INVITATION FOR COMPETITIVE BID

ROMEO

Medical Imaging Cluster Analysis Tool

INDEX

Introduction

Cluster analysis (or clustering, also known as exploratory data analysis) is the task of grouping a set of homogeneous objects in a group of data (clusters). It is a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

In image segmentation, clustering can be used to divide a digital image into distinct regions for border detection or object recognition: each of the pixels in a region are similar with respect to some characteristic or computed property, such as colour, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse.

The clinical value of medical images is also to show different tissues and organs of the body and distinguish normal tissues from abnormal ones, but manual organ segmentation is a tedious process and prone to considerable subjective variability. Cluster analysis plays a crucial role in many medical-imaging applications, by automating or facilitating the delineation of anatomical structures and other regions of interest. In particular, the procedure can accelerate the processing speed and accuracy of the diagnosis, allowing to locate pathological tissues and measure tissue volumes, and it has thus great potential for prevention, prognosis, and treatment.

Applications of clustering algorithms in biomedical research are ubiquitous, with typical examples including gene expression data analysis, genomic sequence analysis, biomedical document mining, ultrasounds, and Magnetic Resonance Imaging (MRI)/Positron Emission Tomography (PET) image analysis.

Different clustering algorithms exist and most of them are freely available online, as source code or integrated in various software, but at the moment most of the programs are

developed for the analysis and visualization of microarray and sequence data, or on simple lists of data (thus 1-dimensional or 2-dimensional data).

At the moment, only PMOD (a commercial PET/MR quantification software) implements cluster analysis on biomedical data, but only as secondary tool within the complete pipeline of analysis of the data. In fact only a particular type of clustering (developed by Orsay research centre mainly for PET data) is integrated within the software, which is dedicated to PET/MR data analysis, thus not considering any other biomedical image. Moreover the software is expensive and, although very powerful, quite complex to learn and to use.

Therefore is clear the necessity of a simple yet complete tool which integrates in the same environment the most common clustering algorithms for the segmentation of medical images, but potentially applicable to all kinds of data of interest.

Hereafter, a schematic representation of the pipeline of analysis of a cluster analysis applied to biomedical image (Figure 1).
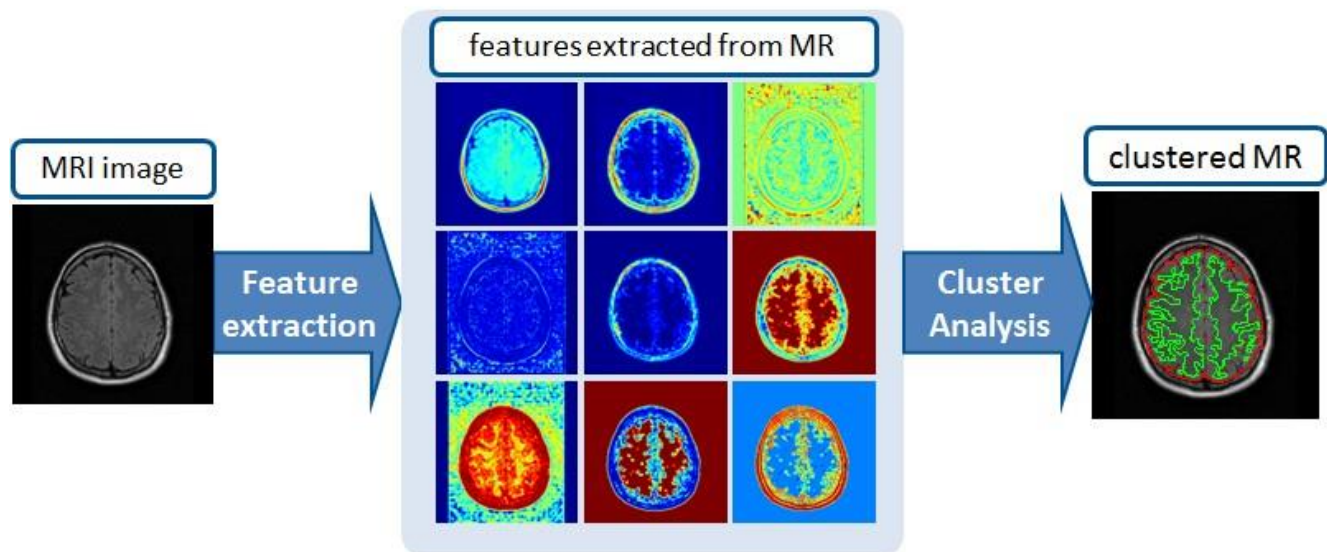

Figure 1: Feature extraction and cluster analysis of a MR image

From the input image (Magnetic Resonance (MR) image) various features of interest are extracted as first step. The cluster analysis is then applied on these features, thus obtaining the output clustered MR image (in this case, two clusters are identified – green and red solid lines).

The Project

Objectives of the bid

The invitation to bid (hereinafter referred to as ITB) is designed to solicit proposals from qualified vendors who will be responsible **to develop a software for the cluster analysis of biomedical images**. The main aim is to offer to the international scientific community a simple but complete and flexible tool to apply clustering algorithms to medical images. This tool has to be as much user friendly as possible in order to allow its utilization by people with limited computer knowledge (e.g.: doctors, biologists, chemists, clinicians,…) but with high interest in cluster analysis.

Project Support

This ITB is supported by the Bioengineering group of the Department of Information Engineering (University of Padova) and by the Institute of Psychiatry, King's College, London. The purchaser group (hereinafter referred to as COM) reserves the right to accept or reject any or all proposals received as a result of the ITB in whole or in part, to negotiate with all qualified bidder, and/or to cancel in whole or in part the ITB if in the best interest of the COM.

Mandatory Requirements

Mandatory requirements are explicitly indicated in the ITB. All the mandatory requirements must be fully satisfied.

Optional Requirements

Optional requirements are explicitly indicated in the ITB. The realization of the optional requirements does not undermine the eligibility of the proposals but it will be considered as evaluation criteria.

Change of requirements

Changes in the mandatory requirements are allowed only in the directions of clear improvements of the COM requests. The COM reserves the right to change the mandatory requirements during all the bid process, i.e. before the proposal submission as well as the during the realization phase.

Documentation

The proposer must provide an operating manual and all the technical documentation necessary for the software utilization. The use of English language is required. All such documents must be included in the delivery of the product.

Warranty and Maintenance

The proposer must be capable of guarantying the correctness of its application. The proposer shall host development, test, and quality assurance environments during and through the project life cycle. This will include system application modification testing as well as conversion data testing.

The proposer must be capable of providing full support for its application. This will include the corrections for the not conformities of the software coming out from the test phase.

The test proceeding will be indicated by the proposer and it will be considered as criteria for the proposal evaluation. The COM reserves the right to reject the test proceeding indicated by the proposer and suggests alternative testing solutions.

The proposer must be capable of ensuring the future expandability of its application. The characteristics of expandability of the software will be considered as mandatory criteria for the proposal evaluation.

Discussion

The COM reserves the right to seek clarifications from bidder so as to assure a full understanding of the proposed activities and processes. Eligible bidder will be accorded fair and equal treatment with respect to any opportunity for discussions and revisions concerning their proposals.

Notification

During the bid period, all communication between the COM and a proposer shall be between a proposer's representative and the COM contacts.

COM contacts

Gaia Rizzo
Postdoc
Dept. of Information Engineering (DEI)
University of Padova
Via G. Gradenigo 6/A
35131 Padova, Italy
Phone: +39 049 8277758
email: gaia.rizzo@gmail.com

Enea Poletti
Postdoc
Dept. of Information Engineering (DEI)
University of Padova
Via G. Gradenigo 6/A
35131 Padova, Italy
Phone: +39 049 8277758
email: enea.poletti@dei.unipd.it

Mattia Veronese
Research Associate
Institute of Psychiatry
King's college London
DeCrespigny Park, Denmark Hill
SE5 8AF London, UK
Phone: +44(0) 203 228 3048
email: mattiaveronese@gmail.com

Marco Castellaro
PhD student
Dept. of Information Engineering (DEI)
University of Padova
Via G. Gradenigo 6/A
35131 Padova, Italy
Phone: +39 049 8277640
email: castell2@dei.unipd.it

Errors and omissions

The proposer shall not take advantage of any errors and/or omissions in this ITB. The proposer must promptly notify the COM of any errors and/or omissions that are discovered.

Terms and Conditions

Proposals shall conform to all instructions, conditions, and requirements included in the ITB. Prospective bidder are expected to carefully examine all documentation, schedules and requirements stipulated in this ITB, and respond to each requirement in the format prescribed. The proposals will first be examined to determine if all mandatory requirements listed below have been addressed to warrant further evaluation. Proposals not meeting mandatory requirements will be excluded from further evaluation.

This ITB is written conforming the laws for the management of public procurements. The COM referred to that laws for all the missing details concerning the ITB.

No commercial Purposes

The aim of the ITB is NOT to create a software with commercial purposes. Instead proposals will represent prototypal examples which will demonstrate the feasibility of a PET quantification software based on spectral analysis algorithms.

SOFTWARE

Cluster analysis: a short overview

By definition, **cluster analysis** or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). In Figure 2: The result of a cluster analysis is reported an example of result of cluster analysis.
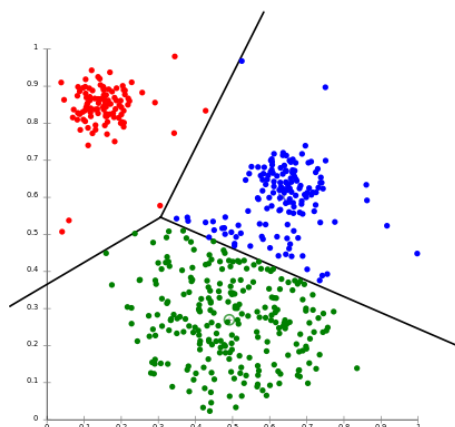
Figure 2: The result of a cluster analysis

Classification methods can be divided in supervised or unsupervised approaches: while the first assigns a predefined class to each datum, the latter finds the classes directly from the data themselves. Clustering falls into the unsupervised approaches.

Clustering serves well to fulfil the objective of biomedical research to extract knowledge from huge amounts of data elements, and in the last decades many clustering algorithms have been developed and applied. According to Pubmed (the US National Library of Medicine National Institutes of Health Search database), there are 42,894 journals and conference

papers using the phrase "cluster analysis" within the paper title, abstract, and keywords in the subject area of Life and Health Sciences, as of June, 2013.

In the lights of above, since it is not feasible to develop a software including all the cluster algorithms available in literature, a key feature of the tool must be *the flexibility*: this will allow the user to integrate the software with new algorithms, based on her/his needs and on the characteristics of the data to be analysed.

General Aspects

Glossary

CLUSTERING ALGORITHM: function that applies a clustering technique to the data (both original image and extracted *Features*);

DATA: the image to be analyzed;

DATABASE: system to manage data archives (*Datasets*);

DATASET: archive of *Data*. It is organized in *Subjects*, each one with one image to analysed. For each dataset one or more protocols can be associated;

FEATURE EXTRACTOR: function that calculates a *Feature* of interest from the data;

FEATURE: information extracted from the image, not present originally;

MASK: a binary image which is used to limit the analysis to the image elements of interest (the *Data* elements whose corresponding *Mask* values are zero are discarded from the analysis);

PROTOCOL: combination of a set of *Feature Extractors* and one *Clustering Algorithm*. When it is applied to the *Dataset*, the execution of the *Protocol* must involve all the *Subjects*.

SUBJECT: a single instance of the *Dataset*, to whom the *Data* belong to.

The list of features and clustering algorithms related to the elaboration of biomedical data will be provided by the COM in a second moment. As explained in the introduction, it is not possible to integrate all the existing algorithms, therefore the list will be limited to some cases

of interest, based on the needs of the COM. It is possible (and suggested) to take advantage of the already developed solutions available online.

Each feature and algorithm will be described in detail and it will be up to the proposer their implementation and integration in the software. The eventual development of new features and algorithms will be up to the user.

The proposer is supposed to design and develop an application which allows the features extraction and the implementation of the clustering algorithms, according to good software engineering practices. Feature extraction and cluster analysis can be performed both independently and sequentially.

It is moreover required the implementation of a macro-recorder and a system for the management and storage of the performed analyses into a custom database.

Mandatory Requirements:

- The software has to be deployed as a stand-alone application;
- The software documentation and online data in the application MUST be written in English language;
- The software has to work on Windows, Linux and Mac platforms (both 32-bit and 64-bit);
- The software must provide a Graphical User Interface (GUI) as a front-end to the application;
- The GUI has to follow an intuitive scheme to be used by expert and not expert ICT users;
- **The software must be expandable both in terms of features and algorithms**: the user her/himself must be able to add an add-in for a new feature/algorithm that can be directly applied by the software.

Structure of the application

The software structure consists in four logical blocks:

1) Dataset generation
2) Protocol generation
3) Execution of the protocol
4) Results visualization

The software includes the management of a database of data and performed analyses.

1) Dataset generation

The user can create a dataset, adding a new subject, or manage an existing dataset, adding or deleting subjects (main field *Subjects*) or protocols (main field *Protocols*). Each subject has one mandatory and one optional field, i.e. *Data* and *Mask* respectively.
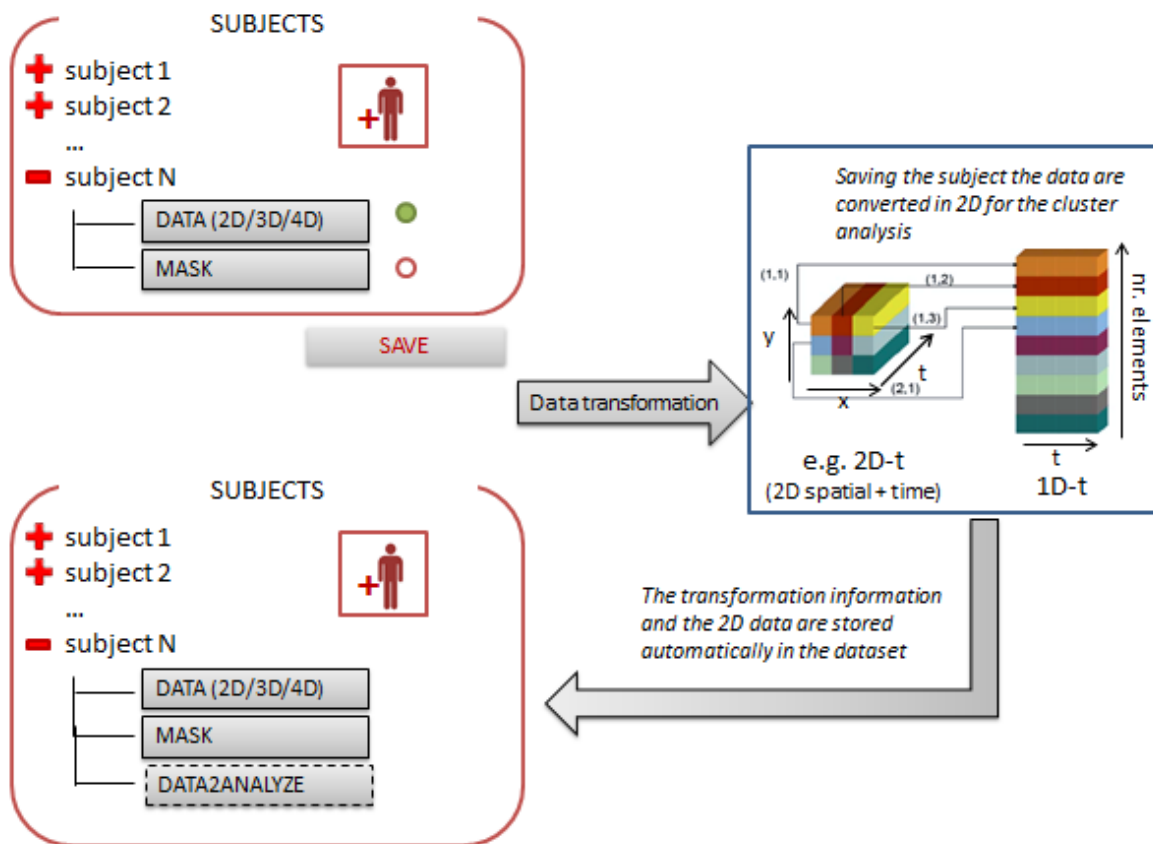

Figure 3: Dataset generation and data transformation

Mandatory requirements

- A protocol cannot be modified after its application
- The user can delete entirely a subject but cannot delete its fields separately
- It is not mandatory that ALL the subjects have a mask. In this case for the subjects without a mask, the analysis will be done on the whole image
- All the subjects of the dataset will share the same data format
- The user can modify only the optional field (*Mask*)
- The user is asked to select a destination folder where the software will create a folder for each subject to store the final results.

Data can be a picture, video, MR, PET, …, with various formats. The images can be time-independent or time-dependent:

Time-independent:

- ✓ 2D: the two dimensions correspond to x-axis and y-axis. It is a single slice, acquired once;
- ✓ 3D: the three dimensions correspond to x-axis, y-axis and z-axis. It is a volume of interest, i.e. multiple slices overlapped, acquired once;

Time-dependent:

- ✓ 2D-t two spatial dimensions (x-axis, y-axis), the 3rd dimension is the time, i.e. there are multiple slices, each acquired in one time frame;
- ✓ 3D-t: 3 spatial dimensions (x-axis, y-axis and z-axis), the 4th dimension is the time, i.e. there are multiple volumes of interest, each acquired in one time frame.

Mask must have the same format as *Data*.

When the dataset is generated, **the software automatically transforms the data** in an internal format (i.e. *Data2analyze*, a 1D vector, or a 2D matrix if time-dependent) for the

following analysis, and *Data2analyze* is stored in the database (Figure 3). The *Data2analyze* format has two purposes: it can be seen as a compression format to save the data (especially when a mask is available) and it represents a consistent format to apply the analysis to. It must be properly codified and maintained for all the subjects and datasets, as it is the key for the protocol library extendibility.

Moreover, in order to allow the change of the mask, *Data2analyze* must represent the original data (including those elements to be eliminated with the use of a mask). Such a way, even if the mask is modified, the software can use the original data directly.

2) Protocol generation

As second step, the user defines the protocol of the analysis. The list of features and algorithms will be given by the COM and must be expandable, i.e. the user can add new features/algorithms integrating new functions. Therefore the input/output scheme of the feature extractor functions must be always the same, in order to make the user able to write new functions.
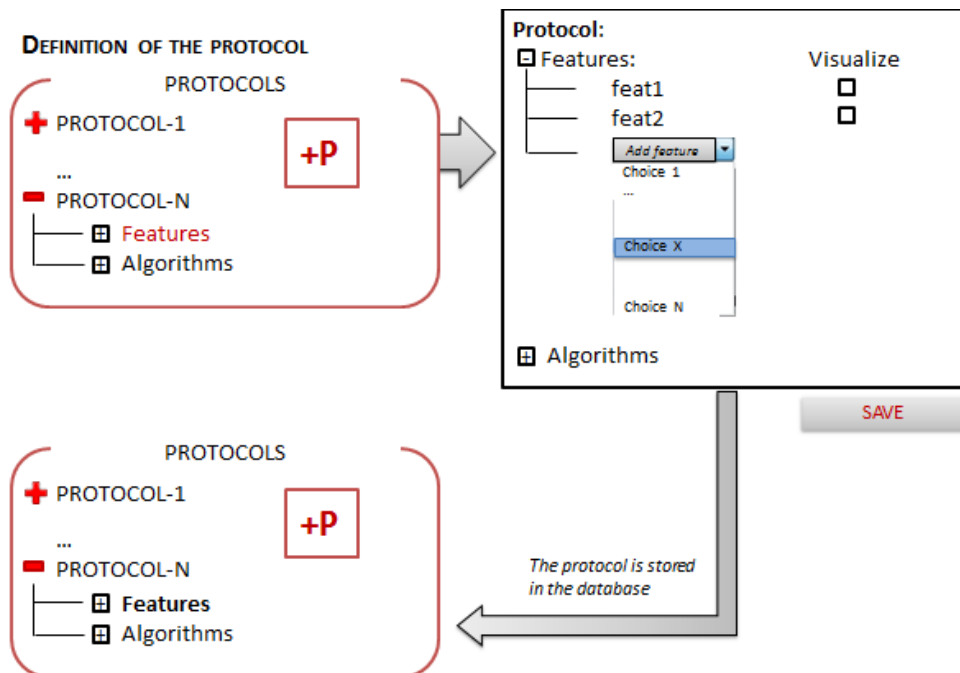

Figure 4: Protocol generation

Mandatory requirements

- Each dataset might have 1 or more protocols
- When a protocol is applied, it is executed for all the dataset subjects: it is not possible to have different protocols for different subjects within the same dataset.
- The features are selected by the user and there is no limit on the number of features that can be selected.
- If required, it is possible to visualize the features calculated in the original space (Figure 4).
- The protocol must be saved and stored in the database.

3) Execution of the protocol

The first part of the analysis consists in the features extraction: the software calculates the features selected in the protocol from the data for each subject of the dataset.

Secondarily, the software performs the cluster analysis applying the algorithms defined in the protocol on the feature matrix for each subject.

At the end, the software can move to results visualization or to a new protocol definition.

Mandatory requirements

- Run for each subject the selected protocol
- For each subject, the features extracted and the cluster results must be saved within the dataset
- The results must also be exported and saved as images with the same format of the input files
- It is necessary to provide to the user the real-time status of analysis (waiting bar,…)
- When the analysis of a subject is finished, the results must be saved even if the entire analysis is not completed

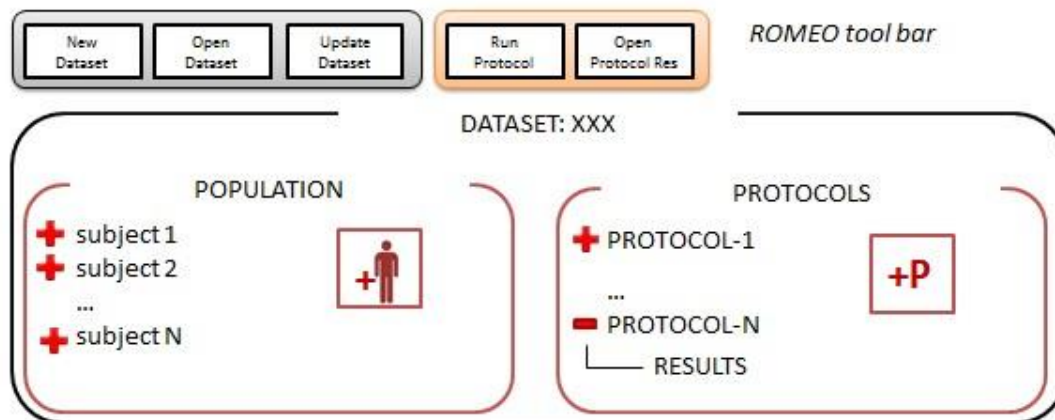An example of GUI for the execution of the protocol and the results visualization is reported in Figure 5.

Figure 5: Run protocol and results visualization (GUI example)

4) Result visualization

When the application of a single protocol is terminated, it is possible to move to the results of each subject. They can be visualized from the software within the *Protocol* field (structured similar to *File Explorer*).

Mandatory requirements

- Visualization must be flexible to the particular characteristics of the data (2D/3D).
- There must be the possibility to export again the results in the same format of the input files (in case the results files in the destination folder are deleted or modified…).

Algorithms

The proposer can choose the programming language to develop the interface of the software with. As regard the internal core of the application (data management and clustering), the COM suggest the use of C/C++ as programming language. In this case the proposer will be supported by various open-source codes performing visualization and segmentation of

images are already available online (such as Image ToolKit, http://www.itk.org/ITK/project/about.html; Visualization ToolKit, http://www.vtk.org/). In particular, ITK includes several of the algorithms required. Alternative programming languages for the internal core can be proposed but the choice has to be motivated.

Project Degree of Freedom

GUI

One critical point of the software regards the necessity to allow its utilization by people with limited computer knowledge. Hence the use of the software has to be as user-friendly as possible. In order to achieve this result the bidder are free to design the best solutions they want, with particular interest to the GUI of the software. Innovative solutions for the GUI will be considered as criteria for the proposal evaluation.

Innovative solutions for the visualization and management of images

Moreover it will be very relevant for the proposal evaluation the way in which will be treated by the software in term of *quality of visualization and easiness of interaction with the software users.*

Particular care will be necessary in the first and last block of analysis (the management and storage of the raw data and of the results).

Additional features clustering algorithms

Suggestion of additional features extractor and cluster algorithms will be welcome but must be approved by the COM prior their integration in the software. This will be considered as criteria for the proposal evaluation.