

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

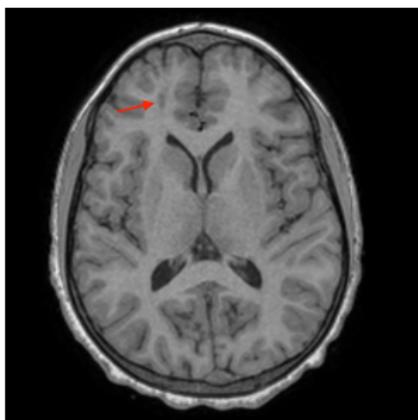
31/03/2023

# Sickle Cell Disease and Silent Cerebral Infarcts

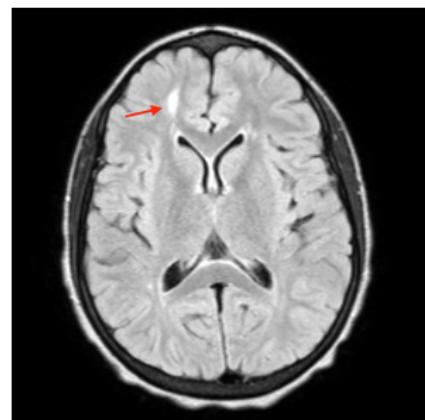
- Sickle Cell Disease (SCD) is a group of inherited red blood cell disorders that causes abnormal hemoglobin.
- Silent Cerebral Infarcts (SCIs) are defined as abnormal MRI of the brain in the setting of a normal neurologic examination.

SCIs appear as:

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



T1W



FLAIR

# Why Segment SCIs?

The impact on the quality of life

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.

# How Segment SCIs?

A Machine and Deep learning approach.

Up to now is **manually** done:

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

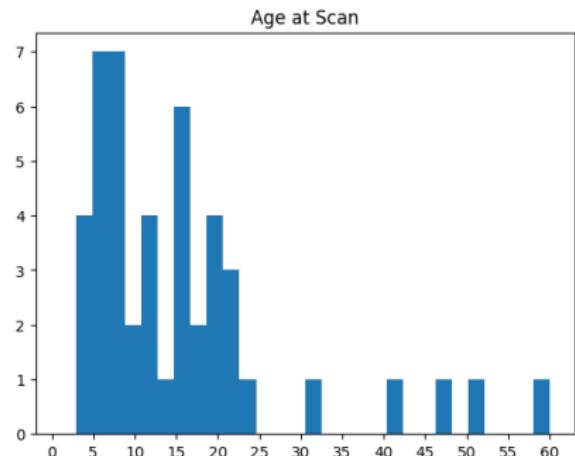
An **automatic** pipeline has been proposed:

- In the contest of the European project *Genomed4All*;
- Core of the pipeline: pre-trained **U-Net** ensemble;
- **Pre-processing** step to provided standardized data;
- **Post-processing** step to refine the segmented labels.

# Data Set

## Head MRI Scans

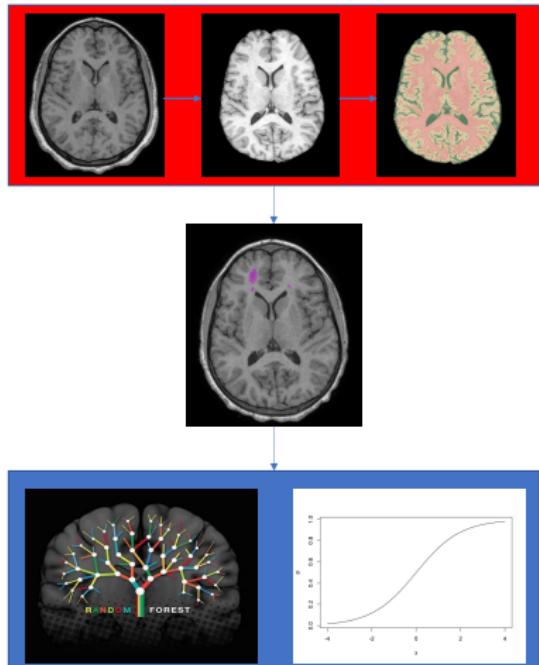
- 57 MRI **T1W** and **FLAIR** head scans;
- Acquired from 2009 to 2020;
- From three different medical centers in Italy;
- High **heterogeneity** in acquisition parameters;
- Mainly **underaged** patients;
- Manual SCIs segmentation for each scan.



# Main Pipeline

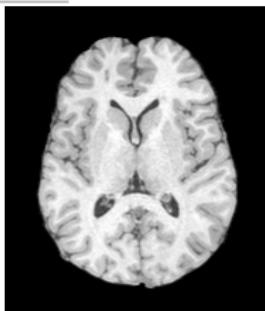
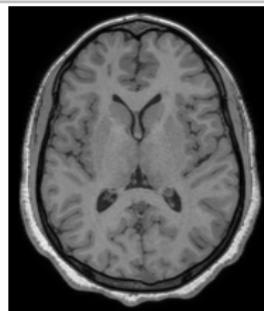
## U-Net ensemble, Pre-Processing and Post-Processing

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

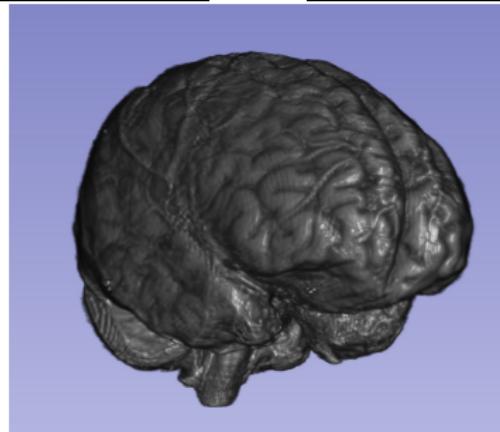


# Brain Extraction

Atlas registration, Brain normalization and Skull Stripping



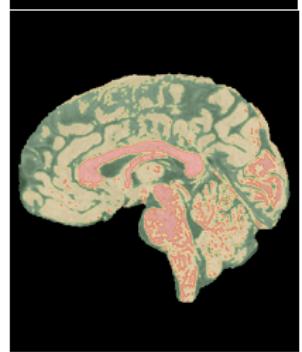
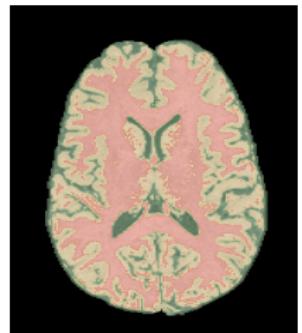
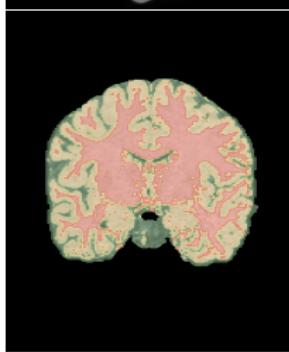
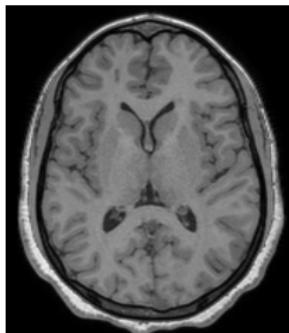
- Atlas Registration;
- Normalization;
- Thresholding;
- Find the Largest Connected Component.



# Tissue Segmentation

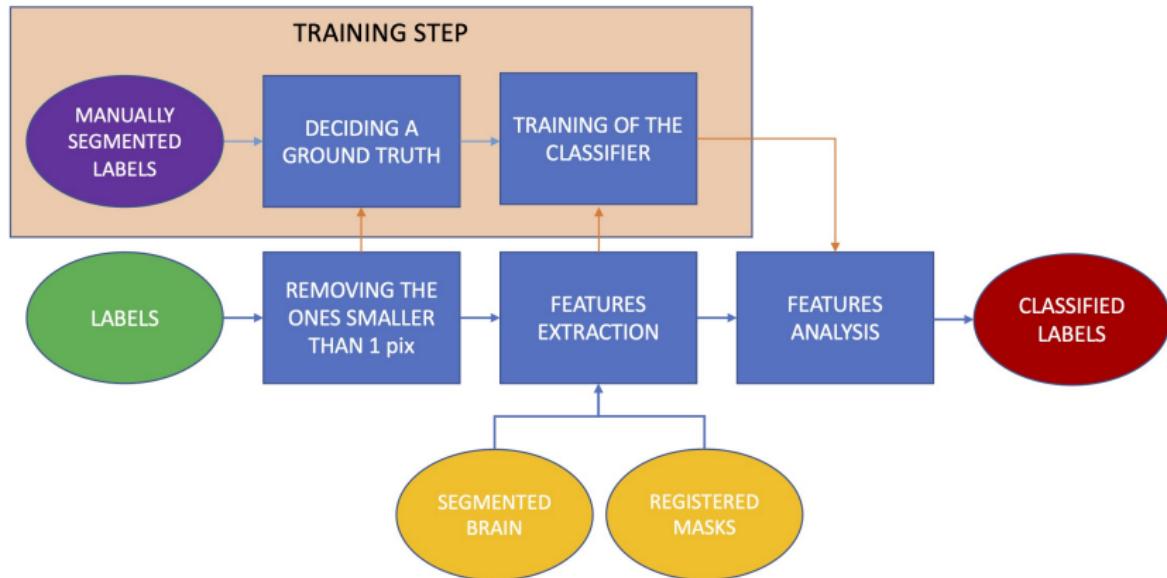
A Gaussian Mixture model approach.

- Find Tissues Mean Values and Weights;
- Define a Gaussian Mixture Model;
- Expectation Maximization;
- Classification.



# Post-Processing

Removals of the False Positives.

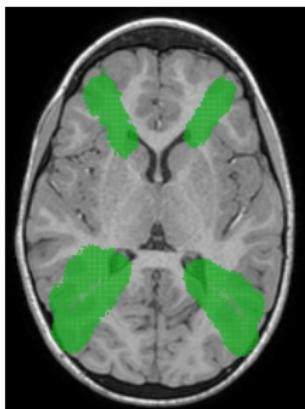


# Features Extraction

Seven different features were extracted for each segmented lesion:

- Lesion probability;
- Label's volume;
- Overlapping with boundary region mask;
- Overlapping with exclusion region mask;
- Probability to belong to each tissue.

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

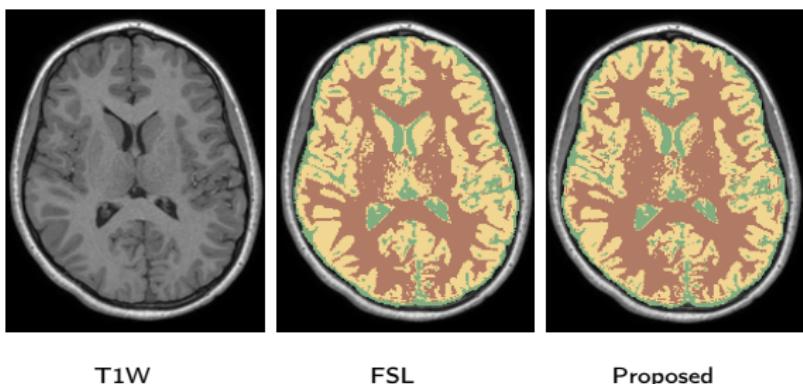
## Pre-Processing

- Comparison with masks obtained with FSL;
- Computed metric: Dice Similarity Coefficient (DSC).

$$DSC = 2 \frac{|X \cap Y|}{|X| + |Y|}$$

### Dice Similarity Coefficient

	Mean	Std. Dev.	Median	IQR
<b>Brain Mask</b>	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



# 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	$1.07 \times 10^5$	$4.58 \times 10^4$	$7.82 \times 10^3$
Rnd F.	0.000	0.146	0.027	0.029	0.615	0.067	0.116

# Conclusions

## Pre-Processing

- Fully automated;
- Good DSC in a comparison with standard used softwares;
- Standardized images provided to U-Net permitting a better SCI segmentation.

## Post-Processing

- Fully automated;
- Improved U-Net results;
- Explainable features and classifiers.

**THANK YOU FOR YOUR  
ATTENTION!**

# 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

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

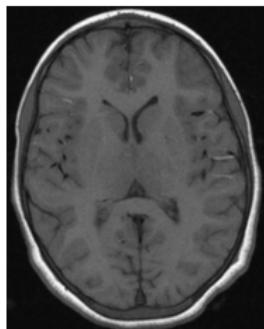
- $X$  : automatic segmented label;
- $Y_i$  : overlapping manual label;
- $n$  : number of overlapping manual labels.

# Results

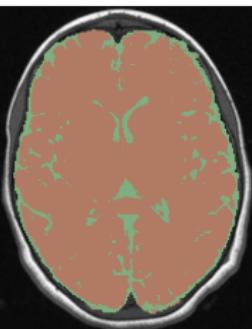
## Pre-Processing

---

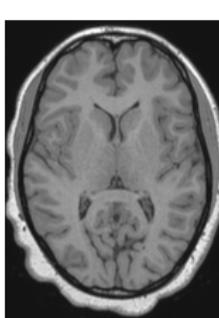
Dice coefficient measures overlapping between two images.



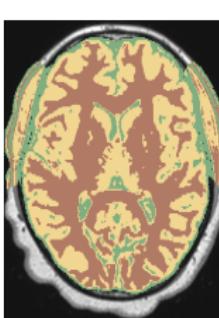
Original Scan



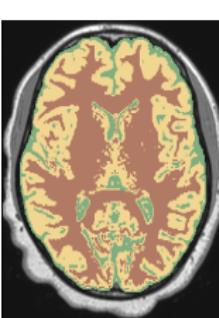
Segmentation



Original Scan



FSL



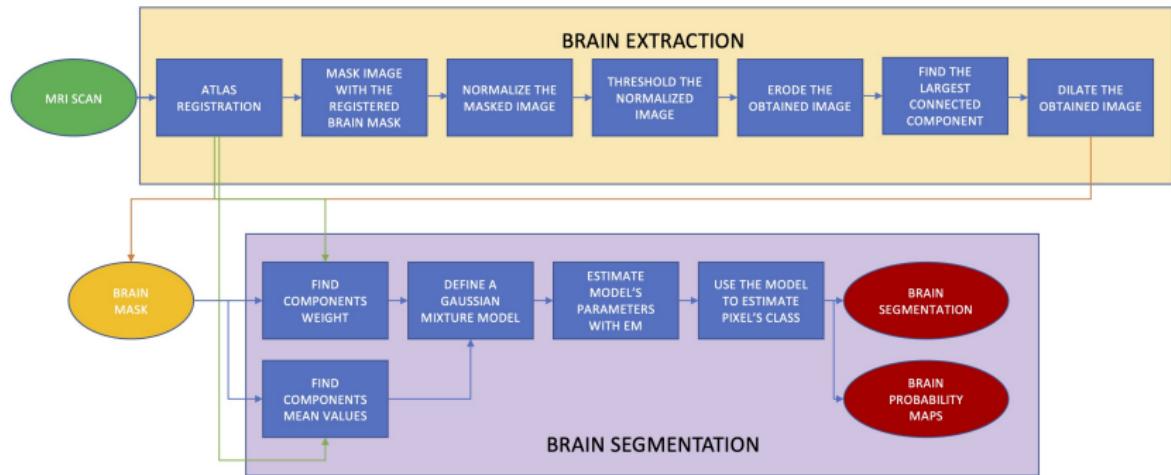
Pipeline

Pipeline Fault

FSL Fault

# Pre-Processing

## Brain Extraction, Normalization, Tissue Segmentation



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

