WIC 2014 / The 2014 International Conference on Brain Informatics and Health
11–14 August 2014 / Warsaw, Poland

SPECIAL SESSION ON
NEUROIMAGING DATA PROCESSING STRATEGIES

K-Surfer: A KNIME Extension for the Management and Analysis of Human Brain MRI FreeSurfer/FSL Data

Alessia Sarica[1], Giuseppe Di Fatta[2], and Mario Cannataro[1]

[1] *Bioinformatics Laboratory, Department of Medical and Surgical Sciences, Magna Graecia University of Catanzaro, Italy*
sarica,cannataro@unicz.it

[2] *School of System Engineering, University of Reading, UK*
g.difatta@reading.ac.uk

University of Reading

Università Magna Græcia di Catanzaro

Summary

- Knowledge Discovery from Neuroimaging
  - Workflow for the Knowledge Discovery from Neuroimaging
- Pre-processing of Neuroimages:
  - FreeSurfer and FSL;
  - Overview on outputs produced by FreeSurfer/FSL;
  - Drawbacks on managing and analysing FreeSurfer/FSL outputs.
- The Konstanz Information Miner (KNIME)
- **K-Surfer**:
  - software architecture and features;
  - nodes and meta nodes.
- A case study using FreeSurfer tutorial data.
- Conclusion and Future Works.

Knowledge Discovery from Neuroimaing

- The processing of brain images has exponentially increased the quantity of data to be analysed.
- A Machine Learning approach has proved necessary where data from neuroimaging are merged with data from multiple sources, as in the case of Alzheimer's disease[1].
- **The aim is to provide the basis for decisions, both immediately and in the long term, especially for the early detection of neurodegenerative diseases.**

[1] Jieping Ye, Teresa Wu, Jing Li, Kewei Chen. Machine Learning Approaches for the Neuroimaging Study of Alzheimer's Disease. *IEEE Computer Society* 2011; 44(4):99-101.

Workflow for the Knowledge Discovery from Neuroimaging

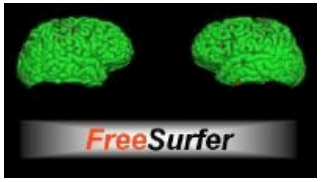
1. Acquisition of structural and/or functional neuroimages;
2. Pre-processing of neuroimages (e.g. Segmentation and reconstruction, tractography).
3. Merging of clinical, genomics or proteomics information (*optional*);
4. Data analysis by of statistical techniques and mining algorithms;

Automatic approach to the pre-processing of Neuroimages

- Although automated pre-processing may introduce some overestimations[2], it is definitely faster than manual segmentation.
- Software for neuroimaging processing could be roughly categorized into monolithic and modular implementations[3].
- **FreeSurfer** and **FSL** are popular examples of the modular approach: their tools and scripts can be used in conjunction for obtaining best results.

[2] Cannataro M, Guzzi PH, Sarica, A. Data mining and life sciences applications on the grid. WIREs Data Mining Knowl Discov, 2013, 3:216238.
[3] D. E. Rex, J. Q. Ma, and A. W. Toga, 'The Loni Pipeline Processing Environment', Neuroimage, 19 (2003), 1033-48.

FreeSurfer and FSL



FreeSurfer is an open-source suite of modules for processing, analysing and visualising human brain MRI:

- Volume based and surface based analysis;
- Tractography, *TRACULA* (TRActs Constrained by UnderLying Anatomy).

The **FMRIB Software Library (FSL)** provides freely available image analysis and statistical tools for fMRI, MRI and DTI brain data.

Outputs produced by FreeSurfer/FSL

- Morphological measures:
  - volumes of 45 brain areas.
  - surface area, average thickness and thickness standard deviation of specific structures of left and right hemispheres (34 for the left and 34 for the right).
- Diffusion indices of white-matter pathways (18 reconstructed tracts):
  - average measures (axial diffusivity, radial diffusivity, mean diffusivity and fractional anisotropy);
  - diffusion values for each voxel of the tract (statistics as a function of position along the trajectory of a tract).

Drawbacks on managing and analysing FreeSurfer/FSL outputs

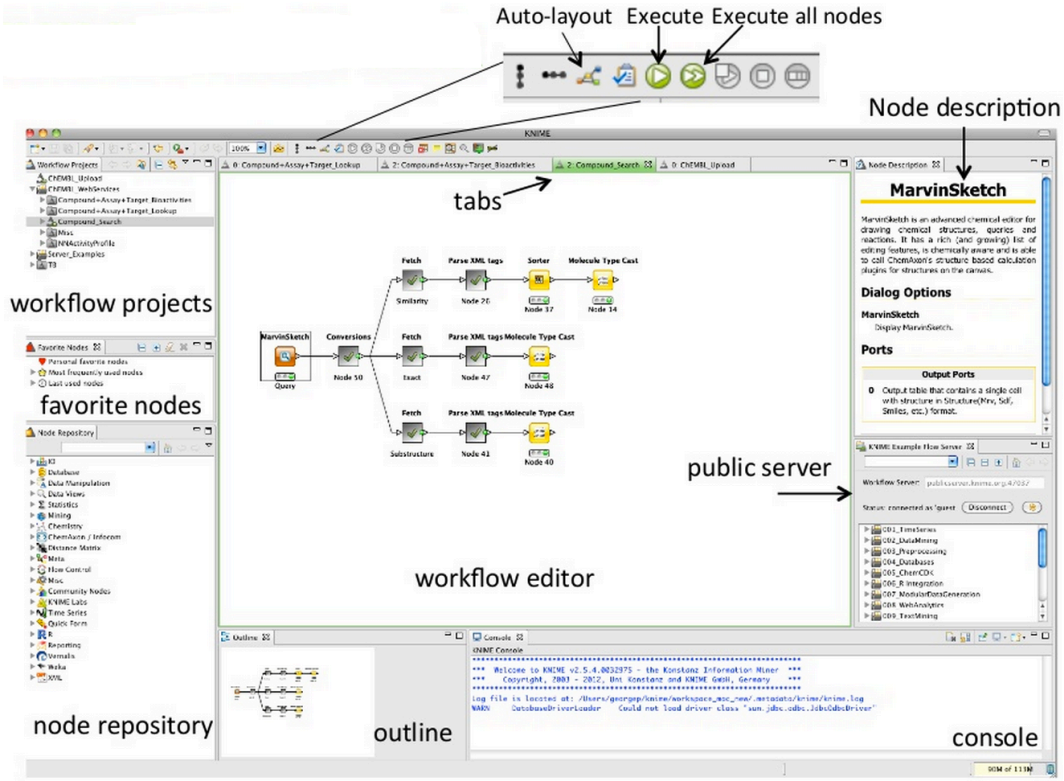
- FreeSurfer/FSL tools and scripts use is not immediate and is suitable for expert users only.
  - E.g. for converting diffusion data into a table format:

```
tractstats2table --inputs $SUBJECTS_DIR/diffusion/hugo/
dpath/lh.ilf_AS_avg33_mni_bbr/pathstats.overall.txt --
overall --tablefile $SUBJECTS_DIR/diffusion /
lh.ilf.hugo.All.table
```

- No joining functionality for volume, thickness, diffusion values and demographical/clinical, genomics and proteomics data.

- **At the moment, the only attempt to give an instrument for importing, managing and analysing FreeSurfer/FSL data into an analytics platform, is K-Surfer, a plugin for the Konstanz Information Miner (KNIME).**

The Konstanz Information Miner



- A tool for data analysis, data manipulation, data visualization, data generating and reporting.
  - free and open–source;
  - one of the most popular data mining tools.
- **Splits each process into small and understandable units (nodes) linked by pipelines.**

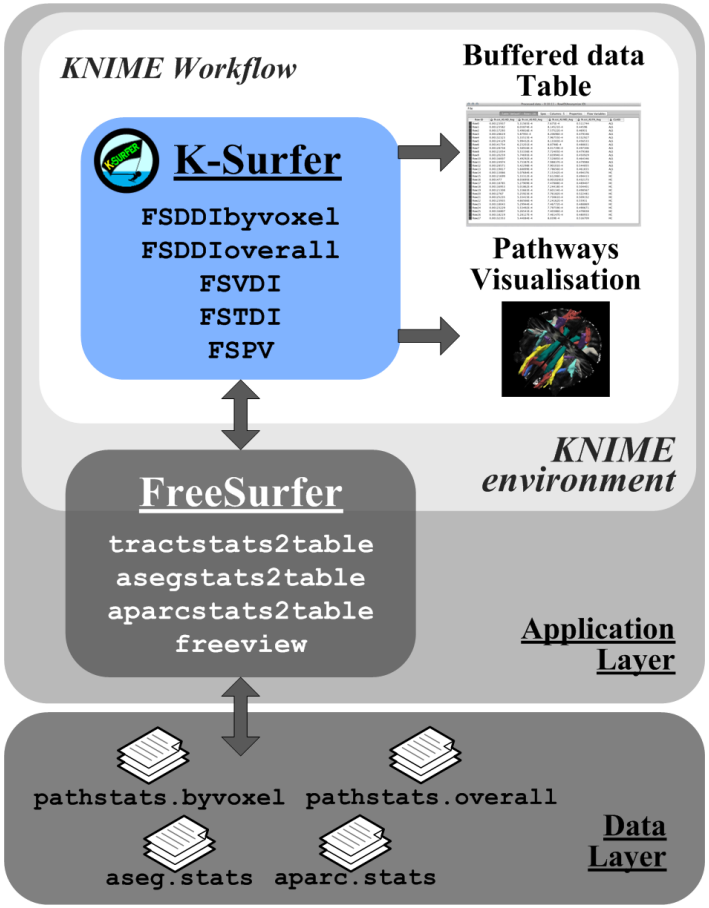
K-Surfer

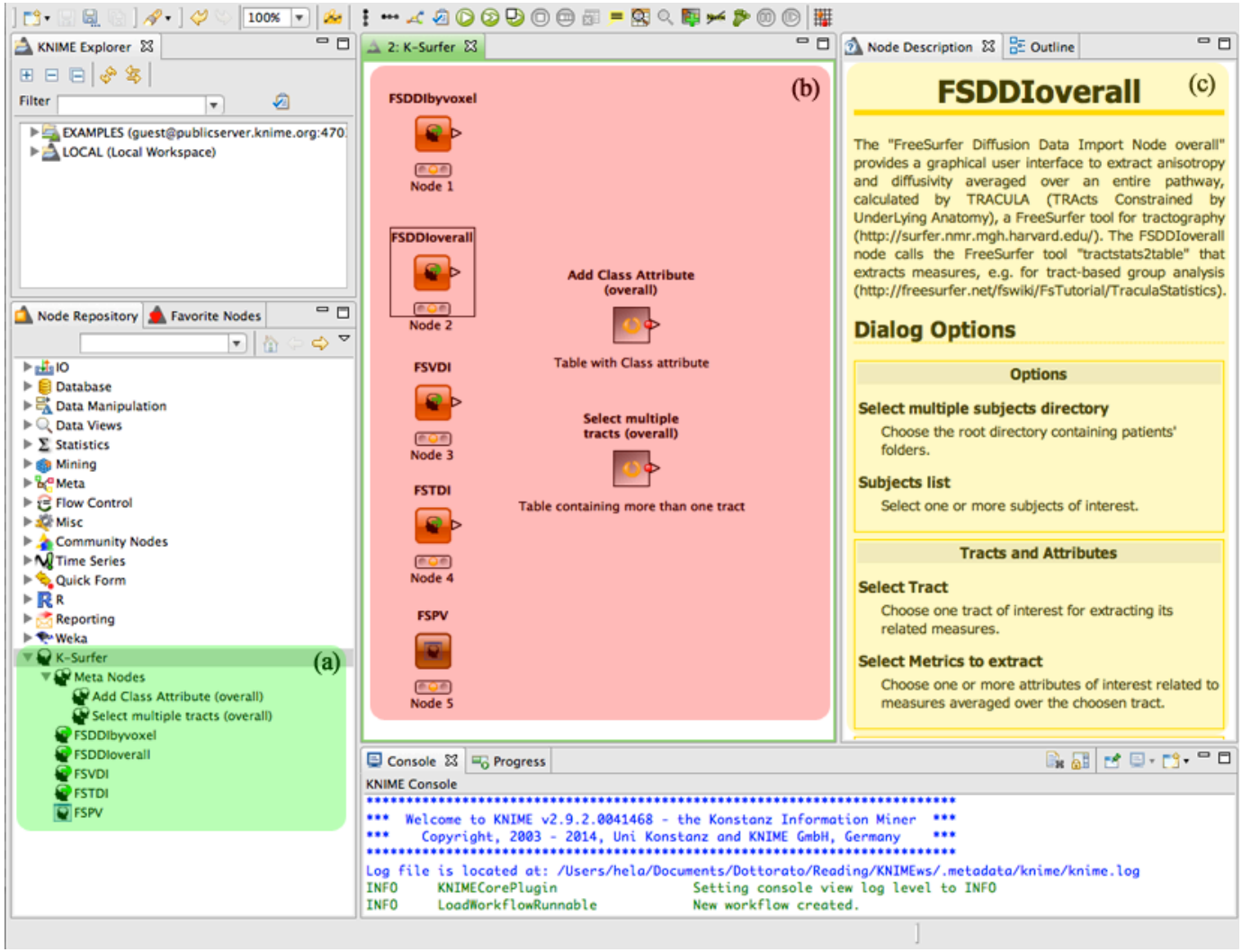
http://sourceforge.net/projects/ksurfer/

- K-Surfer is a novel KNIME plug-in that contributes a number of nodes and meta nodes to integrate 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.
- Its user-friendly nodes configuration dialogues do not require the user to write UNIX shell commands and scripts.

K-Surfer features

- Data from both local and remote machines;
- Uniqueness of row IDs;
- Name conventions for attributes following FreeSurfer specifications;
- New functionalities:
  - importing measure related to one or more brain tracts as a single step;
  - selecting subjects of different studies stored in different directories;
  - merging volume, thickness and diffusion data tables;

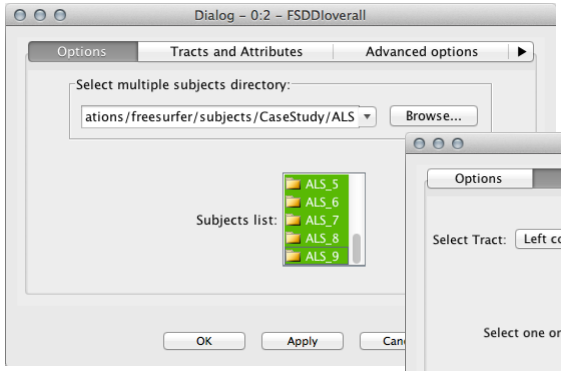




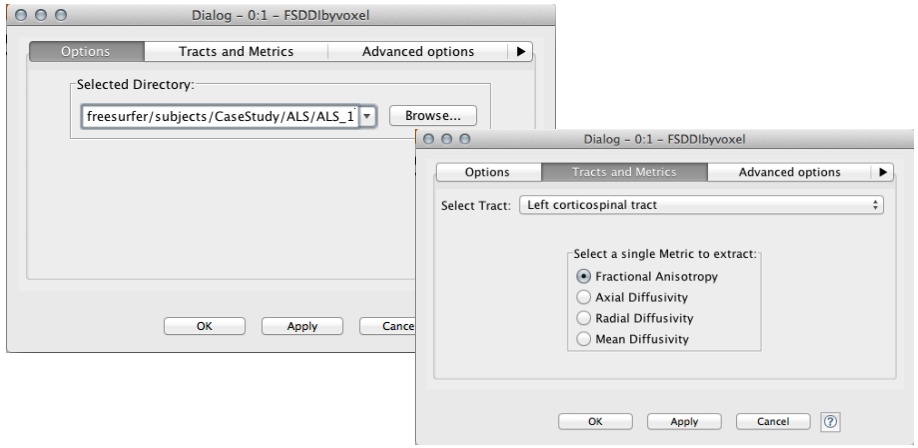
- Screenshot of KNIME workbench (a) K-Surfer nodes category, (b) Workspace containing K-Surfer nodes, (c) Node Description for "*FreeSurfer Diffusion Data Import overall*".

K-Surfer nodes

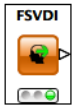
FSDDIoverall – FreeSurfer Diffusion Data Import overall. Extracts anisotropy and diffusivity values averaged over an entire pathway.



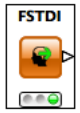
FSDDIbyvoxel - FreeSurfer Diffusion Data Import by voxel. Extracts measures as a function of position along the trajectory of the pathway.



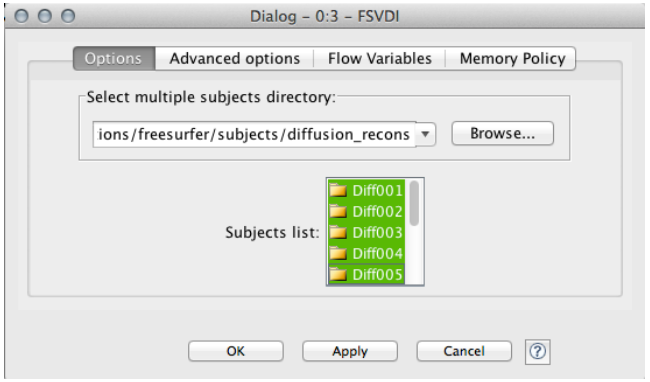
K-Surfer nodes

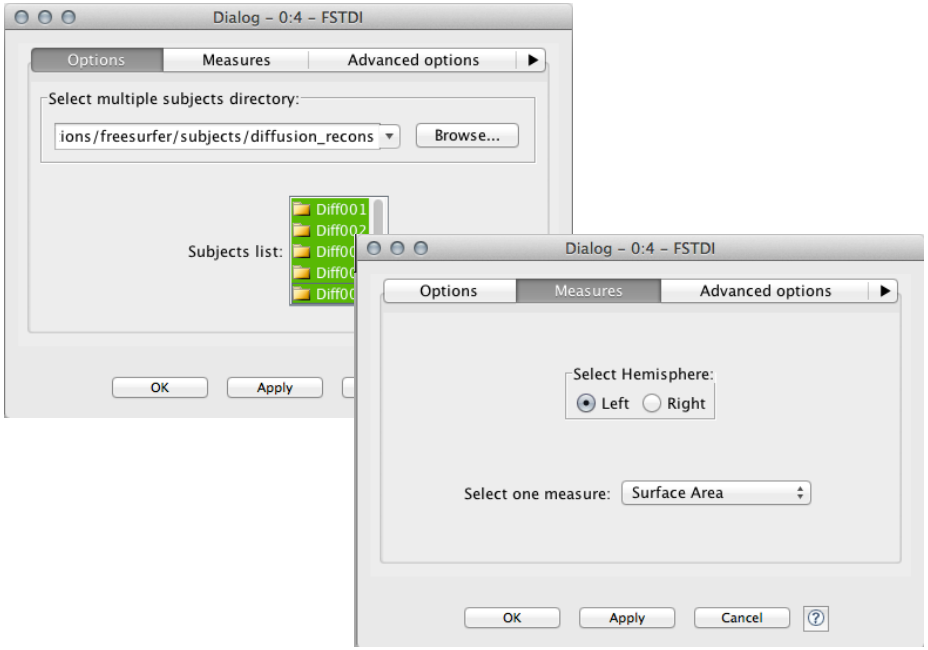


FSVDI - FreeSurfer Volume Data Import. Extracts the volumes [mm^3] of specific structures.

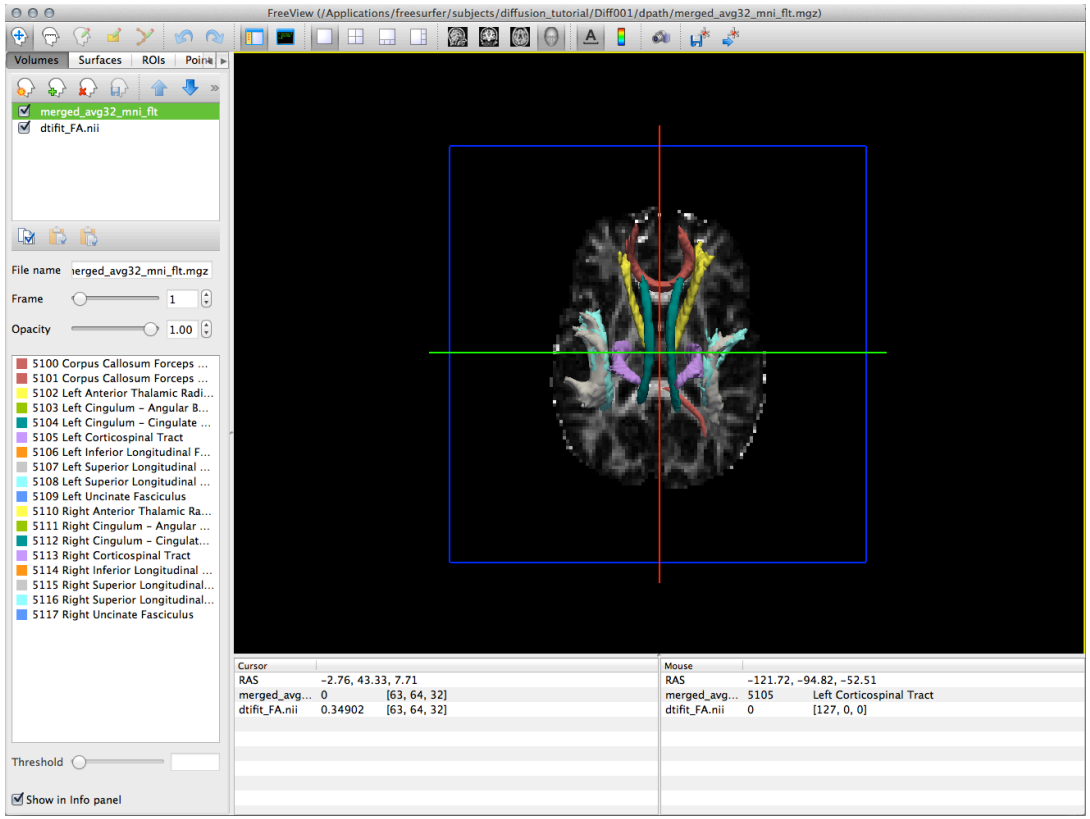


FSTDI - FreeSurfer Thickness Data Import. Extracts the thickness [mm] of specific structures.

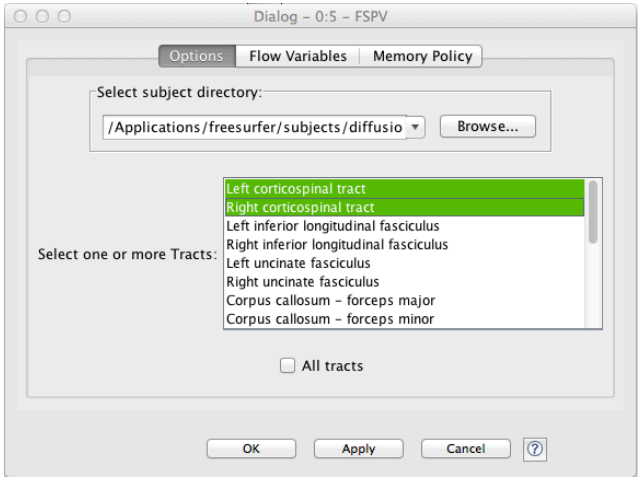




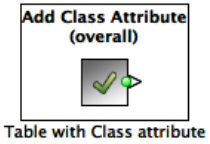
K-Surfer nodes



FSPV - FreeSurfer Pathways Viewer. Visualises the probability distribution of single white-matter pathways or all white-matter pathways.

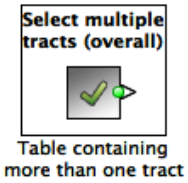


K-Surfer Meta Nodes



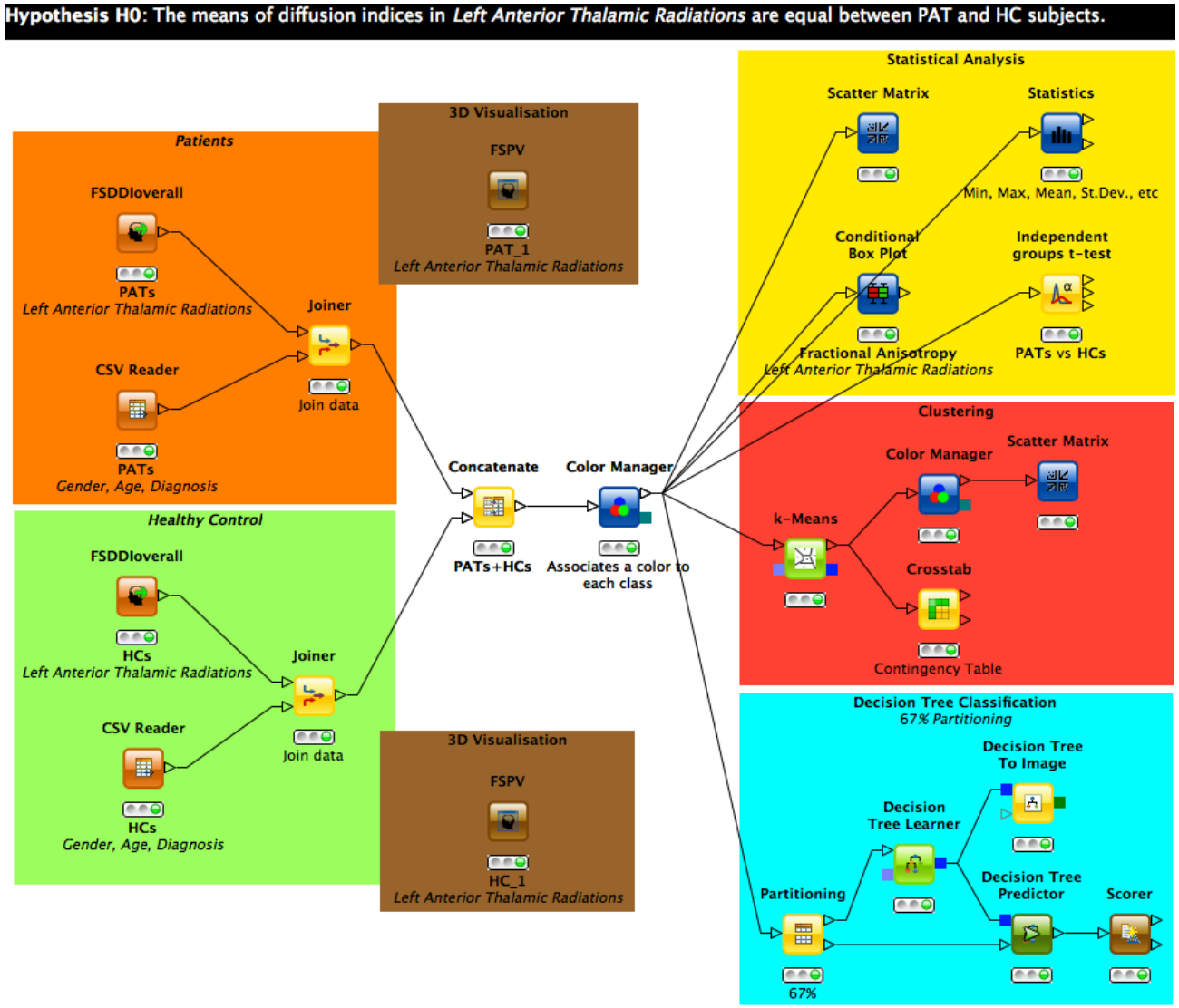
Add Class Attribute (overall).

Uses two FSDDoverall nodes to add to a diffusion data table a new column containing the class attribute. *E.g. For group analysis, the user needs to add class feature, such as healthy control patients, or affected patients*.



Select multiple tracts (overall).

Extracts the diffusion values of more than one tract at once, using the FSDDoverall node.



- Case Study workflow for confuting the null hypothesis that "*the means of diffusion indices in Left Anterior Thalamic Radiations are equal between PAT and HC subjects*".

Conclusion and Future Works

- A main bottleneck in brain MRI data analysis is the overwhelming availability of experimental data and the complexity of the pre-processing and data selection phases.
- K-Surfer automatizes the overall process and helps to reduce time costs and human errors by removing the need for low-level data manipulation.
- As future work:
  - new nodes, for example, for visualising cortical and subcortical segmentation.
  - extension to support data from other tools like Caret or 3D Slicer.

K-Surfer has been successfully used in the contest of the CADDementia challenge 2014

A Sarica, G Di Fatta, G Smith, M Cannataro, D Saddy ***Advanced feature selection in multinominal dementia classification from structural MRI data***

Thanks for the attention.

Alessia Sarica - *sarica@unicz.it*