

School of Science
Department of Physics and Astronomy
Master Degree in Physics

**Development of pre and post-processing
steps to a pipeline aimed to identify silent
cerebral infarcts**

Applicant
Nicolas Biondini

Supervisor
Prof. G. Castellani
Co-Supervisor
Prof. D. Remondini
Dr. R. Biondi

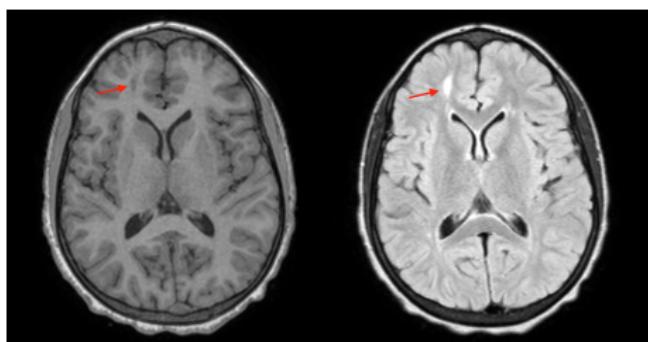
31/03/2023

Sickle Cell Disease and Silent Cerebral Infarcts

- Sickle Cell Disease (SCD) is a group of inherited red blood cell disorders that causes abnormal hemoglobin.
- Silent Cerebral Infarcts (SCIs) are defined as abnormal MRI of the brain in the setting of a normal neurologic examination.

SCIs appear as:

- **Hypointense** regions in T1W images.
- **Hyperintense** regions in T2W and FLAIR images.



Why and How Segment SCIs?

Why?

SCIs could lead to:

- a decrement in general intellectual abilities,
- poor academic achievement,
- poor working abilities
- a minor quality of life.

SCI segmentation is necessary to:

- understand how SCI negatively affects cognition;
- provide a starting point for the identification of potential targets for preventive therapies.

Why and How Segment SCIs?

How?

Up to now is **manually** done:

- made by highly trained and specialized neuroradiologists;
- time consuming;
- influenced by the experience of the operator.

An **automatic** pipeline has been proposed:

- In the contest of the European project *Genomed4All*;
- Core of the pipeline: pre-trained **U-Net** ensemble;
- Necessity of **pre** and **post-processing** steps.

Data Set

Head MRI Scans

- 57 MRI **T1W** and **FLAIR** head scans;
- Acquired from 2009 to 2020;
- From three different medical centers in Italy;
- 51 with SCI evidences, 6 without lesions;
- High **heterogeneity** in both acquisition times and spatial resolution;
- Mainly **underaged** patients;
- Manual SCIs segmentation for each scan.

Axis	Size (pixel)		Spacing (mm)	
	Mean	Std. Dev.	Mean	Std. Dev.
x	256	0	0.85	0.12
y	256	0	0.81	0.14
z	90	132	4.34	1.93

Spatial Resolution of the images

Main Pipeline

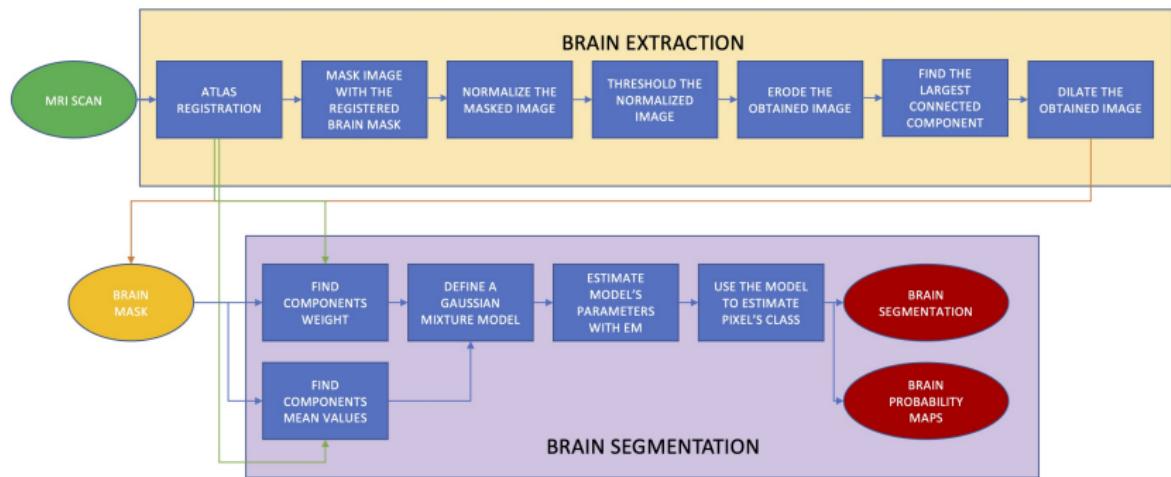
U-Net ensemble, Pre-Processing and Post-Processing

The U-Net ensemble needed standardized data for the segmentation and the obtained labels had to be refined.

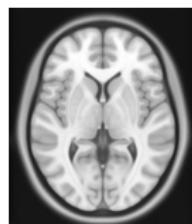
- **Pre-Processing:**
 - Gaussian Normalization;
 - Brain Extraction;
 - Tissue Segmentation;
- **SCIs segmentation;**
- **Post-Processing:**
 - Removals of the smallest labels;
 - Features extraction;
 - Classifiers' training;
 - Classifiers' test.

Pre-Processing

Brain Extraction, Normalization, Tissue Segmentation



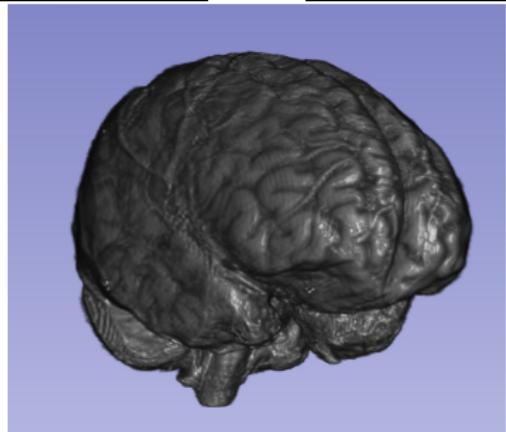
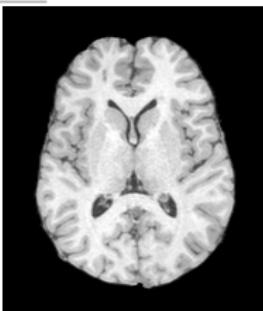
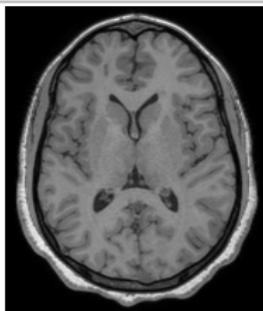
The Pre-Processing pipeline uses an already segmented atlas: ICBM MNI 152 was used.



Brain Extraction

Atlas registration, Brain normalization and Skull Stripping

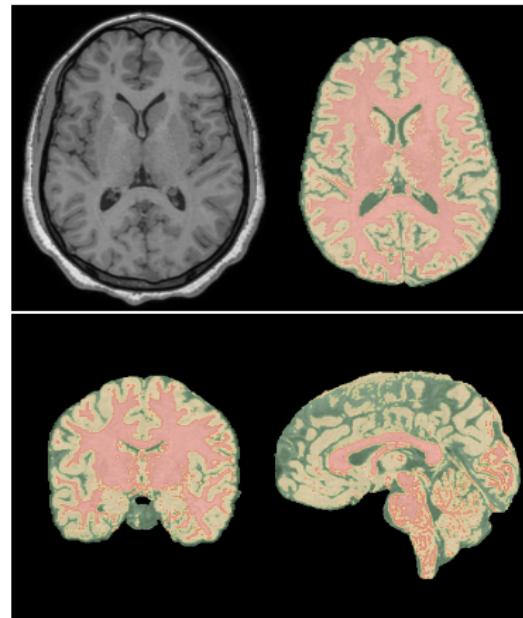
- **Atlas Registration:** The MNI 152 atlas was non rigidly registered on the T1-W image;
- **Normalization:** Using the brain mask of the atlas the image was normalized considering only the masked pixels;
- **Thresholding:** The normalized image was thresholded;
- **Find the Largest Connected Component:** The connected components of the images were counted and only the largest one was kept.



Tissue Segmentation

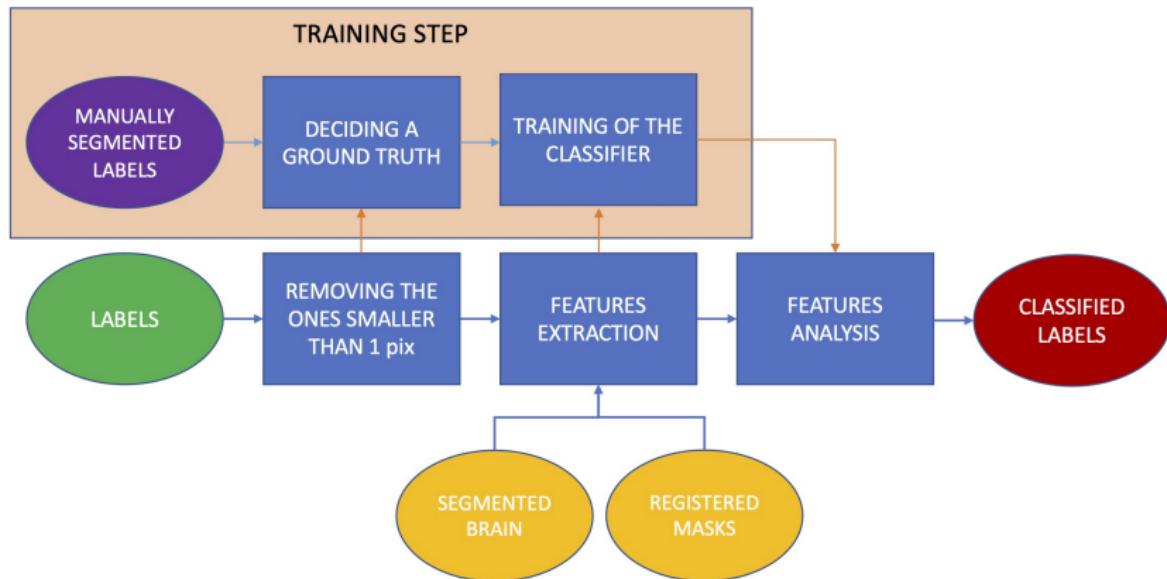
A Gaussian Mixture model approach.

- **Find Tissues Mean Values and Weights:** Starting from the atlas' partial volume maps compute the mean pixel value and the fraction of pixels belonging to each tissue;
- **Define a Gaussian Mixture Model:** A mixture of three Gaussian distributions is defined starting from the parameters found;
- **Expectation Maximization:** An expectation maximization algorithm finds the best parameters of the Gaussian mixture model;
- **Classification:** Each pixel is classified.



Post-Processing

Removals of the False Positives.

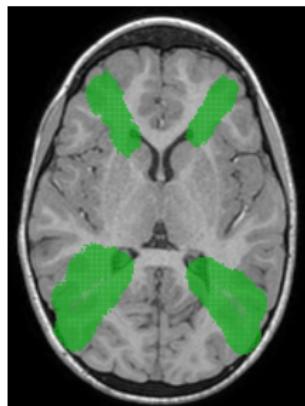


Features Extraction

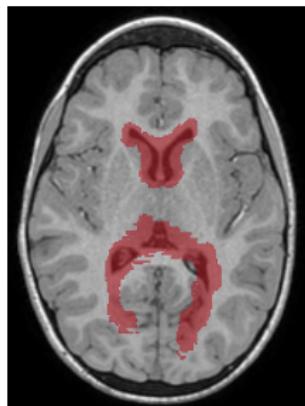
Seven different features were extracted for each segmented lesion:

- **Lesion probability:** The probability in output from the U-Net ensemble;
- **Label's volume:** The total volume of the label;
- **Overlapping with boundary region mask:** A region where is more probable to find SCIs;
- **Overlapping with exclusion region mask:** A region that can easily give false positives;
- **Probability to belong to each tissue:** The probability for a mask to belong to each tissue using the segmentation obtained in pre-processing step.

The boundary and exclusion masks manually segmented on the MNI 152 atlas by expert clinicians.



Boundary.



Exclusion.

Training

Classifiers, Data Sets, Data Processing

Classifiers

Three different classifiers were implemented:

- Decision Tree;
- Logistic Regression;
- Random Forest.

Data Sets

The data set was previously divided in:

- 51 scans: Training;
- 6 scans: Test.

The division kept in account heterogeneity for:

- Center of provenience;
- Scan's resolution;
- Presence of lesions.

Data Processing

The training data set was highly unbalanced: 3.7% considered true.

- SMOTE resampling;
- Random undersampling;
- Robust Features Scaler.

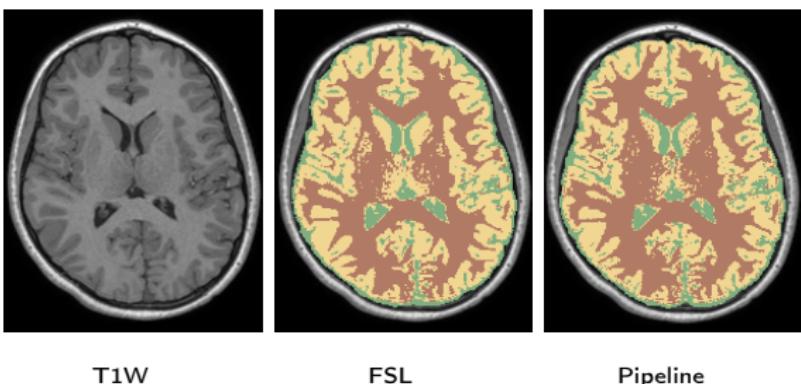
Results

Pre-Processing

Dice Similarity Coefficient

	Mean	Std. Dev.	Median	IQR
Brain Mask	0.87	0.12	0.93	0.11
WM	0.78	0.19	0.83	0.19
GM	0.67	0.30	0.80	0.28
CSF	0.66	0.24	0.79	0.22

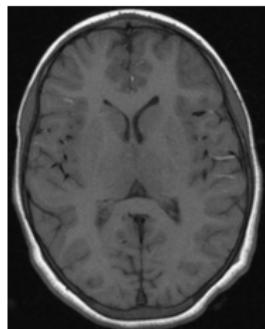
- Comparison with masks obtained with FSL;
- Computed metric: Dice Similarity Coefficient;



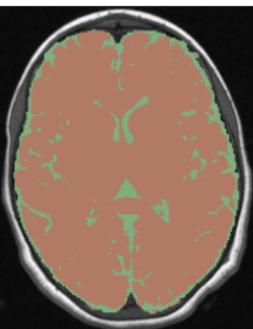
Results

Pre-Processing

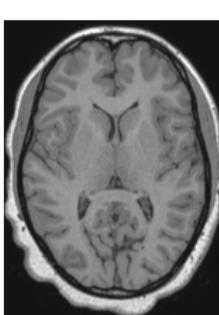
Dice coefficient measures overlapping between two images.



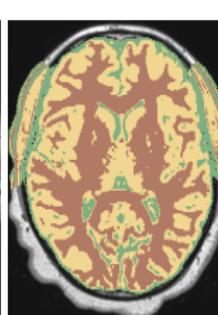
Original Scan



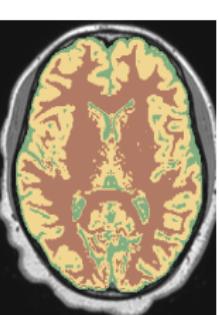
Segmentation



Original Scan



FSL



Pipeline

Pipeline Fault

FSL Fault

Results

Post-Processing

Classifiers Performances

	Bal.	Acc.	AUC	ROC	AUC	PR	Precision	Recall
Decision Tree	0.74		0.81		0.45		0.38	0.96
Logistic regression	0.68		0.83		0.61		0.32	0.96
Random Forest	0.74		0.90		0.74		0.38	0.97

Features' Importance

	Prob.	Size	Boundary	Exclusion	WM	GM	CSF
Dec. Tree	0.000	0.0133	0.032	0.034	0.659	0.042	0.098
Log. reg.	0.00	0.0883	0.236	-1.50	1.07×10^5	4.58×10^4	7.82×10^3
Rnd F.	0.000	0.146	0.027	0.029	0.615	0.067	0.116

Conclusions

Pre-Processing

- Fully automated;
- Good agreement with standard used softwares;
- Possibility to improve the tissue segmentation.

Post-Processing

- Fully automated;
- Improved U-Net results.

Ground Truth Decision

Modified Jaccard Score

To decide if a automatically found label is to considered true it was necessary to:

- Define an overlapping metric;
- Measure the overlapping of a label with the manually segmented ones;
- Decide a threshold.

Modified Jaccard Score

$$score(X) = \frac{|X \cap (Y_1 \cup Y_2 \cup \dots \cup Y_n)|}{|X \cup (Y_1 \cup Y_2 \cup \dots \cup Y_n)|}$$

- X : automatic segmented label;
- Y_i : overlapping manual label;
- n : number of overlapping manual labels.