UK Symposium on Knowledge Discovery and Data Mining 2016

Organised by BCS SGAI - The Specialist Group on Artificial Intelligence

Data-driven Knowledge Discovery of Atypical Brain States

Dr. Giuseppe Di Fatta

Associate Professor of Computer Science

Director MSc Advanced Computer Science

University of Reading

G.DiFatta@reading.ac.uk

http://www.personal.reading.ac.uk/~sis06gd/

London, 21 March 2016

Outline

- Motivations
  - Dementia and Alzheimer's Disease (AD)
- The Human Brain
  - Gross Anatomy
  - Structural Magnetic Resonance Imaging (MRI)
- Data-driven Knowledge Discovery
  - Pre-processing:
    - Feature generation
    - Data manipulation
  - Feature selection
  - Classification
- Conclusions

Human Brain Awareness Week

- Every year in March, many organizations promote activities such as open days at neuroscience labs, exhibitions and lectures about the brain, to raise awareness and attention.

- Importance of early diagnosis of neurodegenerative diseases
  - Quality of life can be significantly improved.
  - Mental disorders are one of the five most costly conditions for medical spending
- Dementia is the most expensive condition for medical and social cost
  - Still it receives only a fraction of the amount spent on research into other health conditions.
  - It is not a normal part of aging.
  - Need for understanding causes to find better therapies and cure.

Alzheimer's Disease (AD)

- AD accounts for 60 to 80 percent of dementia cases.
- More than 40 million people worldwide suffer from AD, 150 million by 2050.
- It is a neurodegenerative disease that slowly and progressively destroys brain cells.
  - German neurologist Alois Alzheimer in 1907 first described the symptoms and identified the brain anomalies associated with it.

- AD affects memory, behaviour and cognitive function leading to confusion, changes of mood and disorientation.
- AD is most often diagnosed in people over 65 years. Early symptoms are mild cognitive difficulties (up to 8 years before), such as short term memory loss.
- In the late stage it causes a general deterioration in health and is terminal. Life expectancy following diagnosis is 3-9 years.


NHS

Alzheimer's disease

Overview | Real stories | Clinical trials | Community

Alzheimer's disease | Symptoms | Causes | Diagnosis | Treatment

Introduction

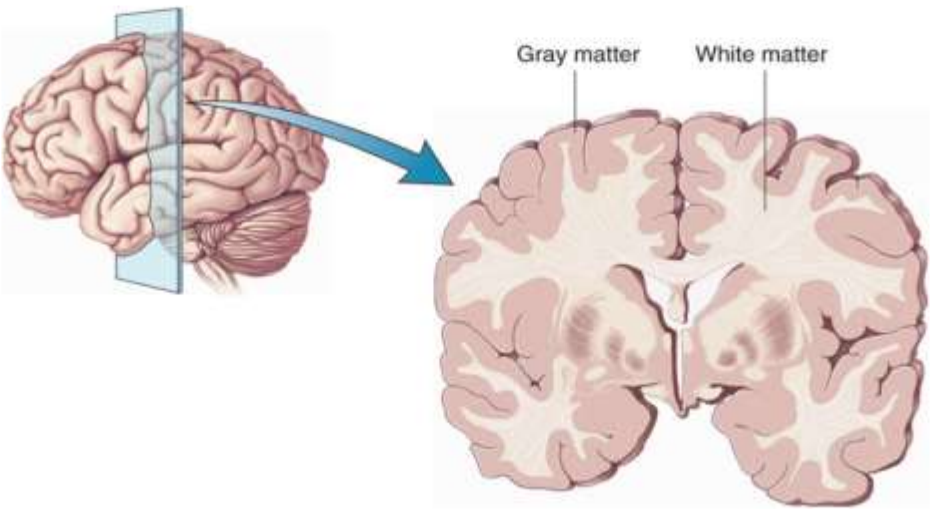
Alzheimer's disease is the most common type of dementia, affecting almost 500,000 people in the UK.

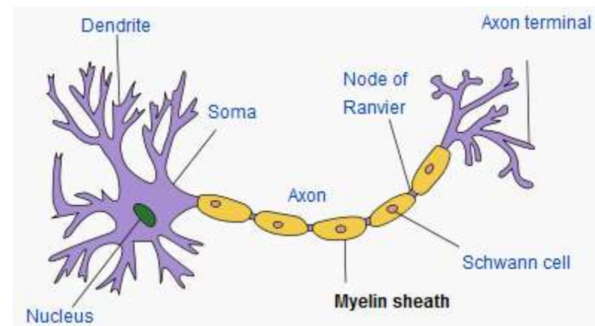
The term "dementia" describes a loss of mental ability associated with gradual death of brain cells.

The exact cause of Alzheimer's disease is unknown

The Human Brain

- The cerebral cortex (grey matter) is the outer layer of neural tissue
  - Grey folded tissue that consists mainly of <u>cell bodies</u>.
  - It is primarily associated with processing and cognition.
  - It has a stratified structure.

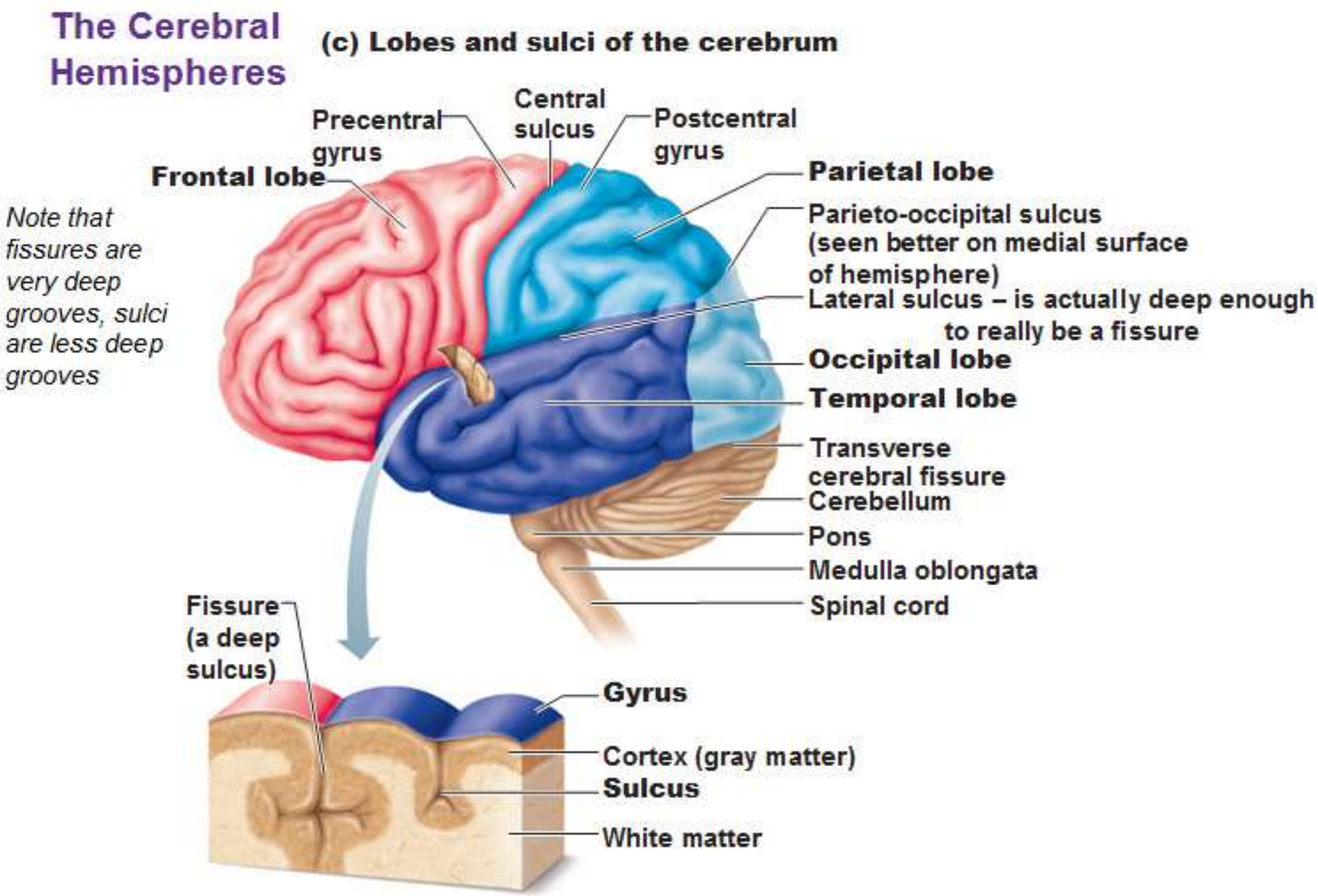




- The subcortical white matter
  - White matter that consists mainly of long-range <u>neuronal axons</u>.
    - It forms the deep parts of the brain and the superficial parts of the spinal cord.
    - It modulates the distribution of action potentials, i.e. communication between different brain regions.
    - Myelin is a fatty <u>white</u> substance that surrounds the axon, forming an electrically insulating layer.

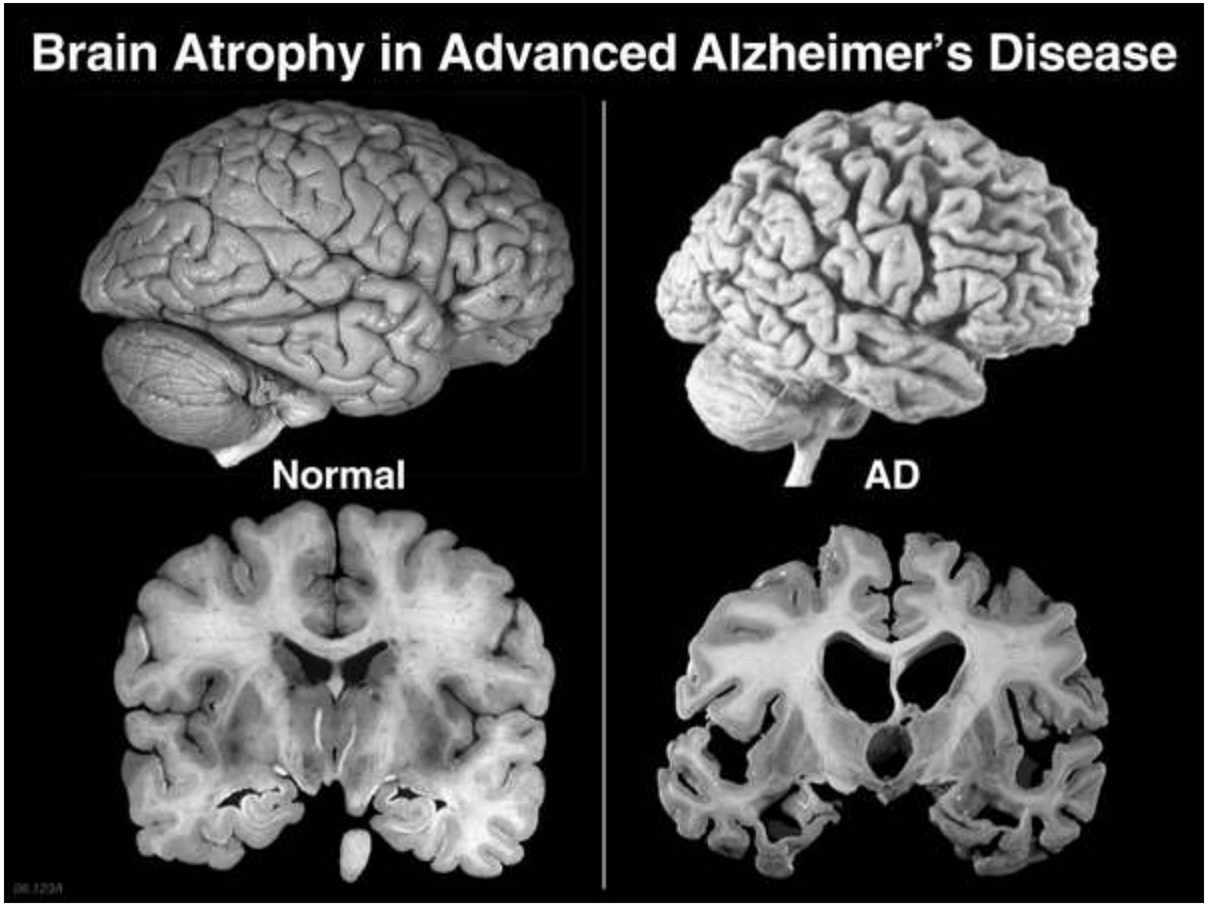
The Human Brain

- Gross anatomy
  - Lobes, cortical and subcortical structures
  - Sulci and Gyri



Alzheimer's Disease (AD)

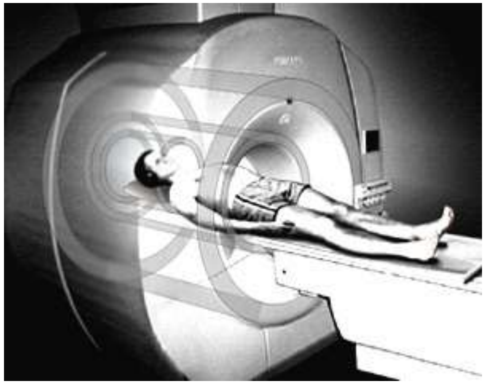
University of Reading

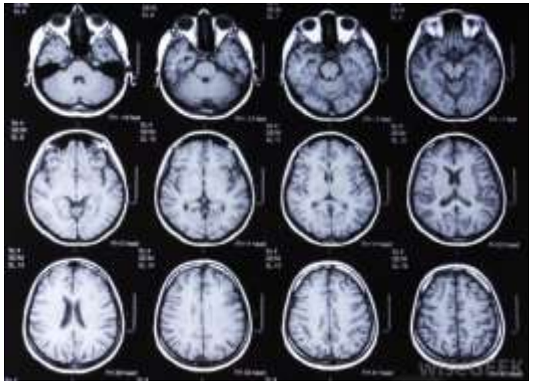


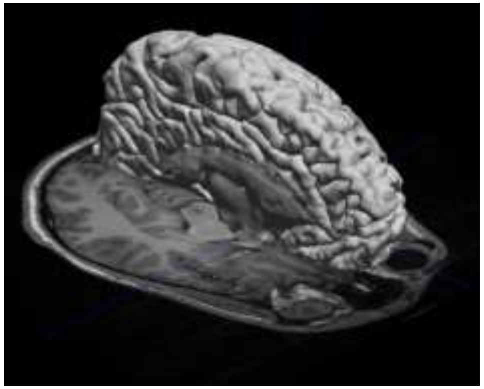
Magnetic Resonance Imaging (MRI)

University of Reading

- MRI scanners use strong magnetic fields, radio waves, and field gradients to form images of the body.
- A scan generates a highly detailed view of tissues within the body.
- Structural MRI: a high resolution 3D image is generated.

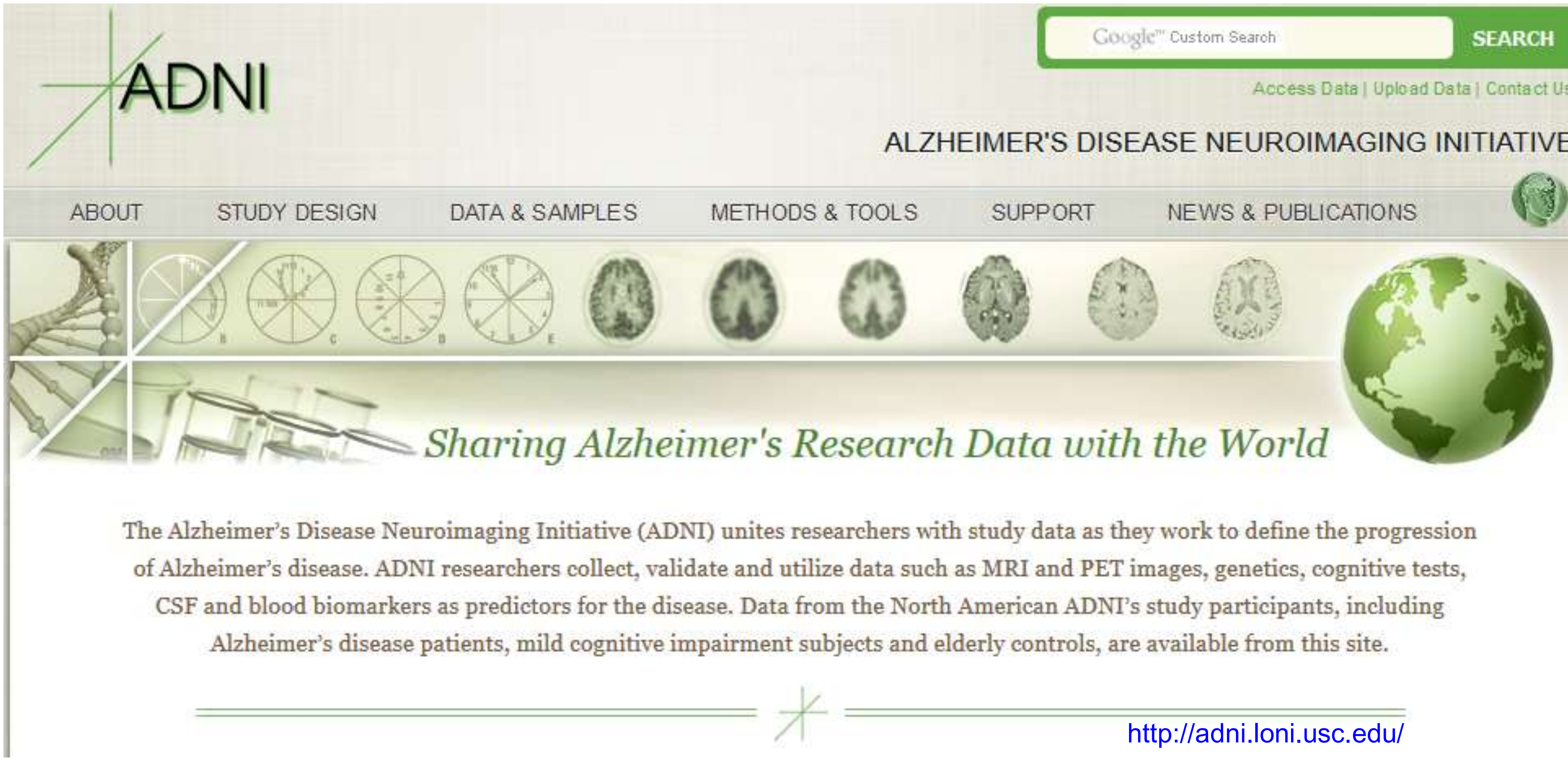






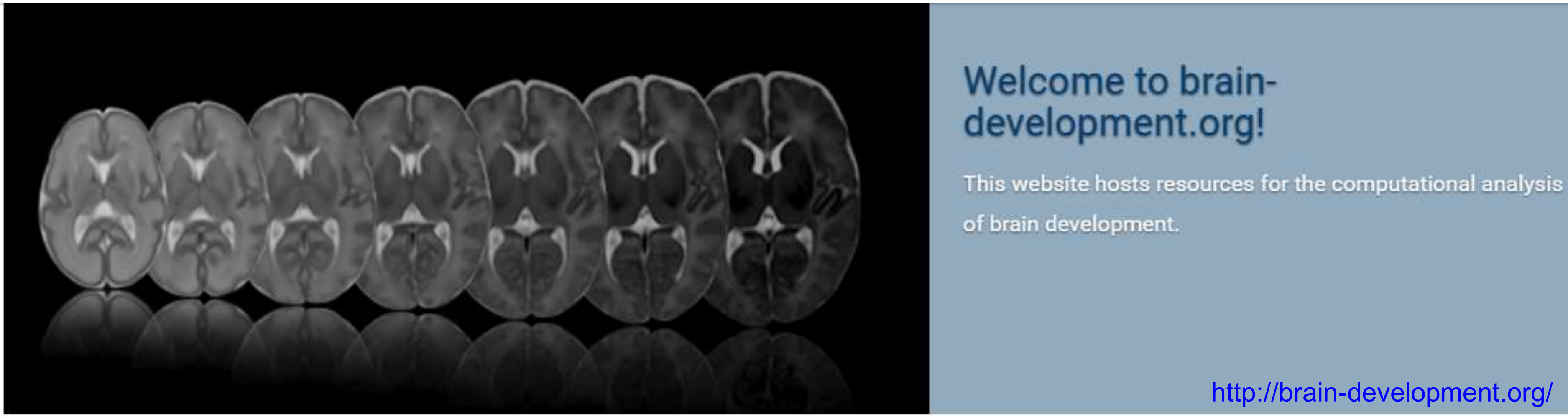
- Brain MRI data are more frequently collected and made publicly available.
- Examples of publicly available data sets include…

Alzheimer's Disease Neuroimaging Initiative



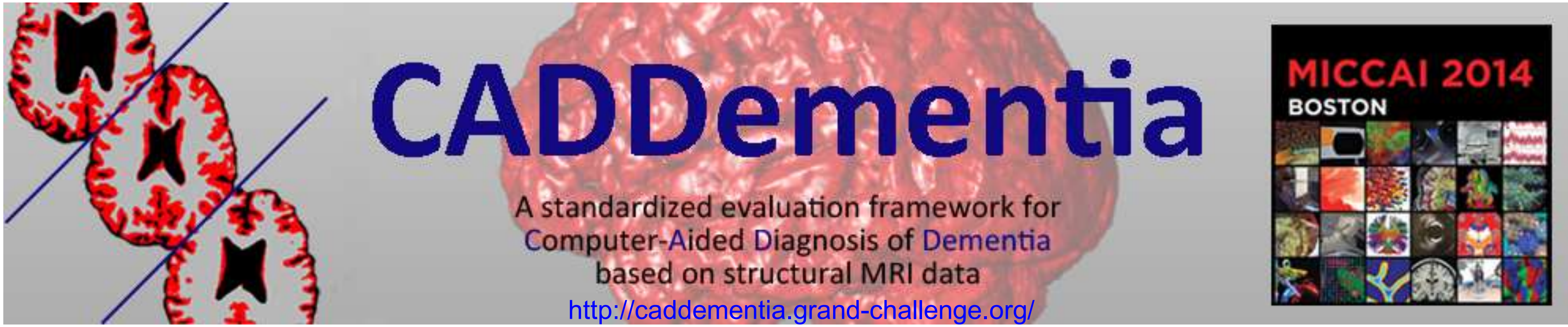
http://adni.loni.usc.edu/

Information eXtraction from Images



- Information eXtraction from Images (IXI) dataset
  - nearly 600 MR images from normal, healthy subjects
  - The data has been collected at three different hospitals in London:
    - Hammersmith Hospital
    - Guy’s Hospital
    - Institute of Psychiatry

CADDementia



- Computer-aided diagnosis methods based on structural brain MRI may improve the early diagnosis of dementia, but frequently the methods are optimized for specific data sets.
- CADDementia is a “grand challenge” workshop to objectively validate and compare methods for computer-aided diagnosis based on structural brain MRI scans.
- A multi-center data set including scans of 384 subjects.
  - Erasmus MC (EMC), Rotterdam, The Netherlands;
  - VU University Medical Center (VUmc), Amsterdam, The Netherlands;
  - University of Porto / Hospital de São João (UP), Porto, Portugal

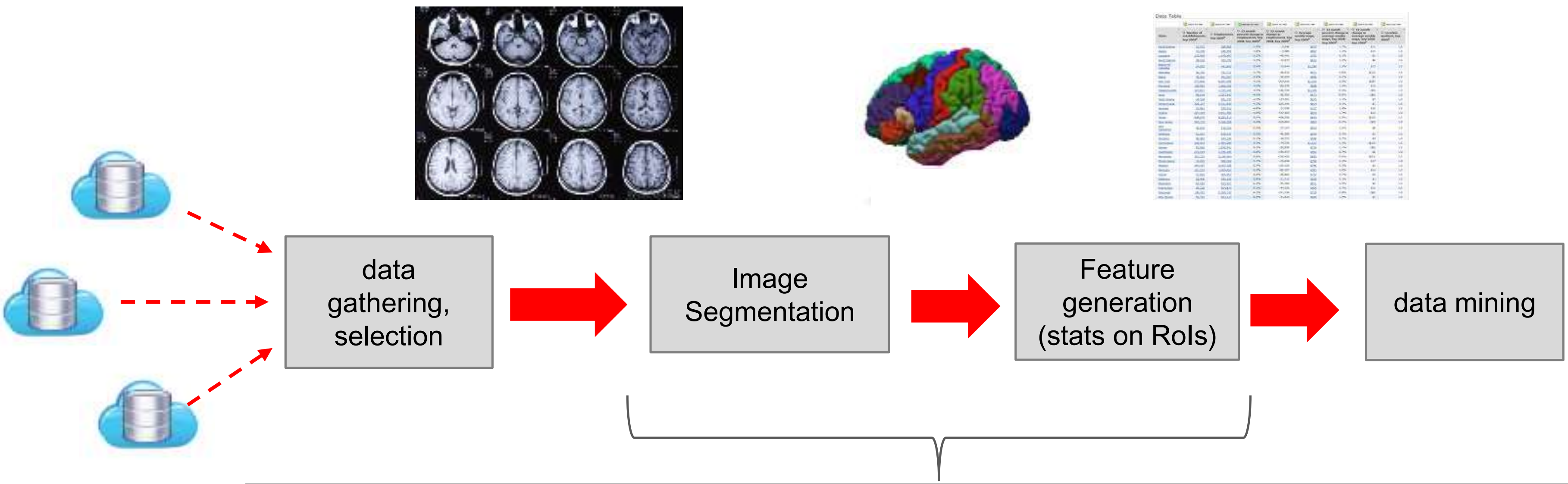
Predictive Task

- The data sets contain structural MRI scans of the participants, demographic information and a class label (diagnosis):
  - Alzheimer's Disease (AD)
  - Mild Cognitive Impairment (MCI)
    - It is not a type of dementia.
    - Brain function syndrome with similar symptoms to AD (lesser extent)
    - Recognised as a risk factor for developing AD
  - Healthy Controls (HC)
    - Participants without a dementia syndrome

❑ Predictive task

  - Classify whether a subject is healthy or has MCI or AD based on a structural MRI scan.

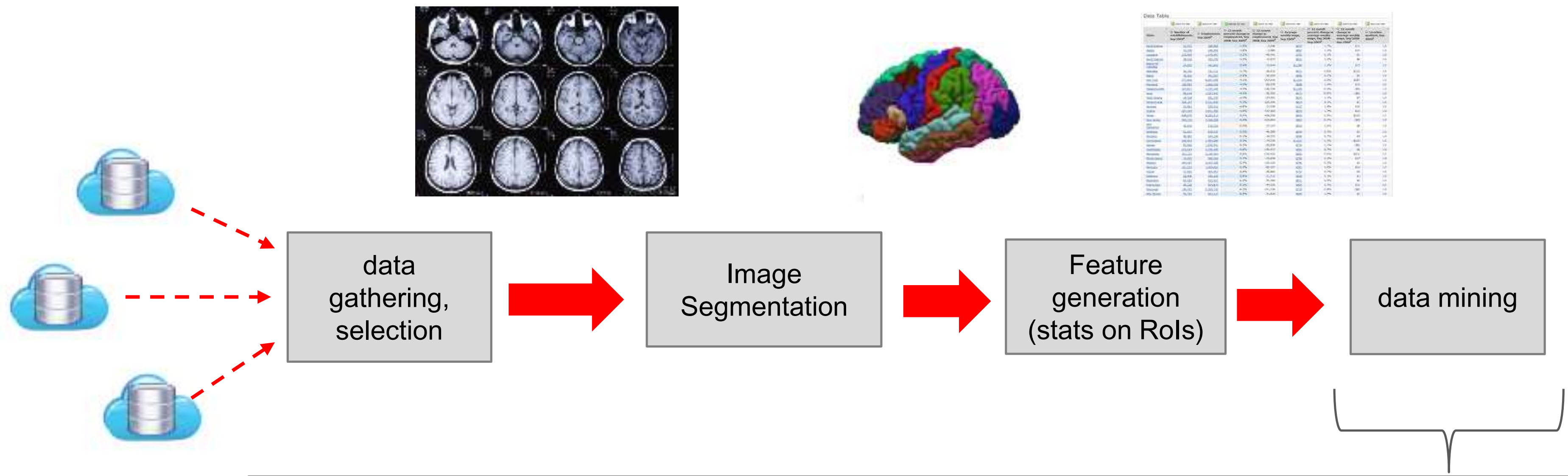
Data Workflow



[Pre-processing]

- A single MRI scan file typically takes 30-40 MB.
- After pre-processing the files generated for a single image take 350-500 MB.
- We used MRI data of ~500 subjects: ~200GB.
- Pre-procesing of a single image takes ~8h → ~168 days in total.
- Data parallel approach used to speed up computation → ~10 days.

Data Workflow



Data Mining Predictive Task

- [Feature selection]
- Generate a predictive model from a training set: classification method
- Estimate accuracy of the model on a test set
- Use cross-validation to obtain a reliable estimate of the performance

Pre-processing Tools

University of Reading

- **FreeSurfer** is an open source software suite for processing and analyzing human brain MRI images. (http://surfer.nmr.mgh.harvard.edu/)
  - Image Registration, Subcortical Segmentation, Cortical Surface Reconstruction, Cortical Segmentation, Cortical Thickness Estimation, etc.

- The **FMRIB Software Library (FSL)** provides image analysis and statistical tools for fMRI, MRI and DTI brain imaging data. (http://fsl.fmrib.ox.ac.uk/)

Feature Generation

A total of 361 attributes:

- Morphological measures of specific regions of interests of left and right hemispheres:
  - 138 cortical thickness measures (mean and standard deviation)
  - 70 cortical areas
  - 68 cortical volumes
  - 16 hippocampal subfield volumes
  - 66 volumes of subcortical structures
- age, gender, diagnosis

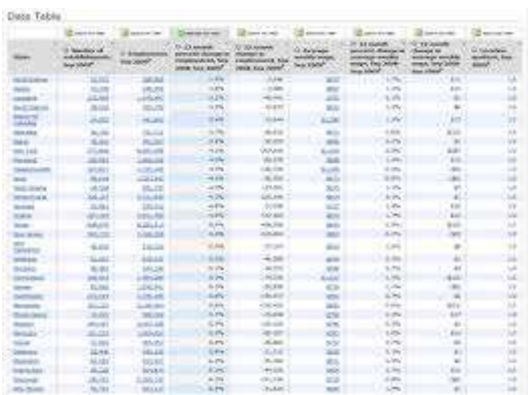
volume data

area data

thickness data

diffusion data

demographical/clinical data

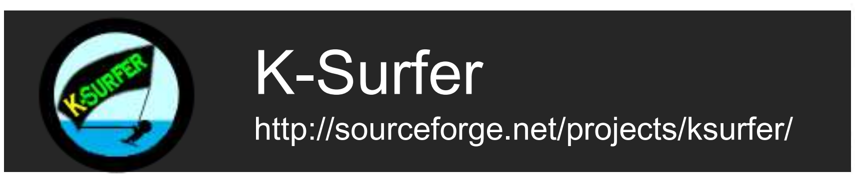


Data Manipulation

- FreeSurfer/FSL tools are based on UNIX shell scripts and generate a large number of files for a single image.
- Data manipulation may be a limiting factor for fast prototyping.

➢ Adopting an advanced analytics platform
  - KNIME, the Konstanz Information Miner

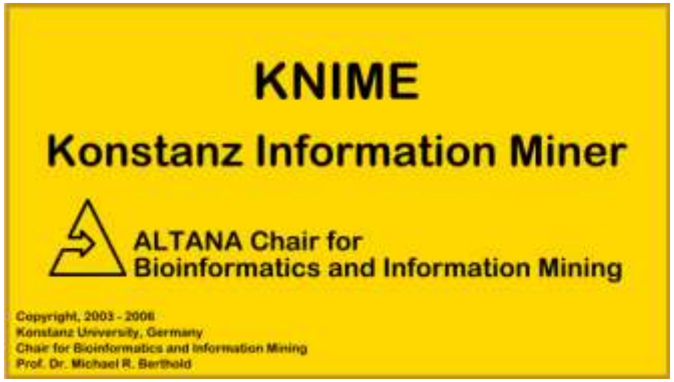
➢ Introducing K-Surfer, a plugin for KNIME
  - a tool for importing, managing and analysing FreeSurfer/FSL data in KNIME



Advanced Analytics Platforms

Gartner 2016 Magic Quadrant for Advanced Analytics Platforms

(source: https://www.gartner.com/doc/reprints?ct=160210&id=1-2YEVU8U&st=sb)

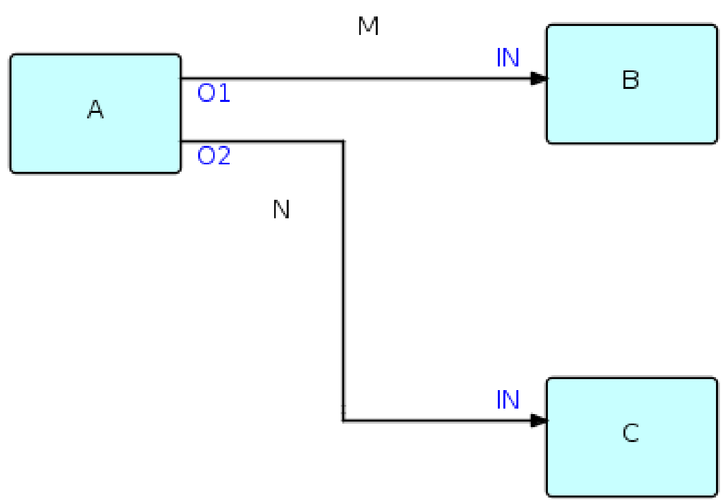


- A platform for data analysis, data manipulation, data visualization and reporting
  - Originally developed at the Department of Computer and Information Science, University of Konstanz, Germany
  - Under continuous evolution and extension
    - 1st release in April 2006
    - ver.3 released in Dec. 2015

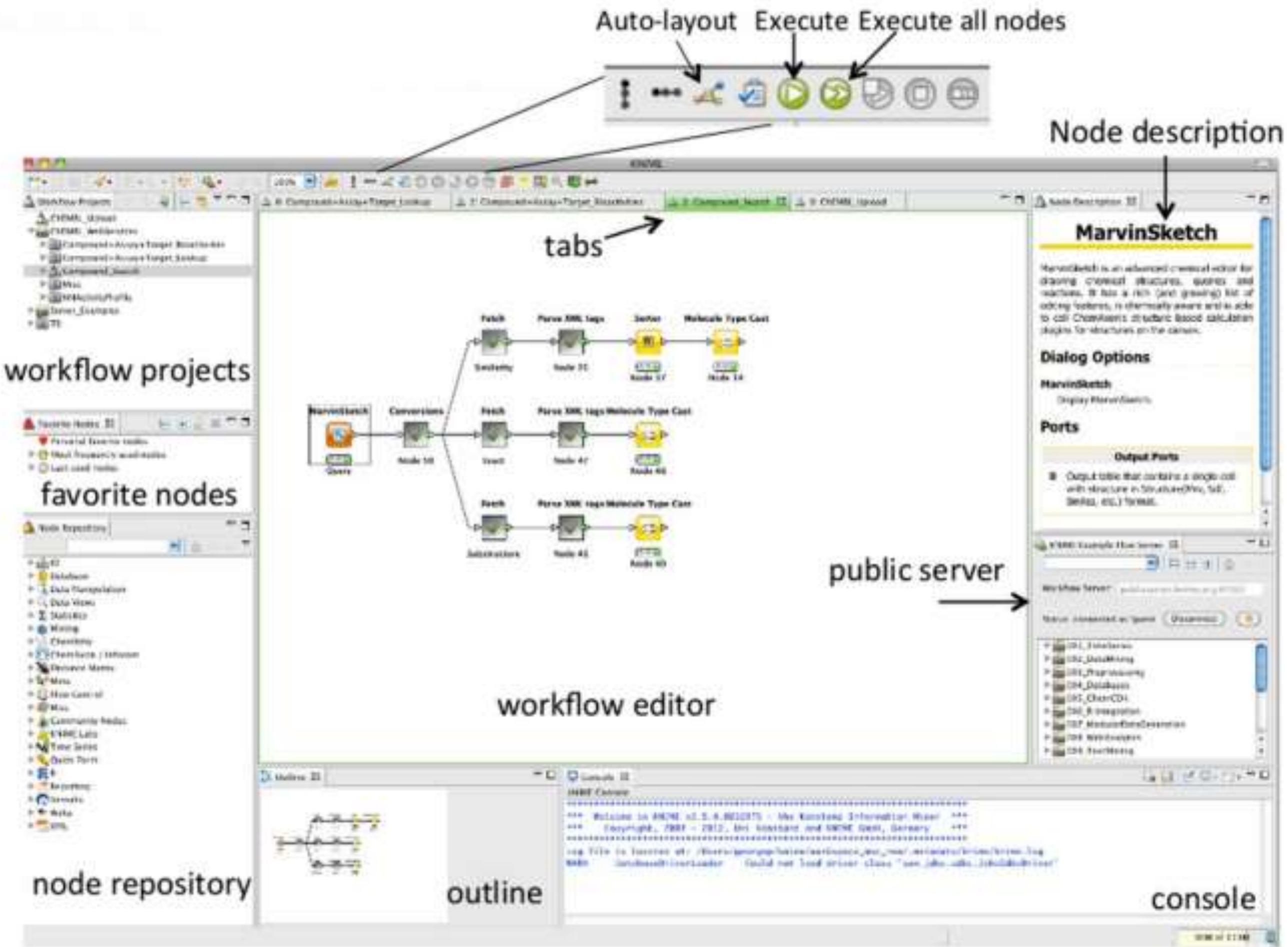
- based on Eclipse, free and open source
- flexibility, openness and ease of integration with other tools

Flow-based Programming

- Flow-based Programming (FBP) is a programming paradigm that defines applications as networks of "black box" processes, which exchange data across predefined connections by message passing, where the connections are specified externally to the processes.
- These black box processes can be reconnected endlessly to form different applications without having to be changed internally. FBP is thus naturally component-oriented.



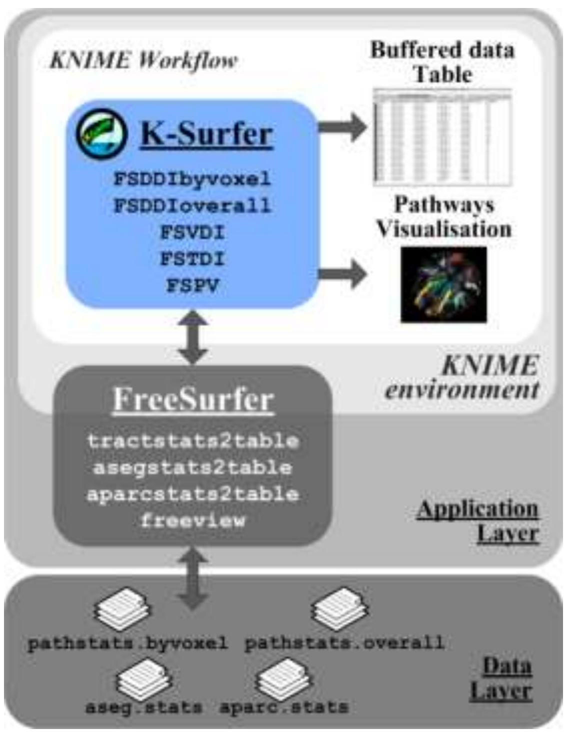
KNIME Workflow

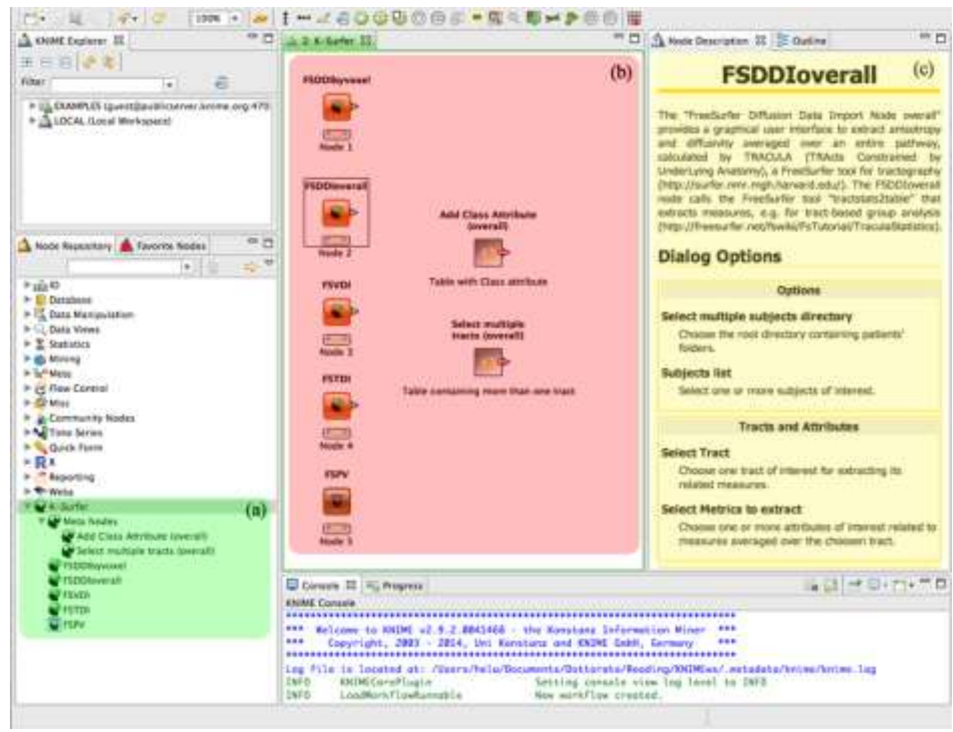


K-Surfer

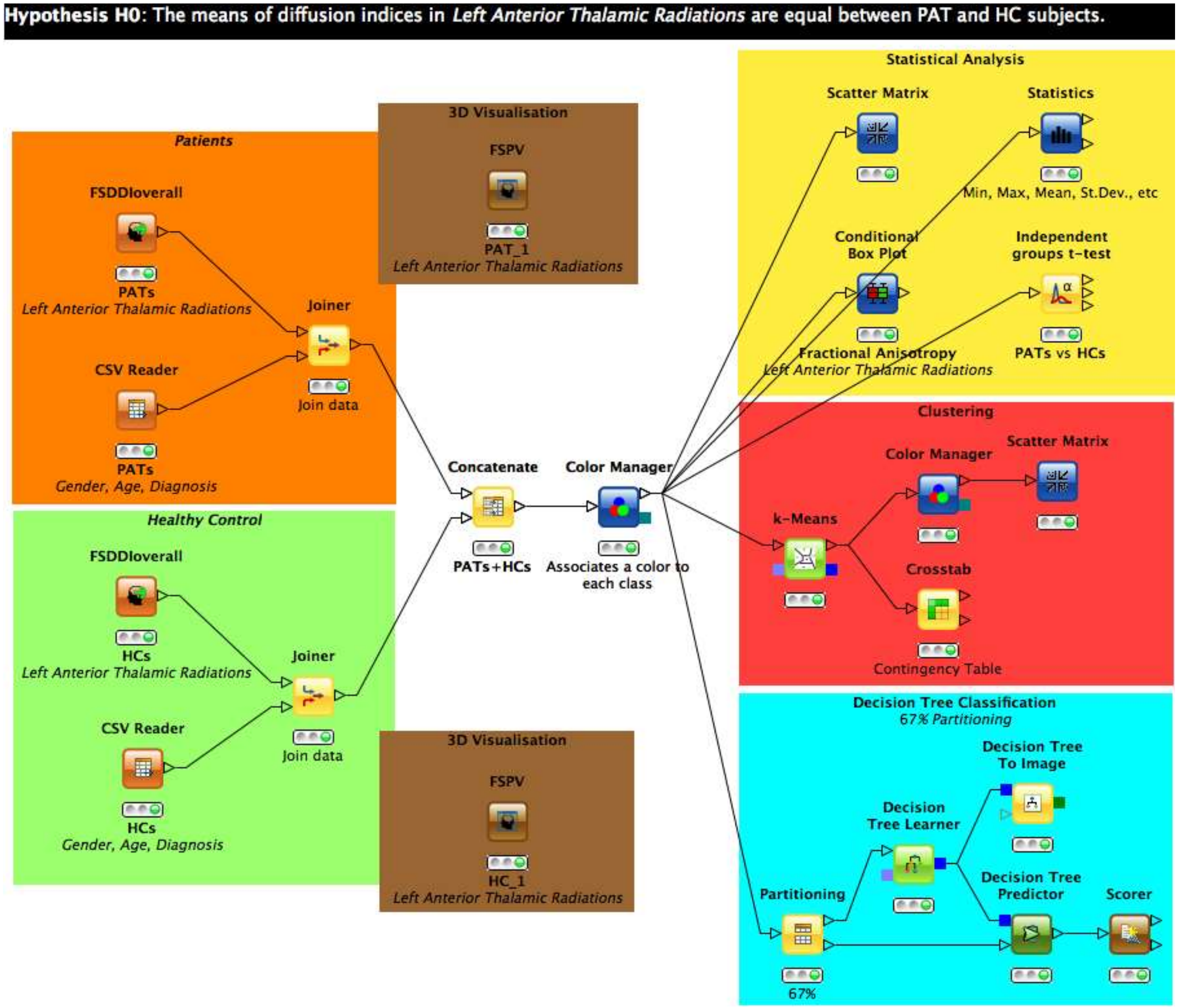
University of Reading

- K-Surfer is a KNIME plug-in that contributes a number of nodes and meta-nodes to import FreeSurfer and FSL data into KNIME.
- K-Surfer simplifies the importing of multi-dimensional data for group analysis based on the volume, thickness and diffusion data of neuroimages.





Example of Data Workflow with K-Surfer



CADDementia Challenge 2014

University of Reading

- K-Surfer was successfully used in the CADDementia challenge 2014.
  - The results are available at: http://caddementia.grand-challenge.org/results_all/

Original results table presented on 18 September 2014 at the MICCAI workshop (*Bron et al., 2015*).

Algorithm	Rank accuracy	Accuracy [CI] (%)	TPF_{CN} [CI] (%)	TPF_{MCI} [CI] (%)	TPF_{AD} [CI] (%)	Rank AUC	AUC_{all} [CI] (%)	AUC_{CN} [CI] (%)	AUC_{MCI} [CI] (%)	AUC_{AD} [CI] (%)	Date	Documentation
Soerensen-equal	1.0	63.0 [57.9 - 67.5]	96.9 [92.9 - 99.2]	28.7 [21.3 - 37.4]	61.2 [51.6 - 69.8]	1.5	78.8 [75.6 - 82.0]	86.3 [81.8 - 89.3]	63.1 [56.6 - 68.3]	87.5 [83.4 - 91.1]	15/06/2014	Paper | Wiki | Pres.
Soerensen-optimized	2.0	59.9 [54.8 - 64.7]	70.5 [62.8 - 77.8]	41.0 [33.3 - 50.0]	68.9 [59.6 - 77.2]	1.5	78.8 [75.5 - 82.1]	86.3 [81.9 - 89.3]	62.7 [56.8 - 68.4]	86.7 [82.3 - 90.4]	15/06/2014	Paper | Wiki | Pres.
Wachinger-enetNorm	3.0	59.0 [54.0 - 63.6]	72.1 [63.4 - 79.2]	51.6 [43.5 - 61.3]	51.5 [41.5 - 61.2]	4.0	77.0 [73.6 - 80.3]	83.3 [78.5 - 87.0]	59.4 [52.9 - 65.5]	88.2 [83.8 - 91.4]	15/06/2014	Paper | Wiki | Pres.
Ledig-ALL	4.0	57.9 [52.5 - 62.7]	89.1 [83.7 - 93.8]	41.0 [32.4 - 49.6]	38.8 [30.7 - 50.0]	5.0	76.7 [73.6 - 79.8]	86.6 [82.7 - 89.8]	59.7 [53.3 - 65.1]	84.9 [79.7 - 88.7]	15/06/2014	Paper | Wiki | Pres.
Moradi	5.0	57.6 [52.3 - 62.4]	57.4 [48.7 - 66.1]	59.8 [51.3 - 68.1]	55.3 [46.7 - 65.2]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Franke	6.0	56.2 [50.8 - 61.3]	58.9 [50.4 - 67.5]	43.4 [34.8 - 51.7]	68.0 [58.8 - 77.1]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Ledig-CORT	7.5	55.1 [49.7 - 59.9]	68.2 [60.5 - 76.0]	45.1 [35.3 - 53.4]	50.5 [41.2 - 60.5]	12.0	73.7 [69.9 - 77.2]	79.6 [75.0 - 84.2]	58.9 [52.9 - 64.9]	82.4 [76.7 - 87.3]	15/06/2014	Paper | Wiki | Pres.
Sensi	7.5	55.1 [50.0 - 60.2]	71.3 [63.6 - 78.8]	40.2 [31.2 - 49.6]	52.4 [42.7 - 62.0]	11.0	73.8 [70.2 - 77.5]	81.7 [77.1 - 85.8]	55.0 [48.8 - 61.0]	83.9 [78.8 - 87.7]	15/06/2014	Paper | Wiki | Pres.
Ledig-GRAD	9.5	54.0 [48.9 - 59.3]	87.6 [81.7 - 92.6]	37.7 [29.3 - 47.5]	31.1 [22.4 - 40.4]	6.0	75.4 [72.4 - 78.6]	85.6 [81.5 - 88.9]	60.3 [53.9 - 66.5]	81.7 [76.3 - 86.1]	15/06/2014	Paper | Wiki | Pres.
Wachinger-step1	9.5	54.0 [48.8 - 59.0]	68.2 [60.2 - 75.4]	41.0 [31.9 - 50.9]	51.5 [42.2 - 61.1]	8.0	74.6 [70.8 - 78.1]	79.1 [73.5 - 83.1]	55.0 [48.5 - 61.4]	89.2 [85.3 - 92.3]	15/06/2014	Paper | Wiki | Pres.
Abdulkadir	12.5	53.7 [48.3 - 58.2]	45.7 [37.0 - 53.6]	65.6 [56.1 - 73.0]	49.5 [39.4 - 58.8]	3.0	77.7 [74.2 - 81.0]	85.6 [81.4 - 89.0]	59.9 [54.1 - 66.4]	86.7 [82.3 - 90.3]	15/06/2014	Paper | Wiki | Pres.
Sarica	12.5	53.7 [48.3 - 58.8]	65.9 [57.4 - 74.2]	39.3 [30.0 - 48.2]	55.3 [44.9 - 64.9]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Wachinger-step1Norm	12.5	53.7 [48.6 - 58.8]	63.6 [54.9 - 71.9]	47.5 [38.4 - 56.6]	48.5 [39.6 - 59.1]	9.5	74.3 [70.5 - 77.9]	79.3 [74.1 - 83.5]	55.5 [48.5 - 61.6]	87.7 [83.7 - 91.1]	15/06/2014	Paper | Wiki | Pres.
Wachinger-step2	12.5	53.7 [47.5 - 58.8]	66.7 [58.1 - 74.1]	38.5 [30.1 - 48.1]	55.3 [45.5 - 65.0]	13.0	72.7 [68.9 - 76.4]	79.3 [74.0 - 83.5]	51.9 [45.3 - 58.7]	86.5 [81.9 - 90.3]	15/06/2014	Paper | Wiki | Pres.
Ledig-MBL	15.0	53.4 [47.7 - 57.9]	82.9 [76.0 - 88.7]	43.4 [35.1 - 52.9]	28.2 [20.2 - 37.4]	7.0	75.2 [72.0 - 78.1]	82.5 [77.8 - 86.0]	57.3 [50.9 - 63.6]	86.4 [81.4 - 89.9]	15/06/2014	Paper | Wiki | Pres.
Wachinger-man	16.0	53.1 [47.7 - 57.9]	61.2 [53.5 - 69.6]	60.7 [51.7 - 70.0]	34.0 [25.7 - 44.7]	9.5	74.3 [70.9 - 77.9]	80.6 [75.7 - 84.9]	56.3 [49.7 - 63.0]	86.1 [81.7 - 90.0]	15/06/2014	Paper | Wiki | Pres.
Eskildsen-ADNI1	17.5	52.0 [46.6 - 56.8]	65.1 [56.9 - 73.2]	32.0 [24.1 - 40.9]	59.2 [49.5 - 68.3]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Eskildsen-FACEADNI1	17.5	52.0 [46.9 - 57.1]	65.1 [56.6 - 73.1]	36.1 [28.1 - 45.5]	54.4 [44.6 - 63.6]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Eskildsen-Combined	19.0	51.1 [45.5 - 56.2]	64.3 [56.2 - 72.3]	35.2 [27.1 - 44.3]	53.4 [43.0 - 62.9]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Dolph	20.0	49.7 [44.6 - 54.8]	84.5 [77.9 - 90.4]	23.0 [16.4 - 31.2]	37.9 [28.9 - 47.3]	17.0	63.0 [59.6 - 67.2]	66.2 [61.3 - 70.3]	55.4 [50.0 - 60.0]	65.8 [60.6 - 71.3]	15/06/2014	Paper | Wiki | Pres.
Routier-adni	21.0	49.2 [43.5 - 54.2]	94.6 [89.8 - 97.7]	11.5 [6.2 - 17.7]	36.9 [27.4 - 46.5]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Eskildsen-FACEADNI2	22.5	48.3 [43.2 - 53.4]	48.8 [40.5 - 57.4]	42.6 [33.9 - 51.3]	54.4 [45.5 - 64.0]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Routier-train	22.5	48.3 [42.9 - 53.4]	48.1 [39.8 - 56.9]	21.3 [14.8 - 29.0]	80.6 [72.2 - 87.3]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Eskildsen-ADNI2	24.5	47.7 [42.1 - 52.8]	59.7 [51.2 - 68.4]	38.5 [29.9 - 47.3]	43.7 [33.7 - 53.8]	-	-	-	-	-	15/06/2014	Paper | Wiki | Pres.
Ledig-VOL	24.5	47.7 [42.1 - 52.8]	66.7 [57.1 - 74.1]	36.9 [28.9 - 45.9]	36.9 [28.6 - 47.2]	14.0	68.4 [64.5 - 72.5]	75.7 [70.3 - 81.0]	50.1 [44.1 - 56.4]	79.0 [73.3 - 83.5]	15/06/2014	Paper | Wiki | Pres.
Amoroso	26.0	46.9 [41.5 - 52.3]	67.4 [58.5 - 75.2]	42.6 [33.6 - 51.1]	26.2 [18.3 - 35.4]	15.0	67.2 [63.3 - 71.3]	73.4 [67.8 - 78.7]	56.0 [49.7 - 61.9]	72.3 [66.2 - 77.5]	15/06/2014	Paper | Wiki | Pres.
Tangaro	27.0	46.6 [41.0 - 51.4]	68.2 [60.2 - 76.5]	37.7 [29.2 - 46.3]	30.1 [21.7 - 39.0]	16.0	67.1 [63.2 - 71.0]	73.1 [67.8 - 78.0]	52.6 [45.9 - 58.6]	75.8 [70.2 - 80.6]	15/06/2014	Paper | Wiki | Pres.
Cardenas-Pena	28.0	39.0 [33.9 - 43.8]	50.4 [41.5 - 59.1]	28.7 [21.6 - 38.5]	36.9 [27.4 - 46.8]	18.0	55.9 [51.2 - 59.9]	57.8 [51.6 - 63.4]	50.0 [43.9 - 57.1]	59.8 [53.5 - 65.7]	15/06/2014	Paper | Wiki | Pres.
Smith	29.0	32.2 [27.4 - 36.7]	48.1 [39.6 - 57.1]	20.5 [13.9 - 28.3]	26.2 [18.3 - 35.0]	19.0	50.4 [46.7 - 54.6]	54.1 [48.0 - 60.0]	50.6 [45.0 - 57.1]	46.6 [40.0 - 53.6]	15/06/2014	Paper | Wiki | Pres.

Ranking:
- 11/29 multi-nominal accuracy
- 8/29 for AD classification accuracy

Feature Selection

- Feature selection can reduce dimensionality, thus mitigating computational performance issues and improving the classification accuracy.

1. Advanced Feature Selection composed by five stages (5S-Combo)
    - Combining several methods and performing binary and multinominal classification: Correlation filter, Random Forest filter, SVM wrapper

2. Genetic Algorithms (GA)
    - Exploiting the ability of GA to explore large search spaces efficiently

Advanced Feature Selection

- Binary classifiers: HCvsMCI, HCvsAD, ADvsMCI
- Multinominal OVO classification

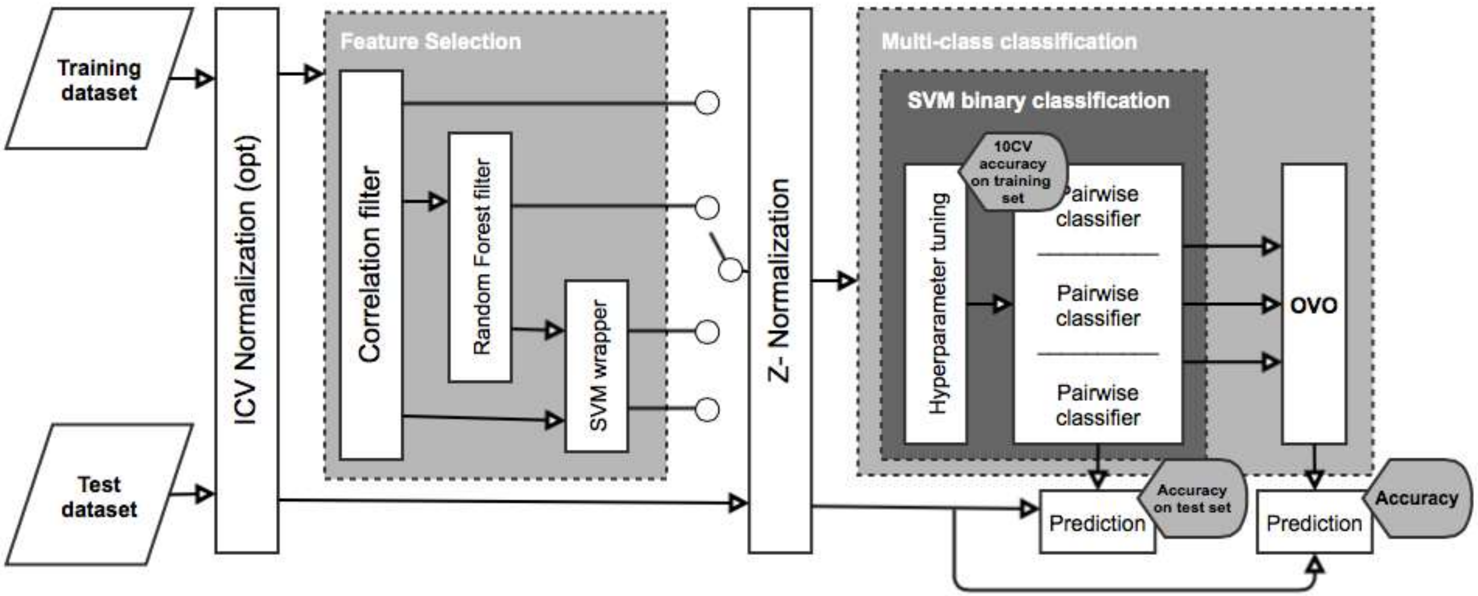


Table 2. Results of the binary classification: training performed on D2 and accuracy estimated on D2 (D2/D2) and D1 (D2/D1). Best results on D1 are indicated in bold.

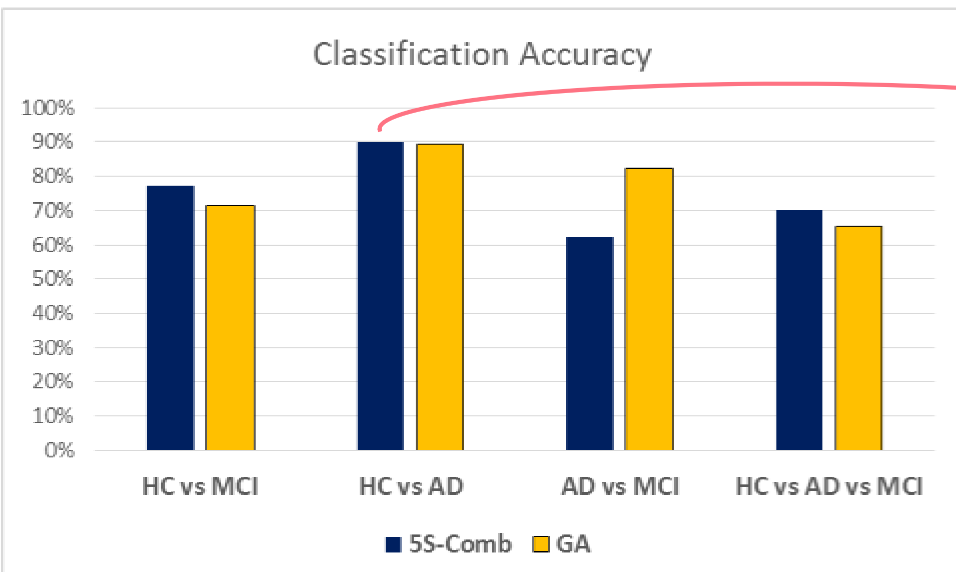
	Feature selection method																							
	Corr. filter $\lvert r\rvert<0.90$						*Random Forest filter*						*SVM wrapper*						*Random Forest filter + SVM wrapper*					
	ICV			*No ICV*			*ICV*			*No ICV*			*ICV*			*No ICV*			*ICV*			*No ICV*		
	Nr. feat.	D2/D2 10xval	D2/D1 holdout	Nr. feat.	D2/D2 10xval	D2/D1 holdout	Nr. feat.	D2/D2 10xval	D2/D1 holdout	Nr. feat.	D2/D2 10xval	D2/D1 holdout	Nr. feat.	D2/D2 10xval	D2/D1 holdout	Nr. feat.	D2/D2 10xval	D2/D1 holdout	Nr. feat.	D2/D2 10xval	D2/D1 holdout	Nr. feat.	D2/D2 10xval	D2/D1 holdout
HCvsMCI	133	71.4%	42.9%	140	67.9%	**66.7%**	71	71.4%	42.9%	41	65.7%	52.4%	133	71.4%	42.9%	140	67.9%	66.7%	30	77.1%	42.9%	41	65.7%	52.4%
HCvsAD	133	90%	42.9%	140	81.4%	95.2%	109	86.4%	42.9%	95	82.1%	**95.2%**	133	90%	42.9%	140	81.4%	95.2%	109	86.4%	42.9%	95	82.1%	95.2%
ADvsMCI	133	58.6%	50%	139	58.6%	83.3%	47	60%	50%	44	62.1%	**88.9%**	133	58.6%	50%	139	58.6%	83.3%	47	60%	50%	44	62.1%	88.9%

Feature Selection Results

University of Reading

1. Advanced Feature Selection composed by five stages (5S-Combo),
2. Genetic Algorithms (GA)

- Both tested with
  - Multinominal classification task
  - SVM with radial basis filter kernel
  - ADNI data
  - 10-fold cross validation



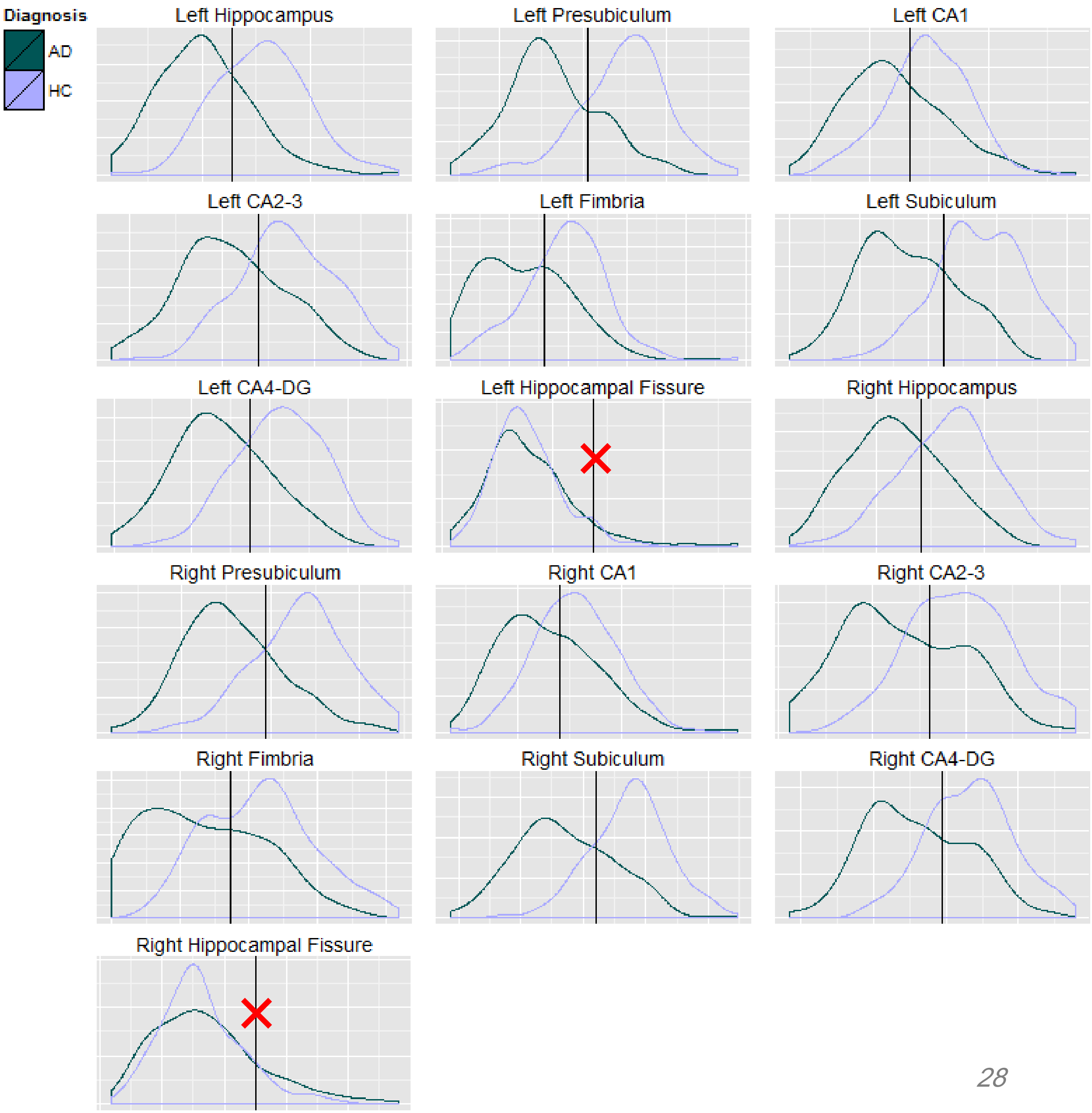
SVM used 133 features

LDA and Probability-based Classifiers

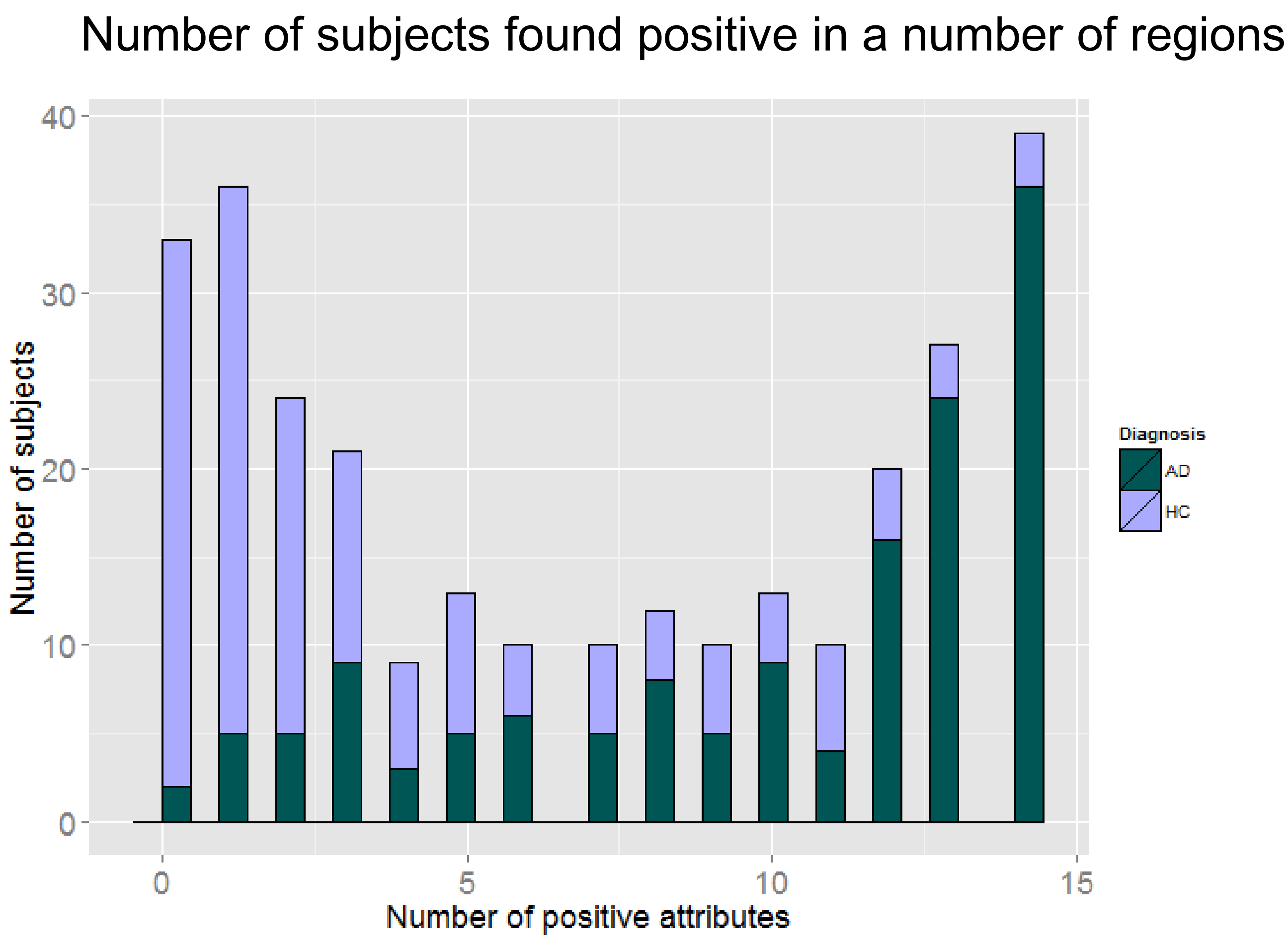
- PDF of volumes of 16 hippocampal subfields across subjects: HC vs AD

- X-axis: volume measure v
- Y-axis: P(AD | v) and P(HC | v)

- Thresholds determined by Linear Discriminant Analysis (LDA)

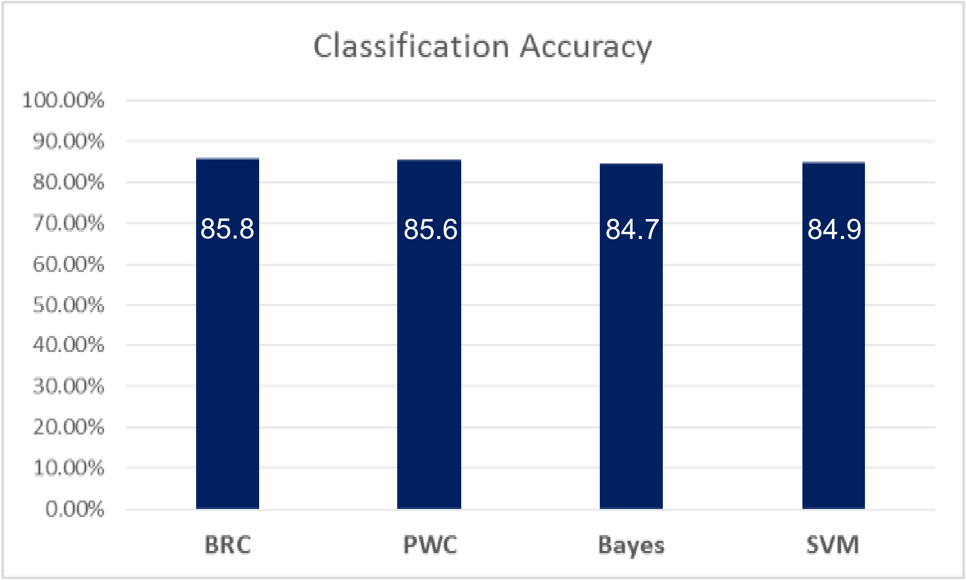


LDA and Probability-based Classifiers



Classification Methods

- Classification methods:
  1. Binary Region Classifier (BRC)
  2. Probability Weight-based Classification (PWC)
  3. Bayesian Classifier (Bayes)
  4. Support Vector Machine (SVM)
- 10-fold cross validation (avg accuracy over ten 10-f xval trials)



Conclusions

- Some predictive models also provide a descriptive representations which is a desirable plus.
  - Black-boxes approaches may get a good/better accuracy but they do not help understanding, especially in high-dimensional spaces.
- With same pre-processing, feature selection and normalisation most classification methods are equivalent in terms of accuracy.
- The hippocampus subfields are good predictors as already known in the literature.
- Isolating MCI cases is a difficult task
  - MCI is primarily a clinical diagnosis
- Automatic data manipulation helped to save time and to focus on the data mining challenges.

Current and Future Work

- An extensive study of normalisation techniques
- R-Surfer, an R package equivalent to K-Surfer
- Integration of supervised and unsupervised techniques
  - Data-driven identification of subject subgroups may help
    - to generate more accurate subgroup-specific predictive models
    - to distinguish among distinctive progressions (longitudinal studies) of regional atrophy to help selecting better treatments
    - ultimately to shed light on the causes

Collaborations and Publications

Many thanks to the colleagues involved in these activities:

- Prof. Doug Saddy, Director of Centre for Integrative Neuroscience and Neurodynamics, University of Reading, UK
- Prof. Mario Cannataro, Department of Medical and Surgical Sciences, School of Informatics and Biomedical Engineering, University 'Magna Graecia' of Catanzaro, Italy
- Dr. Alessia Sarica, Istituto di Bioimmagini e Fisiologia Molecolare (IBFM), National Research Council (CNR), Catanzaro, Italy
- Alexander Spedding, PhD student, Department of Computer Science, University of Reading, UK




Publications:

- A. Spedding, G. Di Fatta and M. Cannataro, **"A Genetic Algorithm for the Selection of Structural MRI Features for Classification of Mild Cognitive Impairment and Alzheimer's Disease"**, Workshop on Machine learning in decision making for biomedical applications, IEEE Int.l Conference on Bioinformatics and Biomedicine (BIBM), Nov. 9-12, 2015, Washington D.C., pp. 1566-1571.
- A. Spedding, G. Di Fatta and J.D. Saddy, **"An LDA and Probability-based Classifier for the Diagnosis of Alzheimer's Disease from Structural MRI"**, Workshop on High Performance Bioinformatics and Biomedicine (HiBB), the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Nov. 9-12, 2015, Washington D.C., pp. 1404 - 1411.
- E.E. Bron et al., **"Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge"**, NeuroImage, an Elsevier Journal of Brain Function, 1 May 2015, 111:562-79.
- A. Sarica, G. Di Fatta, G. Smith, M. Cannataro, D. Saddy, **"Advanced Feature Selection in Multinominal Dementia Classification from Structural MRI Data"**, in the Proc. of the Workshop on Computer-Aided Diagnosis of Dementia based on structural MRI data (CADDementia), Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI), Sept. 18, 2014, Boston, MA, USA
- A. Sarica, G. Di Fatta, M. Cannataro, **"K-Surfer: A KNIME extension for the management and analysis of human brain MRI FreeSurfer/FSL Data"**, Int.l Conf. on Brain Informatics and Health, 11-14 August 2014, Warsaw, Poland, Springer LNCS V.8609, 2014, pp. 481-492.
- M. Berthold, N. Cebron, F. Dill, G. Di Fatta, T. Gabriel, F. Georg, T. Meinl, P. Ohl, C. Sieb, B. Wiswedel, **"KNIME: the Konstanz Information Miner"**, Proc. 4th Annual Industrial Simulation Conference (ISC), Palermo, Italy, June 5-7, 2006, pp.58-61.