (EEG) signals.

Doecke et al. [24] used blood protein levels measured in plasma. They performed a rigorous feature selection on 180 input protein features. The procedure used 2 independent data sets and four different feature selection methods per each set. The features that were most commonly selected by the methods of each set and were common between sets were used for classification purposes with NB, Random Forest (RF) and SVM. Most selected features were identified as commonly associated with AD. Carcinoembryonic antigen stood out since it was never associated with AD before.

Ray et al. [64] examined 120 proteins using Significance Analysis of Microarrays (SAM). SAM is a statistical method for determining whether changes

| Authors | Method | Classes | Accuracy |
|---|---|---|---|
| Patil et al. [57] | Artificial Neural Networks were used for classification. | healthy demented | > 95% |
| Payan et al. [58] | Sparse autoencoders and 3D-CNN were used for classification. | NC MCI AD | 95% (AD vs. NC) 87% (AD vs. MCI) 92% (MCI vs. NC) |
| Liu et al. [48] | An ensemble based method using Sparse representation-based classifier as base classifier. | NC MCI AD | 90% (AD vs. NC) 87% (MCI vs. NC) |
| Lopez et al. [50] | PCA transformed features were feed into Neural network and SVM. | NC AD | 90% |
| Gray et al. [36] | Longitudinal data was used in combination with SVM. | NC MCI AD | 88% (AD vs. NC) 81% (MCI vs. NC) |
| Cho et al. [16] | PCA transformed data was feed into LDA for classification. | NC AD | 87% |
| Ewers et al. [28] | Logistic Regression was used for predicting disease while Cox Regression was used to model time to conversion from MCI to AD. | NC MCI AD | 86% (AD vs. NC) 62% (MCI vs. AD) |
| Chu et al. [17] | Four different feature selection methods were compared using SVM for classification. | NC MCI AD | 85% (AD vs. NC) 67% (MCI vs. NC) |
| Dyrba et al. [26] | SVM and NB were applied upon the DTI dataset. | healthy demented | 83% (SVM) 75% (NB) |
| Challis et al. [13] | Bayesian Gaussian process logistic regression was used for classification. | NC MCI | 80% |
| Huang et al. [39] | Custom multilayer classifier that was able to incorporate spatial information of raw MRI voxels. | NC AD | 79% |
| Moradi et al. [53] | Features were selected by using sparse logistic regression. After that, instances were classified with hierarchical classification framework that was consisted of multiple binary classifiers. | healthy MCI Alzheimer | 77% |
| Burge et al. [11] | Causal relationships were extracted by using DBN. Based on those relationships, neural–anatomical networks were used for classification. | healthy demented | 73% |
| Meszlényi et al. [51] | Convolutional neural network for functional connectome classification was used to classify fMRI scans. | NC MCI | 72% |
| Gomar et al. [34] | ANOVA and t-tests were used to select significant features that were later used to build Logistic Regression model. | MCI AD | 72% |
| Szenkovits et al. [72] | Features are selected with new mGA method and 1-NN classifier is used for its evaluation. | NC MCI | 63% |

**Table 2:** An overview of classification methods and accuracies that were obtained from MRI data.

in proteins are statistically significant. It carries out a variant of t-test for each

protein to measure its relationship to the response variable. They identified 18 proteins that were significantly different between AD and NC. Their biological analysis showed that they belong to two independent regulatory networks connecting them. The pathways that connect the proteins define their interdependence. Any dysfunction of such pathway prevents normal generation of proteins and can indicate early onset of a disease like AD.

Llano et al. [49] compared performance of multiple methods for feature selection and classification tasks. They analysed prediction power of proteins for classification of AD vs. NC using ANCOVA and t-tests. The significant predictors were additionally subjected to one of the feature selection algorithms: RF, Partial Least Squares, Bagging and Simulated Annealing. The optimal subset was then used for building classification model using one of the classification algorithms: Diagonal LDA, RF, SVM, ANN, Partial Least Squares, Bagging and $k$NN. The best classification performance was achieved for all optimal subsets using the RF model.

EEG signals have been, however, used more rarely. Informativeness in their raw form is questionable due to a high amount of noise, and sometimes information is also lost due to a compression process [10]. Authors of [61] measured EEG signals on 19 different positions of the skull and transformed them into frequency domain using the Fourier transform. Each of the 19 transformed signals was divided into 5 frequency sections and average frequencies were calculated. Additional non-linear features were constructed by estimating correlation using Takens-Ellner method. All features were joined and used as an input for Discriminant Analysis, $k$NN and ANN. Best performance was achieved using ANN with AD vs. NC classification accuracy reaching 92%. Additional analysis of frequency sections showed that brain activity significantly dropped for subjects with AD.

### 2.4 Classification using data fusion

Integration of multi source data calls for a more complex algorithm to extract the available information. Data fusion methods that solve these problems can be divided into three main groups – early, intermediate and late integration methods – depending on the point where the fusion takes place within the supervised learning framework. Early integration transforms all data sources into one table that can be treated as single data set. Intermediate integration relies on algorithms that can take the data sources and fuse them on the level of a predictive model. In late integration a predictive model is constructed for each of the data sources. Final prediction is determined by joining predictions of all models by model weighting.

Zhang et al. [80] fused MRI, PET and CSF biomarkers using methods of early and intermediate data integration. They proposed a method for multimodal multi-task learning that could perform regression and classification at the same time. The core of the method comprised a SVM with kernel that integrates multimodal data. Early integration was done with simple data con-

catenation that produced one table. The table was then fed into the SVM. Both methods were used for classification of AD vs. NC and MCI vs. NC and achieved better results than simple concatenation.

R. Gray et al. [35] used intermediate integration for fusing MRI, PET, CSF and gene data. They built the RF model for each modality. The models were used to calculate similarities between samples. This produced similarity matrix for each of the modalities that were linearly combined to assemble manifold representation. The result was used to build final RF model for predicting disease diagnosis. The proposed method was significantly better than models using only one of the modalities.

Suk et al. [71] fused MRI and PET data with the use of Deep Boltzmann Machine (DBM), which is an extension of an ANN called Restricted Boltzmann Machine. A DBM is constructed by stacking multiple RBMs in a hierarchical manner. Their proposed method selected significant voxel patches from the images and fed them into the DBM. The DBM extracted relevant features from each modality and fused them on its lowest level. The output was then used in a hierarchical framework proposed by Liu et al. [19] with a SVM as a weak classifier. Data fusion improved classification accuracy of framework for both classification tasks AD vs. NC and MCI vs. NC, which reached 95% and 85%, respectively. Another valuable contribution was confirmation of brain regions commonly associated with AD.

Authors of [60] used late data integration method to fuse information of multiple EEG channels. Their proposed method uses ensemble for each modality. Final decision is based on adjusted weighted majority voting. The prediction accuracy of fused channels using the proposed method was 79%. It was significantly higher than both, the accuracy of single channel predictions and the accuracy of channel data concatenated in one table.

### 2.5 Comparative analyses of classification algorithms

To explore how different statistical and machine-learning algorithms contribute to early diagnosis of dementia, researchers conducted comparative analyses. In this section we present papers that made comparisons between different classification algorithms.

Chen et al. [15] made an experiment with seven different supervised learning approaches: naive Bayes, Bayesian-network classifier with inverse tree structure, decision tree, support vector machine, multiple-layer perceptron, discriminant analysis and logistic regression. These experiments were carried out on structural MRI scans features of patients with very mild dementia and patients serving as healthy controls. Features were extracted using image processing tools that provided insight into regional brain volumes, regional gray-matter volumes, etc. Final results of different approaches turned out to depend on selected feature sets, but the overall impression was that they were not significantly different from each other. The most notable result, however, was the

accuracy of over 85%. That once again proved that dementia can be detected in the early phase.

Williams et al. [75] wanted to improve accuracy of classification of mild cognitive impairment and dementia. In order to do that, they experimented with: naive Bayes, C4.5 decision tree, back-propagation neural network and support vector machine classifiers. On top of that they used a wrapper feature selection technique and proved that proper selection of features can improve classification accuracy on the studied problem. Data that was used here was collected across two studies [67,66]. Their experiments showed that the naive Bayes classifier reached the highest accuracy, while the support vector machine classifier reached the lowest accuracy. The best obtained accuracies in this paper were slightly above 80%. The problem of low accuracies for mild cognitive impairment class remained unresolved.

Even though many Alzheimer's disease classification approaches were applied and examined in many different papers, those approaches could not be easily mutually compared due to variety of datasets. In order to conduct reliable comparison of different approaches, Cuingnet et al. [21] applied different classification experiments on a single dataset from the ADNI database. They tested three basic feature extraction approaches: voxel-based, vertex-based and ROI-based. Voxel-based approach implies that the features are defined at the level of the MRI voxel. Features of each voxel represent the probability of different tissue classes (grey matter, white matter and cerebrospinal fluid). In the second category, i.e. vertex-based approach, features are defined at the vertex-level on the cortical surface [23,62]. Finally, ROI-based approaches capture features of different regions of interest, and they are limited only to hippocampus [20,19,18,33]. Besides different feature extraction techniques, authors also analyzed different classification tasks. They analyzed three binary classification tasks: control subjects (CN) vs. patients with AD, CN vs. mild cognitive disease (MCI) patients who had converted to AD within 18 months (MCI converters - MCIc), and MCIc vs. MCI patients who had not converted to AD within 18 months (MCI non-converters - MCInc). Results of those experiments showed that the highest recall values (in range 59% - 82%) were obtained for classification task CN vs. AD, while the lowest recall values were obtained for classification task MCInc vs. MCIc (in range 0% - 70%). Additional thing to note is that CN vs. AD, and CN vs. MCIc classifiers perform better if the features are extracted with voxel-based approaches, while the MCInc vs. MCIc classifier performs better with ROI-based feature extraction approaches.

### 2.6 Unconventional classification approaches

As it can be seen in the previous sections, many authors devoted their research to exploration of different classification settings upon similar data, in order to achieve the same task – detection of subjects with cognitive diseases. Now we would like to introduce a few papers that are at some extent different.

The Clock Drawing Test (CDT [5]) is used in order to detect cognitive dysfunctions. During the test, subjects are required to draw the numbers on a clock face and set the hands to a time given by the medical examiner. Bennasar et al. [9] performed an interesting research in which they automated analysis of clock drawings in order to differentiate subjects between three levels of dementia: normal, mild cognitive impairment and moderate/severe dementia. The research was based on 604 clock drawings provided by the Memory Clinic at the Llandough Hospital in Cardiff, UK. Authors extracted 47 visual features from these drawings and used them to train and test a classifier. Feature set was comprised of the majority of the features employed in the most common CDT scoring systems, as well as new geometrical features that were based on additional data analysis. For classification, the authors used a cascade classifier that is actually a concatenation of multiple classifiers. Three binary classifiers that were combined within cascade classifier were: Classifier 1 that discriminates drawings into normal+ and abnormal cases, Classifier 2 that takes abnormal cases from Classifier 1 and differentiates the MCI diagnosis from severe and moderate dementia, and Classifier 3 that takes normal+ cases from Classifier 1 and decides if the case is normal or functional (without dementia, but with depression and anxiety). For each of these three classifiers the authors examined different classification algorithms (k-nearest neighbors, least squares Support Vector Machine and Random Forest) and presented comparative results. Finally, in this research the highest accuracy that was achieved for ternary classification was around 78%, while for the single-stage classification accuracy was around 71%.

Another interesting approach was presented by Gerrard et al. [32] who used textual speech transcripts in order to detect subjects with semantic dementia. Samples of connected speech are elicited by showing the picnic picture to the subject, together with instruction: "Have a look at the picture, tell me what you see, and try to talk in sentences." Dataset had 32 semantic dementia patients and 10 healthy controls. In this research the authors used naive Bayes Gaussian and naive Bayes multinomial classifiers. They represented speech transcripts as a bag-of-words and extracted most relevant features according to the information gain. Binary classification obtained by this setting achieved accuracy of above 90%.

Buza et al. [12] also based their research on patients' speech capabilities in order to estimate the severity of Parkinson's disease. The most commonly used scale for Parkinson's disease severity is the Unified Parkinson's Disease Rating Scale (UPDRS). The authors' main aim was to estimate UPDRS score of Parkinson's disease patients by using biomedical voice recordings. Evaluation was performed by measuring performance of 6 different neural networks using the mean absolute error (MAE) and root-mean-square error (RMSE). Final results suggested that hubness-aware artificial neural network systematically improves UPDRS estimation accuracy.

In some cases, areas and domains collected data contain a lot of missing values. They often have a high influence on classification accuracy. Zaffalon et al. [79] propose a way to deal with incomplete data by proposing the Naive

Credal Classifier [77,78] which generalizes the discrete Naive Bayes classifier. Naive Credal Classifier takes into account imprecision that could be a consequence of a too small sample size, missing data, etc. In case of uncertainty, this classifier assigns multiple classes to a given instance. In such a way it overcomes problems that arise due to existence of incomplete data. The authors tested their idea on data obtained from the Cognitive Drug Research computerized assessment system. Data of this system were not collected with the specific purpose of analysis, so they have a significant amount of missing values. Binary classification experiments (healthy vs. demented) showed that this classifier achieves accuracy of 94.77% on the subset of instances where it predicted a single class (this subset consists of 90.3% of all instances), while the Naive Bayes classifier achieves accuracy of 92.41% on the entire test set and accuracy of 70.37% on the subset of instances for which the Naive Credal classifier predicted multiple different classes.

Galili et al. [31] chose a completely unconventional approach to classification of subjects to disease categories. The main principle behind their idea was the distrust in the original disease diagnosis. Using the clinical measurements and the originally assigned disease diagnosis they grouped examples into 10 categories by performing the k-means clustering. Afterward, they assigned a new classification label to examples and built a classification tree using the CART algorithm.

### 2.7 Prediction and early detection of cognitive diseases

One of the most valuable and hard tasks is to detect a cognitive disease even before symptoms arise. At the moment when symptoms start to manifest, patient can already suffer from serious and incurable brain damage. To predict dementia, Klein et al. [46] experimented with a very specific database of MRI brain scans of 29 subjects. These subjects were asymptomatic at the time of scanning, but were diagnosed as having dementia within 0.7 to 5 years after the scan. Besides these 29 subjects, the database also contained 29 healthy controls. A $k$NN classifier was ran upon this data, achieving classification accuracy of 81%. This is a very promising result, since it confirms a possibility of a very early detection of dementia.

Besides focusing on supervised learning with classification, there are also other machine learning approaches that helped obtain important information about cognitive diseases. For example, Joshi et al. [43] determined the major risk factors for development of the Alzheimer's disease, vascular dementia and Parkinson's disease. The authors used the chi-square attribute evaluation scheme with ranker search method [14,40], along with four other attribute evaluation schemes. The obtained rankings of the attributes are given in Table 3. The results showed that genes, diabetes, age, smoking and stroke, seem to be the most influential attributes for this problem.

| | Chi-Square (Weights) | GainRatio (Weights) | InfoGain (Weights) | ReliefF (Weights) | Symmetrical Uncert (Weights) |
|---|---|---|---|---|---|
| 1 | Genes (0.85) | Diabetes (0.91) | Genes (15.79) | Genes (12.08) | Genes (6.85) |
| 2 | Smoking (0.74) | Genes (0.89) | Age (14.85) | Diabetes (10.67) | Diabetes (6.64) |
| 3 | Age (0.72) | Stroke (0.88) | Diabetes (10.85) | Stroke (10.68) | Stroke (6.62) |
| 4 | Diabetes (0.71) | Age (0.72) | Stroke (10.66) | Smoking (8.96) | Smoking (6.02) |
| 5 | Stroke (0.68) | Smoking (0.71) | Smoking (9.01) | Age (6.82) | Age (6.00) |
| 6 | Family_Hist (0.63) | Family_Hist (0.63) | Family_Hist (7.65) | Family_Hist (5.63) | Family_Hist (5.59) |
| 7 | Alcohol (0.48) | Alcohol (0.51) | Alcohol (5.49) | Alcohol (4.89) | Alcohol (5.29) |
| 8 | LDL (0.32) | Hist_heart _Disease (0.31) | Hypertension (4.06) | Hypertension (4.12) | Hypertension (4.08) |
| 9 | Hypertension (0.12) | Hypertension (0.22) | Hist_heart _Disease (1.09) | Hist_heart _Disease (2.09) | Hist_heart _Disease (2.85) |
| 10 | Hist_heart _Disease (0.09) | LDL (0.18) | LDL (1.01) | LDL (1.16) | LDL (1.85) |
| 11 | BMI (0.06) | BMI (0.06) | BMI (0.09) | BMI (0.06) | BMI (0.11) |

**Table 3:** Results of various attribute evaluation methods on the major risk factors for Alzheimer's disease, vascular dementia and Parkinson's disease. [43]

Miller et al. [52] used the OASIS dataset to train and test Support Vector Machine classifier which had a task of classifying very mild to moderate dementia. They obtained the accuracy of 85%.

Hosseini-Asl et al. [38] used a deep 3D convolutional neural network (3D-CNN) that is built upon a 3D convolutional autoencoder, with the final goal of predicting the Alzheimer's disease (AD). Interesting part of their research is fine-tuning of fully connected upper neural network layers for each task-specific AD classification. They ran experiments upon the ADNI dataset and achieved accuracy of 77% for ternary classification.

Zhang et al. [81] used MRI scans based on eigenbrain in order not only to detect patients with the Alzheimer's disease, but also to detect brain regions that are related to AD. They detected 30 brain regions that are related to the disease: Anterior Cingulate, Caudate Nucleus, Cerebellum, Cingulate Gyrus, Claustrum, Inferior Frontal Gyrus, Inferior Parietal Lobule, Insula, Lateral Ventricle, Lentiform Nucleus, Lingual Gyrus, Medial Frontal Gyrus, Middle Frontal Gyrus, Middle Occipital Gyrus, Middle Temporal Gyrus, Paracentral Lobule, Parahippocampal Gyrus, Postcentral Gyrus, Posterial Cingulate, Precentral Gyrus, Precuneus, Subcallosal Gyrus, Sub-Gyral, Superior Frontal Gyrus, Superior Parietal Lobule, Superior Temporal Gyrus, Supramarginal Gyrus, Thalamus, Transverse Temporal Gyrus, and Uncus. In this research the authors used the Support Vector Machine classifier on the OASIS database.

Patients with the Parkinson's disease have a high risk for developing the dementia. Morales et al. [54] dedicated their research to detecting dementia particularly in patients with the Parkinson's disease. They used Bayesian network classifiers in order to classify subjects into one of the next three classes: PDCI (cognitively intact patients with Parkinson's disease), PDMCI (Parkinson's disease patients with mild cognitive impairment) and PDD (Parkinson's disease patients with dementia). Dataset consists of MRI scans of 45 subjects (16 PDCI, 15 PDMCI and 14 PDD). The highest obtained accuracy for the mentioned ternary classification was 70%. Besides that, the authors noted that the most relevant variables related to dementia in PD were cerebral white matter and volumes of the lateral ventricles and hippocampi.

## 2.8 Summary of approaches and data sources

Fig. 2 contains average classification accuracies calculated on results of all presented papers. The figure confirms that MCI classification task is the most challenging of all.

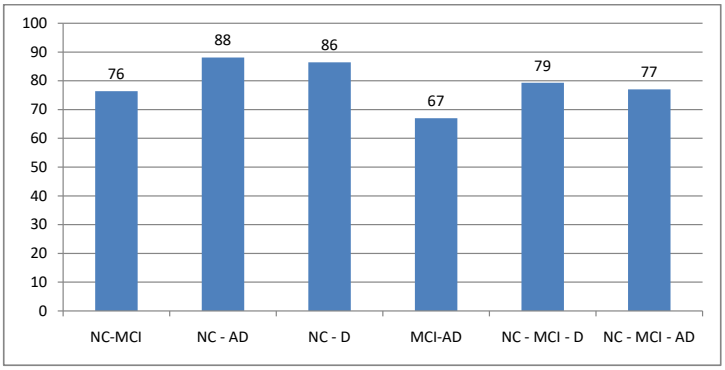


**Fig. 2:** Average classification accuracies of different classification tasks calculated on results of all presented papers.

Table 4 summarizes all methods and data sources used in the papers from previous subsections. Table rows correspond to methods and the columns correspond to data sources. Table cells contain references to papers that use corresponding method and data source. Enumeration of methods is explained in table caption.

| Method | Demographic data | Medical data | Neuropsychological tests | BOMC | FAQ | MMSE | ICP | CERAD | Wechsler Memory | WAIS-R | CDT | Voice recordings | ApoE | CSF biomarkers | Blood plasma proteins | EEG | eTIV | Atlas scaling factor | nWBV | MRI | MRI ROIs | MRI voxels | MRI - Cortical thickness | PET | PET ROIs | PET voxels | DTI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | [22] [68] | [22] [68] | | | | | | [9] | | | | [49] | [61] | | | | [46] | [72] | | | | | | |
| 2 | [69] [70] [24] [75] | [70] | [75] | [22] [68] | [22] [68] [69] | [69] [70] | [69] | [70] | [70] | [70] | | [32] | | | [24] | | | | | [15] | | | | | | | [26] |
| 3 | [24] [75] | | [75] | | | | | | | | [9] | | | [80] | [24] [49] | | | | | [21] [52] [11] [13] [15] [81] | [80] [51] | [71] [48] [17] | | | [36] [80] | [71] [50] | [26] |
| 4 | [69] [70] | [70] | | [22] [68] | [22] [68] [69] | [69] [70] | [69] | [70] | [70] | [70] | | | | | | | | | | | | | | | | | |
| 5 | [24] | [31] [43] | [43] | | | | | | | | [9] | | [35] [31] | [35] [31] | [24] [49] | | | | | | [35] [31] | | | | [31] | [35] | |
| 6 | [75] | | [75] | | | | | | | | | | | | | | | | | [15] | | | | | | | |
| 7 | | [43] | [43] | | | | | | | | | | | | [49] | | | | | | | | | | | | |
| 8 | | | [79] | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | | | | | | | [37] | | | | [37] | |
| 10 | | | | | | | | | | | [9] | | | | | | | | | [53] | | | | | | | |
| 11 | | | | | | | | | | | | | | | | | | | | [53] | | | | | | | |
| 12 | | | | | | | | | | | | | | | | | | | | | | [71] | | | | [71] | |
| 13 | | | | | | | | | | | | | | | | | | | | | | [48] | | | | | |
| 14 | [73] [34] | [34] | [73] [28] | | | | | | | | | | [73] [34] | [34] [28] | | | | | | [15] [13] | [34] [28] | | | | | | |
| 15 | | | [28] | | | | | | | | | | | [28] | | | | | | | [28] | | | | | | |
| 16 | | [43] | [43] | | | | | | | | | | | | | | | | | | | | | | | | |
| 17 | [69] | | | | [69] | [69] | [69] | | | | | | | | | | | | | | | | | | | | |
| 18 | [70] | [70] [31] | | | | [70] | | [70] | [70] | [70] | | | [31] | [31] | | | | | | | [31] | | | | [31] | | |
| 19 | [73] [34] | [34] | [73] | | | | | | | | | | [73] [34] | [34] | | | | | | | [34] | | | | | | |
| 20 | [73] | | [73] | | | | | | | | | | [73] | | | | | | | | | | | | | | |
| 21 | | | | | | | | | | | | | | | | | | | | | [47] | | | | | | |
| 22 | [75] | [43] | [75] [43] | | | | | | | | | [12] | | | [49] | [61] | | | | | [51] | | | | [45] | [50] | |
| 23 | [57] | | | | | [57] | | | | | | | | | | | [57] | [57] | [57] | [57] [15] | | | | | | | |
| 24 | | | | | | | | | | | | | | | | | | | | [54] | | | | | | | |
| 25 | | | | | | | | | | | | | | | | | | | | [15] | | | | | | | |
| 26 | | | | | | | | | | | | | | | | | | | | [38] [58] | [51] | | | | | | |
| 27 | | | | | | | | | | | | | | | | | | | | [11] | | | | | | | |
| 28 | | | | | | | | | | | | | | | | | | | | [11] | | | | | | | |
| 29 | | [31] | | | | | | | | | | | [31] | [31] | | | | | | | [31] | | | | [31] | | |
| 30 | | [43] | [43] | | | | | | | | | | | | | | | | | | | | | | | | |
| 31 | [75] | | [75] | | | | | | | | | | | | | | | | | | | | | | | | |
| 32 | [24] | | | | | | | | | | | | | | [24] | | | | | | [72] | | | | | | |
| 33 | [73] | | [73] | | | | | | | | | | [73] | | | | | | | | | | | | | | |
| 34 | | | | | | | | | | | | | | | | | | | | | | [17] | | | | | |
| 35 | | | | | | | | | | | | | | | [49] | | | | | | | | | | | | |
| 36 | | | | | | | | | | | | | | | [49] | | | | | | | | | | | | |
| 37 | | | | | | | | | | | | | | | | | | | | | | | [16] | | | [50] | |
| 38 | | | | | | | | | | | | | | | [49] | [61] | | | | [15] | | | [16] | | [45] | [50] | |
| 39 | | | | | | | | | | | | | | [80] | | | | | | | [80] | | | | [80] | | |

| Method | Demographic data | Medical data | Neuropsychological tests | BOMC | FAQ | MMSE | ICP | CERAD | Wechsler Memory | WAIS-R | CDT | Voice recordings | ApoE | CSF biomarkers | Blood plasma proteins | EEG | eTIV | Atlas scaling factor | nWBV | MRI | MRI ROIs | MRI voxels | MRI - Cortical thickness | PET | PET ROIs | PET voxels | DTI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | | | | | | | | | | | | | | [80] | | | | | | | [80] [72] [51] | | | | [80] | | |
| 41 | | | | | | | | | | | | | | | | | | | | | | [17] | | | | | |
| 42 | [34] | [34] | | | | | | | | | | | [34] | [34] | [49] | | | | | | [34] | [48] [37] [71] [17] | | | | [37] [71] | |
| 43 | [34] | [34] | | | | | | | | | | | [34] | [34] | | | | | | | [34] | | | | | | |
| 44 | | | | | | | | | | | | | | | [64] | | | | | | | | | | | | |
| 45 | | | | | | | | | | | | | | | [64] | | | | | | | | | | | | |
| 46 | | | | | | | | | | | | | [35] | [35] | | | | | | | [35] | | | | | [35] | |
| 47 | | [43] | [43] | | | | | | | | | | | | | | | | | | | | | | | | |
| 48 | | | | | | | | | | | | | | | | | | | | [38] [58] | | | | | | | |
| 49 | | | | | | | | | | | | | | | | | | | | [81] | | | | | | | |
| 50 | | | | | | | | | | | | | [35] | [35] | | [60] | | | | | [35] | | | | | [35] | |

**Table 4:** Overview of all data sources and methods that were used in the presented papers. The methods are represented by numbers 1, 2, ..., 50, as follows: **1** – kNN; **2** – NB; **3** – SVM; **4** – C4.5; **5** – RF; **6** – Decision tree; **7** – Bagging; **8** – NCC; **9** – LPBoost; **10** – Cascade classification; **11** – Low density separation; **12** – Deep Boltzmann Machine; **13** – RPSE_SRC; **14** – Logistic regression; **15** – Cox regression; **16** – BF tree; **17** – FOCL; **18** – CART; **19** – ANOVA; **20** – ANCOVA; **21** – MANOVA; **22** – NN; **23** – MLP; **24** – BNC; **25** – BNCIT; **26** – CNN; **27** – DBN; **28** – GNBN; **29** – K-means clustering; **30** – Filter feature selection; **31** – Wrapper feature selection; **32** – Custom feature selection; **33** – Intercorrelation matrix; **34** – ROI filtering; **35** – Partial least squares; **36** – Simulated annealing; **37** – Principal component analysis; **38** – Discriminant analysis; **39** – Multi-task feature selection; **40** – Lasso feature selection; **41** – Recursive feature elimination; **42** – T-test; **43** – Chi-squared test; **44** – SAM; **45** – PAM; **46** – Multidimensional scaling; **47** – Decision table; **48** – 3D convolutional autoencoder; **49** – Eigenbrains; **50** – Data fusion.

## 3 Frequently used datasets

A lot of experiments that are mentioned and used in the analyzed papers are based on private and unpublished datasets collected by authors of those papers which however are not publicly available. Nevertheless, some datasets have been made publicly available for the scientific community. This section gives a brief overview of such public datasets that are frequently used for the purpose of applying different machine learning algorithms in a domain of cognitive diseases.

A well-known repository is UCI ADRC, which provides three different datasets: minimum dataset (MDS) [8], uniform dataset (UDS) [55,7,74], and neuropathology dataset (NP) [8]. All these datasets contain basic patient information as well as scores of different tests. MDS contains 250 instances and 67 features, UDS contains 460 instances and 725 attributes, and NP contains 520 instances and 85 attributes.

There are also OASIS datasets [4] from the Washington University Alzheimer's Disease Research Center, dr. Randy Buckner at the Howard Hughes Medical Institute (HHMI) at the Harvard University, the Neuroinformatics Research Group (NRG) at the Washington University School of Medicine, and the Biomedical Informatics Research Network (BIRN). These datasets are freely available to the scientific community, and they consist of MRI brain scans and the scores from different tests and questionnaires.

Alzheimers Disease Neuroimaging Initiative (ADNI) [2] also provides datasets that are relevant to cognitive diseases studies. It 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), some private pharmaceutical companies and some non-profit organizations, as a 60 million US dollars, 5-year public-private partnership. Their goal was to test whether MRI data, PET data, biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. It also contains various kind of patients' data, such as clinical data, genetics data, MRI image data, etc.

AddNeuroMed [1] study began as a public-private partnership for biomarker discovery in Alzheimer's Disease. Data for this longitudinal study was collected from around the Europe during the period from 2006 to 2008. Participating institutions came from Italy, Finland, Greece, United Kingdom, Poland and France. The study enrolled a total of 781 subjects, however not all with complete data at each assessment. Datasets contain cognitive assessments, clinical data, MRI, genetic data and blood biomarkers.

The Australian Imaging, Biomarker and Lifestyle (AIBL) [27] flagship study of aging is a study to improve understanding of development of Alzheimer's Disease. Its focus lies in early diagnosis, identification of factors that may predict onset of AD which includes biomarkers, cognitive characteristics, health and lifestyle factors and possible prevention mechanisms. Their datasets include cognitive assessments, clinical data, blood and CSF biomarkers, lifestyle data, MRI and PET scans.

Datasets from all these sources either already contain various features or they can be extracted using tools such as Freesurfer, Mindboogle, etc. Clinical datasets contain personal information about each patient. This usually includes demographic data (sex, age, years of education, etc.), family history (diseases in family, data about relatives, etc.), sensory and motory tests, and cognitive assessments. Usually several tests are performed to obtain the most accurate possible disease diagnosis. The most common tests are Mini Mental State exam (MMSE), Neuropsychiatric Inventory (NPI), Sensory and Motory Impairment test, Clinical Dementia Rating (CDR), Cambridge Mental Disorders of the Elderly Examination (Camdex), Health related quality of life test (DEMQOL), etc. Some of these tests, especially those related to the quality of life, are taken by both caretaker and caregiver.

Blood and cerebrospinal fluid biomarkers datasets contain quantity measurements of proteins in respective fluid. The most commonly measured pro-

tein levels are tau and amyloid beta since their abnormal accumulation in brains is the reason for atrophy of the brain.

Neuroimaging datasets usually contain raw brain scans. There are three different feature extraction approaches that are commonly used. The first one is voxel-based and uses voxel intensities as immediate features. The second is vertex-based and produces cortical thickness, sulcal depth, or cortical surface area, cortical surface curvature and convexity as features. The last one takes well-known brain regions of interest and calculates features for larger formations of voxels like volumes and thicknesses.

## 4 Open challenges in the field

The today's biggest, yet unresolved challenge is that the classification of cognitively impaired patients and patients with mild dementia is still not reliable enough. Different authors tried to cope with this problem, but none of them achieved strong breakthroughs. The past research does, however, reveal that data obtained from the MRI scans is more suitable for this problem than data obtained from questionnaires, as they seem to be better predictors for the early detection of cognitive diseases.

The improvement in the field would be to introduce machine learning systems into everyday medical practice and with significant number of patients health records for a large number of patients. Many papers reported significant results, but, as far as we know, none of them reported outcome of a real system that actually helps clinical staff in their everyday practice and decision support.

The majority of researches are dedicated to diagnosing diseases, which is undoubtedly a very important task. Nevertheless, a revolutionary opportunity to use machine learning system as a treatment planning assistant tool still remains fully unexplored. This topic is also an open problem that should be explored more deeply in the future.

## 5 Conclusion

Cognitive diseases are affecting large number of people. Given that they are incurable, it is of a high importance to detect them in their early stage. Much effort has been devoted to improving diagnostics of cognitive diseases, and machine learning techniques provide one way of dealing with this problem. In this paper we have presented various machine learning techniques and approaches that deal with the problem of improving classification accuracy of patients with cognitive diseases, which has been a goal for the past 20 years.

We have shown that promising results can be achieved on various kind of input data. Machine learning community has mostly experimented with questionnaire data and with MRI data. The goal of those papers was not only to improve classification accuracy, but also to point out some interesting and

useful facts about underlying processes (e.g. to determine the best classification algorithm, to detect relevant and irrelevant features, etc.). Besides these conventional kind of data, many experiments were conducted upon datasets that are not primarily used together with cognitive diseases, such as blood sample data and EEG signals. Results have shown that even such data can contain valuable information for predicting cognitive diseases.

Data fusion became a useful technique for knowledge acquisition in cases when two or more related datasets are available. We presented research papers that used data fusion of datasets, which successfully led to an increase of the classification accuracy. Many research papers are also devoted to comparative analyses of different machine learning algorithms upon cognitive diseases data. Besides comparing the machine learning approaches we also presented the most commonly used datasets.

Even though a lot of valuable results in this field have already been obtained, there is still a room for improvements. Classification of patients with mild dementia is still not reliable. In addition to that, as far as we know, none of the papers reported usage and outcomes of a machine learning expert systems in the everyday practice.

We hope that this overview paper can be a valuable starting point for many researchers who would either like to start or continue with a further exploration of this field. Besides that, numerous methods that are presented in this paper, can be inspiration for some new, unexplored ideas that would improve quality of automated diagnostics, prediction, and treatment planning for cognitive diseases.

## 6 Compliance with Ethical Standards

**Funding:** This work is supported by the bilateral project between Slovenia and Serbia: BI-RS/16-17-047, "Intelligent computer techniques for improving medical detection, analysis and explanation of human cognition and behavior disorders" between the Ministry of Education, Science and Technological Development of the Republic of Serbia and the Slovenian Research Agency. B. Bratić, V. Kurbalija and M. Ivanović also thank the Ministry of Education, Science and Technological Development of the Republic of Serbia for additional support through project no. OI174023, "Intelligent techniques and their integration into wide-spectrum decision support".

**Conflict of Interest:** The authors declare that they have no conflict of interest.

**Ethical approval:** This article does not contain any studies with human participants or animals performed by any of the authors.

## References


1. The addneuromed study. https://www.synapse.org/#!Synapse:syn2790911/wiki/235387
2. Alzheimers disease neuroimaging initiative (ADNI) database. http://adni.loni.usc.edu
3. Challenge on computer-aided diagnosis of dementia based on structural MRI data. http://caddementia.grand-challenge.org/home
4. Open access series of imaging studies (OASIS) database. http://www.oasis-brains.org
5. Agrell, B., Dehlin, O.: The clock-drawing test. Age and ageing **27**(3), 399–404 (1998)
6. Aha, D.W., Kibler, D., Albert, M.K.: Instance-based learning algorithms. Machine learning **6**(1), 37–66 (1991)
7. Beekly, D.L., Ramos, E.M., Lee, W.W., Deitrich, W.D., Jacka, M.E., Wu, J., Hubbard, J.L., Koepsell, T.D., Morris, J.C., Kukull, W.A., et al.: The National Alzheimer's Coordinating Center (NACC) database: the uniform data set. Alzheimer Disease & Associated Disorders **21**(3), 249–258 (2007)
8. Beekly, D.L., Ramos, E.M., Van Belle, G., Deitrich, W., Clark, A.D., Jacka, M.E., Kukull, W.A., et al.: The National Alzheimer's Coordinating Center (NACC) database: an Alzheimer disease database. Alzheimer Disease & Associated Disorders **18**(4), 270–277 (2004)
9. Bennasar, M., Setchi, R., Hicks, Y., Bayer, A.: Cascade classification for diagnosing dementia. In: Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on, pp. 2535–2540. IEEE (2014)
10. Birvinskas, D., Jusas, V., Martisius, I., Damasevicius, R.: Fast DCT algorithms for EEG data compression in embedded systems. Computer Science and Information Systems **12**(1), 49–62 (2015)
11. Burge, J., Clark, V.P., Lane, T., Link, H., Qiu, S.: Bayesian classification of FMRI data: Evidence for altered neural networks in dementia. University of New Mexico, Tech. Rep. TR-CS-2004-28 (2004)
12. Buza, K., Varga, N.Á.: Parkinsonet: estimation of updrs score using hubness-aware feedforward neural networks. Applied Artificial Intelligence **30**(6), 541–555 (2016)
13. Challis, E., Hurley, P., Serra, L., Bozzali, M., Oliver, S., Cercignani, M.: Gaussian process classification of alzheimer's disease and mild cognitive impairment from resting-state fmri. NeuroImage **112**, 232–243 (2015)
14. Chan, K.C.: A statistical technique for extracting classificatory knowledge from databases. Knowledge discovery in databases pp. 107–123 (1991)
15. Chen, R., Herskovits, E.H.: Machine-learning techniques for building a diagnostic model for very mild dementia. Neuroimage **52**(1), 234–244 (2010)
16. Cho, Y., Seong, J.K., Jeong, Y., Shin, S.Y., Initiative, A.D.N., et al.: Individual subject classification for alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage **59**(3), 2217–2230 (2012)
17. Chu, C., Hsu, A.L., Chou, K.H., Bandettini, P., Lin, C., Initiative, A.D.N., et al.: Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage **60**(1), 59–70 (2012)
18. Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., Benali, H., Garnero, L., Colliot, O.: Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus **19**(6), 579–587 (2009)
19. Chupin, M., Hammers, A., Liu, R.S., Colliot, O., Burdett, J., Bardinet, E., Duncan, J.S., Garnero, L., Lemieux, L.: Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. Neuroimage **46**(3), 749–761 (2009)
20. Colliot, O., Chételat, G., Chupin, M., Desgranges, B., Magnin, B., Benali, H., Dubois, B., Garnero, L., Eustache, F., Lehéricy, S.: Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. Radiology **248**(1), 194–201 (2008)
21. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., Initiative, A.D.N., et al.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. neuroimage **56**(2), 766–781 (2011)

22. Datta, P., Shankle, W.R., Pazzani, M.: Applying machine learning to an alzheimer's database. In: Artificial Intelligence in Medicine: AAAI-96 Spring Symposium, pp. 26–30 (1996)
23. Desikan, R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., Greve, D.N., Salat, D.H., Buckner, R.L., et al.: Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain **132**(8), 2048–2057 (2009)
24. Doecke, J.D., Laws, S.M., Faux, N.G., Wilson, W., Burnham, S.C., Lam, C.P., Mondal, A., Bedo, J., Bush, A.I., Brown, B., et al.: Blood-based protein biomarkers for diagnosis of alzheimer disease. Archives of neurology **69**(10), 1318–1325 (2012)
25. DUDA/HART: Pattern classification and scene analysis. John Wiley. (1973)
26. Dyrba, M., Ewers, M., Wegrzyn, M., Kilimann, I., Plant, C., Oswald, A., Meindl, T., Pievani, M., Bokde, A.L., Fellgiebel, A., et al.: Robust automated detection of microstructural white matter degeneration in Alzheimer's disease using machine learning classification of multicenter DTI data. PloS one **8**(5), e64925 (2013)
27. Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., et al.: The australian imaging, biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease. International Psychogeriatrics **21**(4), 672–687 (2009)
28. Ewers, M., Walsh, C., Trojanowski, J.Q., Shaw, L.M., Petersen, R.C., Jack, C.R., Feldman, H.H., Bokde, A.L., Alexander, G.E., Scheltens, P., et al.: Prediction of conversion from mild cognitive impairment to alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiology of aging **33**(7), 1203–1214 (2012)
29. Fillenbaum, G.G., Heyman, A., Wilkinson, W.E., Haynes, C.S.: Comparison of two screening tests in Alzheimer's disease: The correlation and reliability of the mini-mental state examination and the modified blessed test. Archives of Neurology **44**(9), 924–927 (1987)
30. Folstein, M.F., Folstein, S.E., McHugh, P.R.: "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research **12**(3), 189–198 (1975)
31. Galili, T., Mitelpunkt, A., Shachar, N., Marcus-Kalish, M., Benjamini, Y.: Categorize, cluster, and classify: a 3-c strategy for scientific discovery in the medical informatics platform of the human brain project. In: International Conference on Discovery Science, pp. 73–86. Springer (2014)
32. Garrard, P., Rentoumi, V., Gesierich, B., Miller, B., Gorno-Tempini, M.L.: Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. Cortex **55**, 122–129 (2014)
33. Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.S., Niethammer, M., Dubois, B., Lehéricy, S., Garnero, L., et al.: Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage **47**(4), 1476–1486 (2009)
34. Gomar, J.J., Bobes-Bascaran, M.T., Conejero-Goldberg, C., Davies, P., Goldberg, T.E., Initiative, A.D.N., et al.: Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to alzheimer disease in patients in the alzheimer's disease neuroimaging initiative. Archives of general psychiatry **68**(9), 961–969 (2011)
35. Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D., Initiative, A.D.N., et al.: Random forest-based similarity measures for multi-modal classification of alzheimer's disease. NeuroImage **65**, 167–175 (2013)
36. Gray, K.R., Wolz, R., Heckemann, R.A., Aljabar, P., Hammers, A., Rueckert, D., Initiative, A.D.N., et al.: Multi-region analysis of longitudinal fdg-pet for the classification of alzheimer's disease. Neuroimage **60**(1), 221–229 (2012)
37. Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C., Initiative, A.D.N., et al.: Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage **48**(1), 138–149 (2009)

38. Hosseini-Asl, E., Gimel'farb, G.L., El-Baz, A.: Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. CoRR **abs/1607.00556** (2016). URL http://arxiv.org/abs/1607.00556
39. Huang, M., Yang, W., Feng, Q., Chen, W., Initiative, A.D.N., et al.: Longitudinal measurement and hierarchical classification framework for the prediction of alzheimer's disease. Scientific reports **7** (2017)
40. Imam, I., Michalski, R., Kerschberg, L.: Discovering attribute dependence in databases by integrating symbolic learning and statistical analysis techniques. In: Proceeding of the AAAI-93 Workshop on Knowledge Discovery in Databases, Washington DC (1993)
41. Ishihara, S., Force, R.A.A.: Ishihara tests for colour blindness. Sydney: Shephard & Newman (1943)
42. Jones, D.K., Leemans, A.: Diffusion tensor imaging. Magnetic Resonance Neuroimaging: Methods and Protocols pp. 127–144 (2011)
43. Joshi, S., Shenoy, P.D., GG, V.S., Venugopal, K., Patnaik, L.: Classification of neurodegenerative disorders based on major risk factors employing machine learning techniques. International Journal of Engineering and Technology **2**(4), 350 (2010)
44. Jurica, P., Mattis, S., Leitten, C.: Dementia Rating Scale-2: DRS-2. Psychological Assessment Resources (2001). URL https://books.google.si/books?id=c5iKjwEACAAJ
45. Kippenhan, J.S., Barker, W.W., Pascal, S., Nagel, J.H., Duara, R.: Evaluation of a neural-network classifier for pet scans of normal and alzheimer's disease subjects (1992)
46. Klein, S., Loog, M., van der Lijn, F., den Heijer, T., Hammers, A., de Bruijne, M., van der Lugt, A., Duin, R.P., Breteler, M.M., Niessen, W.J.: Early diagnosis of dementia based on intersubject whole-brain dissimilarities. In: Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on, pp. 249–252. IEEE (2010)
47. Li, J., Wang, Z.J., McKeown, M.J.: A framework for group analysis of fmri data using dynamic bayesian networks. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5991–5994 (2007). DOI 10.1109/IEMBS.2007.4353713
48. Liu, M., Zhang, D., Shen, D., Initiative, A.D.N., et al.: Ensemble sparse classification of alzheimer's disease. NeuroImage **60**(2), 1106–1116 (2012)
49. Llano, D.A., Devanarayan, V., Simon, A.J., (ADNI, A.D.N.I., et al.: Evaluation of plasma proteomic data for alzheimer disease state classification and for the prediction of progression from mild cognitive impairment to alzheimer disease. Alzheimer Disease & Associated Disorders **27**(3), 233–243 (2013)
50. López, M., Ramírez, J., Górriz, J.M., Álvarez, I., Salas-Gonzalez, D., Segovia, F., Chaves, R., Padilla, P., Gómez-Río, M., Initiative, A.D.N., et al.: Principal component analysis-based techniques and supervised classification schemes for the early detection of alzheimer's disease. Neurocomputing **74**(8), 1260–1271 (2011)
51. Meszlényi, R.J., Buza, K., Vidnyánszky, Z.: Resting state fmri functional connectivity-based classification using a convolutional neural network architecture. Frontiers in neuroinformatics **11**, 61 (2017)
52. Miller, V.A., Erlien, S., Piersol, J.: Support vector machine classification of dimensionally reduced structural mri images for dementia. arXiv preprint arXiv:1406.6568 (2014)
53. Moradi, E., Gaser, C., Huttunen, H., Tohka, J., Alzheimerr's: MRI based dementia classification using semi-supervised learning and domain adaption, pp. 65–73 (2014)
54. Morales, D.A., Vives-Gilabert, Y., Gómez-Ansón, B., Bengoetxea, E., Larrañaga, P., Bielza, C., Pagonabarraga, J., Kulisevsky, J., Corcuera-Solano, I., Delfino, M.: Predicting dementia development in Parkinson's disease using bayesian network classifiers. Psychiatry Research: NeuroImaging **213**(2), 92–98 (2013)
55. Morris, J.C., Weintraub, S., Chui, H.C., Cummings, J., DeCarli, C., Ferris, S., Foster, N.L., Galasko, D., Graff-Radford, N., Peskind, E.R., et al.: The uniform data set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. Alzheimer Disease & Associated Disorders **20**(4), 210–216 (2006)
56. Murphy, K.P.: Dynamic bayesian networks. Probabilistic Graphical Models, M. Jordan **7** (2002)
57. Patil, M., Yardi, A.: MLP classifier for dementia levels. International Journal of Modeling and Optimization **2**(5), 567 (2012)

58. Payan, A., Montana, G.: Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. arXiv preprint arXiv:1502.02506 (2015)
59. Pfeffer, R., Kurosaki, T., Harrah, C., Chance, J., Filos, S.: Measurement of functional activities in older adults in the community. Journal of gerontology **37**(3), 323–329 (1982)
60. Polikar, R., Topalis, A., Parikh, D., Green, D., Frymiare, J., Kounios, J., Clark, C.M.: An ensemble based data fusion approach for early diagnosis of alzheimer's disease. Information Fusion **9**(1), 83–95 (2008)
61. Pritchard, W.S., Duke, D.W., Coburn, K.L., Moore, N.C., Tucker, K.A., Jann, M.W., Hostetler, R.M.: Eeg-based, neural-net predictive classification of alzheimer's disease versus control subjects is augmented by non-linear eeg measures. Electroencephalography and clinical Neurophysiology **91**(2), 118–130 (1994)
62. Querbes, O., Aubry, F., Pariente, J., Lotterie, J.A., Démonet, J.F., Duret, V., Puel, M., Berry, I., Fort, J.C., Celsis, P., et al.: Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain **132**(8), 2036–2047 (2009)
63. Quinlan, J.R.: C4. 5: programs for machine learning. Elsevier (2014)
64. Ray, S., Britschgi, M., Herbert, C., Takeda-Uchimura, Y., Boxer, A., Blennow, K., Friedman, L.F., Galasko, D.R., Jutel, M., Karydas, A., et al.: Classification and prediction of clinical alzheimer's diagnosis based on plasma signaling proteins. Nature medicine **13**(11), 1359–1362 (2007)
65. Roher, A.E., Weiss, N., Kokjohn, T.A., Kuo, Y.M., Kalback, W., Anthony, J., Watson, D., Luehrs, D.C., Sue, L., Walker, D., et al.: Increased A$\beta$ peptides and reduced cholesterol and myelin proteins characterize white matter degeneration in Alzheimer's disease. Biochemistry **41**(37), 11080–11090 (2002)
66. Schmitter-Edgecombe, M., Parsey, C., Cook, D.J.: Cognitive correlates of functional performance in older adults: comparison of self-report, direct observation, and performance-based measures. Journal of the International Neuropsychological Society **17**(05), 853–864 (2011)
67. Schmitter-Edgecombe, M., Woo, E., Greeley, D.R.: Characterizing multiple memory deficits and their relation to everyday functioning in individuals with mild cognitive impairment. Neuropsychology **23**(2), 168 (2009)
68. Shankle, W.R., Datta, P., Dillencourt, M., Pazzani, M.: Improving dementia screening tests with machine learning methods. Alzheimer's Research **2**(3) (1996)
69. Shankle, W.R., Mani, S., Pazzani, M.J., Smyth, P.: Detecting very early stages of dementia from normal aging with machine learning methods. In: Conference on Artificial Intelligence in Medicine in Europe, pp. 71–85. Springer (1997)
70. Shanklea, W., Mani, S., Dick, M.B., Pazzani, M.J.: Simple models for estimating dementia severity using machine learning. Studies in health technology and informatics (1), 472–476 (1998)
71. Suk, H.I., Lee, S.W., Shen, D., Initiative, A.D.N., et al.: Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage **101**, 569–582 (2014)
72. Szenkovits, A., Meszlényi, R., Buza, K., Gaskó, N., Lung, R.I., Suciu, M.: Feature selection with a genetic algorithm for classification of brain imaging data. In: Advances in Feature Selection for Data and Pattern Recognition, pp. 185–202. Springer (2018)
73. Tierney, M., Szalai, J., Snow, W., Fisher, R., Nores, A., Nadon, G., Dunn, E., George-Hyslop, P.S.: Prediction of probable alzheimer's disease in memory-impaired patients a prospective longitudinal study. Neurology **46**(3), 661–665 (1996)
74. Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N.R., Chui, H., Cummings, J., DeCarli, C., Foster, N.L., Galasko, D., et al.: The Alzheimer's disease centers' uniform data set (UDS): The neuropsychological test battery. Alzheimer disease and associated disorders **23**(2), 91 (2009)
75. Williams, J.A., Weakley, A., Cook, D.J., Schmitter-Edgecombe, M.: Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia. In: Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence (2013)
76. Wu, X., Zhu, X., Wu, G.Q., Ding, W.: Data mining with big data. IEEE transactions on knowledge and data engineering **26**(1), 97–107 (2014)
77. Zaffalon, M.: Statistical inference of the naive credal classifier. In: ISIPTA, vol. 1, pp. 384–393 (2001)

78. Zaffalon, M.: The naive credal classifier. Journal of statistical planning and inference **105**(1), 5–21 (2002)
79. Zaffalon, M., Wesnes, K., Petrini, O.: Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. Artificial Intelligence in Medicine **29**(1), 61–79 (2003)
80. Zhang, D., Shen, D., Initiative, A.D.N., et al.: Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. NeuroImage **59**(2), 895–907 (2012)
81. Zhang, Y., Dong, Z., Phillips, P., Wang, S., Ji, G., Yang, J., Yuan, T.F.: Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. Frontiers in Computational Neuroscience **9**, 66 (2015)