



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/>



**University of
Reading**

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

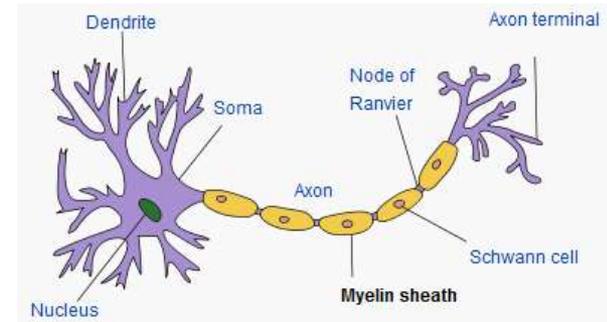
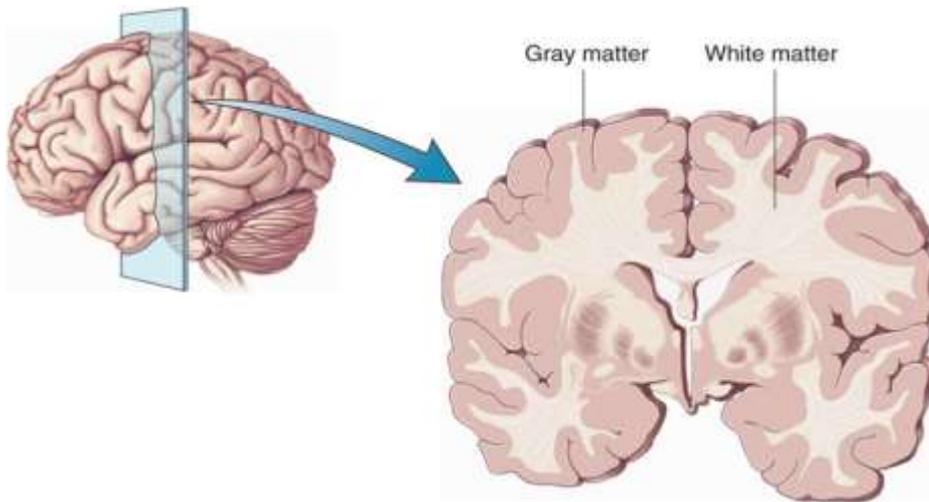
Alzheimer's disease is the most common type of dementia, affecting almost 500,000 people in the UK.

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

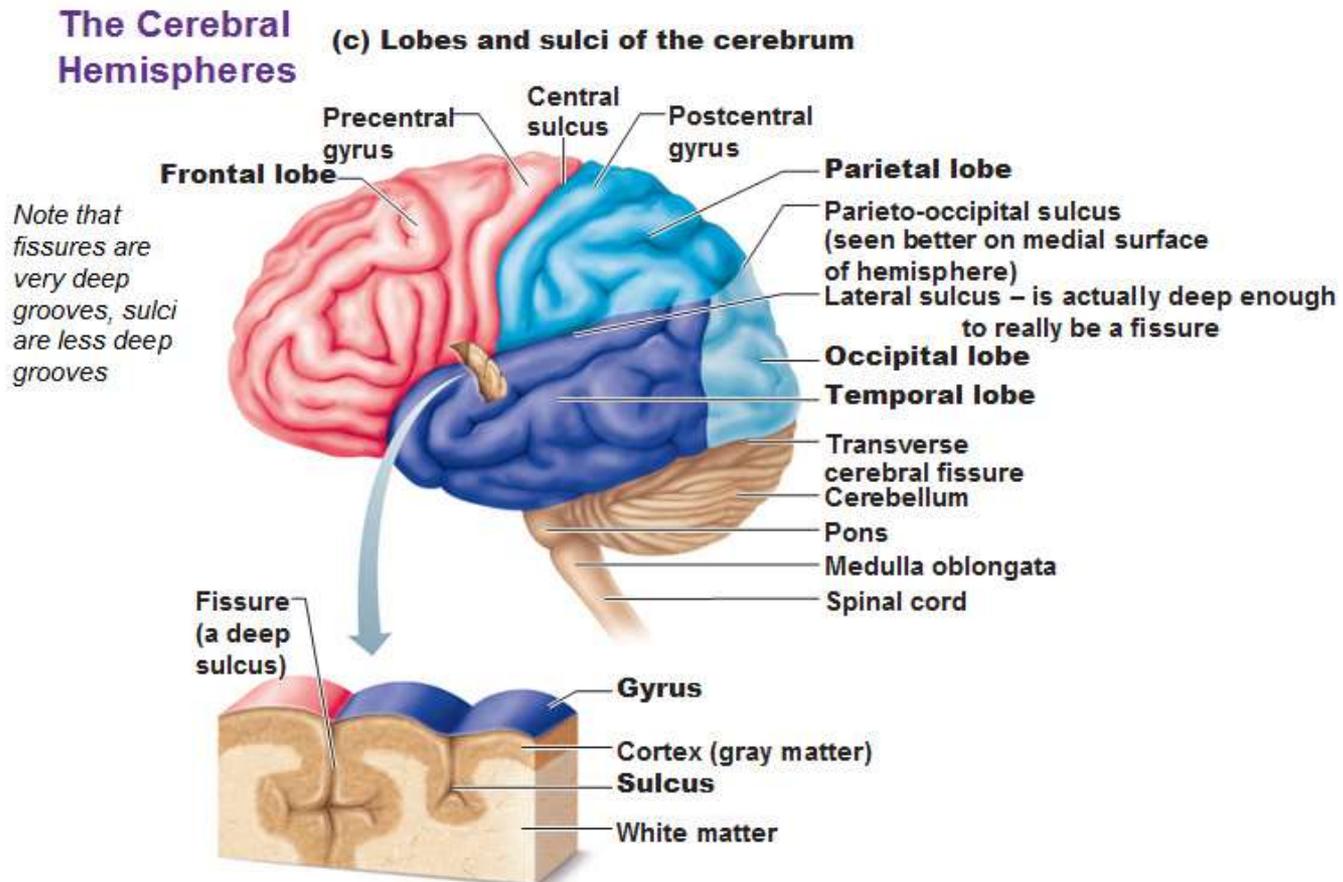
- The **cerebral cortex (grey matter)** is the outer layer of neural tissue
 - Grey folded tissue that consists mainly of cell bodies.
 - It is primarily associated with processing and cognition.
 - It has a stratified structure.



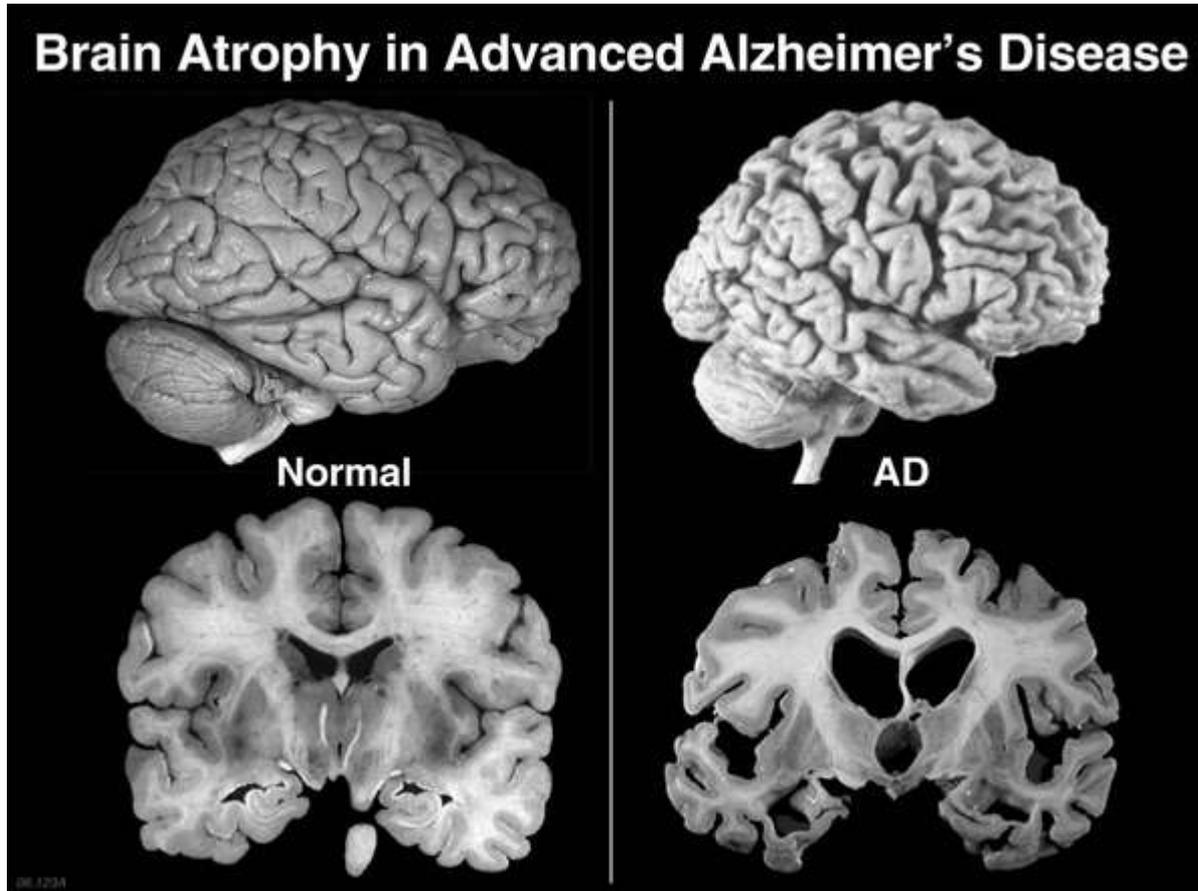
- The **subcortical white matter**
 - White matter that consists mainly of long-range neuronal axons.
 - It forms the deep parts of the brain and the superficial parts of the spinal cord.
 - It modulates the distribution of action potentials, i.e. communication between different brain regions.
 - Myelin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

The Human Brain

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

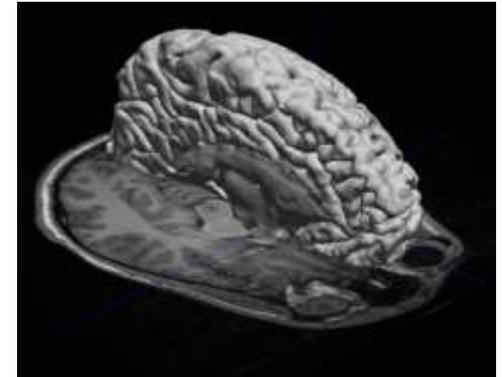
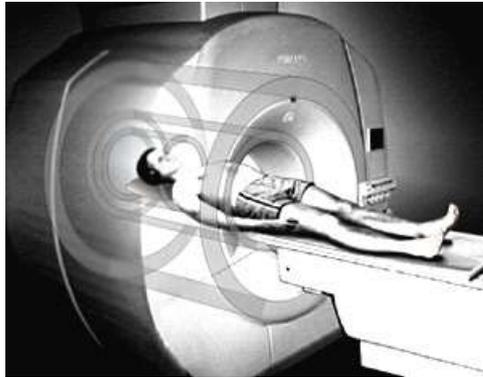


Alzheimer's Disease (AD)



Magnetic Resonance Imaging (MRI)

- 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...



Google™ Custom Search

SEARCH

[Access Data](#) | [Upload Data](#) | [Contact Us](#)

ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE

[ABOUT](#)

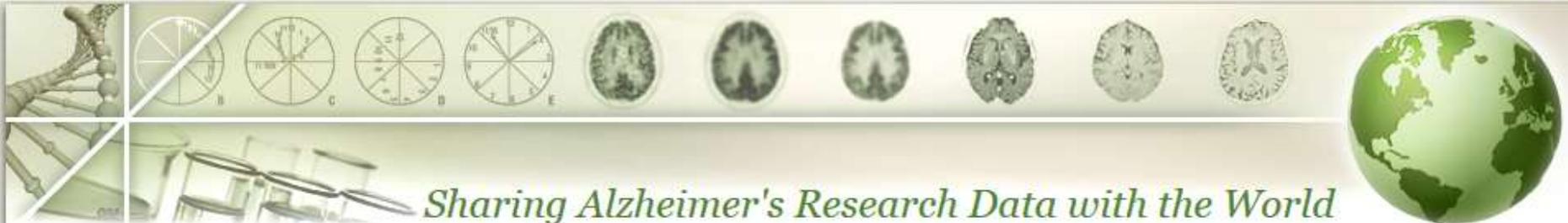
[STUDY DESIGN](#)

[DATA & SAMPLES](#)

[METHODS & TOOLS](#)

[SUPPORT](#)

[NEWS & PUBLICATIONS](#)



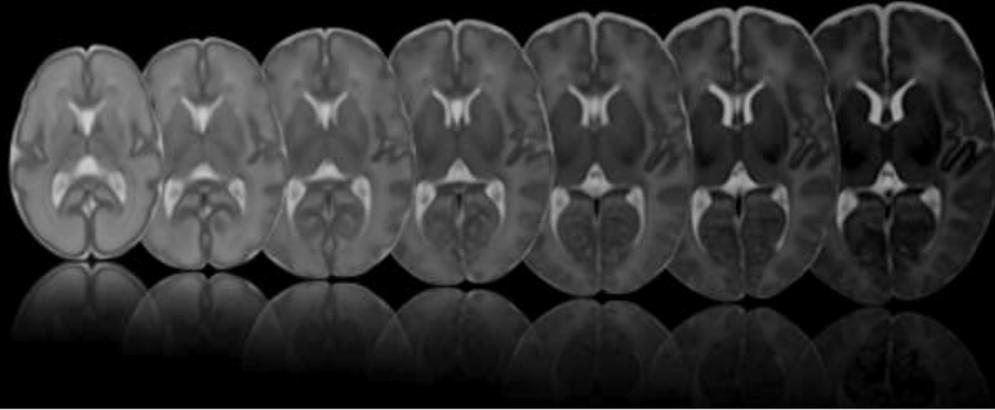
Sharing Alzheimer's Research Data with the World

The Alzheimer's Disease Neuroimaging Initiative (ADNI) unites researchers with study data as they work to define the progression of Alzheimer's disease. ADNI researchers collect, validate and utilize data such as MRI and PET images, genetics, cognitive tests, CSF and blood biomarkers as predictors for the disease. Data from the North American ADNI's study participants, including Alzheimer's disease patients, mild cognitive impairment subjects and elderly controls, are available from this site.



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

Information eXtraction from Images

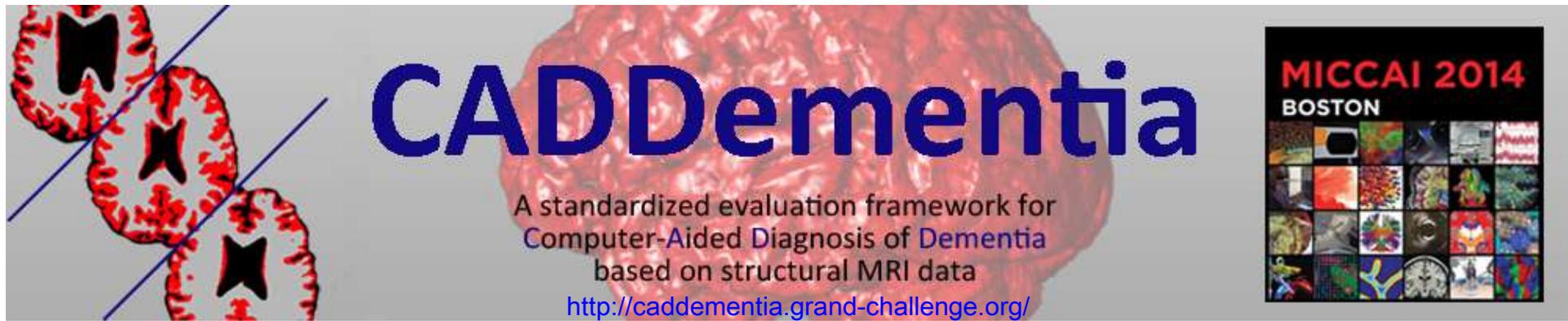


Welcome to brain-
development.org!

This website hosts resources for the computational analysis
of brain development.

<http://brain-development.org/>

- 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

A standardized evaluation framework for
Computer-Aided Diagnosis of Dementia
based on structural MRI data

<http://caddementia.grand-challenge.org/>



- 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

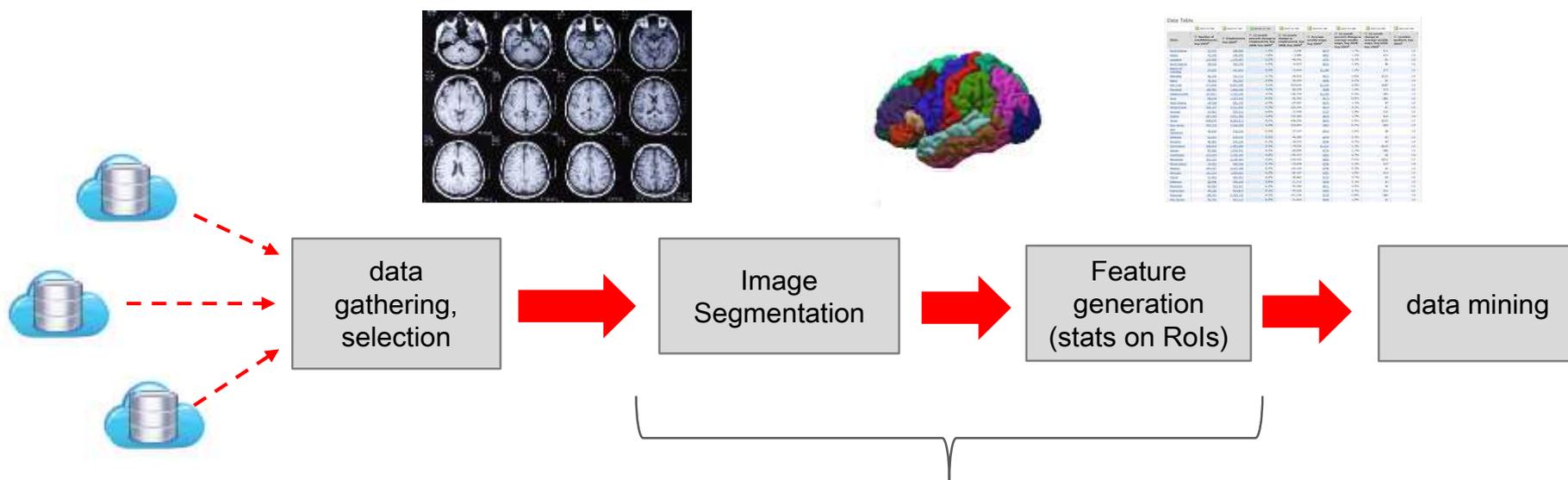
Predictive Task

- 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.

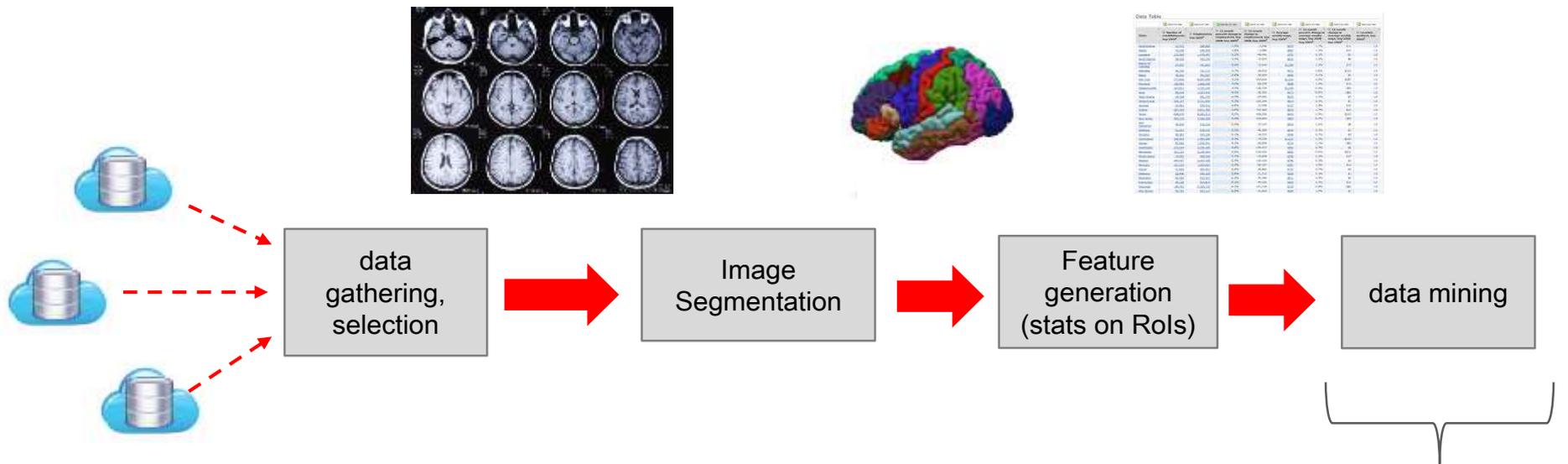
Data Workflow



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-processing of a single image takes ~8h → ~168 days in total.
- Data parallel approach used to speed up computation → ~10 days.

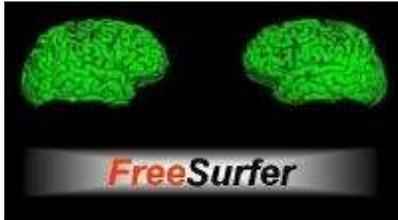
Data Workflow



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



- **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.

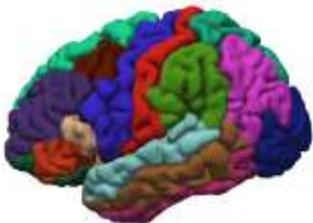


- The **FMRIB Software Library (FSL)** provides image analysis and statistical tools for fMRI, MRI and DTI brain imaging data. (<http://fsl.fmrib.ox.ac.uk/>)

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



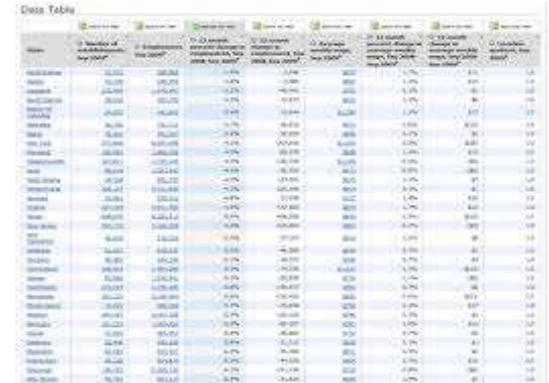
volume data

area data

thickness data

diffusion data

demographical/clinical data



Area	Volume (mm³)	Area (mm²)	Thickness (mm)	Diffusion (mm²)	Demographical/Clinical
ACC	10,000	100,000	1.0	0.001	0.0
AMYG	15,000	150,000	1.5	0.002	0.0
ANG	20,000	200,000	2.0	0.003	0.0
ANT	25,000	250,000	2.5	0.004	0.0
BAUD	30,000	300,000	3.0	0.005	0.0
BAUN	35,000	350,000	3.5	0.006	0.0
BAUN	40,000	400,000	4.0	0.007	0.0
BAUN	45,000	450,000	4.5	0.008	0.0
BAUN	50,000	500,000	5.0	0.009	0.0
BAUN	55,000	550,000	5.5	0.010	0.0
BAUN	60,000	600,000	6.0	0.011	0.0
BAUN	65,000	650,000	6.5	0.012	0.0
BAUN	70,000	700,000	7.0	0.013	0.0
BAUN	75,000	750,000	7.5	0.014	0.0
BAUN	80,000	800,000	8.0	0.015	0.0
BAUN	85,000	850,000	8.5	0.016	0.0
BAUN	90,000	900,000	9.0	0.017	0.0
BAUN	95,000	950,000	9.5	0.018	0.0
BAUN	100,000	1,000,000	10.0	0.019	0.0
BAUN	105,000	1,050,000	10.5	0.020	0.0
BAUN	110,000	1,100,000	11.0	0.021	0.0
BAUN	115,000	1,150,000	11.5	0.022	0.0
BAUN	120,000	1,200,000	12.0	0.023	0.0
BAUN	125,000	1,250,000	12.5	0.024	0.0
BAUN	130,000	1,300,000	13.0	0.025	0.0
BAUN	135,000	1,350,000	13.5	0.026	0.0
BAUN	140,000	1,400,000	14.0	0.027	0.0
BAUN	145,000	1,450,000	14.5	0.028	0.0
BAUN	150,000	1,500,000	15.0	0.029	0.0
BAUN	155,000	1,550,000	15.5	0.030	0.0
BAUN	160,000	1,600,000	16.0	0.031	0.0
BAUN	165,000	1,650,000	16.5	0.032	0.0
BAUN	170,000	1,700,000	17.0	0.033	0.0
BAUN	175,000	1,750,000	17.5	0.034	0.0
BAUN	180,000	1,800,000	18.0	0.035	0.0
BAUN	185,000	1,850,000	18.5	0.036	0.0
BAUN	190,000	1,900,000	19.0	0.037	0.0
BAUN	195,000	1,950,000	19.5	0.038	0.0
BAUN	200,000	2,000,000	20.0	0.039	0.0
BAUN	205,000	2,050,000	20.5	0.040	0.0
BAUN	210,000	2,100,000	21.0	0.041	0.0
BAUN	215,000	2,150,000	21.5	0.042	0.0
BAUN	220,000	2,200,000	22.0	0.043	0.0
BAUN	225,000	2,250,000	22.5	0.044	0.0
BAUN	230,000	2,300,000	23.0	0.045	0.0
BAUN	235,000	2,350,000	23.5	0.046	0.0
BAUN	240,000	2,400,000	24.0	0.047	0.0
BAUN	245,000	2,450,000	24.5	0.048	0.0
BAUN	250,000	2,500,000	25.0	0.049	0.0
BAUN	255,000	2,550,000	25.5	0.050	0.0
BAUN	260,000	2,600,000	26.0	0.051	0.0
BAUN	265,000	2,650,000	26.5	0.052	0.0
BAUN	270,000	2,700,000	27.0	0.053	0.0
BAUN	275,000	2,750,000	27.5	0.054	0.0
BAUN	280,000	2,800,000	28.0	0.055	0.0
BAUN	285,000	2,850,000	28.5	0.056	0.0
BAUN	290,000	2,900,000	29.0	0.057	0.0
BAUN	295,000	2,950,000	29.5	0.058	0.0
BAUN	300,000	3,000,000	30.0	0.059	0.0
BAUN	305,000	3,050,000	30.5	0.060	0.0
BAUN	310,000	3,100,000	31.0	0.061	0.0
BAUN	315,000	3,150,000	31.5	0.062	0.0
BAUN	320,000	3,200,000	32.0	0.063	0.0
BAUN	325,000	3,250,000	32.5	0.064	0.0
BAUN	330,000	3,300,000	33.0	0.065	0.0
BAUN	335,000	3,350,000	33.5	0.066	0.0
BAUN	340,000	3,400,000	34.0	0.067	0.0
BAUN	345,000	3,450,000	34.5	0.068	0.0
BAUN	350,000	3,500,000	35.0	0.069	0.0
BAUN	355,000	3,550,000	35.5	0.070	0.0
BAUN	360,000	3,600,000	36.0	0.071	0.0
BAUN	365,000	3,650,000	36.5	0.072	0.0
BAUN	370,000	3,700,000	37.0	0.073	0.0
BAUN	375,000	3,750,000	37.5	0.074	0.0
BAUN	380,000	3,800,000	38.0	0.075	0.0
BAUN	385,000	3,850,000	38.5	0.076	0.0
BAUN	390,000	3,900,000	39.0	0.077	0.0
BAUN	395,000	3,950,000	39.5	0.078	0.0
BAUN	400,000	4,000,000	40.0	0.079	0.0
BAUN	405,000	4,050,000	40.5	0.080	0.0
BAUN	410,000	4,100,000	41.0	0.081	0.0
BAUN	415,000	4,150,000	41.5	0.082	0.0
BAUN	420,000	4,200,000	42.0	0.083	0.0
BAUN	425,000	4,250,000	42.5	0.084	0.0
BAUN	430,000	4,300,000	43.0	0.085	0.0
BAUN	435,000	4,350,000	43.5	0.086	0.0
BAUN	440,000	4,400,000	44.0	0.087	0.0
BAUN	445,000	4,450,000	44.5	0.088	0.0
BAUN	450,000	4,500,000	45.0	0.089	0.0
BAUN	455,000	4,550,000	45.5	0.090	0.0
BAUN	460,000	4,600,000	46.0	0.091	0.0
BAUN	465,000	4,650,000	46.5	0.092	0.0
BAUN	470,000	4,700,000	47.0	0.093	0.0
BAUN	475,000	4,750,000	47.5	0.094	0.0
BAUN	480,000	4,800,000	48.0	0.095	0.0
BAUN	485,000	4,850,000	48.5	0.096	0.0
BAUN	490,000	4,900,000	49.0	0.097	0.0
BAUN	495,000	4,950,000	49.5	0.098	0.0
BAUN	500,000	5,000,000	50.0	0.099	0.0
BAUN	505,000	5,050,000	50.5	0.100	0.0
BAUN	510,000	5,100,000	51.0	0.101	0.0
BAUN	515,000	5,150,000	51.5	0.102	0.0
BAUN	520,000	5,200,000	52.0	0.103	0.0
BAUN	525,000	5,250,000	52.5	0.104	0.0
BAUN	530,000	5,300,000	53.0	0.105	0.0
BAUN	535,000	5,350,000	53.5	0.106	0.0
BAUN	540,000	5,400,000	54.0	0.107	0.0
BAUN	545,000	5,450,000	54.5	0.108	0.0
BAUN	550,000	5,500,000	55.0	0.109	0.0
BAUN	555,000	5,550,000	55.5	0.110	0.0
BAUN	560,000	5,600,000	56.0	0.111	0.0
BAUN	565,000	5,650,000	56.5	0.112	0.0
BAUN	570,000	5,700,000	57.0	0.113	0.0
BAUN	575,000	5,750,000	57.5	0.114	0.0
BAUN	580,000	5,800,000	58.0	0.115	0.0
BAUN	585,000	5,850,000	58.5	0.116	0.0
BAUN	590,000	5,900,000	59.0	0.117	0.0
BAUN	595,000	5,950,000	59.5	0.118	0.0
BAUN	600,000	6,000,000	60.0	0.119	0.0
BAUN	605,000	6,050,000	60.5	0.120	0.0
BAUN	610,000	6,100,000	61.0	0.121	0.0
BAUN	615,000	6,150,000	61.5	0.122	0.0
BAUN	620,000	6,200,000	62.0	0.123	0.0
BAUN	625,000	6,250,000	62.5	0.124	0.0
BAUN	630,000	6,300,000	63.0	0.125	0.0
BAUN	635,000	6,350,000	63.5	0.126	0.0
BAUN	640,000	6,400,000	64.0	0.127	0.0
BAUN	645,000	6,450,000	64.5	0.128	0.0
BAUN	650,000	6,500,000	65.0	0.129	0.0
BAUN	655,000	6,550,000	65.5	0.130	0.0
BAUN	660,000	6,600,000	66.0	0.131	0.0
BAUN	665,000	6,650,000	66.5	0.132	0.0
BAUN	670,000	6,700,000	67.0	0.133	0.0
BAUN	675,000	6,750,000	67.5	0.134	0.0
BAUN	680,000	6,800,000	68.0	0.135	0.0
BAUN	685,000	6,850,000	68.5	0.136	0.0
BAUN	690,000	6,900,000	69.0	0.137	0.0
BAUN	695,000	6,950,000	69.5	0.138	0.0
BAUN	700,000	7,000,000	70.0	0.139	0.0
BAUN	705,000	7,050,000	70.5	0.140	0.0
BAUN	710,000	7,100,000	71.0	0.141	0.0
BAUN	715,000	7,150,000	71.5	0.142	0.0
BAUN	720,000	7,200,000	72.0	0.143	0.0
BAUN	725,000	7,250,000	72.5	0.144	0.0
BAUN	730,000	7,300,000	73.0	0.145	0.0
BAUN	735,000	7,350,000	73.5	0.146	0.0
BAUN	740,000	7,400,000	74.0	0.147	0.0
BAUN	745,000	7,450,000	74.5	0.148	0.0
BAUN	750,000	7,500,000	75.0	0.149	0.0
BAUN	755,000	7,550,000	75.5	0.150	0.0
BAUN	760,000	7,600,000	76.0	0.151	0.0
BAUN	765,000	7,650,000	76.5	0.152	0.0
BAUN	770,000	7,700,000	77.0	0.153	0.0
BAUN	775,000	7,750,000	77.5	0.154	0.0
BAUN	780,000	7,800,000	78.0	0.155	0.0
BAUN	785,000	7,850,000	78.5	0.156	0.0
BAUN	790,000	7,900,000	79.0	0.157	0.0
BAUN	795,000	7,950,000	79.5	0.158	0.0
BAUN	800,000	8,000,000	80.0	0.159	0.0
BAUN	805,000	8,050,000	80.5	0.160	0.0
BAUN	810,000	8,100,000	81.0	0.161	0.0
BAUN	815,000	8,150,000	81.5	0.162	0.0
BAUN	820,000	8,200,000	82.0	0.163	0.0
BAUN	825,000	8,250,000	82.5	0.164	0.0
BAUN	830,000	8,300,000	83.0	0.165	0.0
BAUN	835,000	8,350,000	83.5	0.166	0.0
BAUN	840,000	8,400,000	84.0	0.167	0.0
BAUN	845,000	8,450,000	84.5	0.168	0.0
BAUN	850,000	8,500,000	85.0	0.169	0.0
BAUN	855,000	8,550,000	85.5	0.170	0.0
BAUN	860,000	8,600,000	86.0	0.171	0.0
BAUN	865,000	8,650,000	86.5	0.172	0.0
BAUN	870,000	8,700,000	87.0	0.173	0.0
BAUN	875,000	8,750,000	87.5	0.174	0.0
BAUN	880,000	8,800,000	88.0	0.175	0.0
BAUN	885,000	8,850,000	88.5	0.176	0.0
BAUN	890,000	8,900,000	89.0	0.177	0.0
BAUN	895,000	8,950,000	89.5	0.178	0.0
BAUN	900,000	9,000,000	90.0	0.179	0.0
BAUN	905,000	9,050,000	90.5	0.180	0.0
BAUN	910,000	9,100,000	91.0	0.181</	

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



K-Surfer

<http://sourceforge.net/projects/ksurfer/>



PRODUCTS / RESOURCES / PARTNERS / COMMUNITY / USE CASES / ABOUT

Open for Innovation

Navigate complex data with the agility and freedom that only an open platform can bring

[Learn More](#)

<https://www.knime.org>

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>)

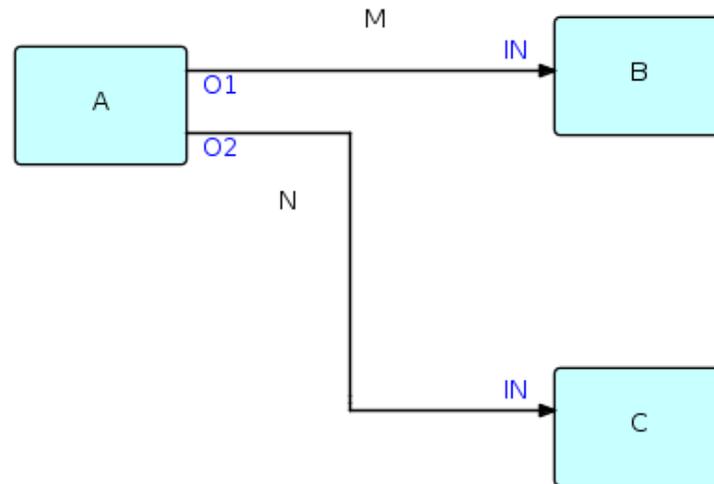


- A platform for data analysis, data manipulation, data visualization and reporting
 - Originally developed at the Department of Computer and Information Science, University of Konstanz, Germany
 - Under continuous evolution and extension
 - 1st release in April 2006
 - ver.3 released in Dec. 2015
- based on Eclipse, free and open source
- flexibility, openness and ease of integration with other tools

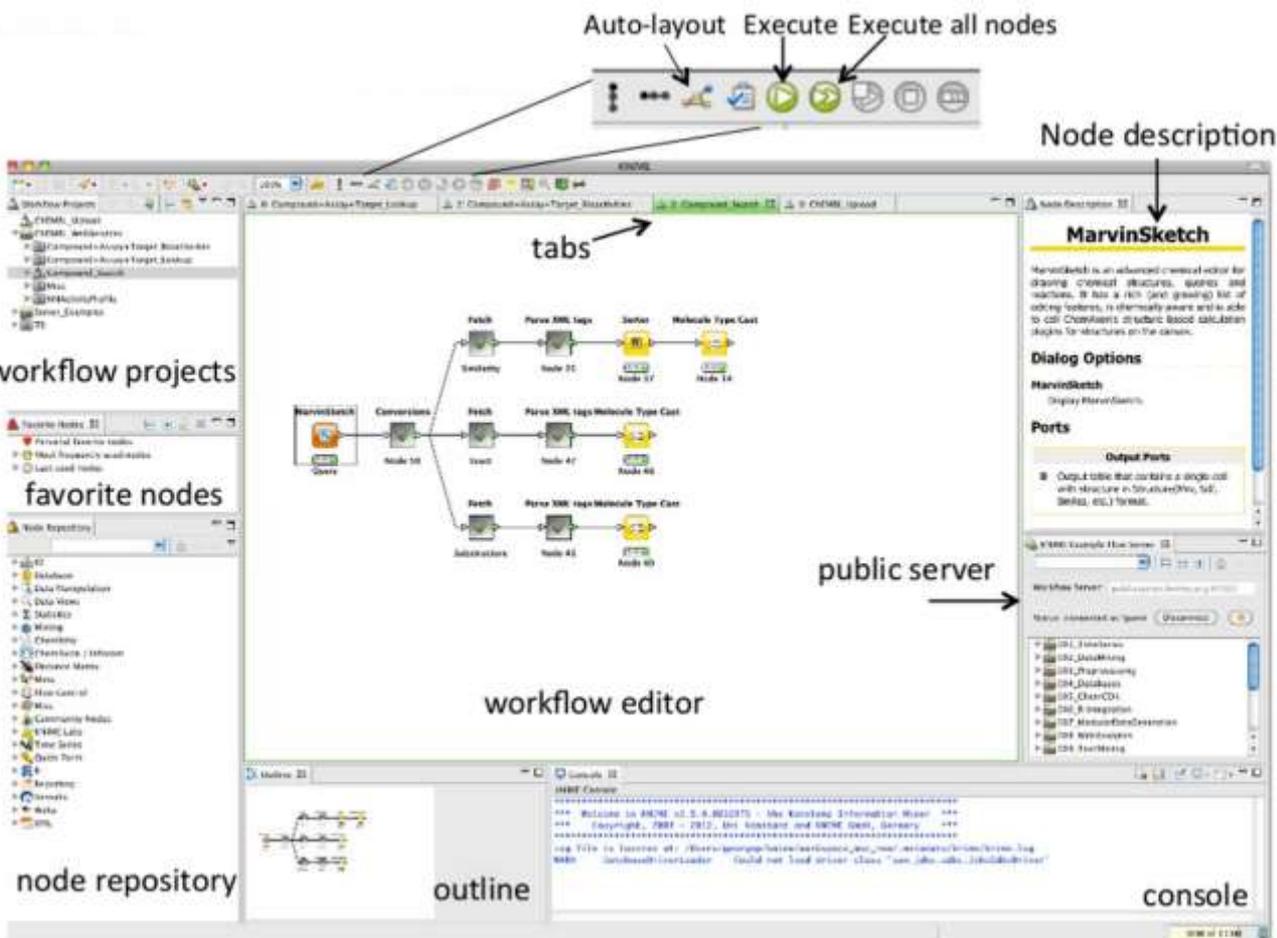


Flow-based Programming

- 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



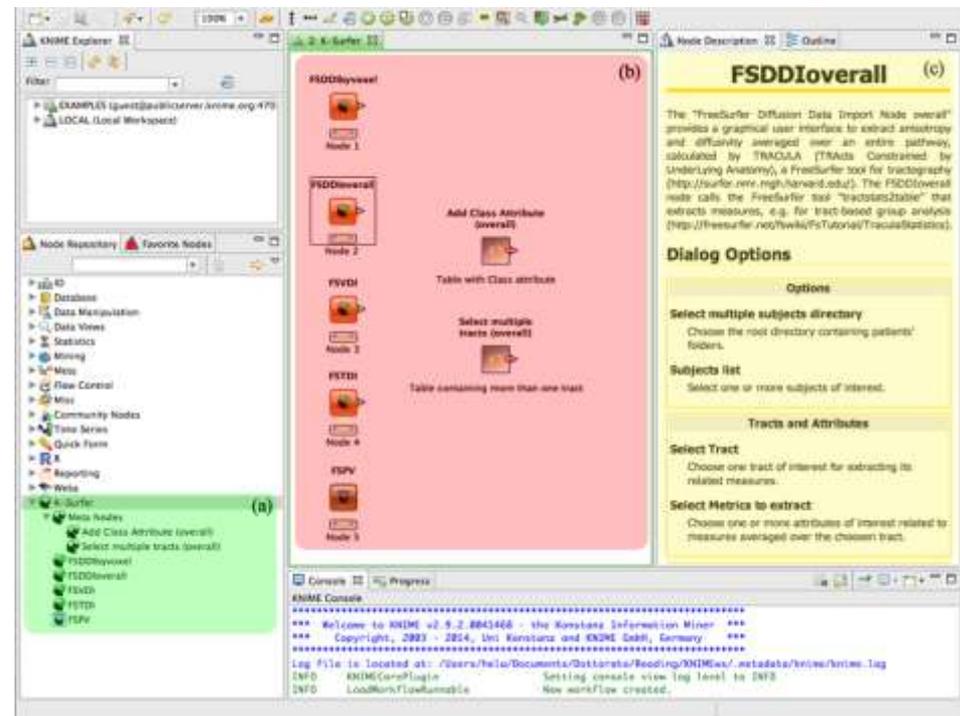
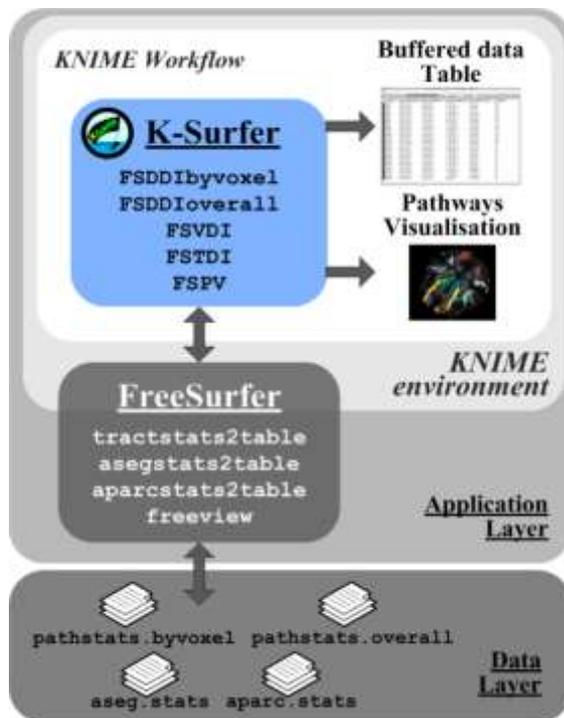
The image shows the KNIME software interface with several components labeled:

- Auto-layout**: Points to the 'Auto' icon in the top toolbar.
- Execute**: Points to the 'Execute' icon in the top toolbar.
- Execute all nodes**: Points to the 'Execute All' icon in the top toolbar.
- Node description**: Points to the 'Node Description' panel on the right, which is currently showing the 'MarvinSketch' node.
- tabs**: Points to the 'Tabs' icon in the top toolbar.
- workflow projects**: Points to the 'Workflow Projects' pane on the left.
- favorite nodes**: Points to the 'Favorite Nodes' pane on the left.
- node repository**: Points to the 'Node Repository' pane on the left.
- workflow editor**: Points to the central workspace containing a workflow diagram with nodes like 'Parse SMILES', 'MolVec', and 'MolVec Type Cast'.
- public server**: Points to the 'Public Server' section in the 'Node Description' panel.
- outline**: Points to the 'Outline' pane at the bottom left.
- console**: Points to the 'Console' pane at the bottom right, which displays execution logs.



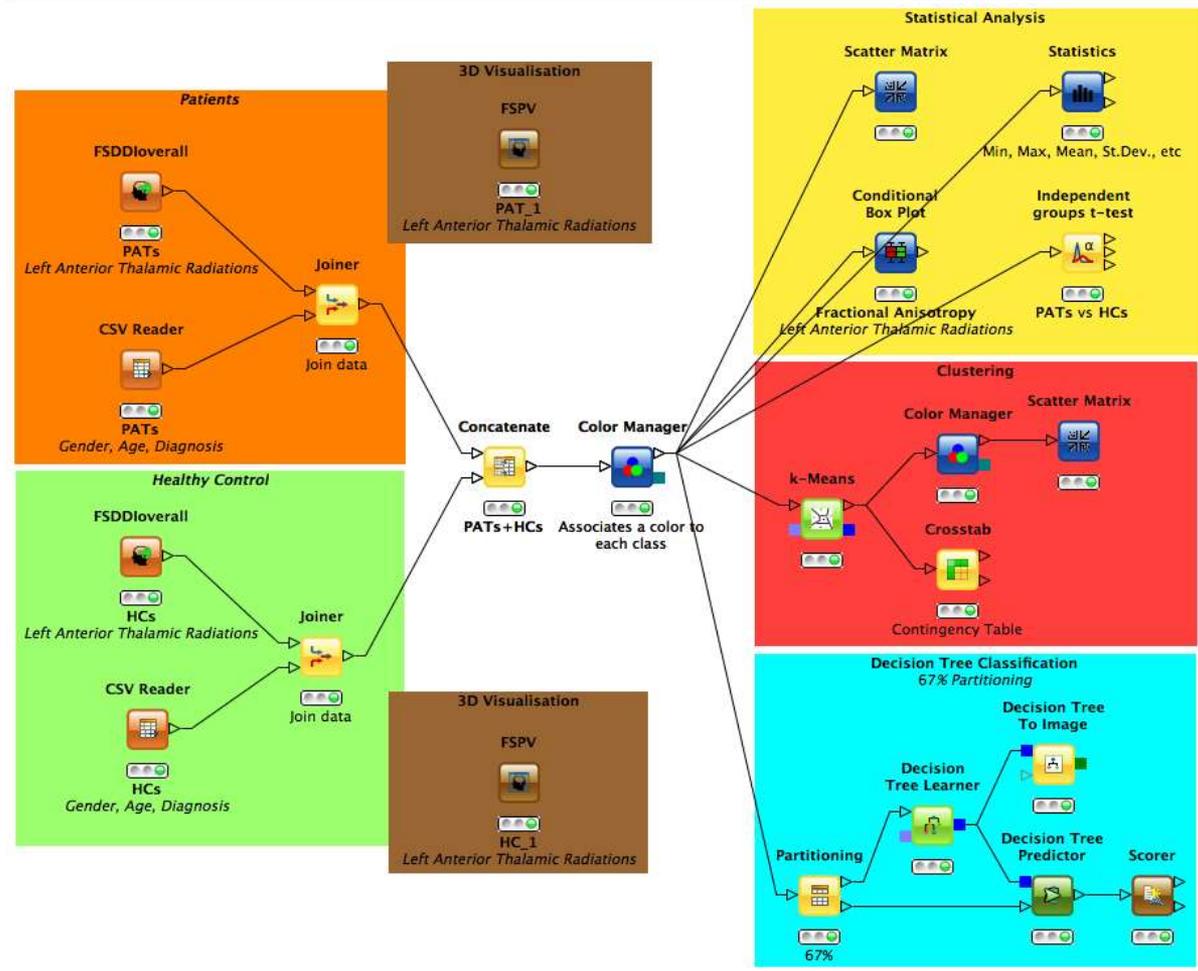
K-Surfer

- 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

Hypothesis H0: The means of diffusion indices in Left Anterior Thalamic Radiations are equal between PAT and HC subjects.



CADDementia Challenge 2014

- 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 _{Ch} [CI] (%)	TPF _{MCI} [CI] (%)	TPF _{AD} [CI] (%)	Rank AUC	AUC _{All} [CI] (%)	AUC _{Ch} [CI] (%)	AUC _{MCI} [CI] (%)	AUC _{AD} [CI] (%)	Date	Documentation
Soerensen-equal	1.0	63.0 [57.9 - 67.5]	66.9 [62.9 - 69.2]	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-evalNorm	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.
Franka	8.0	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 [48.9 - 59.0]	88.2 [80.2 - 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.
Abdulkadr	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]	38.3 [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-ADN1	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-FACEADN1	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]	85.8 [80.6 - 91.3]	15/06/2014	Paper Wiki Pres.
Router-admi	21.0	49.2 [43.5 - 54.2]	94.6 [89.8 - 97.7]	11.5 [6.2 - 17.7]	30.9 [27.4 - 46.5]	-	-	-	-	-	15/06/2014	Paper Wiki Pres.
Eskildsen-FACEADN2	22.5	48.3 [43.2 - 53.4]	48.6 [40.5 - 57.4]	42.6 [33.9 - 51.3]	54.4 [45.5 - 64.0]	-	-	-	-	-	15/06/2014	Paper Wiki Pres.
Router-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-ADN2	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	46.0 [41.5 - 52.3]	87.4 [85.5 - 75.2]	43.6 [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 - 57.1]	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]	50.7 [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]	29.2 [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 [40.9 - 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 multinomial 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
- Multinomial OVO classification

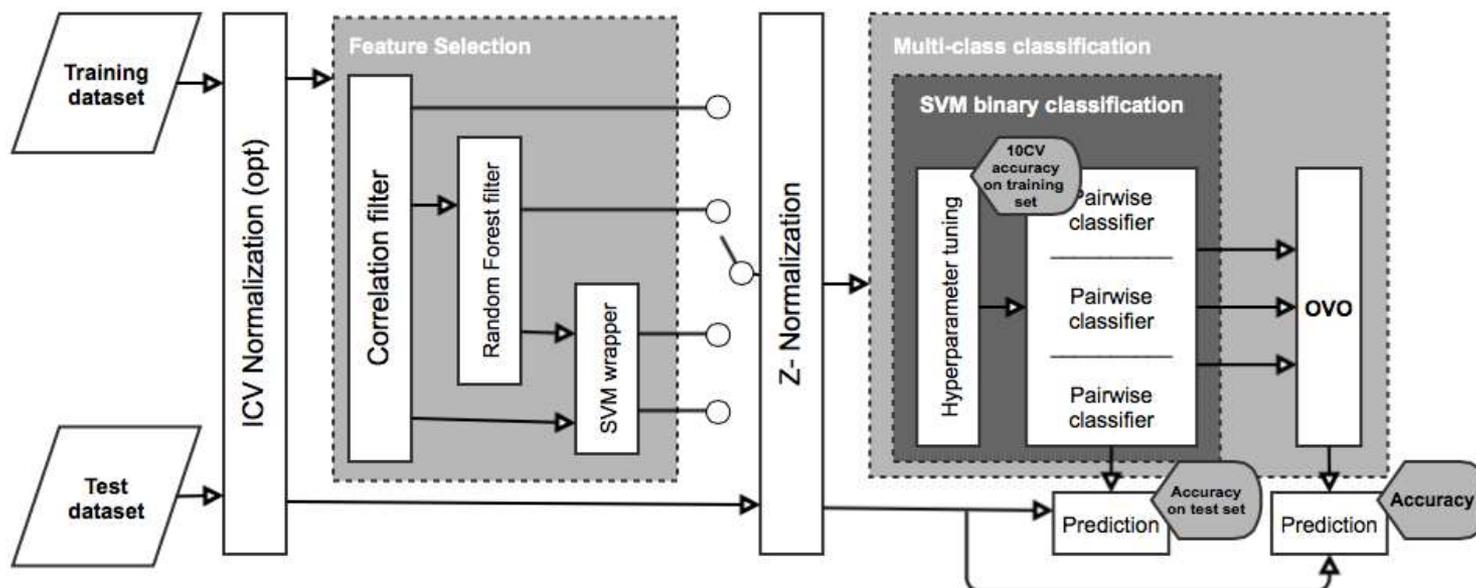
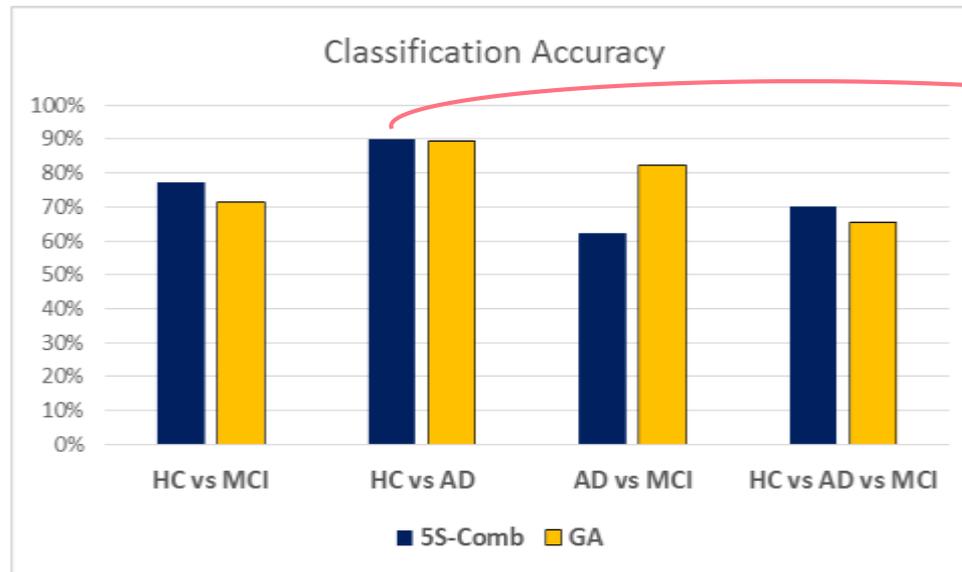


Table 2. Results of the binary classification: training performed on D2 and accuracy estimated on D2 (D2/D2) and D1 (D2/D1). Best results on D1 are indicated in bold.

	Feature selection method																							
	Corr. filter $ r < 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	D2/D1	Nr. feat.	D2/D2	D2/D1	Nr. feat.	D2/D2	D2/D1	Nr. feat.	D2/D2	D2/D1	Nr. feat.	D2/D2	D2/D1	Nr. feat.	D2/D2	D2/D1						
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

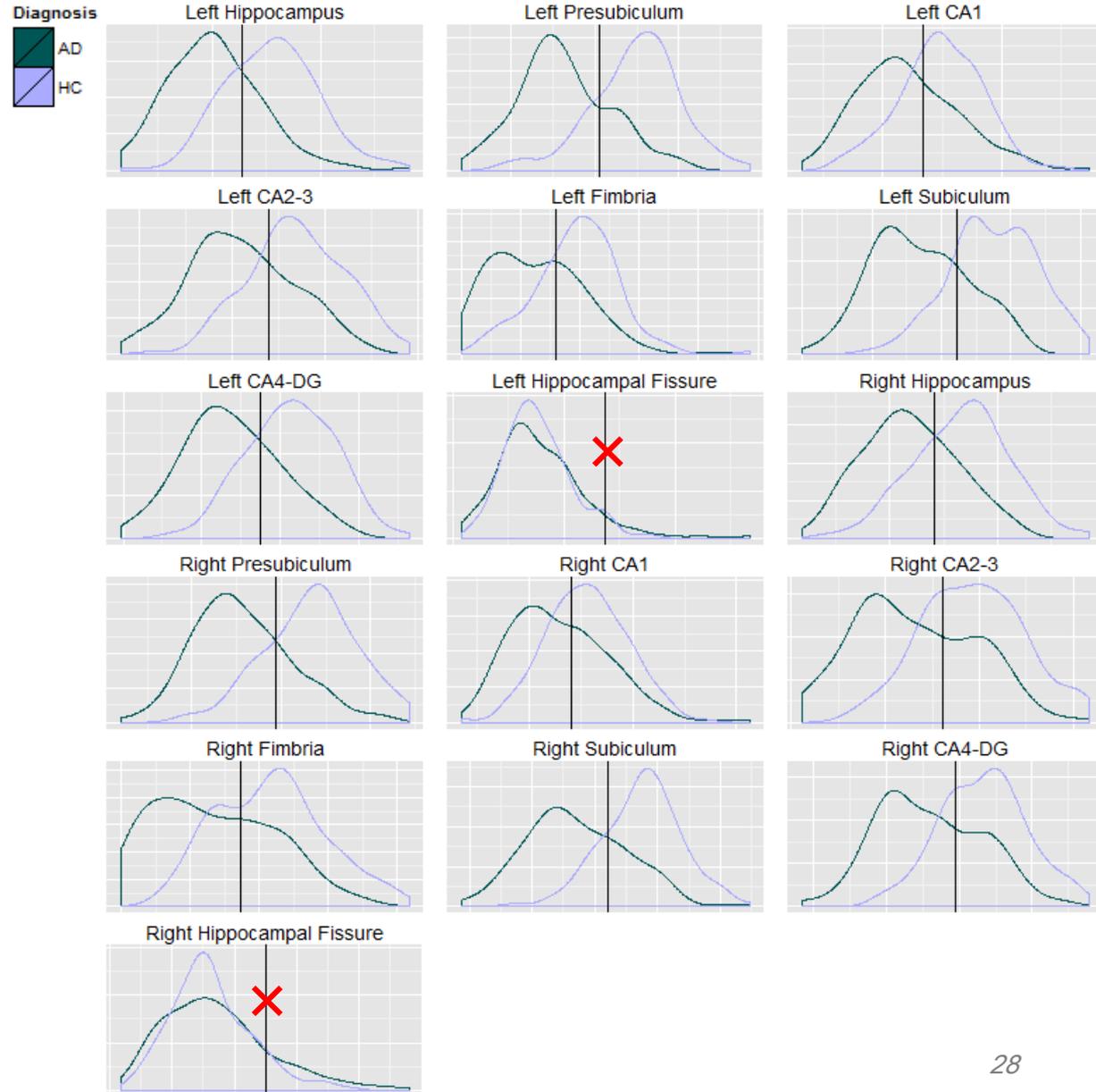
1. Advanced Feature Selection composed by five stages (**5S-Combo**),
 2. Genetic Algorithms (**GA**)
- Both tested with
 - Multinomial classification task
 - SVM with radial basis filter kernel
 - ADNI data
 - 10-fold cross validation



SVM used 133 features

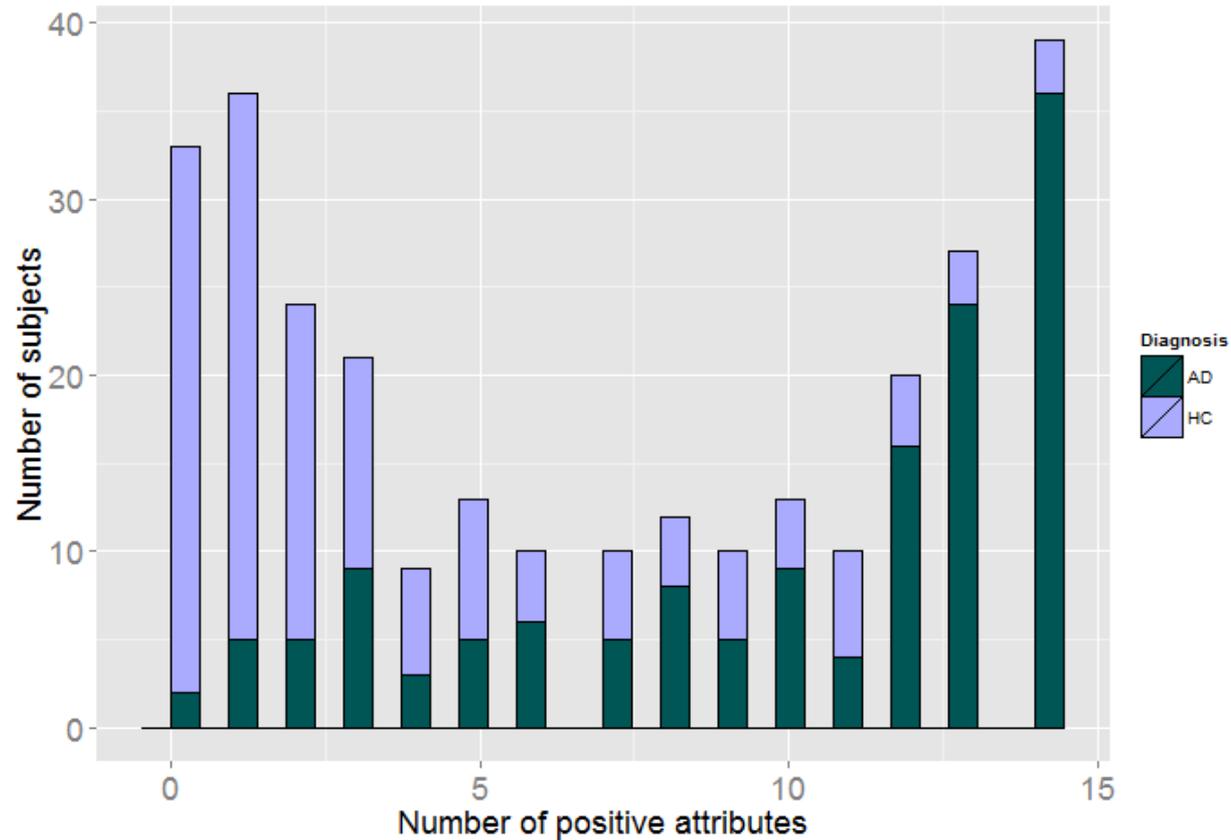
LDA and Probability-based Classifiers

- PDF of volumes of 16 hippocampal subfields across subjects: HC vs AD
- X-axis: volume measure v
- Y-axis: $P(\text{AD} | v)$ and $P(\text{HC} | v)$
- Thresholds determined by Linear Discriminant Analysis (LDA)



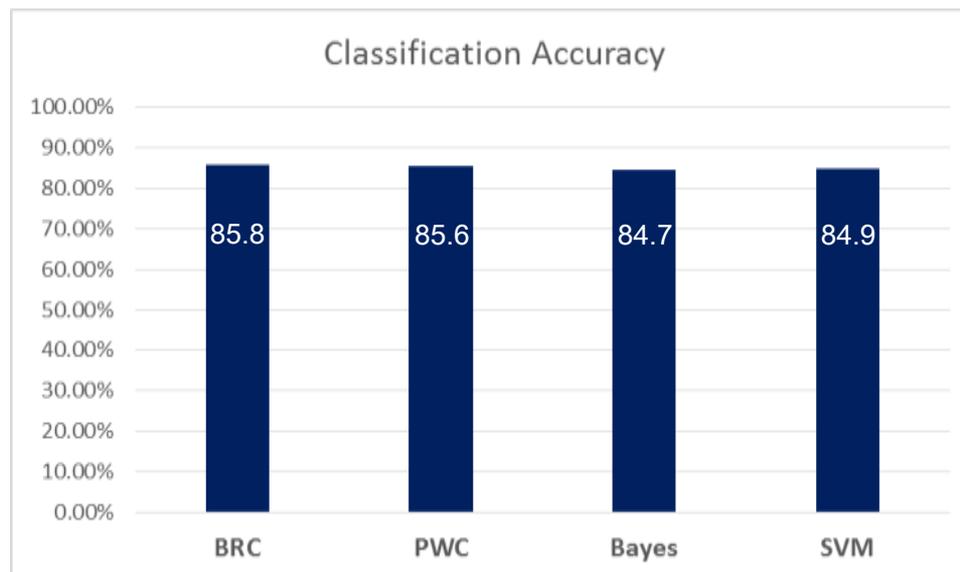
LDA and Probability-based Classifiers

Number of subjects found positive in a number of regions



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 Multinomial 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.