MRI analysis for Hippocampus segmentation on a distributed infrastructure

S. Tangaro[1*], N. Amoroso[1,2], M. Antonacci[2], M. Boccardi[3], M. Bocchetta[3,4], A. Chincarini[6], D. Diacono[1], G. Donvito[1], R. Errico[2,7], G. B. Frisoni[3,5], T. Maggipinto[1,2], A. Monaco[1], F. Sensi[6], A. Tateo[2], R. Bellotti[1,2], for the Alzheimer's Disease Neuroimaging Initiative[a]

[1] INFN, Sezione di Bari, Bari, Italy. [2]Università degli studi di Bari, Dipartimento di Fisica, Bari, Italy. [3]LENITEM Laboratory of Epidemiology, Neuroimaging, and Telemedicine - IRCCS Centro S. Giovanni di Dio - FBF, Brescia, Italy. [4]Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy. [5]Memory Clinic and Laboratoire de Neuroimagerie du Vieillissement (LANVIE), University Hospitals and University of Geneva, Geneva, Switzerland. [6]INFN, Sezione di Genova, Genova, Italy. [7]Università degli studi di Genova, Italy.
*Corresponding author. E-mail: sonia.tangaro@ba.infn. Address: Sezione INFN di Bari, Via Orabona 4, 70125, Bari, Italy

Abstract

Medical image computing raises new challenges due to the scale and the complexity of the required analyses. Medical image databases are currently available to supply clinical diagnosis. For instance, it is possible to provide diagnostic information based on an imaging biomarker comparing a single case to the reference group (controls or patients with disease). At the same time many sophisticated and computationally intensive algorithms have been implemented to extract useful information from medical images. Many applications would take great advantage by using scientific workflow technology due to its design, rapid implementation and reuse. However this technology requires a distributed computing infrastructure (such as Grid or Cloud) to be executed efficiently. One of the most used workflow manager for medical image processing is the LONI pipeline (LP), a graphical workbench developed by the Laboratory of Neuro Imaging (http://pipeline.loni.usc.edu).

In this article we present a general approach to submit and monitor workflows on distributed infrastructures using LONI Pipeline, including European Grid Infrastructure (EGI) and Torque-based batch farm.

In this paper we implemented a complete segmentation pipeline in brain magnetic resonance imaging (MRI). It requires time- consuming and data-intensive processing and for which reducing the computing time is crucial to meet clinical practice constraints. The developed approach is based on web services and can be used for any medical imaging application.

I. INTRODUCTION

Neuroscience is an exponentially growing research field, as shown by the recent worldwide funding initiatives, such as the Human Brain Project[1] and the Human Connectome Project[2]. As a consequence, the availability of huge volumes of data driven by neuroscience researches poses two fundamental problems. On the one hand, as also stated by the previous mentioned research projects, this data explosion cannot be fully exploited because of the lack of effective strategies to integrate data processing and management. On the other hand, new data mining techniques and sophisticated analysis algorithms, which should deal now with big data-sets, require the use of distributed infrastructures and dynamic strategies, to deal with the computational burdens and the run-time requirements [1][2][3]. This is particularly true for neuroimaging which naturally involves high dimensional data. For instance, brain magnetic resonance imaging (MRI) plays a fundamental role for detection of neurodegeneration such as the Alzheimer's disease (AD). In particular, the hippocampal volume has been recognized as a diagnostic marker by the revised diagnostic criteria for AD [4]. The manual segmentation on MRI is widely considered the gold signature for the hippocampal volume, however automated segmentation tools are mandatory because manual segmentation is a time-consuming task. It is possible to provide diagnostic information based on the hippocampal atrophy comparing a single case with the reference group (controls or patients with disease). The implementation of user-friendly tools allowing the community to share and access the biomarker information is essential to improve the diagnostic accuracy. Feasibility studies demonstrated the potentiality of Grid tools for medical applications [5][6][7][8].

Nevertheless, distributed computing approaches needs to overcome several hurdles to spread into a community, those of neuroscientists, which finds many difficulties to set up advanced experiments, lacking of computational training and skills. This need of simplicity is, for example, the main goal of important international initiatives such as neuGRID [9][10]. This paradigm of simplicity, but also dynamic efficiency is nowadays represented by workflows technologies [11]. Workflow technologies are emerging as the dominant approach to coordinate groups of distributed services; in particular this is true for Grid computing services. The basic philosophy of this approach is that if a client (another service or an end user) makes a call to a remote server,

it should not to be related with the inner protocols (e.g. its programming language) to take advantage of its functionalities. This is the approach pursued by workflows. The main idea is that each service is independent from the others. This will offer a great degree of flexibility and scalability for evolving applications. Although the concept of service-oriented architectures is not new, this paradigm has been widely adopted through the Web services approach, which makes use of a suite of standards such as eXtensible Markup Language (XML), Web Services Description Language (WSDL) and Simple Object Access Protocol (SOAP) to facilitate service interoperability [12].

One of the most used workflow manager for medical image processing is the LONI pipeline (LP) [13], a graphical workbench developed by the Laboratory of Neuro Imaging[3] with the goal to manage and execute neuroimaging processing algorithms. The LP is a simple and efficient computing solution to problems of organization, handling and storage of intermediate data as well as for processing data and performing modular analysis. However several requirements must be fulfilled to run the LP environment on distributed infrastructures. In particular the user has to be sure that the available computing infrastructure is compatible with LP and this can represent a strong constraint, because different open source resource managers can be used and are often preferred. The framework proposed in this work allows the users to exploit the local resources available through a batch queue system, e.g. Portable Batch System (PBS), and the remote resources provided by the Grid.

As a case study a fully automatic pipeline for hippocampal segmentation is presented here; this choice is driven by two considerations, the significant role this problem plays in clinical research [14] and the computational burden it requires, which allows to consider it as a good candidate for benchmarking.

II. THE COMPUTATIONAL INFRASTRUCTURE

The Data Center *Bari Computer Center for Science*[4] *(BC^2S)* is a computing infrastructure realized to tackle different use cases from different research teams involved in, for instance, astroparticle physics, nuclear physics, medical physics, statistics, bioinformatics, theoretical physics. The available computing nodes are able to provide up to 5000 CPU cores and about 1.8 PByte of

storage. The operating system used by the Data Center is Linux, Scientific Linux[5] (SL) 5 deployed by Fermilab[6] and CERN[7] and various other laboratories and universities around the world with the main purpose to have a common install base for the various experiments. Debian distribution is available too. The storage system is based on Lustre[8], a distributed file-system, that is able to improve performances in reading data and to simplify the users activities through the adoption of just one distributed file system providing a standard posix interface. Torque provides the batch system capabilities and Maui manage the job execution.

Concerning the use of the Grid resources, in our data analysis the jobs were submitted to Grid through the Workload Management System (WMS) service offered to the users of the *biomed* Virtual Organization. This allows to use about two hundreds of Computing Elements and the great abundance of resources, this is important in order to the reduce of the average job response time. However it has some drawbacks: the execution environment is not completely controlled due to the heterogeneity of the resources (pre-installed software, supported data transfer protocols, etc.). Therefore we conducted preliminary tests aimed at assessing the main adverse events that could significantly lower the job success rate and impact on the overall performances. The system configuration (e.g. black list of Grid sites that do not satisfy specific software requirements, maximum number of retries in case of job failure, connection and upload/download timeouts, etc.) was tuned on the basis of the tests results in order to maximize the job success rate and the performances.

Fallback and recovery solutions were implemented for the storage operations. Input data reading and output data writing are very critical phases, being the major job failure factors. In this regard, multiple data transfer protocols (srm, gridftp, http) and multiple storage elements were used in our implementation. It is worthwhile to note that the described tuning can be done dynamically: the system is able to take charge of these changes without stopping or restarting the job submission operations.

III. METHODS

The proposed approach has been designed to take advantage of distributed computing resources

available both in a local batch farm and in a Grid-based infrastructure. Depending on the requirements, in our framework the user is allowed to configure the execution environment for each task in the LP workflow: data-intensive jobs should be run on the local batch farm in order to reduce the latency caused by data transfers, whereas compute-intensive jobs can be submitted to the Grid, composed of about 300 sites geographically distributed around the world.

A. Data Management and Security

Data management is a critical issue when dealing with distributed systems. Usually, applications need computing power but in most cases the requirements for data sharing, storage and transfer are also compelling. This is particularly true for medical image analysis that requires efficient and reliable file storage and transfer. Moreover, non technical issues concerning privacy problems tend to minimize data replication and transfers.

These issues have been addressed from two different perspectives. From the algorithmic point of view, all the information needed by our segmentation algorithm, meanly based on voxels intensity, is extracted from the input data loaded by the end user and then anonymously stored or transferred to the processing nodes. On the other side, for this specific application no clinical data (about the patient's sex, age, etc.) are retained or used and therefore we are not directly concerned about privacy policies management. Yet, in future we will handle these sensitive data: for this reason, we paid particular attention to protecting data, using security protocols and addressing known vulnerabilities.

A strong authentication protocol, based on X.509 certificate, is required for all the operations performed on the Grid infrastructure (e.g. job submission and data management). This provides a good level of data protection. All the jobs submitted to the local batch farm are executed using the standard authentication mechanism based on Unix rules and permissions. To access the data to be analyzed we have exploited the standard Data Management tool offered by the EGI Grid infrastructure that implements a fast and reliable data transfer service.

We are using Storage Elements based on Lustre file-system and StoRM implementing the standard interfaces of the Site Recovery Manager (SRM) layer [15]. This configuration is particularly suitable since allows users to access the data stored on the storage elements from

both the local farm (using Lustre posix-compliant interfaces) and from the Grid infrastructure (using SRM interfaces).

Furthermore, VOMS (Virtual Organization Membership Service) is used, giving to people different privileges according to their roles and group membership. However, despite these high level features, the site admin of the grid site where the data are, could still access to data. In some cases, this is still a problem. So we tested a procedure that erases all the sensible data from the images that we process. So the proposed procedure is:

. we firstly erase the patient data (name, surname, etc)

. than we store on the grid

In this way even if a site administrator is able to read the files there are not way to retrieve sensible informations about patients.

B. Loni pipeline on Distributed Computing infrastructures

Web services provide a solution to a complex problem, i.e. the orchestration of a group of services to achieve a shared task or goal. Workflows are the glue for joining together distributed services, which are owned and maintained by different organizations.

The success of contemporary computational neuroscience depends on the large amounts of heterogeneous data, powerful computational resources and dynamic web-services. To provide an extensible framework for interoperability of these resources in neuroimaging LP exploits a decentralized infrastructure, where data, tools and services are linked via an external inter-resource mediating layer whose backbone schema is formed by standard XML. The pipeline environment does not require an application-programming interface: its graphical user interface has been originally programmed as a lightweight Java 1.3 environment.

The LONI Pipeline is widely used by the scientific community since it has proved to be a convenient and powerful tool. In particular XML resource description facilitates the integration of disparate resources and provides a natural and comprehensive mechanism to support data provenance. It also enables the broad dissemination of resource metadata descriptions via web-services and the constructive utilization of multidisciplinary expertise by experts, novice users and trainees.

We have developed a general approach to submit and monitor LP workflows on distributed infrastructures. The proposed framework is based on a meta-scheduler, the Job Submission Tool (JST) [16], that is able to submit jobs to different computing architectures, exposing to the end users only a simple Web Service interface based on the Representational State Transfer (REST) protocol (Figure 1).

The JST architecture consists of two main components: 1) the Front-end, that provides a set of web service methods to submit jobs, monitor them and retrieve their output; 2) the Back-end, that takes care of the job execution or submission to the Torque/BPS batch system or to the EGI Grid infrastructure. The Front-end hides the complexity of the underlying layer, developed in Java and based on Apache Tomcat (it exploits MySQL RDBMS).

Since LP supports SOAP-based web services, we needed to add a SOAP layer to the JST that originally exposed only REST interfaces.

Each workflow module is submitted and executed on the infrastructure that better fits the requirements in terms of the computational time needed for the job execution and the input data size: for example, short jobs that need to access large amounts of data will be executed on the local batch farm, as this will increase the available bandwidth to read the input data. Each module executes bash scripts that transparently handle the submission and monitoring of farm/grid jobs that, in turn, execute the application code. Our implementation of the LONI modules encapsulates the management of the asynchronous communication with the JST Web Service to submit and monitor the jobs, as shown Figure 2.

IV. RESULTS

In this study we have developed and implemented on the described distributed computing infrastructure a fully automated pattern recognition system for accurate and reproducible segmentation of the hippocampus on structural MRI. The system has been validated on a mixed cohort of 56 T1-weighted structural brain MR images performed at the Città di Brescia Hospital (Brescia, Italy) with a 1.0-Tesla Philips Gyroscan (Philips, Eindhoven, The Netherlands), using a standardized protocol (gradient echo 3D technique, TR = 20 ms, TE = 5 ms, flip angle = 30°, field of view = 220 mm, acquisition matrix of 256 × 256, and slice thickness of 1.3 mm). Images were

then reconstructed with a voxel size of 1.00x1.50x1.00 mm^3.

The algorithm consists of three processing levels:

1- Registration of all brains to the Montreal Neurological Institute (MNI) standard template (ICBM152) using 12- parameter affine-registration to capture the global shape of the hippocampus, according to the methods described in [17][18][19].

2- Feature extraction: all voxels included in the previously selected volume were characterized by 315 features computed from local information such as image intensity, voxel positions, Haar-like filters, and selected Haralick features, according to the methods described in [20][21]. A feature file has typical dimensions of 150 MB this is an important aspect both for the storage and the upload on worker nodes.

3- Voxel classification: a Random Forest algorithm was used to classify voxels as belonging or not belonging to the hippocampus. In order to improve the classification performance, a training subset was selected through the use of the Pearson's correlation coefficient between the test image and the training dataset (active learning) according to the methods described in [19].

In the subsequent validation phase, using a leave-one-out approach, the results were compared with images hand- labeled by an expert of neuroanatomy.

The proposed method showed reliable results comparing automatic and manual segmentation (Dice similarity index = 0.81 ± 0.03, recall = 0.85 ± 0.05, relative overlap = 0.68 ± 0.05) at state of art demonstrating its plausibility for accurate, robust and reliable segmentation of hippocampus. The proposed classification approach may be suitable for large-scale research studies where the hippocampal volume is an important biomarker: in the first instance for AD but potentially also for other brain disorders in which the hippocampus plays a relevant pathogenetic role.

The most relevant steps of the analysis and the main parameters are represented in Figure 3.

The implementation of the segmentation pipeline into an LP workflow requires the following operations: each workflow module is implemented as a bash script that transparently handles the submission and monitoring of farm/grid jobs through the JST Web Services. The job retrieves the application executable and the input data that have to be accessible using standard protocols like http(s), gftp, ftp, executes the application code and stores the output data in configurable

storage elements. The MATLAB code has been compiled with the Matlab Compiler Runtime (MCR) that allows sharing Matlab programs as standalone applications. The Matlab Compiler Runtime (MCR) is freely provided from MathWorks to running Matlab code as a standalone application in distributed framework. The -nojvm compile option has been used to disable the java virtual machine in order to speed-up our non-graphical applications.

The training of the active classification is the least computational intensive module, but the most data intensive since it needs to access to all the images composing the database. Therefore, it is candidate to run on the local farm where the input data can be read without significant latency. Moreover, this module can rely on the Matlab software pre- installed on the local farm; the availability of the same software environment cannot be guaranteed on all the Grid worker nodes where the other modules can run. Therefore, in the case of the Grid execution, the wrapper script is in charge to download the MCR package from a configured repository and to install it before executing the application.

A single segmentation run needs in average 80 minutes on our local workstation with AMD Phenom 9950 Quad-Core Processor and 2 GB RAM. The setup of the reference workstation with 2 GB Ram allows to solve this 3D image processing problems as showed in figure 4 where the distribution of RAM used from all pipeline module on 400 segmentation is represented. Also, in figure 5 the RAM distribution averagely employed for the overall processing pipeline is shown. In figure 5 we observe that the RAM used is less than 2 GB. This result is comparable with manual segmentation. Since medical databases consist in average of hundreds of images, their sequential processing on a local workstation cannot provide results in an acceptable time (e.g. for screening programs). The advantages of using grids can be better appreciated when processing large amount of data in parallel. Therefore we have performed a proof-of-concept test using an independent database of 400 images obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).

Using our distributed framework we completed 400 hippocampal segmentation in about 7 hours, whereas on our local workstation it would require a processing time of about 22 days.

We also investigated the overall running times varying with the size of the image database. In figure 6 the curves for the Grid, the local farm and the workstation are plotted. The Grid 'inertia'

due to job submission and match making operations performed by the WMS translates into an initial latency that can be further dampened by increasing the number of parallel workflow executions. After this initial step the advantages of the Grid execution are evident since we obtained the 90% of the segmentations after less than 7 hours. Also figure 6 shows the results collected executing 400 parallel workflows (with about 2400 submitted jobs in total): the run-time reduction with the Grid implementation allowed to produce results in a reasonable time with respect to the application execution as a sequential process on limited resources.

Finally, we present a quantitative assessment of the segmentation performance. As no ground truth segmentation is provided for the ADNI database, we estimated the segmentation accuracy by using the hippocampal volumes to detect the Alzheimer disease presence. In fact, the 400 scans we used to evaluate the feasibility of the proposed framework consisted of 94 Alzheimer disease (AD), 217 mild cognitivie impairment (MCI) and 89 healthy control (CTRL) subjects. The following Figure 6 shows the boxplot of left hippocampal volumes, as reterieved by our algorithm.

V. DISCUSSION AND CONCLUSION

The proposed algorithm compares well with other state-of-the- art methodologies. It is well known that a sound comparison is very difficult to perform, several differences can affect the segmentation results, including data acquisition, scan or segmentation protocol differences[1]. Accordingly, the hippocampal measurements obtained with this processing pipeline were used as structural features (with different methodologies [22][23][24]) to predict Alzheimer in several international challenges focused on the early prediction of Alzheimer's disease, such as the Alzheimer's Disease Big Data DREAM Challenge 1 [9] [23]. This challenge in particular generated a significant interest. More than 500 participants from around the world registered to participate and at the end of the challenge 1,157 and 1,478 submissions were made for the two problems of the challenge. Thirty-two teams submitted final results and were included in the final community phase for the publication of the results. As a remark, it should be kept in mind that medical imaging databases, especially for multi-center studies, require an intensive processing. In particular, further developments of the proposed methodology should include more sophisticated image enhancement techniques [25][26].

The workflow manager technology allows the end user to easily integrate these improvements, run the already available workflows and eventually develop new analysis pipelines. Moreover, the usage of a distributed infrastructure offers a huge amount of computational power, enabling the parallelization of the pipelines processing. Time-consuming and data-intensive applications, such as the hippocampal segmentation, can be executed in reasonable times in order to meet clinical practice constraints.

In this study the LONI Pipeline processing environment was used, a user-friendly and efficient workflow manager for complex analysis pipeline. In particular, the presented method allows to use the LP with Torque batch system and at the same time to use grids: the implemented framework, in fact, is based on a meta-scheduler exposing RESTful and SOAP interfaces that can enable basically any workflow manager to exploit complex computational infrastructure in an easy and is highly configurable way.

As a case study we adopted this framework to efficiently and robustly segment the hippocampus. The hippocampal volume is a well known marker of neurodegenerative processes and consequently used, for example, for the Alzheimer disease diagnosis. This processing pipeline has been successfully employed in several international challenges.

Finally, we developed a web service/cloud solution to allow users to exploit an already available workflow changing only the configuration parameters and the input files by the link https://recasgateway.ba.infn.it. In this way it is possible for the users to execute some analysis without the need of specific expertise in Grid/cloud management.

REFERENCES

[1] Bron E. E. et al. Standardized evaluation of methods for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. NEUROIMAGE 111 (2015) pp. 562–579.

[2] Tangaro S et al. Feature Selection based on Machine Learning in MRIs for Hippocampal Segmentation. Computational and Mathematical Methods in Medicine (2015), ISSN: 1748-670X

[3] Amoroso N et al. Hippocampal unified multi-atlas network (HUMAN): protocol and scale validation of a novel segmentation tool. Physics in medicine and biology, vol. 60 (2015), pp. 8851-8867

[4] Frisoni GB Structural imaging in the clinical diagnosis of alzheimer's disease: problems and tools. Journal of Neurology, Neurosurgery & Psychiatry; 70(6) (2001), pp. 711–8.

[5] Bellotti R et al. Distributed medical images analysis on a Grid infrastructure. Future Generation Computer Systems; 23(3) (2001) pp. 475 – 84.

[6] Tangaro S et al. MAGIC-5: an Italian mammographic database of digitised images for research. La radiologia medica, 113(4), (2008) pp. 477-485.

[7] Cerello P et al. GPCALMA: a grid-based tool for mammographic screening. Methods of Information in Medicine (2005) 44(2): pp. 244–8.

[8] Bellotti R et al. The Magic-5 Project: Medical Applications on a Grid Infrastructure Connection. IEEE Nuclear Science Symposium Conference Record (2004), 3 pp. 1902-6.

[9] Frisoni G B et al. Virtual imaging laboratories for marker discovery in neurodegenerative diseases. Nature Reviews Neurology (2011) 7 pp. 429-38.

[10] Redolfi A et al. Grid infrastructures for computational neuroscience: the neuGRID example. Future Neurology (2009) 4: pp. 703-22.

[11] Dinov I et al. Efficient, Distributed and Interactive Neuroimaging Data Analysis Using the LONI Pipeline. Front Neuroinformatics (2009) 3: 22.

[12] Barker A and Hemert J Scientific Workflow: A Survey and Research Directions. In Parallel Processing and Applied Mathematics (2008) 4967: pp. 746–53.

[13] Rex DE et al. The LONI Pipeline Processing Environment. NeuroImage (2003) 19(3): pp. 1033–48.

[14] Frisoni G B, The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology

6 (2) (2010): pp. 67-77

[15] Donvito G et al. Interactive Access and optimization of a CMS computing farm. Nuclear Physics B, Proceedings supplements (2011) 215(1): pp. 82-4.

[16] Donvito G et al. The bioVeL Project: Robust phylogenetic workflows running on the GRID. 2013. http://pos.sissa.it/archive/conferences/162/029/EGICF12- EMITC2_029.pdf

[17] Calvini P et al. Automatic analysis of medial temporal lobe atrophy from structural MRIs for the early assessment of Alzheimer disease. Medical Physics (2009) 36(8): pp. 3737–47.

[18] Inglese P et al. Multiple RF classifier for the hippocampus segmentation: method and validation on EADC-ADNI Harmonized Hippocampal Protocol. Physica Medica 31 (8) (2015), pp. 1085-1091, ISSN: 1724- 191X

[19] Tangaro S et al. Automated voxel-by-voxel tissue classification for hippocampal segmentation: Methods and Validation. Physica Medica: European Journal of Medical Physics, vol. 30 (8) (2014), pp. 878–887.

[20] Maglietta R et al. Random forest classification for hippocampal segmentation in 3D MR images. Proceeding on the 12th International Conference on Machine Learning and Applications ICMLA'13 (2013).

[21] Tangaro S et al. Active Learning Machines for Automatic Segmentation of Hippocampus in MRI, proceeding on Industrial Conference on Data Mining, ICDM (2013).

[22] Tangaro S et al. MIND-BA: fully automated method for Computer- Aided Diagnosis of Dementia based on structural MRI data. Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data (2014), pp. 119–128.

[23] Allen G I et al. 2016. Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease. Alzheimer's & Dementia (in press).

[24] Amoroso N et al. PRISMA-CAD: Fully automated method for Computer-Aided Diagnosis of Dementia based on structural MRI data. Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data (2014), pp. 16–23.

[25] Greco A et al. Improving automatic segmentation of tissue-targeted nanoparticles on echographic images. Proc Medical Measurements and Applications Proceedings (MeMeA), IEEE International Workshop (2011), pp. 488-491

[26] De Caro L et al. Deconvolution by finite-size-source effects of x-ray phase-contrast images. Medical physics, 38(4) (2011), pp. 1951-1961.

FOOTNOTES:

1 www.humanbrainproject.eu

2 www.humanconnectomeproject.org

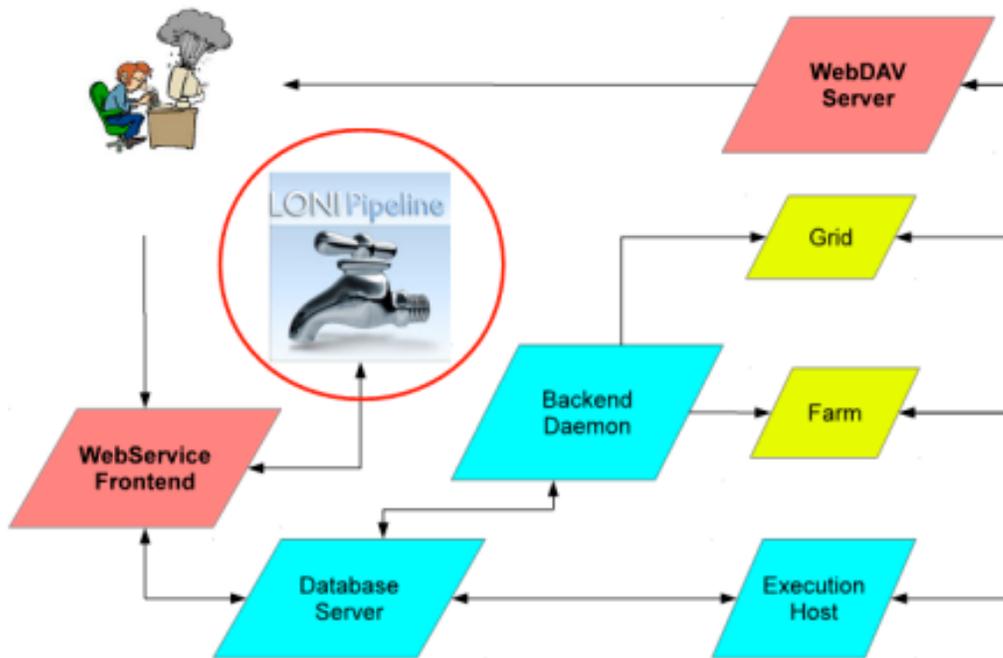3 http://pipeline.loni.usc.edu

4 http://recas.ba.infn.it/recas1/index.php/recas-ba/datacenter

5 https://www.scientificlinux.org/

6 http://www.fnal.gov/

7 http://home.web.cern.ch/

8 http://wiki.lustre.org

9 https://www.synapse.org/#!Synapse:syn2290704/wiki/60828

FIGURES

Fig. 1: The workflow implementation. In reddish diamonds there are the input/ouput modules, the backend analysis modules are represented by turquoise diamonds. To emphasize the possibility to dynamically choose the local farm or the Grid infrastructures, these backend modules are shown in yellow.
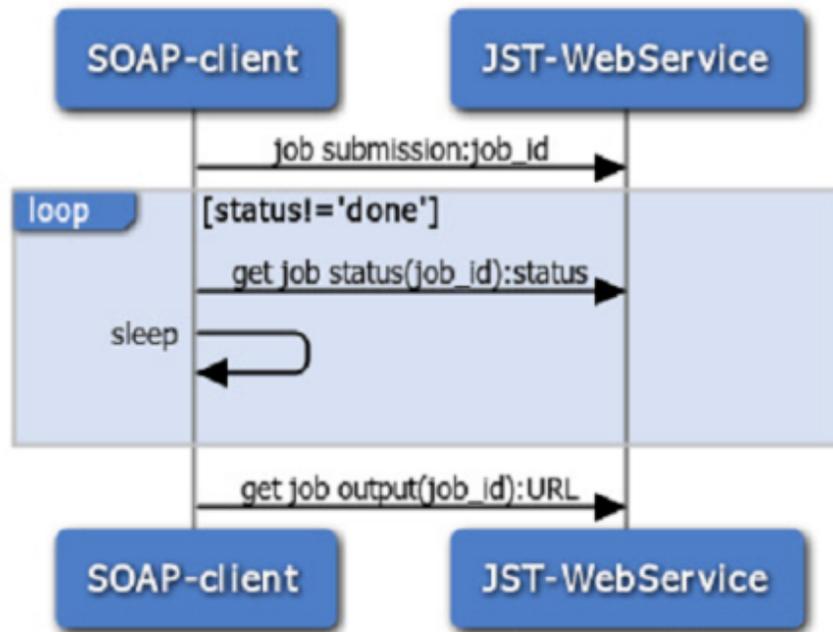
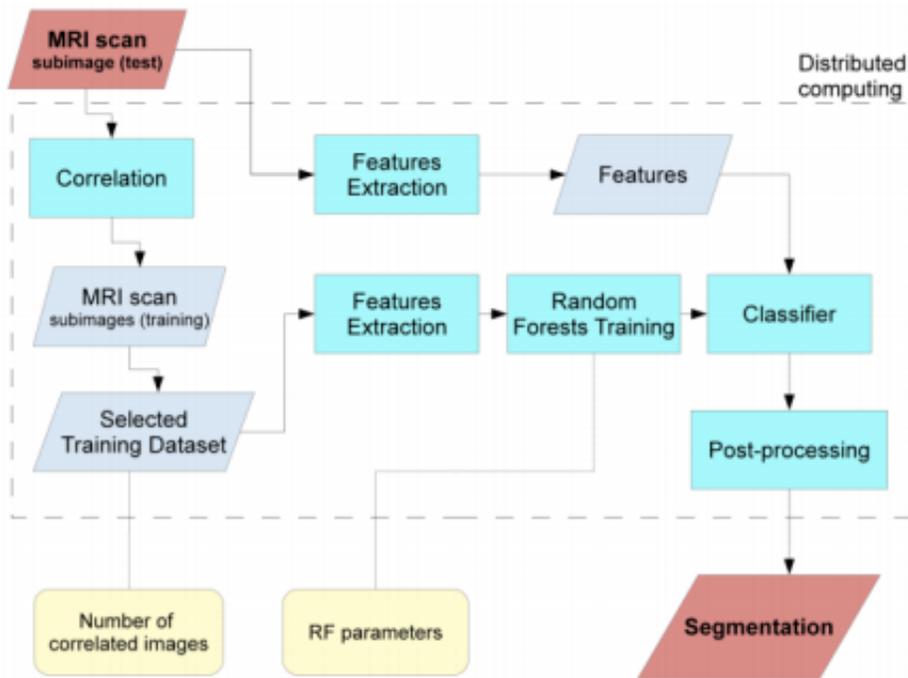Fig. 2: Web service call sequence implemented in the LP modules.

Figure 3: A representation of the hippocampus segmentation algorithm, parameters and critical aspects are shown in round boxes, processing steps in rectangles and files in diamonds. The distributed computing is enclosed by a dotted line while the end user interface is shown in reddish diamonds.
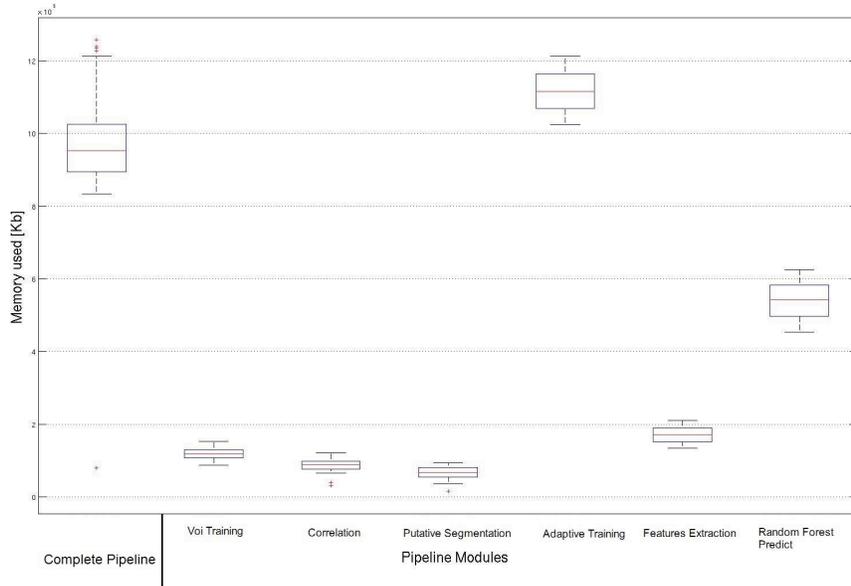
Figure 4: Distribution of RAM used on 400 segmentations for both complete pipeline and for each modules of the pipeline (all less than 2 GB). Boxes have lines at the lower quartile, median, and upper quartile values, with whiskers extending to 1.5 times the inter-quartile range. Outliers are indicated by a plus sign.
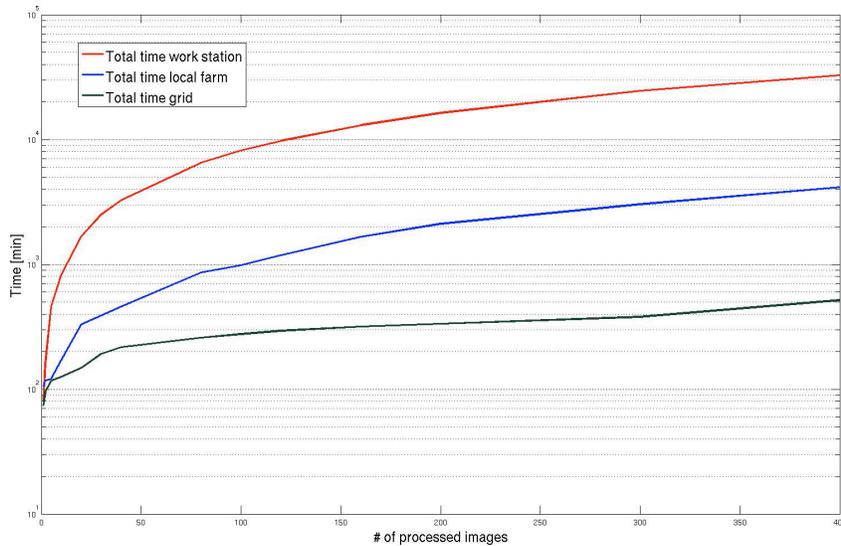

Figure 5: This figure shows the overall running times versus the size of the database of images for the Grid, the local farm and the workstation.
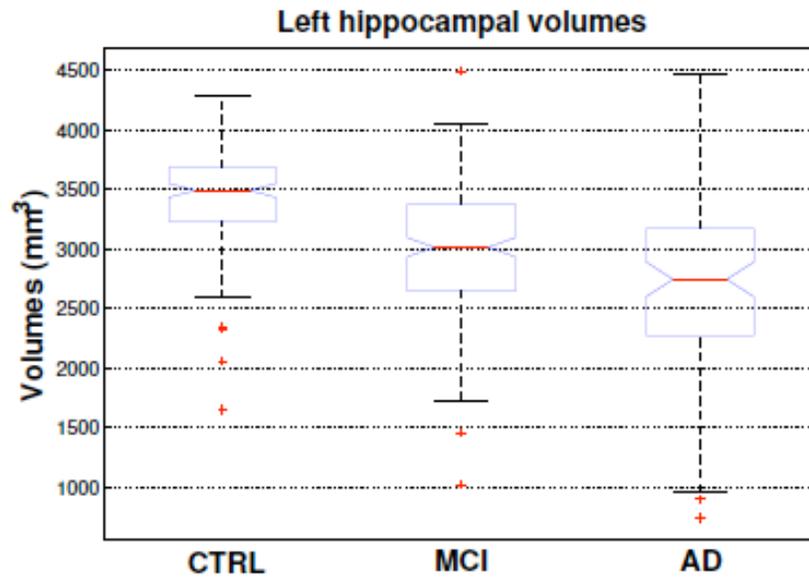
Figure 6: This figure shows how the hippocampal volumes allow to separate the three populations.