

VIRTUAL GASTROINTESTINAL TRACT VIGOR++

Scenario and Validation

Goals of VIGOR++

Authors: P.J. Schueffler, C.S. Ong, H. Hatzakis, F.M. Vos

Project no. 270379
Instrument: Collaborative Project (STREP) / CP-FP-INFISO
Thematic Priority: FP7-ICT-2009-6
Start date of project: 1 February 2011
Duration: 36 months
Date of submission Deliverable:
Classification: Confidential/Restricted/Program Participants only/Public

Disclaimer:

The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. The opinions expressed in the document are of the authors only and in no way reflect the European Commission's opinions.

Summary:

Crohn's disease endoscopic index of severity (CDEIS) defines the reference standard for quantifying disease severity in patients with Crohn's disease. MRI is a non-invasive alternative increasingly used to assess disease activity. Compared to endoscopy and histopathology, radiologists correctly grade frank disease activity in approximately 90% of patients, while mild disease and remission is correctly graded in approximately 60% [HBSS09]. The algorithm developed in this project is expected to have a similar accuracy for frank disease activity and preferably a better accuracy for mild and inactive disease activity. In addition, our approach should enable a finer grained diagnosis (the CDEIS score) from MR images, when compared to current manual approaches which classify mild vs. severe.

When medical treatment fails, surgery is considered. For surgical treatment the location of the disease and the length of the involved bowel parts are important. At present, the reading of the length of these parts by radiologists is relatively cumbersome and data are not available on the accuracy. An algorithm that correctly identifies the location of diseased bowel parts with a reasonable accurate indication of the length (within 5 cm) would be an important step forward.

The iterations of the development cycle

The following anatomical structures, also called segments, are considered in the study on Crohn's disease to determine the degree of disease activity:

1. Ileum,
2. Sigmoid and left colon,
3. Transverse,
4. Right colon,
5. Rectum.

The CDEIS score as a measure for the disease severity ranges between 0-44 (see Figure 1). The CDEIS score is composed of partial scores which grade various aspects of Crohn's disease. Clinicians estimate the CDEIS score to assess the disease severity.

Semi-automatic estimation of the CDEIS score involves two learning tasks:

1. Segmentation of diseased bowel wall areas after manual identification of a suspected area by a radiologist.
2. Classification of disease severity, i.e., identifying remission and mild vs. frank patients.

The validation study should determine how precisely the CDEIS score with its sub-scores can be automatically estimated.

CDEIS						
	Ileum	Sigmoid and left colon	Transverse colon	Right Colon	Rectum	
Deep ulcerations (0 if none; 12 points if present)						Total 1+
Superficial ulcerations (0 if none; 6 points if present)						Total 2+
Surface involved by disease (cm ²)						Total 3+
Surface involved by ulcerations (cm ²)						Total 4=
Ulcerated stenosi present in segment: 3						
Non-ulcerated stenosis present in segment: 3						
CDEIS score per segment	Total over ileum	Total over sigmoid and left colon	Total over transverse colon	Total over right colon	Total over rectum	
						TOTAL 1-4: A
Number of segments explored (1-5)						N
Total A / n						TOTAL B
If ulcerated stenosis is present anywhere: 3						C
If non-ulcerated stenosis is present anywhere: 3						D
Total B+C+D = CDEIS score						

Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Gut. 1989 Jul;30(7):983-989.

Figure 1: Compartments of the CDEIS score

The CDEIS classification in VIGOR++ is based on the MRI scans. These are 3D scalar images of the abdomen, particularly visualizing the terminal ileum and colon. One MRI scan consists of 40-100 images (the slices) sized 400x400 pixels each. Several such image volumes are acquired, each showing a different contrast by applying different MR imaging sequences. Particularly, time series scans are acquired (DCE - dynamic contrast enhanced), in which a patient is scanned for several minutes after contrast injection. The temporal resolution of the DCE scans is one volume per second for 3T scanners and approximately one volume per 5 seconds at 1.5T. These DCE volumes will be registered by methods developed by TU Delft. The benefit of registration lies in the fact that spatial correspondence is achieved, i.e. a region of interest (e.g. bowel segment or diseased area) in one image volume corresponds to the same region in the next volume. Consequently, regions of interest may be labeled in only one MRI volume in one point of time. Additionally, images of the other sequences will also be registered to each other. All these images can be downloaded from the Biotronics3D platform in DICOM format. For each patient, 3D patches will be obtained that determine the location of registered bowel segments. This information is provided as .nii (nifti) files. There will be a matching between the nifti files and the DICOM files.

The acquisition protocol is defined in the VIGOR++ protocol endorsed by the medical ethical committees at AMC and UCLH.

Goals

The radiologists have 2 main tasks:

1. Identify the disease location (by segment) and length of the affected bowel part. One point to make here is that it is very hard to manually measure the length of a lesion on MRI in a patient with a large amount of such lesions.
2. Assign disease activity (mild, moderate, severe [also called frank])

It is more important to identify and grade each diseased bowel part rather than measure its exact length. Ideally, one wants to take the length into account and monitor the length of the diseased bowel part during treatment, but presumably this will be difficult. When surgery is considered, this point becomes more important.

A secondary goal of the VIGOR++ project is to develop a combined activity score using all the available patient data, e.g. CDEIS, CDAI (activity index), CRP etc.

In general, the objective of VIGOR++ is to support these tasks computationally in a semi-automated fashion. Our goals are therefore to:

1. Identify on a bowel segment, in a semi-automated manner, the bowel wall and lesions. Semi-automated means that the radiologist points roughly at the location of a region of interest, after which the computer automatically segments the region. An automated size-estimation of the affected region might be favorable, but is not mandatory.
2. Predict the CDEIS_{local} score for this bowel segment.
3. Predict the CDEIS for the whole MRI scan.

The current gold standard is the CDEIS score obtained by the gastroenterologist who performs a colonoscopy. Assisting radiologists to estimate a CDEIS score from MR imaging is one of the major goals of the VIGOR++ project.

Fully automated object detection (detection of diseased regions without user interaction) is a very hard task in the context of MR imaging and is beyond the scope VIGOR++. Clearly, such a system would be desirable but not mandatory. Nevertheless, in VIGOR++, we are discussing a labeling system for MR images, in which a semi-automatic algorithm would detect diseased regions based on already manually identified regions (see section “The Semi-Automated Labeling Pipeline”).

Benefits of Machine Learning

- By applying a computational method, we aim to obtain, from the MRI images, a predictor capable of a finer diagnosis scale (the CDEIS score) as compared to the gross scale (mild vs. frank disease) obtained by the radiologist.

- Furthermore, since the method is automated, we would obtain more objective and reproducible results (c.f. computational pathology survey¹).

The Classification Pipeline

The pipeline for processing the MRI scan will be as follows:

1. An MRI scan is loaded into the Biotronics3D system. Comprehensive viewing techniques and annotation techniques are already available. Particularly, regions of interest (e.g. diseased regions) will be roughly marked by a radiologist. Each such region will be manually assigned to a bowel segment (ileum, sigmoid and left colon, transverse, right colon, rectum). The locations marked are those in which the radiologist identifies Crohn's disease activity.
2. Based on the coarse annotation by the radiologist, a segmentation algorithm will segment the regions of interest more accurately. The segmentation comprises the segmentation of bowel wall in each such area to yield full 3D regions of interest (3DROI). Some segmentation algorithms are recently developed by project partner ZIB. Also, superpixel-segmentation is available. *Optionally, the length/size of the segment is calculated (not mandatory).*
3. From each 3D region thus determined (3DROI), representative features are extracted. The features might comprise histogram based intensity features, as well as features derived from the DCE scans. The feature extraction methods are mainly provided by TUDelft (but potentially also by ETHZ).
4. Based on the features obtained from all 3DROI's in a bowel segment, a classifier (Random Forest or Support Vector Machine) will first (a) predict a score/label for the regions itself and then (b) the CDEIS disease severity score of the bowel segment (CDEIS_{local}). The classifier will be developed by ETHZ. To develop a classifier immediately (and not to wait until step 3. of the pipeline is fully developed) ETHZ will directly extract features from manually annotated 3DROIs.
5. Based on the predicted scores of all bowel segments, the patient's total CDEIS score is predicted (see figure 1). This can be done by combining the predicted CDEIS_{local} score per bowel segment to calculate the total CDEIS or by directly computing the CDEIS with a trained estimator. Also, prediction of other scores, such as the Harvey Bradshaw index (single label per patient) may be possible.

Training of the classifier

To train the classifier of step (4a) that estimates a severity score for a given bowel wall segment, we will use features and annotations provided by TUDelft in step 3. These comprise e.g. features based on DCE scans.

Initially, however, ETHZ will develop prototype classifiers that are trained on features that ETHZ will derive themselves from manually annotated training data. Therefore, comprehensively annotated MRI scans will be used that are as such available in Biotronics 3DNet Suite system (acquired/annotated by AMC/UCL) as part of WP2. The MRI scans are provided as DICOM files and the annotations are provided as XML files. With that, pixel- and voxel-wise information will be extracted, features will be calculated

¹ Thomas J. Fuchs and Joachim M. Buhmann: Computational pathology: Challenges and promises for tissue analysis. Computerized Medical Imaging and Graphics, Vol. 35, 2011

and a classifier will be learned (e.g. Random Forest or Support Vector Machine). Single regions of interests within a MRI might not only be binary classified into diseased or healthy, but also be scored to a certain extend. This might facilitate the regression of the $CDEIS_{local}$ score in step (4b).

Step (4b) involves a regression task, estimating the $CDEIS_{local}$ score for a bowel segment. Excel tables are available with local CDEIS scores for each bowel segment for training. These annotations, together with the data from (4a), will be used to train a regressor to predict the $CDEIS_{local}$ score for the bowel segments. These manually assigned $CDEIS_{local}$ scores are obtained by a gastroenterological examination at AMC or UCL. However, not all bowel segments are available for each patient. Usually, there are only 2-3 segments per patient. We have to deal with both missing labels and missing examples.

Since the classifiers depend on the training data, the accuracy/precision of the predictions could change over the course of the project as more annotation become available.

The Semi-Automated Labeling Pipeline

This section discusses the computational assistance for annotating MRI images. Although this is not a primary goal for VIGOR++, a semi-automated annotation approach would be useful addition to the toolkit and the project as a whole. The manual MRI annotation could be dramatically facilitated by proposals of regions with similar patterns as annotated by the radiologist. We suggest following pipeline:

1. A MRI scan is loaded in the Biotronics3D system. One slice of the 3D object is shown.
2. *Optionally: the shown image is automatically segmented with the help of superpixel algorithms or other segmentation algorithms from ZIB.*
3. The radiologist starts annotating the image. Either s/he uses the pre-segmented areas from (2) for labeling diseased/normal tissue, or s/he uses the system of the classification pipeline for annotations (“The Classification Pipeline” step (2)).
4. After annotating several regions (at least one of two classes: normal tissue and suspected regions), a classifier automatically assigns other regions a probability for the classes. Other regions might be these of (2) or a voxel-wise screening through the whole image, in case of (2) was not performed.
In this case, the classifier selects other regions that might be annotated similarly by the radiologist. The labeled regions of the radiologist remain unaffected.
 - a. Dependent on the computational power, this step can be done on-the-fly (while the radiologist is annotating), or batch wise (the radiologist clicks a button).
 - b. Voxel-wise information of the probability for lesions or wall enhancements will be visualized in the Biotronics3D system and will be used by domain experts for labeling new data.
5. After the computer made its selections, the annotator reviews the given annotations. Step (4) and (5) are repeated until the user confirms all annotations.

Validation Procedure for Classifier

Each of the above-mentioned classifiers is validated in a cross-validation (CV) fashion. In CV, a part of the original training data is systematically omitted for training. After training with the reduced training set, the omitted data are predicted and the classification accuracy of the predicted data (or the mean squared error for regression, respectively) is saved. This procedure is then systematically repeated (say 100 times) with different parts of the original data as training and prediction set, each time. The mean of all saved misclassification errors then forms the final CV-error. This validation scenario works for the classification of the segmented regions of interest within the MRIs.

For voxel-wise classification problems, such as they arise in object detection tasks (e.g. in “Semi-Automated Labeling Pipeline” step (4)), the validation has a slightly different nature, since most samples (i.e. shifting windows) will have no labels (background), and so, the data are not *fully labeled*. For training, we will get labeled samples for both diseased bowel wall (positive label) as well for normal bowel wall (negative label). “Sample” means annotated regions in the MRI scans. The rest of the images are not annotated. Labeling all images fully by human experts would be too time consuming and cumbersome. The task is then to find positive or negative regions in the unlabeled parts of new MRI scans. Here, the validation would be a precision and recall measure (F-Score) rather than accuracy: the validation procedure consists of assessing the power of the classifier to detect diseased areas (rather than voxel-wise accuracy). A similar validation scenario has been seen in the automatic nucleus detection in tissue micro array images.

Generally, it has to be kept in mind that all labels, that we get from the domain experts, such as radiologists and medicals, are subjectively obtained in best manner. For example, we have a consensus CDEIS score from several gastroenterologists (AMC: 3; UCLH: 2) for each patient. Note however that we may not obtain the gold standard score for the whole bowel if there are difficulties in the colonoscopy, for example due to blockages in the colon. Some bowel segments may also not be able to be imaged due to collapsed bowel during the MR sequence.

Since the underlying “real and biological ground truth” of diagnostic segmentation in MRI scans or CDEIS scores for patients is even manually hard to assess, a main requirement for an automatic approach with a classifier in this context is not so much to perform with a higher accuracy, but rather to be more objective and reproducible in repeated or even in high throughput assessments.

Validation Targets for Pipeline

Besides the validation of the classifiers itself, a validation of the whole pipeline is needed to establish the success of the VIGOR++ project. For this, it must be clear, what the goals of the project are. Concerning the goals in this deliverable, we can establish three validation targets. We should focus on the intrinsic capabilities of the software informed by but not in comparisons to radiologist performance:

1. Achieve the highest possible correlation between the software derived score of activity (CDEIS) and the endoscopic score (main target).
2. Achieve accurate segmentations to support the radiologists’ tasks.

3. Achieve high predictive power for a segmented region of interest in MRI scans. The segmented regions in the MRI data in the Biotronics3D system serve as validation data.

The consortium has been having ongoing discussions on the above topics and these targets may be revised or extended based on their outcomes.

Glossary

In this deliverable, we use the following terms:

Segmented region, Region Of Interest (ROI)	The delineation of a region in the MRI. This can be 2D or 3D.
Label of a segment	Categorization of a segmented region into diseased or healthy.
Bowel segment	Ileum, left/right colon, transverse colon, sigmoid colon or rectum.
Annotation	Region of interest together with its label and bowel segment.