Improving Alzheimer's disease classification by performing data fusion with vascular dementia and stroke data

Zoran Bosnić[a], Brankica Bratić[b], Mirjana Ivanović[b], Marija Semnic[c,d], Iztok Oder[a], Vladimir Kurbalija[b], Tijana Vujanić Stankov[c,d], Vojislava Bugarski Ignjatović[c,d]

[a]University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia;
[b]University of Novi Sad, Faculty of Sciences, Novi Sad, Serbia;
[c]University of Novi Sad, Faculty of Medicine, Novi Sad, Serbia;
[d]Clinical Centre of Vojvodina, Clinic of Neurology, Novi Sad, Serbia.

ARTICLE HISTORY
Compiled August 31, 2020

ABSTRACT
Improvement of prediction accuracy and early detection of the Alzheimer's disease is becoming increasingly important for managing its impact on lives of affected patients. Many machine learning approaches have been applied to support the diagnosis and prediction of this illness. In this paper we propose an approach for improving the Alzheimer's disease classification accuracy by using data fusion of several independent clinical datasets. Data fusion was performed twofold: 1) by enriching attributes of the base dataset with the attributes of the secondary dataset and 2) by enriching the examples set of the base dataset with the examples of the secondary dataset. In both cases the missing values (for newly added attributes and/or examples) were predicted by using linear regression for numeric and naive Bayes classifier for nominal attributes. We experimented on three data sources: on a dataset of Alzheimer's disease impaired patients, on a dataset of patients with vascular dementia, and on a dataset of patients who have been affected by a stroke. We fused these datasets with different data fusion approaches and analysed the improvement in classification accuracy as well as the quality of the fused attributes. The experiments indicated that we obtained an increase of classification accuracy on the fused dataset compared with the accuracy obtained from individual dataset.

KEYWORDS
Data fusion; machine learning; Alzheimer's disease; vascular dementia; stroke.

1. Introduction

Machine learning is nowadays used in many medical fields (Benzebouchi, Azizi, Ashour, Dey, & Sherratt, 2019; Saâdaoui et al., 2015). However, medical data is usually hard to collect and it can take years to reach their sufficient amount to perform experiments, analyses and research (this usually does not hold for the imaging data (Qin, Chen, Zhang, & Chai, 2018) and for the cases when specialised systems for data collection are implemented (Dai, Fu, Dai, & Lu, 2017)). The problem of collecting data is additionally aggravated by privacy polices, which limit data collection, sharing and

CONTACT Emails: zoran.bosnic@fri.uni-lj.si, brankica.bratic@dmi.uns.ac.rs, mira@dmi.uns.ac.rs, marija.semnic@mf.uns.ac.rs, iztok.oder@gmail.com, kurba@dmi.uns.ac.rs, tijana.vujanic-stankov@mf.uns.ac.rs, vojislava.bugarski-ignjatovic@mf.uns.ac.rs

distribution. Medical datasets are most commonly collected for investigating a single medical phenomenon (e.g., an illness), making them rarely available to researchers from other research fields. The broader availability of data could increase data usability and re-usability, since the same patients analyses are performed for different medical studies. Data for a single patient can be enriched in two distinctive ways: either by supplementing the patient's attributes with additional attributes for the same patient from the second medical database, or by inducing the new missing attributes for the single patient based on properties of the most similar other patients from the second database – a process called *data fusion*.

Alzheimer's disease is a type of dementia that usually affects elderly people. Due to the fact that we are currently living in the era of population ageing, public health impact of this disease is rather high. In 2015, more than 15 million family members of Alzheimer's disease patients' provided an estimated $221 billion for caregiving, and the number is increasing (Alzheimer's Association, 2016). A cure for the Alzheimer's disease does not exist, but effective interventions in early phase of the illness may significantly slow its progression, improving patient's life and lowering the caregiving costs (Roberson & Mucke, 2006). As a consequence, a lot of effort has been put into increasing the rate of early detections of the illness. One direction of achieving this goal relies on advanced methods in the field of artificial intelligence and machine learning.

In this work we aim to improve the Alzheimer's disease classification accuracy by applying data fusion of a base Alzheimer's disease dataset with two other datasets: a dataset of patients that suffer from vascular dementia, and a dataset of patients who had a stroke. Although all the three datasets contain data for completely different patients, they share a distinct number of common attributes, which we used for performing data fusion. Data fusion was performed in two different ways: 1) by enriching attributes of the base dataset with the attributes of the secondary dataset, and 2) by enriching examples of the base dataset with the examples of the secondary dataset. Addition of new attributes and/or examples implies a lot of missing values. The missing values are then predicted with the models that are built upon temporary datasets made particularly for this purpose. Finally, we use linear regression for predicting numeric, and naive Bayes for predicting nominal, missing values.

The paper is organised as follows. Section 2 overviews the state of the art for Alzheimer's disease prediction and data fusion. In Section 3 we give detailed insights into datasets and methodology we used. Section 4 presents results of all experiments. Conclusions and directions of further research are given in Section 5.

2. Related work

In this section we present relevant related research in the fields of the Alzheimer's disease prediction with machine learning and data fusion.

2.1. Alzheimer's disease prediction

Prediction and early detection of Alzheimer's disease are important goals in medicine. One way of reaching them is to develop highly accurate machine learning predictive models that can learn from various kinds of collected data. Bratić, Kurbalija, Ivanović, Oder, and Bosnić (2018) made an overview of such methodologies and some of them are described in the following.

Tierney et al. (1996) used a battery of neuropsychological tests for predicting onset of Alzheimer's disease. After selecting most relevant attributes, they obtained a model with accuracy of 89%. Magnetic resonance imaging (MRI) and positron emission tomography (PET) data are also suitable for Alzheimer's disease prediction. Liu, Zhang, Shen, and Alzheimer's Disease Neuroimaging Initiative (2012) suggested using the ensembles with sparse representation-based classifier (SRC) as a weak classifier on subsamples of patches from the raw MRI. Classification accuracy of their method reached up to 90.8%. Dyrba et al. (2013) achieved accuracy of 83% by using Support Vector Machine (SVM) and the Diffusion Tensor Imaging (DTI) dataset. Kippenhan, Barker, Pascal, Nagel, and Duara (1992) split PET voxels into 67 brain regions, and extracted metabolism activity features. By using neural networks on their dataset they obtained a performance comparable to that of an independent expert who classified samples based on examining only PET images.

Longitudinal MRI data is interesting to analyse as well, since it can provide valuable information on how the disease progresses. Huang, Yang, Feng, Chen, and Alzheimer's Disease Neuroimaging Initiative (2017) proposed a hierarchical classification method that builds multiple multilevel classifiers upon longitudinal MRI data. Their method outperformed a base classifier by having an accuracy of 79%. Gray et al. (2012) showed that using data extracted from baseline and 12-month follow-up MRI could increase classification accuracy. Their model achieved the accuracy of 88%. Jiji (2018) performed a volumetric analysis of anatomical components of brain with multiclass particle swam optimisation technique (MPSO) in order to detect the stage of Alzheimer's disease.

Many researchers used protein data in order to predict or detect the Alzheimer's disease. Ray et al. (2007) examined 120 proteins using significance analysis of microarrays (SAM), and they managed to identify 18 proteins that were significantly different between patients with Alzheimer's disease and healthy controls. This could indicate an early onset of the Alzheimer's disease. Llano, Devanarayan, Simon, and Alzheimer's Disease Neuroimaging Initiative (2013) analysed prediction power of proteins for classification of Alzheimer's disease using the analysis of covariance (ANCOVA) and t-test. Doecke et al. (2012) used blood protein levels measured in plasma as an input to feature selection algorithms. Most of the frequently selected features were the ones that are related to the Alzheimer's disease; however, carcinoembryonic antigen feature stood out since it has never been associated with the Alzheimer's disease before.

Electroencephalography (EEG) signals have also been used in detection and prediction of the Alzheimer's disease, although more rarely. Pritchard et al. (1994) extracted features from EEG signals measured on 19 different positions of the skull, and achieved accuracy of 92% by using neural networks. By additional analysis of frequency sections they showed that brain activity significantly drops for subjects with Alzheimer's disease.

Besides aiming to increase the classification accuracy, Y. Zhang et al. (2015) also managed to detect 30 brain regions that are related to the disease. Hinrichs et al. (2009) proposed the use of the Linear Programming Boosting for predicting Alzheimer's disease by using MRI. 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 Alzheimer's disease.

In order to increase general predictability of the Alzheimer's disease, it is important to deepen domain knowledge about the illness. Machine learning can help by proving some already known facts, or can even help obtaining the new knowledge.

2.2. Data fusion

Data for a certain machine learning problem can be represented in different modalities and can come from different sources. To integrate their information into a joint representation and facilitate successful learning, data fusion methods are required. Such sources usually do not contain an explicit primary key that would facilitate easy merging of data, but require advanced data fusion methods.

Past research has tackled data fusion with three different strategies: by applying early, intermediate and late fusion. The *early* data fusion is the most simple process that exploits the availability of a common primary key that enables us to merge records of two independent datasets. The *intermediate* fusion inputs data of different modalities into a machine learning algorithm, which extracts relevant data characteristics (attributes) and uses them for learning (D. Zhang et al., 2011). Finally, within the *late* data fusion, the final outputs (predictions) of machine learning algorithms, that have been separately applied to each of the modalities, are combined into the final prediction (Gray et al., 2013).

An example of applying early and intermediate data fusion was done by D. Zhang et al. (2011), who used three different data modalities: region volumes in MRI images, average regional voxel intensities from PET scans, and biomarkers from cerebrospinal fluid. They applied different SVM kernels to each modality and combined their weighted average into the final SVM. They compared the obtained results with performance of SVM models on each separate modality and on models that perform early data fusion (merging of data). The results showed that the first approach (multi-kernel SVM) generated significantly better results than the other two approaches. Several different studies also applied SVM and other kernel based techniques for the fusion of multiple modalities (Hinrichs, Singh, Xu, & Johnson, 2011; Kohannim et al., 2010; Sun, Qiao, Lelieveldt, & Staring, 2018; Ye et al., 2008).

Gray et al. (2013) also implemented the intermediate data fusion for data of different modalities (MRI data, PET data, cerebrospinal fluid measurements and genetic data). For each modality they learned a separate random forest model and linearly combined their predictions into a *manifold representation*. Their results showed that the model that learns from the combined data achieves higher performance than models for individual modalities. Bi, Cai, Wang, and Liu (2019) proposed a multimodal random forest (MRF) method to distinguish AD from healthy individuals based on neuroimaging and genetic data. The proposed approach tries to construct optimal fusion features, which are then used for the selection of abnormal brain regions and genes. Furthermore, the authors presented a novel machine learning framework of data fusion, classification, feature selection, and disease-causing factor extraction.

As expected, many other machine learning/classification/diagnosis techniques were used for AD diagnosis on fused data. Ortiz, Fajardo, Gorriz, Ramírez, and Martínez-Murcia (2014) performed a fusion of multimodal image (MRI and PET) data by combining Sparse Representation Classifiers. They report accuracies of up to 95% which clearly outperform the classification accuracy obtained using single-modality images. Walhovd et al. (2010) applied multi-method stepwise logistic regression analysis to integrate multiple modalities (MRI, PET, CSF). Again data fusion approach yielded better results than single-data approach, although it is concluded that MRI and PET were more predictive than CSF. Westman, Muehlboeck, and Simmons (2012) used 60 variables from MRI, PET and CSF data for orthogonal projections to latent structures (OPSL) multivariate analysis. Their combined model accomplished 91.8% accuracy compared to 81.6% for CSF measures and 87.0% for MRI measures alone.

Suk, Lee, Shen, and Alzheimer's Disease Neuroimaging Initiative (2014) used MRI and PET images to extract groups of voxels that are relevant for predicting the Alzheimer's disease. They applied the multimodal Deep Boltzmann Machine to extract the attributes from the input data and feed them into the neural network. The features were then fed into the ensemble with SVM as a weak classifier (Liu et al., 2012). Again, models built on multimodal data outperformed the models built on only one modality. As expected, most of the recent research mainly focus on novel and advanced technologies like deep learning (Kim & Lee, 2018; Ning et al., 2018; Suk & Shen, 2013) where neuroimaging data is fused with other types of data.

An example of late integration was applied by Polikar et al. (2008) to fuse data collected with multiple EEG electrodes. They proposed an algorithm that used an ensemble of classifiers for each of the modalities, and joined their predictions using weighted majority voting. The models that utilised data fusion outperformed models built on a single modality. A similar approach was applied by Parikh et al. (2005). The authors applied data fusion of data recorded from the Pz and Cz electrodes of the EEG, since they believed that these electrodes contained complimentary information, for early diagnosis of Alzheimer's disease. The EEG data was further analysed using multi-resolution wavelet analysis which generated multiple classifiers. These classifiers were then combined through a weighted majority voting.

Since the neuroimaging techniques (like MRI and PET) have proved to be a powerful support in AD diagnosis (Teipel et al., 2015) a great majority of papers combined these techniques with some other data sources to perform a more accurate diagnosis. Additionally, a few researches combined several different EEG datasets for successful diagnosis. On the other hand, our research tries to exploit data from several unrelated sources and sub-domains (AD, vascular dementia and stroke) to improve the accuracy of AD diagnosis, while not relying overwhelmingly on neuroimaging data. To the best of our knowledge there is no existing research which fuses data from different neurology subfields for the purpose of AD diagnosis.

3. Methods

In this section we present datasets that were used in the paper and provide a description of the data fusion approach.

3.1. Datasets

In this paper we used three different medical datasets. The datasets contain medical information gathered from three PhD theses conducted at the Medical Faculty of the University of Novi Sad. Research studies were carried out at the Neurology Clinic of the Clinical Centre of Vojvodina and Vojvodina Institute of Oncology, Center for Diagnostic Imaging. Each dataset describes a different medical domain, as follows:

- Dataset of **patients impaired with Alzheimer's disease** (in the following denoted with A). This dataset was used as a base dataset, which is enriched with data fusion techniques by using the other two datasets. The dataset contains data about 85 patients, of which 29 are healthy controls, 27 are patients with amnestic mild cognitive impairment, and 29 are patients with Alzheimer's disease due to NINCDS-ADRDA criteria (McKhann et al., 1984). Demographic criteria for the study were people aged 60 to 85, with minimum 12 years of formal education.

ACE III	–	Addenbrooke cognitive examination III	**MMSE**	–	Mini mental status examination
BDAE	–	Boston diagnostic aphasia examination	**MTA**	–	Mediotemporal atrophy score
BDI	–	Beck depression inventory	**NRS**	–	Neurobehavioural rating scale
BNT	–	Boston naming test	**PA**	–	Parietal atrophy
ESS	–	European stroke scale	**PSMS**	–	Physical self-maintenance scale
EXIT25	–	The executive interview test-25	**QoL**	–	Quality of life
FAB	–	Frontal assessment battery	**RAVLT**	–	Rey auditory verbal learning test
FBI	–	Frontal behavioural inventory	**ROCF**	–	Rey-Osterrieth complex figure
GCA	–	Global cortical atrophy score	**TMT**	–	Trail making test
GDS	–	Geriatric depression scale	**WAIS**	–	Wechsler adult intelligence scale
HAMD	–	Hamilton depression rating scale	**WCST**	–	Wisconsin card sorting test
HIS	–	Hachinski ischaemic stroke scale	**WMH**	–	White matter hyperintensities
HVO	–	Hooper visual organisation	**WMS**	–	Wechsler memory scale
IADL	–	Instrumental activities of daily living			

Table 1.: The list of acronyms that are used in the tables 2, 3, 4, 5, 6 and 7.

1. Gender	10. WCST – Correct response	19. RAVLT – A7	28. Category fluency score
2. Age	11. WCST – Failure to maintain set	20. RAVLT – Recognition A	29. MMSE – Total
3. Years of education	12. WCST – Conceptual level responses	21. RAVLT – Recognition B	30. Verbal forward digit span
4. Arterial hypertension	13. WCST – Perseverative response	22. RAVLT – A6-A5	31. Verbal backward digit span
5. Diabetes	14. WCST – Perseverative errors	23. TMT – A	32. Visual forward digit span
6. Obesity	15. WCST – Nonperseverative errors	24. TMT – B	33. Visual backward digit span
7. Smoking	16. WCST – Total errors	25. Verbal fluency S	34. ROCF Copy
8. Alcohol	17. RAVLT – Total A1-A5	26. Verbal fluency K	35. ROCF immediate recall
9. WCST – Number of categories	18. RAVLT – A6	27. Verbal fluency L	36. ROCF 45 minutes recall

*Table 2.: The list of attributes that exist in all the three datasets (**A, D and S**).*

1. Cardiovascular diseases	5. WCST – Total reponses	9. RAVLT – A3	12. RAVLT – B1
2. Heredity	6. SSS – Total	10. RAVLT – A4	13. HAMD – Total score
3. WAIS- Digit symbol coding	7. RAVLT– A1	11. RAVLT – A5	14. IADL – Total
4. WCST – Trials to complete first category	8. RAVLT – A2		

*Table 3.: The list of attributes that exist only in datasets **A and D**.*

1. Marital status	3. WCST – Errors	4. BDI – Total	5. WMS – Mental control – attention/concentration index
2. BNT – Total			

*Table 4.: The list of attributes that exist only in datasets **A and S**.*

This research was conducted at the Neurology Clinic of the Clinical Centre of Vojvodina in Novi Sad, Serbia and Institute of Oncology in Sremska Kamenica in the period from January 2016 to December 2017. The type of the patient is given in a target class variable. There is a total of 142 attributes (including class), which represent patients' demographics, diagnostic data and scores of various psychological tests.

- Dataset of **patients with vascular dementia** (denoted with D). This dataset includes data about 90 people aged 50 to 80, divided into two groups. The first group, consisted of 50 people, are patients diagnosed with probable vascular dementia based upon NINDS-AIREN (Román et al., 1993) and SCADDTC (Chui et al., 1992) criteria. Second group is a control group consisted of 30 people with the mini-mental state exam (MMSE) score between 28 and 30 who had no cognitive decline. Research was carried out at the Neurology Clinic of the Clinical Centre of Vojvodina in Novi Sad, Serbia, from January 2004 until November 2007. The dataset has 115 attributes containing patients' demographics and

1. Memory impairment	23. Ischaemic heart disease	45. QoL – Caregiver	67. ACE III – Attention
2. Main symptoms	24. Coronary artery disease	46. NPI – Total	68. ACE III – Memory
3. Time orientation impairment	25. Heart rhythm disorders	47. NPI – Delusions	69. ACE III – Fluency
4. Praxia	26. Hypothyroidism	48. NPI – Hallucinations	70. ACE III – Language
5. Gnosia	27. Prior depression	49. NPI – Dysphoria	71. ACE III – Visuospatial
6. Visuospatial impairment	28. Head injury	50. NPI – Anxiety	72. MMSE – Orientation to time
7. Calculation	29. Physical inactivity	51. NPI – Agitation/aggression	73. MMSE – Orientation to place
8. Attention	30. No Focal neurological impairment	52. NPI – Euphoria	74. MMSE – Registration
9. Decision making	31. Focal neurological impairment – pyramidal signs	53. NPI – Disinhibition	75. MMSE – Attention and Calculation
10. Functionality	32. Focal neurological impairment – sensibility	54. NPI – Irritability/lability	76. MMSE – Recall
11. Behavioural and psychological symptoms	33. Focal neurological impairment – cerebelar symptoms	55. NPI – Apathy	77. MMSE – Naming
12. Depression	34. Focal neurological impairment – extrapyramidal signs - tremor	56. NPI – Aberrant Motor behaviour	78. MMSE – Repetition
13. Hallucinations	35. Focal neurological impairment – extrapyramidal signs – rigor	57. HVO – Total	79. MMSE – Complex command
14. Delusions	36. Focal neurological impairment – extrapyramidal signs – bradykinesia	58. BNT – Phonemic cues	80. MMSE – Comprehension
15. Agitation	37. Focal neurological impairment – extrapyramidal signs – postural instability	59. BNT – Semantic cues	81. MMSE – Sentence
16. Emotional lability	38. Focal neurological impairment – extrapyramidal signs – involuntary movements	60. EXIT25 – Total	82. MMSE – Copying
17. Behavioural disinhibition	39. Focal neurological impairment – extrapyramidal signs – chorea	61. FAB – Total	83. GCA score
18. Socially unacceptable behaviour	40. Focal neurological impairment – extrapyramidal signs – athetosis	62. WAIS – Similarities	84. PA Score
19. Frequent falls	41. Focal neurological impairment-extrapyramidal signs – gait disturbances	63. WAIS – Block design test	85. MTA score
20. Losses of consciousness	42. Focal neurological impairment – extrapyramidal signs – disinhibition signs	64. TMT – A errors	86. WHM score
21. Fluctuation of cognition	43. Focal neurological impairment-Incontinence	65. TMT – B errors	87. Alzheimer's disease
22. Evidence of stroke	44. QoL – Patient	66. ACE III – Total	

Table 5.: The list of attributes that exist only in dataset $\boldsymbol{A}$.

scores of psychological tests.

- Dataset of **patients who had a stroke** (denoted with S). This dataset contains 70 examples, including 40 acute ischaemic stroke patients, aged 45-78, and 30 healthy controls. The acute ischaemic stroke diagnosis was established based on clinical symptoms and neuroradiological correlates obtained with brain computerised tomography (CT). The study was conducted at the Neurology Clinic of the Clinical Centre of Vojvodina in Novi Sad, Serbia, in the period from May 2007 to September 2008. The dataset contains 72 attributes which represent patients' demographics and scores of psychological and stroke related tests. Value of the target class variable determines whether a patient had a stroke.

All the above datasets share a number of attributes that could be used as a basis for data fusion, as follows:

- A and D share 50 attributes, which makes 35% of all attributes in A,
- A and S share 41 attributes, which makes 28% of attributes in A.

Lists of common and dataset-specific attributes are shown in Tables 2, 3, 4, 5, 6 and 7. The acronyms used in the tables are presented in Table 1.

Note that although the datasets share a subset of common attributes, they contain data for different patients. Furthermore, the patients are not denoted with unique IDs

1. WCST – % Perseverative errors	18. HAMD – Weight loss	34. NRS – Total score	50. NRS – Suspiciousness
2. RAVLT – Recognition	19. HAMD – Insight	35. NRS – Inattention/reduced alertness	51. NRS – Fatigability
3. HAMD – Depressed mood	20. HAMD – Diurnal variation I	36. NRS – Somatic concern	52. NRS – Hallucinations
4. HAMD – Feelings of guilt	21. HAMD – Diurnal vaiation II	37. NRS – Disorientation	53. NRS – Motor retardation
5. HAMD – Suicide	22. HAMD – Depersonalization, derealization	38. NRS – Anxiety	54. NRS – Unusual thought content
6. HAMD – Insomnia: Early in the night	23. HAMD – Paranoid symptoms	39. NRS – Expressive deficit	55. NRS – Blunted affect
7. HAMD – Insomnia: Middle of the night	24. HAMD – Obsessive and compulsive symptoms	40. NRS – Emotional withdrawal	56. NRS – Excitement
8. HAMD – Insomnia: Early hours of the morning	25. IADL – Ability to use telephone	41. NRS – Conceptual disorganisation	57. NRS – Poor planning
9. HAMD – Work and activities	26. IADL – Shopping	42. NRS – Disinhibition	58. NRS – Mood lability
10. HAMD – Retardation	27. IADL – Food preparation	43. NRS – Guilt feelings	59. NRS – Tension
11. HAMD – Agitation	28. IADL – Housekeeping	44. NRS – Memory deficit	60. NRS – Comprehension deficit
12. HAMD – Psychic anxiety	29. IADL – Laundry	45. NRS – Agitation	61. NRS – Speech articulation defect
13. HAMD – Somatic anxiety	30. IADL – Mode of transportation	46. NRS – Inaccurate insight	62. NRS – Fluent aphasia
14. HAMD – Somatic symptoms gastrointestinal	31. IADL – Responsibility for own medications	47. NRS – Depressed mood	63. PSMS
15. HAMD – Somatic symptoms general	32. IADL – Ability to handle finances	48. NRS – Hostility/uncooperativeness	64. ESS
16. HAMD – Central symptoms	33. FBI	49. NRS – Decreased initiative/motivation	65. HIS
17. HAMD – Hypochondriasis			

*Table 6.: The list of attributes that exist only in dataset **D**.*

1. National Institute Stroke Scale	9. Ischaemic lesions in grey matter on MR scan	17. Small-vessel stroke	25. Graphesthesia
2. Modified Rankin Scale	10. Ischaemic lesions in white matter on MR scan	18. Large arteries stroke	26. Limb praxia of left arm
3. Solitary ischaemic lesions on MR scan	11. Cortico-subcortical ischaemic lesions on MR scan	19. Ischaemic encephalopathy	27. Limb praxia of right arm
4. Multiple ischaemic lesions on MR scan	12. Acute stroke	20. BDAE 0-repetition	28. Ideational praxia
5. Unilateral ischaemic lesions on MR scan	13. Chronic ischaemic lesions on MR scan	21. BDAE – Complex ideational material	29. Dyspraxia
6. Bilateral ischaemic lesions on MR scan	14. Ischaemic lesions in cerebrum on MR scan	22. BDAE – Comprehension	30. GDS
7. Ischaemic lesions in left cerebral hemisphere on MR scan	15. Ischaemic lesions in cerebellum on MR scan	23. Finger gnosia	31. Stroke
8. Ischaemic lesions in right cerebral hemisphere on MR scan	16. Brain atrophy	24. Stereognosia	

*Table 7.: The list of attributes that exist only in dataset **S**.*

that could enable merging of different datasets even if they contained the data about the same patients.

The goal of this paper is to overcome these obstacles and enrich (fuse) the base dataset A with additional attributes from datasets D and S by utilising machine learning. We hope that additional attributes would improve predictive performance, which we evaluate with our experiments.

3.2. Data fusion

Data fusion is a technique of merging data from multiple datasets into a single dataset. In this paper we propose two approaches to this task: approach for **enriching attributes** of the base dataset and approach for **enriching the examples set** of the base dataset. Both approaches can also be combined together, yielding a dataset with enriched attributes and examples. We proceed by describing both approaches in the following.

Algorithm 1: Outline of the enriching attributes data fusion approach.

input : dataset X and dataset Y
output: dataset XY
XY = copy of X; // initialization of the resulting dataset
/* each iteration of the loop enriches XY with one new attribute from Y */
foreach *att* ∈ *attributes(Y)* \ *attributes(X)* **do**
    /* *filter(D,A)* returns a dataset D having only attributes A */
    Z = filter(Y, att ∪ (attributes(X) ∩ attributes(Y)));
    train a predictive model C on Z, where target is the attribute *att*;
    add *att* to XY; // adds new attributes, values still missing
    /* fill values of *att* for each instance in XY */
    foreach *instance* ∈ *XY* **do**
        instance[att] = predict(C, instance);
    end
end
Return *XY*.

Enriching attributes. The outline of this data fusion approach is given in Algorithm 1. It uses dataset X as the base dataset and dataset Y as a dataset that we would like to fuse with dataset X. The final result of the algorithm is the dataset XY that contains the same examples as X and additional attributes that are specific only to Y. Values of new attributes in X are calculated in the following manner. For each attribute that is specific only for Y (denoted with *att* in the algorithm), a temporary dataset Z is created. Attributes of Z consist of all common attributes of X and Y and the single additional attribute *att*. Then, a predictive model is trained on the dataset Z, having *att* as a target attribute (if *att* is discrete, a classifier is used; if it is continuous, a regression model is used). By computing predictions for instances in XY with the trained predictive model, the procedure yields values of attribute *att* for instances of X.

Enriching number of examples. The second approach is presented in Algorithm 2. The algorithm starts by concatenating all examples from datasets X and Y into the newly created dataset XY that has the same attribute set as X. This means that instances from dataset Y will have empty values for the attributes that exist in X

and do not exist in Y. These missing values are afterwards predicted analogously as within the previous approach, as follows. For each attribute *att* that exists in X and does not exist in Y, a new dataset Z is created. Attribute set of Z consists of all attributes common to X and Y plus the additional attribute *att*. The examples of Z are imported from dataset X. Then, a predictive model is trained on the dataset Z, having *att* as a target attribute. Finally, predictions are calculated for all examples in XY that originate from Y and the obtained predictions are stored as the values of missing attributes for those examples.

Algorithm 2: Outline of the enriching number of examples data fusion approach.

input : dataset X and dataset Y
output: dataset XY

```
/* filter(D,A) returns a dataset D having only attributes A                    */
XY = filter(X ∪ Y, attributes(X));
/* update values of XY's instances from Y, which have missing values for X's attributes
   */
foreach att ∈ attributes(X) \ attributes(Y) do
    Z = filter(X, att ∪ (attributes(X) ∩ attributes(Y)));
    train classifier C on Z, where target class is att;
    /* fill values of att for each instance in XY                               */
    foreach XY's instance that originates from Y do
        instance[att] = predict(C, instance);
    end
end
Return XY.
```

The two data fusion approaches can also be combined, enriching the dataset with new attributes as well as adding new examples. Predictive models that are used to predict missing attribute values can be arbitrarily chosen, considering the type of the target variable.

4. Experiments and results

4.1. Experiments outline

By using three datasets (see Subsect. 3.1) and with the goal to improve the predictive accuracy on Alzheimer's disease dataset (base dataset A), we can perform data fusion either by fusing:

- the base dataset A with the vascular dementia dataset D into $A + D$,
- the base dataset A with the stroke dataset S into $A + S$, or
- the base dataset A with both datasets D and S into $A + D + S$, by first creating $A + D$ and then fusing S into it [1].

The datasets can be fused in three ways: by enriching the set of attributes, by enriching the number of examples, or both. Within our experiments we performed all

[1] The experiments revealed that the different order of fusing, i.e. into $A + S + D$, produces negligible difference in results.

possible data fusions (3 fused datasets × 3 fusion approaches), which resulted in 9 fused datasets.

Since fusing several datasets and supplementing them with a large number of additional attributes can be a complex task in terms of computational complexity, we decided to use simple supervised learning models in this research. As suggested by related work (Gray et al., 2013) and several survey works that describe machine learning applications in medicine (Erickson, Korfiatis, Akkus, & Kline, 2017; Kononenko, 2001; Yoo, Ramirez, & Liuzzi, 2014), we applied linear regression as a predictive model for numerical attributes and Naive Bayes as predictive model for nominal attributes. In addition to alleviating computational complexity, both models also allow interpretation of predicted feature values, should it be required to better understand the dependencies within the domain.

With our experiments we tested if the fused datasets improve classification performance compared to the performance obtained only on the base dataset A. The classification accuracy was obtained by performing 10-fold cross-validation in 10 runs. For datasets obtained by applying the *enrichment of the number of examples*, cross-validation test data always consisted only of examples that originate from the base dataset A (otherwise, the remaining examples in the test set would contain the class attribute that was predicted using the data fusion approach instead of their true class). The folds within the cross-validation process were formed to be the same for all testing datasets, enabling us to further investigate the statistical significance of the results using the paired Wilcoxon signed-rank test.

Prior to computing classification accuracies, we also applied different attribute selection methods to: (1) improve classification performance and (2) to determine how the fused attributes rank compared with the original attributes from the base dataset A (this was also possible when applying the *attribute enrichment* fusion approach).

4.2. Evaluation of fused attributes

We started by applying three attribute selection methods on fused datasets to determine how well the fused attributes rank compared to the base attributes. We used two filter, and one wrapper approach: information gain (denoted with IG), ReliefF (RE) (Robnik-Šikonja & Kononenko, 2003; Slavkov, Karcheska, Kocev, & Dzeroski, 2018), and wrapper subset with naive Bayes classifier (NB). We used filter selection methods to select 2, 4, 8, 16, 32, 64 and 128 best attributes; and the wrapper methods with the following search algorithms: best first (BF), genetic search (Gen), greedy stepwise (GS) and rank search (RS).

The results of the fused attribute evaluation for different fused datasets and fusion approaches are shown in Table 8. The table contains percentages of fused attributes that were selected among the best selected attributes using each attribute selection method and its parameter. The results show that the fused attributes rather frequently appear amongst the best selected attributes and that their relative frequencies reach up to 50%. We can also observe that the fused attributes of dataset A+S seem to be ranked worse than attributes of the other fused datasets.

4.3. Predictive performance evaluation

We applied the following well-known classification algorithms, and evaluated their classification accuracy on the original and the fused dataset: random forest (RF),

Information Gain

dataset	fusion	2	4	8	16	32	64	128
A+D+S	att	50%	25%	13%	25%	34%	44%	35%
A+D	att	50%	25%	13%	25%	31%	36%	25%
A+S	att	50%	25%	13%	6%	13%	16%	16%
A+D+S	comb	0%	0%	25%	19%	19%	33%	45%
A+D	comb	50%	50%	38%	44%	38%	36%	30%
A+S	comb	0%	0%	13%	13%	16%	14%	14%

Relief

dataset	fusion	2	4	8	16	32	64	128
A+D+S	att	0%	0%	13%	13%	22%	34%	48%
A+D	att	0%	0%	13%	6%	19%	31%	44%
A+S	att	0%	0%	13%	6%	9%	14%	15%
A+D+S	comb	0%	0%	13%	13%	34%	36%	46%
A+D	comb	0%	25%	13%	19%	34%	33%	41%
A+S	comb	0%	0%	13%	13%	19%	14%	14%

Wrapper Subset (Naive Bayes)

dataset	fusion	BestFirst	GeneticSearch	GreedyStepwise	RankSearch
A+D+S	att	20% (5)	45% (96)	20% (5)	0% (5)
A+D	att	17% (6)	31% (96)	17% (6)	0% (4)
A+S	att	17% (6)	25% (77)	17% (6)	7% (14)
A+D+S	comb	50% (8)	37% (123)	50% (8)	25% (52)
A+D	comb	50% (4)	28% (100)	50% (4)	40% (35)
A+S	comb	0% (8)	15% (79)	0% (6)	14% (37)

*Table 8.: Percentage of the fused attributes among the best selected attributes for different fused datasets and fusion approaches (*att *– attribute enrichment,* comb *– combination of both approaches). For the wrapper method, the total number of selected attributes is given in the parentheses.*

dataset	fusion	attributes	examples	RF	R	NB	kNN
base dataset	-	142	85	0.87	0.9	0.89	0.51
A+D+S	att	239	85	0.83	0.89	0.9	0.64
A+D	att	208	85	0.86	0.87	0.9	0.59
A+S	att	173	85	0.87	0.88	0.9	0.56
A+D+S	exa	142	245	0.86	0.79	0.66	0.51
A+D	exa	142	175	0.86	0.88	0.74	0.54
A+S	exa	142	155	0.88	0.79	0.75	0.39
A+D+S	comb	239	245	0.8	0.74	0.75	0.59
A+D	comb	208	175	0.85	0.84	0.75	0.56
A+S	comb	173	155	0.85	0.81	0.75	0.48

*Table 9.: Classification accuracies obtained on the base Alzheimer dataset (in the first row) and on different fused datasets (*att *– attribute enrichment,* exa *– example enrichment,* comb *– combination of both approaches). The shades of green colour indicate the increase in classification accuracy compared with the base data set, and the red shades indicate the decrease. Darker shades of both colours indicate statistically significant results.*

RIPPER (R), naive Bayes (NB), and k-nearest neighbours (kNN) for $k = 1$.

4.3.1. Results on all original and fused attributes

The initial results obtained without attribute selection, are shown in Table 9. We can see that we managed to obtain a minimal improvement in classification accuracy in at least one fused dataset using all classifiers except RIPPER, and achieved the best overall accuracy using the naive Bayes classifier on the first data fusion approach. Nevertheless, low classification accuracy is somewhat expected, given that the data fusion notably increased the dimensionality of the problem space by adding additional attributes and/or introduced new examples. The level of noise that was introduced into the fused dataset also depends on the quality of the data fusion approach itself.

dataset	fusion	attr. sel.	RF	R	NB	kNN
A+D+S	att	IG	0/14	14/0	29/29	14/43
		RE	0/14	14/0	43/0	14/14
		NB	50/25	0/0	50/0	75/0
A+D	att	IG	0/14	14/14	29/29	29/43
		RE	0/57	14/14	29/14	14/29
		NB	50/50	0/25	75/25	25/75
A+S	att	IG	43/0	14/14	0/57	29/29
		RE	0/14	0/57	14/14	29/29
		NB	75/0	0/0	50/0	25/0
A+D+S	exa	IG	0/86	0/86	0/100	14/86
		RE	0/71	0/100	14/71	14/86
		NB	0/100	0/100	0/100	0/75
A+D	exa	IG	0/57	29/43	0/100	14/86
		RE	14/57	29/57	29/71	14/71
		NB	50/25	75/0	0/100	0/100
A+S	exa	IG	0/57	14/29	0/100	0/71
		RE	0/86	0/57	0/100	0/86
		NB	0/25	0/25	0/100	0/100
A+D+S	comb	IG	0/100	0/100	0/100	14/71
		RE	29/71	29/71	29/71	29/57
		NB	0/100	0/100	0/100	0/75
A+D	comb	IG	0/43	43/29	0/100	14/71
		RE	14/71	29/71	29/71	29/57
		NB	50/50	0/25	0/100	25/75
A+S	comb	IG	0/100	0/57	0/100	0/57
		RE	0/100	0/86	0/71	0/100
		NB	0/50	0/0	0/100	0/100

Table 10.: Percentages of statistically significant increases/decreases of classification accuracy for different fused datasets, fusion approaches and attribute selection methods (IG - information gain, RE - Relief, NB - naive Bayes). The green colour indicates the prevailing increase of classification accuracy compared with the base data set, and the red indicates the prevailing decrease.

4.3.2. Results with attribute selection

To improve the predictive performance, we applied attribute selection to the fused datasets. Table 10 presents an overall comparison of all applied attribute selection methods by displaying percentages of experiments in which the classification accuracy

has significantly increased/decreased. The total number of experiments for each cell (combination of dataset, fusion approach, attribute selection method, and a classifier) equals to a number of different parameter's values (parameter is number of desired target attributes for filter methods or type of the search algorithm for wrapper methods), each used for a separate run of the attribute selection method (as explained in Subsection. 4.2). The results indicate that the best data fusion approach seems to be attribute enrichment, which gave the best results on the fully fused dataset A+D+S when combined with naive Bayes classifier. The other data fusion approaches did not produce satisfiable results, i.e. adding new examples decreased the classification accuracy.

dataset	fusion	attributes	examples	RF	R	NB	kNN
base dataset	-	142	85	0.96	0.9	0.94	0.9
A+D+S	att	239	85	0.96	0.9	0.98	0.91
A+D	att	208	85	0.94	0.9	0.97	0.9
A+S	att	173	85	0.96	0.92	0.97	0.9
A+D+S	exa	142	245	0.9	0.83	0.77	0.66
A+D	exa	142	175	0.92	0.92	0.83	0.76
A+S	exa	142	155	0.92	0.9	0.9	0.83
A+D+S	comb	239	245	0.85	0.8	0.83	0.79
A+D	comb	208	175	0.93	0.88	0.9	0.82
A+S	comb	173	155	0.9	0.88	0.84	0.79

(a)

dataset	fusion	attributes	examples	RF	R	NB	kNN
base dataset	-	142	85	0.91	0.84	0.88	0.81
A+D+S	att	239	85	0.91	0.86	0.92	0.81
A+D	att	208	85	0.91	0.87	0.9	0.81
A+S	att	173	85	0.92	0.87	0.9	0.89
A+D+S	exa	142	245	0.83	0.74	0.78	0.69
A+D	exa	142	175	0.83	0.84	0.75	0.68
A+S	exa	142	155	0.84	0.75	0.77	0.71
A+D+S	comb	239	245	0.8	0.7	0.81	0.73
A+D	comb	208	175	0.85	0.8	0.78	0.69
A+S	comb	173	155	0.79	0.74	0.76	0.71

(b)

Table 11.: The best classification accuracies obtained by applying attribute selection on the base Alzheimer's dataset and on different fused datasets. Table (a) shows theoretically best accuracies, while the Table (b) shows accuracies obtained by applying the best attribute selection methods that were chosen on independent datasets. The shades of green colour indicate the improvement of classification accuracy compared with the base data set (in the first row), and the red shades indicate the deterioration. Additionally, darker shades of red and green indicate statistically significant results.

Further, Table 11 displays classification accuracies obtained by applying different attribute selection approaches. The first subtable (Table 11a) displays the accuracies that can be obtained by applying the attribute selection method that yields the best results and can therefore be regarded as maximum achievable accuracies, i.e. theoretically best accuracies. Namely, these accuracies might not be relevant for practical usage of the proposed models, since the attribute selection method was not chosen on independent data. For that reason we introduced the second subtable (Table 11b) which displays the accuracies obtained by applying automatic selection of the attribute selection method. The methods were selected by performing the nested inner 10-by-

10-fold cross-validation (the inner loops were used to select the best attribute selection methods and the outer loops to evaluate the classification performance using it). We can notice the correlation between the results in the both subtables, the second table expectedly displaying slightly lower accuracies. Both tables again indicate that the prevailing significant increases of classification accuracy occurred with the attribute enrichment fusion approach and especially with the naive Bayes classifier. This combination of approach and classifier yielded the highest overall classification accuracy of 98 % in Table 11a and 92 % in Table 11b.

The results do not allow us to draw any consistent conclusions about the choice of the fused datasets as the best results differ from one fusion approach to the other. For the *att* approach the best average results were obtained with datasets A+D+S and A+S and the worst with A+D; for the *exa* approach the best results were obtained with A+S and the worst with A+D+S; and for the *comb* approach the best results were obtained with A+D and the worst with A+D+S.

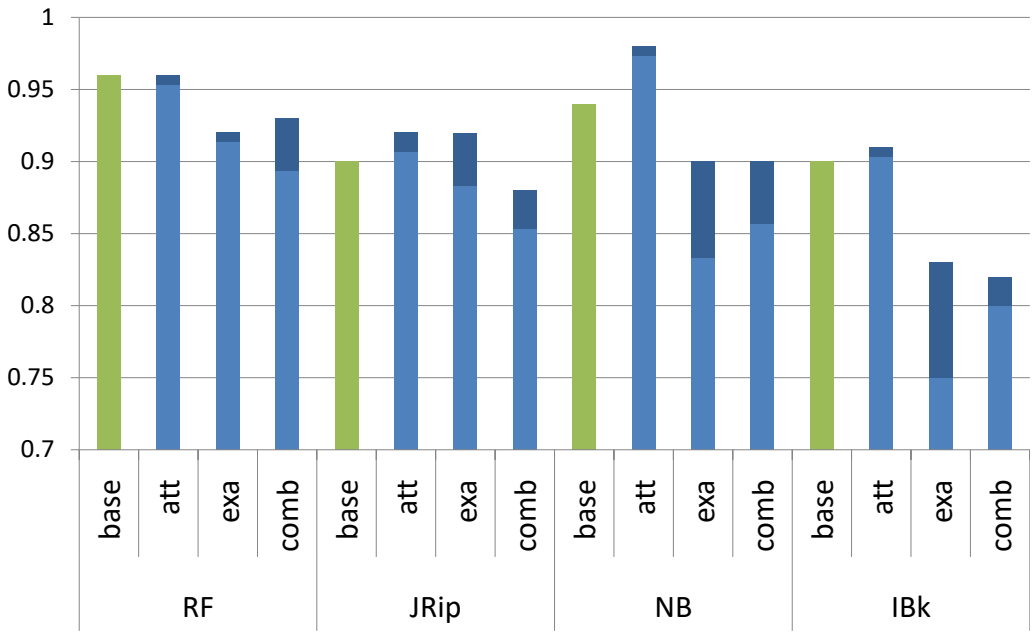

Figure 1.: Average (light blue) and the highest (dark blue) classification accuracies on fused datasets (blue), compared with the accuracy on the base Alzheimer's dataset (green).

Figure 1 displays the average and maximum performance of different data fusion methods/classifiers to enable easier comparison. We can see that the *att* approach performs better on the average with all classifiers except with RF. Note that although the *exa* and *comb* approaches decreased the classification accuracy on the average, the *exa* approach achieved a good performance when used with the RIPPER classifier.

The heterogeneous results seem promising and motivating for further research on how to select the best performing classifier for a given fused dataset and fusion approach.

5. Conclusion

In the paper we proposed three different data fusion approaches, and applied them to a problem of Alzheimer's disease prediction. We operated on three different datasets: dataset of patients impaired with Alzheimer's disease (A), dataset of patients with vascular dementia (D), and dataset of patients who had a stroke (S). By performing attribute enrichment (*att*), example enrichment (*exa*) and the combination of both (*comb*) we expected to achieve an increase in classification accuracy compared to the accuracy on the base dataset.

Since attribute enrichment increases the dimensionality of the problem space, we tackled this problem by applying various attribute selection methods. In our experiments we obtained the best results by applying attribute enrichment data fusion method, especially with the naive Bayes. In this case, the classification accuracy significantly increased to 98 % compared to the base classifier that achieved the 94 % accuracy.

A very diverse range of obtained classification accuracies for different data fusion approaches, fused datasets and classifiers motivates several ideas for our further work, as follows. First, we shall apply the most promising approach identified in this paper on a greater set of problem domains to verify if the conclusions from this paper can be generalised. Secondly, since all of our data fusion approaches assume that the datasets share a sufficient number of attributes, we shall perform further evaluations of the proposed approaches' sensitivity with respect to the number of shared attributes. Further, we shall investigate performance of more complex (non-linear) supervised learning algorithms, including deep neural networks, within the proposed data fusion process. Finally, we shall further investigate how to select the best performing classifier for a given combination of data fusion approach, dataset and classifier.

6. Compliance with Ethical Standards

Funding:

This work is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (projects 451-03-68/2020-14/200125, ON179006), by the Provincial Secretariat of Vojvodina for Higher Education and Scientific Research (project 142-451-2486/2017), and by the bilateral project between Slovenia and Serbia: grant no. 44/2019, "Techniques for inductive learning from a wide range of partially labelled datasets about various diseases" between the Ministry of Education, Science and Technological Development of the Republic of Serbia and the Slovenian Research Agency.

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

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

Alzheimer's Association. (2016). 2016 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *12*(4), 459–509.

Benzebouchi, N. E., Azizi, N., Ashour, A. S., Dey, N., & Sherratt, R. S. (2019). Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis. *Journal of Experimental & Theoretical Artificial Intelligence*, *31*(6), 841–874.

Bi, X.-a., Cai, R., Wang, Y., & Liu, Y. (2019). Effective diagnosis of alzheimer's disease via multimodal fusion analysis framework. *Frontiers in Genetics*, *10*, 976. Retrieved from `https://www.frontiersin.org/article/10.3389/fgene.2019.00976`

Bratić, B., Kurbalija, V., Ivanović, M., Oder, I., & Bosnić, Z. (2018). Machine learning for predicting cognitive diseases: Methods, data sources and risk factors. *Journal of medical systems*, *42*(12), 243.

Chui, H. C., Victoroff, J., Margolin, D., Jagust, W., Shankle, R., & Katzman, R. (1992). Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's Disease Diagnostic and Treatment Centers. *Neurology*, *42*(3), 473–473.

Dai, G., Fu, X., Dai, W., & Lu, S. (2017). Social evaluation of innovative drugs: A method based on big data analytics. *Computer Science & Information Systems*, *14*(3).

Doecke, J. D., Laws, S. M., Faux, N. G., Wilson, W., Burnham, S. C., Lam, C.-P., ... Lifestyle Research Group (2012). Blood-based protein biomarkers for diagnosis of Alzheimer disease. *Archives of neurology*, *69*(10), 1318–1325.

Dyrba, M., Ewers, M., Wegrzyn, M., Kilimann, I., Plant, C., Oswald, A., ... The EDSD study group (2013). 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.

Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine learning for medical imaging. *Radiographics*, *37*(2), 505–515.

Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., Rueckert, D., & Alzheimer's Disease Neuroimaging Initiative. (2013). Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage*, *65*, 167–175.

Gray, K. R., Wolz, R., Heckemann, R. A., Aljabar, P., Hammers, A., Rueckert, D., & Alzheimer's Disease Neuroimaging Initiative. (2012). Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. *Neuroimage*, *60*(1), 221–229.

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M. K., Johnson, S. C., & Alzheimer's Disease Neuroimaging Initiative. (2009). Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage*, *48*(1), 138–149.

Hinrichs, C., Singh, V., Xu, G., & Johnson, S. C. (2011). Predictive markers for ad in a multi-modality framework: An analysis of mci progression in the adni population. *NeuroImage*, *55*(2), 574 - 589. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1053811910015806`

Huang, M., Yang, W., Feng, Q., Chen, W., & Alzheimer's Disease Neuroimaging Initiative. (2017). Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. *Scientific reports*, *7*.

Jiji, G. W. (2018). Identifying stage of alzheimer disease using multiclass particle swarm optimisation technique. *Journal of Experimental & Theoretical Artificial Intelligence*, *30*(6), 911–925.

Kim, J., & Lee, B. (2018). Identification of alzheimer's disease and mild cognitive impairment using multimodal sparse hierarchical extreme learning machine. *Human Brain Mapping*, *39*(9), 3728-3741. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24207`

Kippenhan, J. S., Barker, W. W., Pascal, S., Nagel, J. H., & Duara, R. (1992). Evaluation of a neural-network classifier for PET scans of normal and Alzheimer's disease subjects.

Kohannim, O., Hua, X., Hibar, D. P., Lee, S., Chou, Y.-Y., Toga, A. W., ... Thompson, P. M. (2010). Boosting power for clinical trials using classifiers based

on multiple biomarkers. *Neurobiology of Aging*, *31*(8), 1429 - 1442. Retrieved from http://www.sciencedirect.com/science/article/pii/S0197458010001879 (Alzheimer's Disease Neuroimaging Initiative (ADNI) Studies)

Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, *23*(1), 89–109.

Liu, M., Zhang, D., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2012). Ensemble sparse classification of Alzheimer's disease. *NeuroImage*, *60*(2), 1106–1116.

Llano, D. A., Devanarayan, V., Simon, A. J., & Alzheimer's Disease Neuroimaging Initiative. (2013). 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.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). 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), 939–939.

Ning, K., Chen, B., Sun, F., Hobel, Z., Zhao, L., Matloff, W., & Toga, A. W. (2018). Classifying alzheimer's disease with brain imaging and genetic data using a neural network framework. *Neurobiology of Aging*, *68*, 151 - 158. Retrieved from http://www.sciencedirect.com/science/article/pii/S0197458018301313

Ortiz, A., Fajardo, D., Gorriz, J., Ramírez, J., & Martínez-Murcia, F. (2014, 12). Multimodal image data fusion for alzheimer's disease diagnosis by sparse representation. In (Vol. 207, p. 11-8).

Parikh, D., Stepenosky, N., Topalis, A., Green, D., Kounios, J., Clark, C., & Polikar, R. (2005). Ensemble based data fusion for early diagnosis of alzheimer's disease. In *2005 ieee engineering in medicine and biology 27th annual conference* (p. 2479-2482).

Polikar, R., Topalis, A., Parikh, D., Green, D., Frymiare, J., Kounios, J., & Clark, C. M. (2008). An ensemble based data fusion approach for early diagnosis of Alzheimer's disease. *Information Fusion*, *9*(1), 83–95.

Pritchard, W. S., Duke, D. W., Coburn, K. L., Moore, N. C., Tucker, K. A., Jann, M. W., & Hostetler, R. M. (1994). 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.

Qin, P., Chen, J., Zhang, K., & Chai, R. (2018). Convolutional neural networks and hash learning for feature extraction and of fast retrieval of pulmonary nodules. *Computer Science & Information Systems*, *15*(3).

Ray, S., Britschgi, M., Herbert, C., Takeda-Uchimura, Y., Boxer, A., Blennow, K., ... Wyss-Coray, T. (2007). Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nature medicine*, *13*(11), 1359–1362.

Roberson, E. D., & Mucke, L. (2006). 100 years and counting: prospects for defeating Alzheimer's disease. *science*, *314*(5800), 781–784.

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine learning*, *53*(1-2), 23–69.

Román, G. C., Tatemichi, T. K., Erkinjuntti, T., Cummings, J., Masdeu, J., Garcia, J., ... Scheinberg, P. (1993). Vascular dementia: diagnostic criteria for research studies: report of the NINDS-AIREN International Workshop. *Neurology*, *43*(2), 250–250.

Saâdaoui, F., Bertrand, P. R., Boudet, G., Rouffiac, K., Dutheil, F., & Chamoux, A. (2015). A dimensionally reduced clustering methodology for heterogeneous occupational medicine data mining. *IEEE transactions on nanobioscience*, *14*(7), 707–715.

Slavkov, I., Karcheska, J., Kocev, D., & Dzeroski, S. (2018). HMC-ReliefF: Feature ranking for hierarchical multi-label classification. *Computer Science & Information Systems*, *15*(1), 187–209.

Suk, H.-I., Lee, S.-W., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*, *101*, 569–582.

Suk, H.-I., & Shen, D. (2013). Deep learning-based feature representation for ad/mci classification. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, & N. Navab (Eds.), *Medical image computing and computer-assisted intervention – miccai 2013* (pp. 583–590). Berlin, Heidelberg: Springer Berlin Heidelberg.

Sun, Z., Qiao, Y., Lelieveldt, B. P., & Staring, M. (2018). Integrating spatial-anatomical regularization and structure sparsity into svm: Improving interpretation of alzheimer's disease classification. *NeuroImage*, *178*, 445 - 460. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1053811918304658`

Teipel, S., Drzezga, A., Grothe, M. J., Barthel, H., Chételat, G., Schuff, N., ... Fellgiebel, A. (2015). Multimodal imaging in alzheimer's disease: validity and usefulness for early detection. *The Lancet Neurology*, *14*(10), 1037 - 1053. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1474442215000939`

Tierney, M., Szalai, J., Snow, W., Fisher, R., Nores, A., Nadon, G., ... George-Hyslop, P. S. (1996). Prediction of probable Alzheimer's disease in memory-impaired patients a prospective longitudinal study. *Neurology*, *46*(3), 661–665.

Walhovd, K., Fjell, A., Brewer, J., Mcevoy, L., Fennema-Notestine, C., Hagler, D., ... Dale, A. (2010, 02). Combining mr imaging, positron-emission tomography, and csf biomarkers in the diagnosis and prognosis of alzheimer disease. *AJNR. American journal of neuroradiology*, *31*, 347-54.

Westman, E., Muehlboeck, J.-S., & Simmons, A. (2012). Combining mri and csf measures for classification of alzheimer's disease and prediction of mild cognitive impairment conversion. *NeuroImage*, *62*(1), 229 - 238. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1053811912004521`

Ye, J., Chen, K., Wu, T., Li, J., Zhao, Z., Patel, R., ... Reiman, E. (2008, 08). Heterogeneous data fusion for alzheimer's disease study. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1025-1033.

Yoo, C., Ramirez, L., & Liuzzi, J. (2014). Big data analysis using modern statistical and machine learning methods in medicine. *International neurourology journal*, *18*(2), 50.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*, *55*(3), 856–867.

Zhang, Y., Dong, Z., Phillips, P., Wang, S., Ji, G., Yang, J., & Yuan, T.-F. (2015). 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.