Semi-Automatic Crohn's Disease Severity Estimation on MR Imaging

Peter J. Schüffler[1,*], Dwarikanath Mahapatra[1], Robiel Naziroglu[2], Zhang Li[2], Carl A.J. Puylaert[3], Rado Andriantsimiavona[4], Franciscus M. Vos[2, 3], Doug A. Pendsé[5], C. Yung Nio[3], Jaap Stoker[3], Stuart A. Taylor[5, 6], Joachim M. Buhmann[1]

[1] Dept of Computer Science, ETH Zurich, Universitätstrasse 6, Zurich, Switzerland
[2] Quantitative Imaging Group, TU Delft, Lorentzweg 1, Delft, The Netherlands
[3] Dept of Radiology, AMC, Meibergdreef 9, Amsterdam, The Netherlands
[4] Biotronics3D Ltd, 205 Marsh Wall, London, United Kingdom
[5] Centre for Medical Imaging, UCL, 250 Euston Road, London, United Kingdom
[6] Dept of Radiology, UCLH, 235 Euston Road, London, United Kingdom
* peter.schueffler@inf.ethz.ch

Abstract. Crohn's disease (CD) is a chronic inflammatory bowel disease which can be visualized by magnetic resonance imaging (MRI). For CD grading, several non-invasive MRI based *severity* scores are known, most prominent the MaRIA and AIS. As these scores rely on manual MRI readings for individual bowel segments by trained radiologists, *automated* MRI assessment has been more and more focused in recent research. We show on a dataset of 27 CD patients that *semi-automatically* measured bowel wall thickness (ABWT) and dynamic contrast enhancement (DCE) completely outperform manual scorings: the segmental correlation to the *Crohn's Disease Endoscopic Index of Severity* (CDEIS) of ABWT and DCE is significantly higher (r=.78) than that of MaRIA (r=.45) or AIS (r=.51). Also on a per-patient basis, the models with ABWT and DCE show significantly higher correlation (r=.69) to global CDEIS than MaRIA (r=.46).

Keywords: Computer Vision, Crohn's disease, Crohn's disease severity, MRI.

1 Introduction

Crohn's disease (CD) belongs to the chronic inflammatory bowel diseases. The *severity* of acute CD is an important indicator for different therapeutic strategies and for the documentation of treatment response. One measure for CD severity is the *Crohn's Disease Endoscopic Index of Severity* (CDEIS) [1]: For that, the bowel is virtually partitioned into the five segments *terminal ileum*, *right colon*, *transverse colon*, *left (and sigmoid) colon* and *rectum*. Each segment is individually scored based on ulcerations and diseased surface identified in the segment. The patients' CDEIS is then the mean of the segmental scores plus additional scores for stenosis in the bowel.

Regular endoscopic examinations as they occur for CD patients come with several drawbacks. They are time consuming for the gastroenterologist and invasive for the

patient with the risk of bowel perforation. Further, occurring stenosis can impede the continuation of the uncomfortable examination.

Magnetic resonance imaging (MRI) has therefore been identified as an alternative for CD severity assessment [2-5]. With MRI, the abdomen is non-invasively visualized in a 3D volume. The bowel wall and surface can be inspected regardless of potential stenosis. In 2009, Rimola *et al.* [2, 3] introduced the *Magnetic Resonance Index of Activity* (MaRIA) calculating a segmental CD severity score based on the following features scored in MRI: *wall thickness (mm)*, *relative contrast enhancement (RCE)*, presence of *edema* and *ulcers*. They showed a significant correlation of this score to the CDEIS (Spearman r=.80). In 2012, Steward *et al.* [4] presented an MRI score of the weighted sum of *mural thickness* and *mural T2*. Although this measure was developed to predict the histopathological Acute Inflammation Score (AIS, R squared = .52), it usually correlates to the CDEIS as well. Schüffler *et al.* [5] showed in 2013 that there might be a set of MRI models with even higher correlation to the CDEIS: *Enhancement T1* and *comb sign* were able to improve the correlation of the two scores by 18%.

However, all scores rely on the manual detailed inspection of MRI scans. This can be time consuming and quite subjective. E.g., the inter-observer variability of scored *wall thickness* in MRI has been shown to significantly change with the radiologists' experience: more experienced radiologists had a higher inter-observer agreement [6].

Vos *et al.* [7] and Tielbeek *et al.* [8] therefore had the vision of automatic MRI processing for CD severity assessment. They reviewed computerized techniques to be able to extract relevant MRI features such as wall thickness and contrast enhancement.

In this paper, we describe a novel computer-generated model for CD severity determination incorporating the semi-automatic features *bowel wall thickness* and *dynamic contrast enhancement*. On a dataset of 27 CD patients, we validate these features together with our best manual model derived by an exhaustive search and compare them to the MaRIA and AIS. We illustrate that the semi-automatic features alone are able to predict the CDEIS more accurately than comprehensively manually scored features. The high objectivity and reproducibility of the semi-automatic model is a further clear benefit compared to manual methods.

2 Methods

2.1 MRI Protocol

Twenty-seven CD patients underwent ileocolonoscopy with CDEIS determination and MRI examination at the Academic Medical Center (AMC) Amsterdam, The Netherlands, within one month. MRI was performed with a 3 Tesla scanner (Intera, Philips Healthcare), according to following protocol [9]: Patients fasted for four hours and drank 1,6 L of mannitol solution (2.5%, Osmitrol, Baxter), one hour before the scan. Pre-contrast sequences comprising axial and coronal T2-weighted single-shot FSE sequences with and without fat saturation as well as coronal 3D T1-weighted spoiled gradient-echo sequence with fat saturation were acquired. Patients were then administered

20 mg of butylscopolamine bromide (Buscopan, Boehringer, Ingelheim) for bowel relaxation and 0.2 mL/kg bodyweight of gadobutrol (1.0 mmol/mL; Gadovist, Bayer Schering Pharma) as contrast agent. A dynamic contrast enhancement (DCE-)MRI was performed over 6 min. After a second dose of 20 mg of butylscopolamine, contrast-enhanced axial and coronal 3D T1-weighted spoiled gradient-echo sequences with fat saturation were acquired. All sequences except for the DCE-MRI were used for manual inspection at the 3DNETMEDICAL platform (www.biotronics3d.com). The DCE-MRI was used for automatic feature extraction (see section 2.4).

2.2 Manual MRI scoring

Four radiologists with 1–18 years of experience in abdominal MRI independently scored 12 CD related features on the five individual bowel segments of all 27 patients (*terminal ileum*, *right colon*, *transverse colon*, *left (and sigmoid) colon* and *rectum*). Binary features (absent / present) were *abscess*, *comb sign*, *fistula* and *ulcers*. Categorical features (normal, mild, moderate, marked) were *T1 enhancement*, *length*, *mural T2 signal*, *pattern*, *perimural T2 signal*. Numerical features were *relative contrast enhancement (RCE)* and *wall thickness*. A patient wide binary feature was the presence of enlarged lymph nodes.

Of the 27*5=135 bowel segments, 2 segments had been resected, 5 segments could not be judged by the radiologists due to bad bowel distention and 6 segments could not be assessed by colonoscopy due to stenosis. Therefore, the manual dataset comprises 122 segments * 4 observers = 488 samples with 12 features, each.

2.3 Automatic Bowel Wall Thickness Measurement (ABWT)

Bowel wall thickness has been found to correlate well to the CDEIS [2-4]. While normal wall thickness ranges from 2 to 3mm, diseased wall can expand to over 15mm due to inflammation or lesions. The semi-automatic measuring, as described in [10], starts from the manual indication of a center line in the lumen of a bowel segment (region of interest, ROI). The automatic steps are then [10]: (i) Starting from the center line, segment the inner bowel wall. (ii) Starting from the inner bowel wall, segment the outer bowel wall. (iii) Average the distance between inner and outer bowel wall over the ROI as *Automatic Bowel Wall Thickness* measure (ABWT). Although a ROI is needed to start the computational process, the indication of a (rough) centerline can be performed much faster than the accurate manual measuring of wall thickness throughout all three dimensions. *Fig. 1* illustrates the automatic inner and outer bowel wall segmentation of a typical example. Note that this is a three dimensional segmentation procedure.

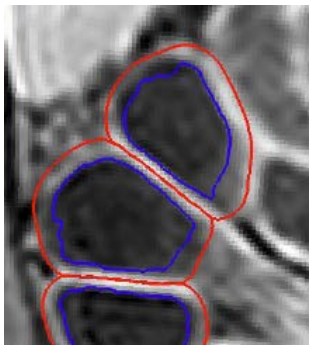

Fig. 1: Example of inner (blue) and outer (red) automatic bowel wall segmentation. The wall thickness is the mean distance between the inner and outer bowel wall.

2.4 Dynamic Contrast Enhancement (DCE)

DCE-MRI monitors the distribution and metabolism of the contrast agent in the bowel wall during 6 min after injection. The idea is to record a higher and faster contrast agent uptake in diseased regions than in normal regions. For this, 450 individual 3D scans are shot at a rate of 0.82 sec per scan (resolution 2.78 x 2.78 x 2.5mm at 227 x 227 x 14px). After DCE-MRI, contrast-enhanced coronal T1-weighted high resolution Isotropic Volume Examination (THRIVE) sequences were acquired at 1.02 x 1.02 x 2mm and 400 x 400 x 100px (see *Fig. 2*).

The DCE feature generation, as described in [11], starts from the manual indication of a diseased ROI. The automatic steps are then [11]: (i) Register the DCE-MRI to the post-contrast THRIVE sequences to remove breathing motion artifacts and to get pixel-wise image correspondence. A more detailed description of the 3D registration can be found in [11]. (ii) Extract the change of the signal intensity in the ROI over time as the *time intensity curve* (TIC). (iii) Fit a bi-exponential model $S(t)$ to the TIC:

$$S(t) = A_1 e^{-\lambda_1 t} - A_2 e^{-\lambda_2 t}$$

where $\boldsymbol{A_1}$ is related to the steepness of the TIC and defines the final *DCE* feature.

An example of DCE-MRI and THRIVE is illustrated in *Fig. 2*. The top left image shows the THRIVE before contrast agent application and DCE-MRI. The top right THIRVE was shot after contrast agent application and DCE-MRI. The bottom row shows two DCE-MRI scans before and during contrast agent uptake. Note the different resolution of the two sequences. The DCE frames are registered to each other, but not yet to the THRIVE post contrast. The diseased ROI indicated by the red arrow has a slightly faster enhancement than normal regions.

Fig. 3 plots two example TICs for a normal (green, lower curve) and a diseased ROI (red, upper curve). Each TIC is modeled by a bi-exponential model (black curves). The TIC is usually significantly steeper for diseased regions than for normal regions [11].

2.5 Model Development

To find the best linear regression model of manual MRI features, we followed an exhaustive search strategy [5]. All 2^{12}-1 = 4095 combinations of manual features were cross-validated on 27 CD patients. For cross-validation, the models were learned on the data of 18 randomly drawn patients and tested on the remaining nine patients. The Spearman rank correlation coefficient between the predicted segmental MRI scores of the test patients and the CDEIS was recorded. This procedure was repeated 50 times with different random patient subsets. The models were then ranked by their median cross-validated correlation to the CDEIS. The best model is subjected for further validation: ABWT and DCE were added to the model to measure the performance gain. Finally, ABWT and DCE were cross-validated as a single “automatic” model in the same procedure as described above.

Feature extraction was performed with MATLAB (2013b). Linear regression modeling and validation was performed with R Statistical language, version 3.0.0.

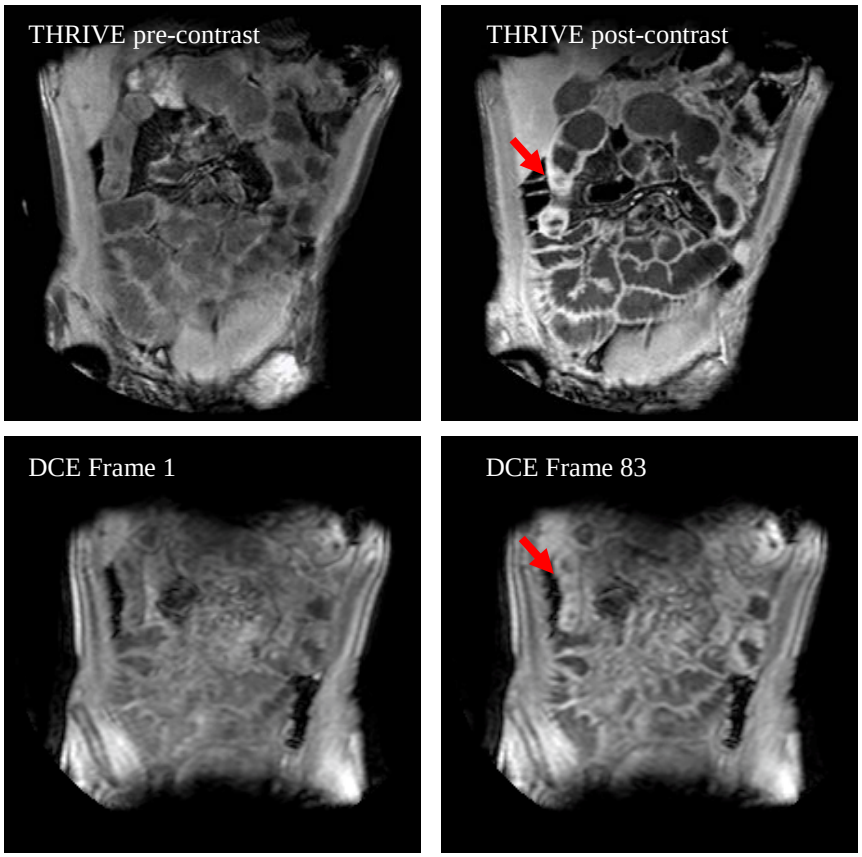

Fig. 2: Typical example of THRIVE and DCE-MRI. ***TOP LEFT:*** *2D slice of a pre-contrast THRIVE scan in our dataset.* ***TOP RIGHT:*** *Corresponding post-contrast THRIVE slice after DCE-MRI (during DCE-MRI, a contrast agent is applied to the patient). CD affected regions show a slightly enhanced signal, as indicated by the red arrow.* ***BOTTOM:*** *Two registered 2D slices of the DCE sequence of the same patient (time frames 1 and 83). 450 such frames are shot during six minutes. The spatial resolution of DCE-MRI is much smaller than of THRIVE.*

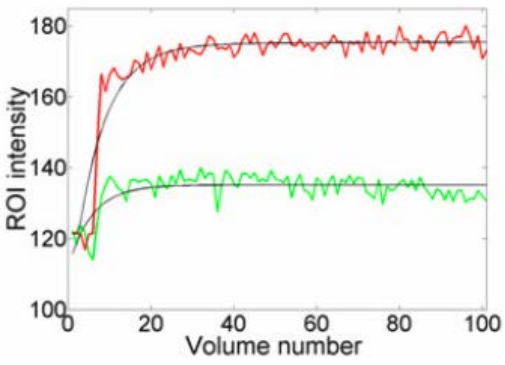

Fig. 3: TIC of a normal bowel segment (green, bottom) and a diseased segment (red, top). The mean MR signal intensity in a given ROI is measured in 100 consecutive DCE frames. At the beginning, the contrast agent is applied to the patient. Diseased regions typically show significantly enhanced drug uptake resulting in steeper curves. The black curves represent the fitted bi-exponential models S(t), whose coefficients A_1 *serve as DCE feature.*

3 Results

3.1 ABWT correlates to CDEIS

ABWT corresponds to the manual measured *wall thickness*, scored by four observers in our dataset. As shown in *Fig. 4* (left), the correlation of *wall thickness* to the CDEIS is r=.44±.08. Each observer is indicated by a specific color. Note the relative high number of segments with a normal CDEIS but thickened bowel wall and vice versa. The right plot displays ABWT versus CDEIS. Here, some samples with a high CDEIS could not be processed (e.g. due poor image quality). However, among the measured cases, there are only few outliers. The overall correlation to the CDEIS is r=.68 without standard deviation since the objective measure is the same for all observers.

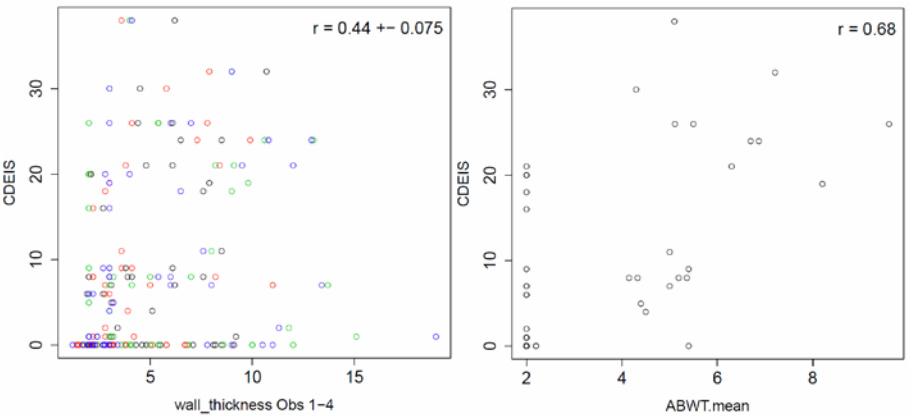

Fig. 4: ***LEFT:*** *Correlation of manually scored wall thickness by four observers to CDEIS. Each observer is denoted by a color.* ***RIGHT:*** *The correlation of ABWT to CDEIS (r=.68) is much higher than that of wall_thickness (r=.44±.08).*

3.2 DCE correlates to CDEIS

DCE best corresponds to *relative contrast enhancement* (RCE), a ratio of the MRI signal in THRIVE post-contrast and pre-contrast images. DCE, in addition, accounts for the "speed" of contrast agent accumulation. In *Fig. 5* becomes apparent that the correlation of RCE to CDEIS is moderate, throughout all observers r=.30±.05. Further, very severe cases might get a comparably low RCE value. Another problem is the difficulty of the complex RCE measuring: In our dataset, 39 samples have negative RCE and three samples have extremely high RCE (400-500). This nicely demonstrates the need of an automated method to facilitate relative contrast enhancement measuring. The DCE feature on the right hand side shows clearly higher correlation to CDEIS (r=.60). Note the high number of severe samples which have not been processed (A_1=0), either due to poor image registration or a mismatch of the field of view of DCE-MRI and THRIVE. This might be improved in future work.

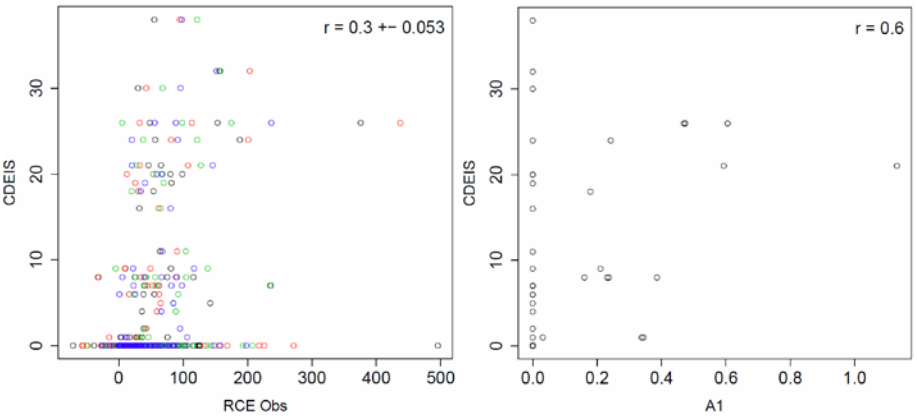

Fig. 5: ***LEFT:*** *Correlation of manually scored relative contrast enhancement (RCE) by four observers to CDEIS.* ***RIGHT:*** *The correlation of DCE (A_1) to CDEIS (r=.60) is much higher than that of RCE (r=.30±.05).*

3.3 Multivariate CDEIS correlation

The best manual model with the highest CDEIS correlation found by our exhaustive search consists of *abscess*, *comb_sign*, *muralT2* and *ulcers*. Note that *muralT2* and *ulcers* have already been identified by Rimola *et al.* [2, 3] and Steward *et al.* [4] as important severity predictors. The median Spearman correlation of this model to the CDEIS is r=.57 (*Fig. 6*, middle box "Best manual"). Indeed, the specific combination of these four features shows significantly higher correlation to CDEIS than the MaRIA (r=.45) or AIS (r=.51) (*Fig. 6*, second and third box). The addition of ABWT or DCE (A_1) to the manual model raises the correlation to the CDEIS significantly to r=.67 and r=.69, respectively (*Fig. 6*, fifth and sixth box). Interestingly, when ABWT and DCE are combined as a stand-alone MRI model, the correlation to CDEIS can even be increased to r=.78 (*Fig. 6*, right box); larger than their univariate CDEIS correlation.

3.4 Global View

We tested our models with the semi-automatic features for *global* CDEIS prediction. A high segmental CDEIS correlation should propagate to a high correlation per patient. A patient's CDEIS is the *mean* of his or her segmental scores plus additional 3 points for non-ulcerated stenosis or ulcerated stenosis. Rimola *et al.* define the global MaRIA as the *sum* of the segmental scores [2, 3].

For testing, we followed a leave-one-patient-out cross-validation procedure: our manual model including ABWT and DCE and the model consisting solely of ABWT and DCE were trained on the data of 26 patients. The predicted segmental scores of the remaining patient were then averaged to the global MRI score. *Fig. 7A* shows the cross-validated global MRI scores of all patients predicted by *abscess*, *comb_sign*, *muralT2*, *ulcers*, ABWT and DCE. Each patient is denoted by a number and each observer by a color. The overall correlation to global CDEIS is r=.66.

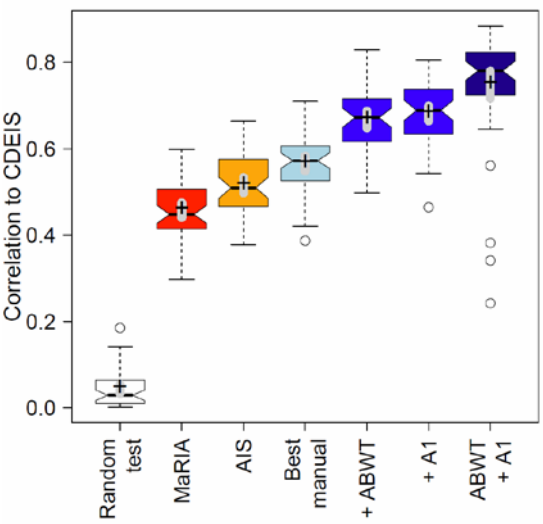

Fig. 6: Spearman correlation to CDEIS of different models. Each box is a 50-fold cross-validated model. Horizontal lines indicate median, cross and bar indicate mean and standard deviation of folds. ***Random test:*** *The best manual model was cross-validated with randomly permuted CDEIS label. The information in the features is not random.* ***MaRIA:*** *The MaRIA on our dataset reaches a segmental correlation of r=.45.* ***AIS:*** *The MRI based AIS has a median correlation of r=.51. Our* ***best manual model*** *(middle, r=.57) can significantly be improved by the two automatic features* ***ABWT*** *(r=.67) or* ***A1*** *(r=69). However, the two automatic features alone,* ***ABWT + A1****, show a superior segmental CDEIS correlation (r=.78).*

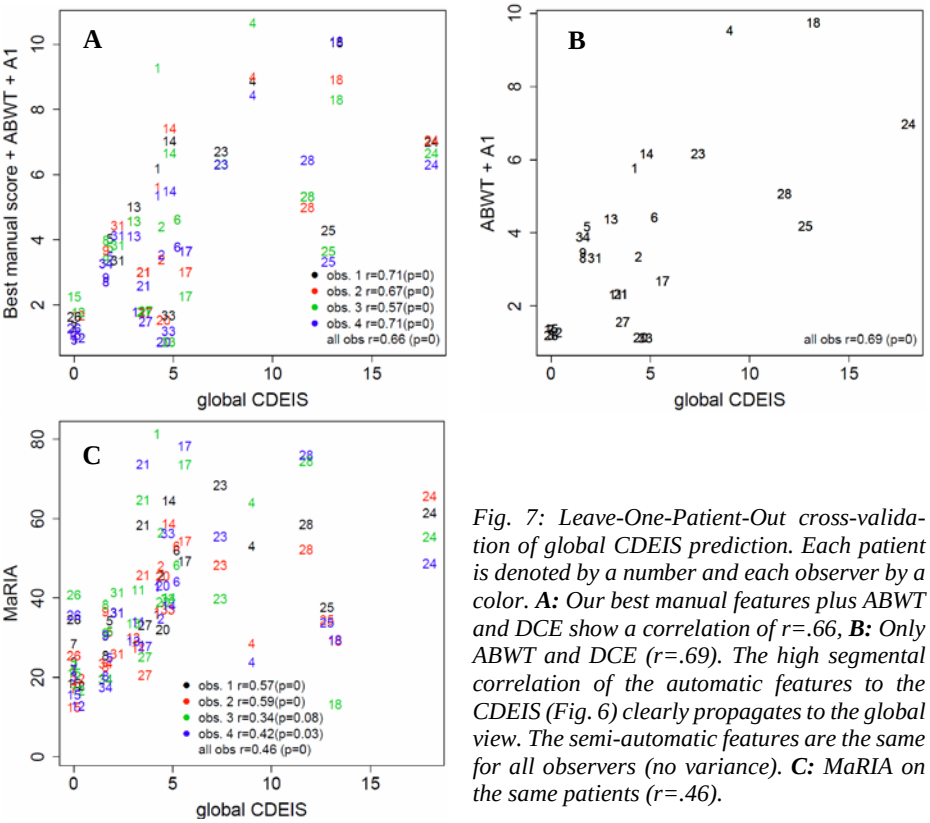

Fig. 7: Leave-One-Patient-Out cross-validation of global CDEIS prediction. Each patient is denoted by a number and each observer by a color. ***A:*** *Our best manual features plus ABWT and DCE show a correlation of r=.66,* ***B:*** *Only ABWT and DCE (r=.69). The high segmental correlation of the automatic features to the CDEIS (Fig. 6) clearly propagates to the global view. The semi-automatic features are the same for all observers (no variance).* ***C:*** *MaRIA on the same patients (r=.46).*

In *Fig. 7B*, the cross-validation of the model consisting of ABWT and DCE is depicted. There is no inter-observer variance for these computer-generated features. The correlation to the CDEIS is r=.69, which is higher than combined with manual features.

As comparison, we illustrate in *Fig. 7C* the relation of the global CDEIS and MaRIA. Again, there is an inter-observer variance due to manual features. The correlation of the two scores ranges from r=.34 (observer 3) to r=.59 (observer 2).

4 Discussion

We emphasize in this paper the potential benefit of computer-read MRI features for CD severity assessment. While most automatic MRI processing methods refer to organ detection and segmentation, the use of automatically extracted clinically relevant features such as bowel wall thickness or DCE for CD severity assessment is completely new.

Automated feature extraction might improve CD severity judgments in two ways. First, it may enable standardized and more objective scorings compared to manual scorings which clearly showed a considerably high inter-observer variance in our experiments. Second, the time of manual MRI processing by physicians can be reduced or replaced by cheaper computer processing time. Especially the measurements of *RCE* and *wall thickness* are time consuming – two features for which we propose computational analogs.

Surely, our semi-automatic features still need manual interaction (e.g. both rely on the indication of ROIs). The *fully* automatic processing of CD MRI will be a topic for future work (e.g. automatic CD detection). The calculation of DCE is especially complex and not successful on *all* bowel segments: DCE-MRI usually has a smaller field of view than THRIVE imaging. Interesting bowel segments could therefore be missed by DCE-MRI, which impedes the DCE feature extraction in these parts. Also, the registration process might fail in some cases due to the complex image structure.

While this might be improved in future work, we have already shown promising results in this paper with automatic features. The models with ABWT and DCE clearly demonstrate a superior correlation to the CDEIS than any manual model. On segment basis, the correlation is improved from 45% (MaRIA) to 78%. On per patient basis, there is an improvement from 46% (MaRIA) to 69%.

5 Conclusion

We demonstrated the clear improvement of MRI based CD severity assessment by the use of computer-aided feature extraction. Semi-automatically measured bowel wall thickness and dynamic contrast enhancement had a higher correlation to the CDEIS than any other manual model in our dataset, including the MaRIA and AIS. While the univariate correlation of the new features to the CDEIS was 60% and 68%, the combination of these two in a linear regression model reaches a correlation of 78%. We propose to validate these new features on further datasets in upcoming studies.

Semi-automatic MRI processing clearly reduces the inter-expert variability observed in conventional manual MRI features. Fully automatic MRI assessment, however,

would require enhanced methods for disease detection, bowel segment detection and feature extraction. The research on automatic MRI processing will significantly facilitate and accelerate MRI inspection and improve our understanding of radiologic signs of Crohn's disease.

Acknowledgments. This study was partly funded from the European Community's Seventh Framework Program (FP7/2007-2013): the VIGOR++ Project (grant agreement nr. 270379).

6 References

1. Mary, J.Y., Modigliani, R.: Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Therapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). Gut 30, 983-989 (1989)
2. Rimola, J., Ordas, I., Rodriguez, S., Garcia-Bosch, O., Aceituno, M., Llach, J., Ayuso, C., Ricart, E., Panes, J.: Magnetic resonance imaging for evaluation of Crohn's disease: validation of parameters of severity and quantitative index of activity. Inflammatory bowel diseases 17, 1759-1768 (2011)
3. Rimola, J., Rodriguez, S., Garcia-Bosch, O., Ordas, I., Ayala, E., Aceituno, M., Pellise, M., Ayuso, C., Ricart, E., Donoso, L., Panes, J.: Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. Gut 58, 1113-1120 (2009)
4. Steward, M.J., Punwani, S., Proctor, I., Adjei-Gyamfi, Y., Chatterjee, F., Bloom, S., Novelli, M., Halligan, S., Rodriguez-Justo, M., Taylor, S.A.: Non-perforating small bowel Crohn's disease assessed by MRI enterography: derivation and histopathological validation of an MR-based activity index. European journal of radiology 81, 2080-2088 (2012)
5. Schüffler, P.J., Mahapatra, D., Tielbeek, J.A.W., Vos, F.M., Makanyanga, J., Pendsé, D.A., Nio, C.Y., Stoker, J., Taylor, S.A., Buhmann, J.M.: A Model Development Pipeline for Crohn's Disease Severity Assessment from Magnetic Resonance Images. In: Yoshida, H., Warfield, S., Vannier, M. (eds.) Abdominal Imaging. Computation and Clinical Applications, vol. 8198, pp. 1-10. Springer Berlin Heidelberg (2013)
6. Tielbeek, J.A.W., Makanyanga, J.C., Bipat, S., Pendse, D.A., Nio, C.Y., Vos, F.M., Taylor, S.A., Stoker, J.: Grading Crohn Disease Activity With MRI: Interobserver Variability of MRI Features, MRI Scoring of Severity, and Correlation With Crohn Disease Endoscopic Index of Severity. Am J Roentgenol 201, 1220-1228 (2013)
7. Vos, F.M., Tielbeek, J.A.W., Naziroglu, R.E., Li, Z., Schüffler, P.J., Mahapatra, D., Wiebel, A., Lavini, C., Buhmann, J.M., Hege, H., Stoker, J., van Vliet, L.J.: Computational modeling for assessment of IBD: To be or not to be? In: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, pp. 3974-3977. (Year)
8. Tielbeek, J.A.W., Vos, F.M., Stoker, J.: A computer-assisted model for detection of MRI signs of Crohn's disease activity: future or fiction? Abdominal imaging 37, 967-973 (2012)
9. Ziech, M.L., Lavini, C., Caan, M.W., Nio, C.Y., Stokkers, P.C., Bipat, S., Ponsioen, C.Y., Nederveen, A.J., Stoker, J.: Dynamic contrast-enhanced MRI in patients with luminal Crohn's disease. European journal of radiology 81, 3019-3027 (2012)
10. Naziroglu, R.E., Van Vliet, L.J., Vos, F.M.: Measuring and quantifying bowel wall thickening for assessing Crohn's disease severity. In: VIGOR++ Workshop 2014. (Year)
11. Li, Z., Tielbeek, J.A.W., Caan, M.W.A., Ziech, M.L.W., Nio, C.Y., Stoker, J., van Vliet, L.J., Vos, F.M.: Expiration Phase Template-based Motion Correction of Free-Breathing Abdominal Dynamic Contrast Enhanced MRI. in submission (2013)