Comparing Endoscopic Imaging Configurations in Computer-Aided Celiac Disease Diagnosis

Michael Gadermayr[1], Hubert Kogler[2], Andreas Uhl[1] and Andreas Vécsei[2]
[1] Department of Computer Sciences, University of Salzburg, Salzburg, Austria
e-mail: mgadermayr@cosy.sbg.ac.at, uhl@cosy.sbg.ac.at
[2] St. Anna Children's Hospital, Department of Pediatrics, Medical University Vienna, Vienna, Austria

***Abstract*— During the past years, significant research has been done on computer aided celiac disease diagnosis based on endoscopic image data. Mostly the aim is to increase the final classification accuracies based on a certain endoscopic data set. In this work, we investigate the impact of different endoscopic devices as well as imaging modalities with distinct impact on the visual properties of the image data. Apart from the obtained classification accuracies, special focus is on a metric measuring the relative performance of computer based methods compared to the visual classification performance of human experts on the respective data sets. Thereby a potential bias due to variations between the data sets can be circumvented. Finally, we can make statements on the performances of computer based methods compared to human experts considering the same data set, and, based on previous knowledge, assumptions on the most appropriate imaging setting in case of computer aided diagnosis can be made.**

***Keywords*— Computer aided diagnosis, celiac disease, feature extraction, narrow-band imaging, white-light imaging, endoscopy;**

I. Introduction

Celiac disease [1], [2], which is commonly known as gluten intolerance, is an autoimmune disorder that affects the small intestine after the introduction of gluten containing nutrition. This disease leads to an inflammatory reaction in the mucosa of the small intestine caused by a dysregulated immune response triggered by ingested gluten proteins of certain cereals. During the course of celiac disease, the mucosa loses its absorptive villi and hyperplasia of the enteric crypts occurs, leading to a strongly diminished ability to absorb nutrients. The overall prevalence [3] of the disease in the USA is approximately one per cent.

Whereas the gold standard for detection is based on biopsies, computer aided diagnosis of celiac disease solely relies on image data captured during duodenoscopies. During the past years, significant research has been done in the field of computer aided celiac disease diagnosis [4], [5], [6], [7], [8], [9], [10]. In these publications, focus is either on obtaining the best possible overall classification rates based on a predefined data set [4], [5], [6], [7], [8], or on selecting appropriate sub-images (patches) from original ones [9], [10]. These patches should ideally exhibit the disease markers (and should not suffer from strong image degradations) in order to achieve again the best accuracies in case of computer based classification.

Due to different imaging configurations, the available image data for computer based classification varies significantly, as different cameras (especially sensors) are deployed and furthermore the imaging modality can be adjusted.

During this work, we compare three different imaging configurations with significant impact on the visual image properties. The first configuration (WL1) relies on the Olympus GIF-Q165 endoscope which is based on white-light imaging (WLI). In WLI, the mucosa of the small intestine is illuminated with a traditional source of (white) light. The second configuration (WL2) relies on the Olympus GIF-H180 endoscope, which provides distinctly improved image quality, and is also based on WLI. The third configuration relies on the Olympus GIF-H180 endoscope and is based on narrow-band imaging (NBI) where the mucosa is illuminated with some specific wavelengths as outlined in the subsequent section.

The main goal of this work is to find out which image data can be classified most accurately in case of computer aided celiac disease diagnosis? One way to perform a comparison is to simply compare the diagnosis performances (i.e. the classification accuracies) obtained with the three different image data sets. However, as we rely on manually extracted preselected patches and manually captured images, a straightforward comparison could contain some bias due to the manual patch extraction and image acquisition stage. Therefore, we additionally compute the relative rates of computer based methods compared to human experts' diagnosis performances based on the respective image data sets. Finally, based on previous knowledge assumptions on the best endoscopic setting for computer aided diagnosis can be done.

This paper is structured as follows. In Sect. II, the difference between the imaging configurations are outlined and a potential bias of a straight-forward accuracy comparison is discussed. In Sect. III, the experimental results are presented and discussed. Section IV finally concludes this paper.

II. Imaging Configurations

All three investigated imaging configurations rely on flexible endoscopy and the modified immersion technique. Flexible endoscopes are equipped with support channels used to extract biopsy specimen, to clean tissue and to inflate the intestine. The modified immersion technique makes use of a support channel to instill water into the duodenal lumen after evacuation of air by suction. Villi, if present, straighten up in water and appear as finger-like moving structures. Experimental evidence has been gained [11] that the visualization of villi utilizing the modified immersion technique corresponds to a

Table 1. Image data sets available for the three different imaging configurations.

Data set	Images (Marsh-0/Marsh-3)	Patients (Marsh-0/Marsh-3)	Device	Modality
WL1	280/280	131/40	Olympus GIF-Q165	WLI
WL2	280/280	85/35	Olympus GIF-H180	WLI
NBI	280/280	88/37	Olympus GIF-H180	NBI

(a) WL1 (b) WL2 (c) NBI

Fig. 1. In the top row image patches based on the three different configurations are shown. In the bottom row magnifications are presented to see the effect of improved imaging technology.

higher positive predictive value in case of manual classification based on visual data. Additionally, the water provides several additional benefits such as a more homogeneous illumination without specular reflections and bubbles. A further benefit of this technique is, that no specialized endoscopic hardware is required.

During experimentation, we evaluate three different imaging configurations, as outlined in Table 1. Figure 1 shows example image data, captured under the three different configurations.

The difference between the WL1 and the WL2 data set is due to the newer endoscopic device (GIF-H180). Using this novel endoscope, the level of noise can be reduced significantly as can be seen especially in the magnified sub-figures on the bottom row. The magnification level of the two endoscopic devices is more or less the same. Most previous work in this field [8], [10], [12], [13], [14], [15], relies on image data captured with the older endoscope (GIF-Q165) or a related model (GIF-N180) exhibiting a similar image quality.

In case of the narrow-band imaging (NBI) data set the same device as in set WL2 is utilized, however, a special illumination method is applied which leads to more emphasized villi structures in the visual data. NBI [16] has been reported to improve the diagnostic accuracy in diverse fields of endoscopy [17]. It is based on specific blue (440 to 460 nm) and green (540 to 560 nm) wavelengths for illumination to enhance the contrast of vascular patterns on the mucosal surface. This method is employed to specifically delineate the outline of the residual villous structures (if present) due to a better visualization of villous height and shape compared to traditional white light endoscopy.

A. *Bias in Data Sets*

Preliminary experimentation showed that a straight forward comparison of classification accuracies based on the different data sets may not provide a fair comparison between the three endoscopic settings used.

Firstly, it is not possible to capture the same mucosal regions (showing the same disease markers) of the same patients with all three imaging modes. Whereas in case of the different illumination modes (WL2 and NBI) this is difficult but theoretically feasible, in case of the different endoscopes (WL1 and WL2) this definitely cannot be done without ethical concerns during medical treatment.

A further problem is that some regions of the endoscopic images often significantly suffer from severe degradations such as blur, overexposure as well as underexposure [9]. Experience indicates that the obtained image quality correlates with the level of expertise of the medical doctor who acquired the visual data.

To solve these issues, this work relies on manually extracted patches (with a size of 128×128 pixels as done in previous work [7], [9], [18]) which have been extracted by experts. The idea is to select regions that are widely free from strong degradations and actually show disease makers which can be effectively used for discrimination between healthy and diseased mucosa.

However, further potential bias can be unintentionally introduced by the experts during the manual selection of the image patches for classification. During manual patch selection, the experts extract sub-images which contain, in their opinion, markers for a potentially successful distinction. If an imaging

configuration delivers images of a lower quality (e.g. the disease markers are not clearly visible), the experts probably extract fewer images. Doing that, the classification accuracies will not represent the overall quality of the original image data. Being aware of this issue, in the following, we mainly do not focus on the absolutely achieved classification accuracies, but we compare the rates of computer based methods with the visual classification performance of human experts using the same image data.

The assumption of a significant bias in the absolute classification rates is supported by the fact that the accuracies obtained with data set WL1 turned out to be on average the highest compared to the other data sets, although the image data (visually) seems to be inferior (compared to the other sets). This is supposed to be because in case of (more noisy) low quality image data, during manual extraction image data with more pronounced disease markers is extracted. We assume that thereby the classification task can even be simplified in case of image data exhibiting a lower quality. Analyzing the relative rates (ratio of computer based accuracies to humans' classification performances), we aim at getting more insight into the requirements for a successful discrimination.

III. Experiments

A. Setup

The test data used for experimentation contains images of the duodenal bulb and the pars descendens taken during endoscopies at the St. Anna Children's Hospital. Prior to automated processing, all images are converted to gray scale images as the additional use of color information did not lead to consistent improvements. This conversion, however, is not done for acquisition of human experts' classification accuracies, because human experts are highly used to color images and the rates could thereby suffer from a gray scale conversion.

In a preprocessing step, texture patches with a fixed size of 128×128 pixels have been manually extracted to get more idealistic data (as done in past research [7], [8] and discussed in Sect. II-A).

To get the ground truth for the texture patches, the condition of the mucosal areas covered by the images has been determined by histological examination of biopsies from corresponding image regions. The severity of the villous atrophy has been classified according to the modified Marsh classification scheme [2]. Although it is possible to theoretically distinguish between several different stages of the disease, we aim in distinguishing between images of patients suffering from celiac disease (Marsh-3) and healthy patients (Marsh-0), as this two classes case is most relevant in practice. Furthermore it has been shown that a proper separation into more classes based on visual markers only is extremely difficult and probably impossible [19].

Our experiments are based on three different balanced data sets, each containing 560 image patches (280 of class Marsh-0 and 280 of class Marsh-3). All overall accuracies presented are based on the mean of 50 random splits. One distinct split divides the data set into an approximately balanced training (80 %) and evaluation set (20 %), restricting images of one patient to be in the same set to avoid any bias.

To get accuracies of human experts for a comparison, three experts (two medical doctors (endoscopists) and one expert in medical imaging) manually annotated all of the data. Thereby, 5040 images have been manually classified which provide a stable basis for further experimentation. In order to get stable data, for further processing the accuracies per expert are calculated and finally these three classification rates are averaged.

The relative classification rates are finally achieved by dividing the accuracies obtained with computer based methods by the mean accuracies of human experts based on the respective image data set.

B. Feature Extraction Methods

For a reasonable comparison, several different well known feature extraction techniques are utilized. Beside state-of-the-art general purpose image representations, we investigate descriptors which have been specifically developed for the problem definition:

- Local Binary Patterns [20] (LBP):
  LBP describe a texture by means of the joint distribution of pixel intensity differences represented by binary patterns. Experimentation is based a radius of two pixels as well as various numbers of neighboring samples (2 (LBP_2), 4 (LBP_4) and 8 (LBP_8)) to gain insight on the impact of specific (not necessarily sensible) configurations.
- Local Ternary Patterns [21] (LTP):
  LTP is a generalization of LBP making the final representation more robust to noise. This is achieved by introducing a different quantization scheme based on three states instead of the binarization applied by LBP. The approach is used with a fixed threshold ($t = 5$), a radius of two pixels and eight circularly aligned neighbors.
- Multi-Fractal Spectrum [22] (MFS):
  First, the local fractal dimension is computed for each pixel using three different types of measures for computing the local density. Finally, the feature vector is built by concatenation of these fractal dimensions.
- Dual-Tree Complex Wavelet Transform [23] (DTCWT):
  This image descriptor is based on fitting a two-parameter Weibull distribution to the wavelet coefficient magnitudes of sub-bands obtained from the dual-tree variant of the complex wavelet transform. Decomposition is performed on five levels.
- Shape Curvature Histogram [13] (SCH):
  This method, which has been specifically designed to deal with celiac disease image data, describes an image as the histogram of contour curvature values. This is done by first selecting contour pixels (by means of edge detection), followed by curvature estimation, based on edge filter responses. Finally all curvatures in contour regions are collected into a histogram.

- Improved Fisher Vectors [24] (IFV):
  Fisher Vectors [25], as well as the next descriptor (VLAD), is a global mid-level image representation that is obtained by pooling local image descriptors. These state-of-the-art standard methods build up and improve the idea of the Bag-of-visual-words approach [26] which has become highly popular in past years. In case of Fisher Vectors, the Gaussian mixture model is used to construct a dictionary, based on a local descriptor. For this local descriptor, we use the well known SIFT (Scale-invariant Feature Transform) [27] feature. The final fisher vector contains information how the parameters of Gaussian mixture model have to be modified to better fit the data. This is done by concatenating the means and the covariance deviation vectors. We use improved fisher vectors [24] which are derivatives based on two ideas. Instead of the linear kernel IFV uses the non-linear Hellinger's kernel which is based on the Bhattacharyya distance. Furthermore, the final feature vector is L^2 normalized.
- Vector of Locally Aggregated Descriptors [28] (VLAD):
  VLAD is technique which is similar to Fisher Vectors. In opposite to Fisher Vectors is does not store any second-order information. Furthermore it uses k-means clustering instead of a Gaussian mixture model to generate the feature vocabulary. The feature vectors finally store information of the difference between the cluster centers and the pooled local descriptors.

For the final classification we apply a linear support vector machine (SVM) which has been often used in previous work on computer aided celiac disease diagnosis [9], [7], [29] and also generally in recent work on texture recognition [30]. The c-value is optimized ($c \in \{2^0, 2^1, ..., 2^{11}, 2^{12}\}$) based on inner cross-validation. The feature vectors are L^2 normalized before classification.

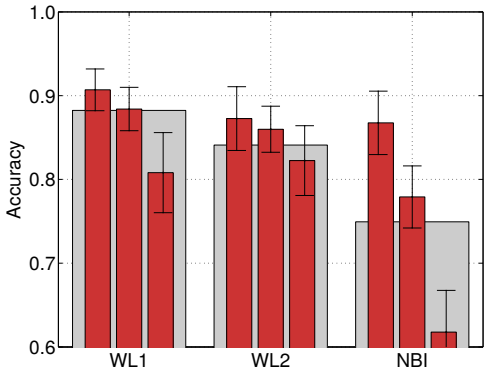


Fig. 2. Classification accuracies of human experts. One narror bar indicates the obtained rates of one certain expert in case of one data set whereas the wide bars represent the means over all experts.

C. Results

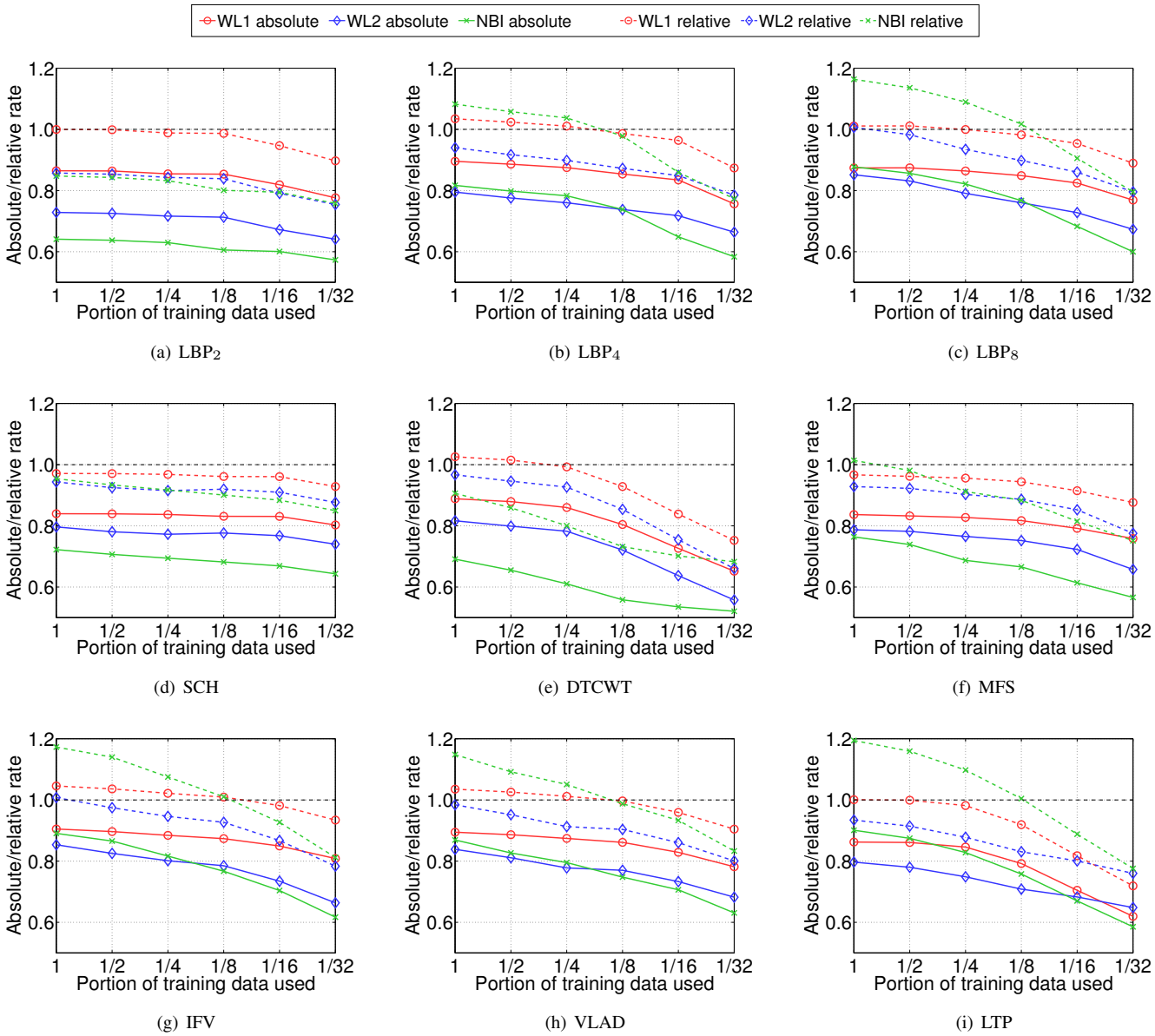
First, Fig. 2 shows the obtained classification performances of the human experts. One bar indicates the obtained rates of one certain expert in case of one data set. The standard deviations are achieved by 50-fold randomly sampling 20% of the classification decisions (as done in case of computer based classification). A distinction by human experts of the WL1 data set seems to be easier whereas the classification task of the NBI data set turned out to be clearly more difficult for the medical doctors. However, we strongly assume that the more elaborated imaging technique does not lead to decreased classification performances of human experts based on unbiased data. Actually it has been shown for NBI [17] that the opposite is the case. Thereby, strong evidence is provided that the distribution of the underlying mucosal structure, which has been captured with different endoscopic configurations, systematically differs (over the three data sets).

In Fig. 3, the main results obtained with computer based systems are shown. Considering absolute accuracies (solid lines) it can be seen that with the WL1 data set, on average the best overall classification rates are obtained. Interestingly, considering this set, the utilized feature extraction method does not have a strong impact and the accuracies are always above 84 %. Even with very few information in the feature vector (see LBP_2), similar accuracies compared to sophisticated image descriptors are achieved. Considering LBP based feature extraction, it can be seen that in case of WL1, the configuration (LBP_2, LBP_4, LBP_8) has a minor impact, which is distinctly different considering the accuracies obtained with the other data sets. The absolute rates observed with the WL2 data set are generally almost constantly lower compared to the WL1 data set. However, we notice a stronger impact of the feature extraction technique used. In case of NBI, the obtained accuracies also significantly vary with different image descriptors (especially if looking at the results of the three LBP configurations). This novel imaging technique does not lead to improved accuracies, however, in general we notice a more distinct positive effect of increased training data.

Based on these considerations, we strongly assume that the problem definition in case of WL1 is easier due to the (biased) manual patch extraction stage. Consequently, in the following focus is on the relative accuracies which are calculated as the fraction on the rates obtained with the computer based methods and the average rates of human experts.

Considering these relative accuracies (dashed lines), we notice that the performances of WL1 and WL2 become more similar (at least apart from the methods LBP_2 and LBP_4 which obviously contain not enough information for classification of the WL2 and the NBI data set). The remaining advantage of WL1 indicates that human experts' classification accuracies suffer more distinctly from the noisy data than computer based methods do. This assumption is supported by previous work [31] which showed that noise not necessarily has a strong impact on computer based classification if the noise remains stable over training and evaluation data.

With WL1 and WL2, in best case the human classification accuracy can be obtained approximately, which is indicated by a relative accuracy of about 1.0. In case of NBI, interestingly, the human accuracies can even be outperformed (relative



(a) LBP_2 (b) LBP_4 (c) LBP_8

(d) SCH (e) DTCWT (f) MFS

(g) IFV (h) VLAD (i) LTP

Fig. 3. Absolute (solid lines) and relative (dashed lines) classification rates achieved with computer-based classification. Whereas the absolute rates are real accuracies (between 0 and 1), the relative rates (calculated as $\frac{\text{absolute rate}}{\text{experts' rate}}$) can be above 1.0 if the computer based method outperforms the human experts.

accuracy > 1.0) with some image description methods. That means, based on NBI data, computer based methods have a relative advantage compared to human expert's visual classification (which is not the case in white-light imaging).

However, so far we are not able to make a statement on the most appropriate imaging modality based on the relative overall classification rates. These data only provides information about the relative performances of computers compared to human experts. Nevertheless, recent work [17] proved that NBI data in general is more appropriate for visual classification of human experts compared to WLI. Consequently, we strongly assume that if having image data from NBI, WL1 and WL2 showing the same mucosal regions, the sophisticated computer based methods (IFV, LTP) are able to deliver the best overall rates utilizing NBI image data. Comparing WL1 and WL2, we assume that computer based methods will not profit as much as human experts do from the improved image quality (of set WL2).

IV. Conclusion

We have investigated the impact of three imaging configurations during endoscopy on the classification performance of computer aided diagnosis methods. Comparing the absolute accuracies, it turned out that the old endoscope (WL1) delivers the best (absolute) accuracies, however, these rates turned out to be biased because of the manual patch extraction stage. Comparing the obtained rates to classification accuracies of human experts, it has been proven that only in case of NBI the computer delivers the more competitive rates compared to

experts. As it has been shown that NBI is more appropriate for a visual classification relying on human experts (in case of unbiased data), we can assume that the combination of NBI and computer based classification leads to the best overall classification rates. Future work could be to select for each configuration variably difficult classification tasks (sub data sets) based on the human experts classification decisions to be able to make an assumption on how computer based methods perform on tasks with differing levels of difficulty.

REFERENCES

[1] M. Marsh, "Gluten, major histocompatibility complex, and the small intestine. a molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue')," *Gastroenterology*, vol. 102, no. 1, pp. 330–354, Jan. 1992.

[2] G. Oberhuber, G. Granditsch, and H. Vogelsang, "The histopathology of coeliac disease: time for a standardized report scheme for pathologists," *European Journal of Gastroenterology and Hepatology*, vol. 11, pp. 1185–1194, Nov. 1999.

[3] Alessio Fasano, Irene Berti, Tania Gerarduzzi, Tarcisio Not, Richard B Colletti, Sandro Drago, Yoram Elitsur, Peter H R Green, Stefano Guandalini, Ivor D Hill, Michelle Pietzak, Alessandro Ventura, Mary Thorpe, Debbie Kryszak, Fabiola Fornaroli, Steven S Wasserman, Joseph A Murray, and Karoly Horvath, "Prevalence of celiac disease in at-risk and not-at-risk groups in the united states: a large multicenter study," *Archives of internal medicine*, vol. 163, pp. 286–92, Feb. 2003.

[4] Edward J. Ciaccio, Christina A. Tennyson, Suzanne K. Lewis, Suneeta Krishnareddy, Govind Bhagat, and Peter Green, "Distinguishing patients with celiac disease by quantitative analysis of videocapsule endoscopy images," *Computer Methods and Programs in Biomedicine*, vol. 100, no. 1, pp. 39–48, Oct. 2010.

[5] E. J. Ciaccio, C. A. Tennyson, G. Bhagat, Lewis S. K., and Green P. H., "Implementation of a polling protocol for predicting celiac disease in videocapsule analysis," *World Journal of Gastrointestinal Endoscopy*, vol. 5, pp. 313–322, 2013.

[6] E. J. Ciaccio, C. A. Tennyson, G. Bhagat, Lewis S. K., and Green P. H., "Use of basis images for detection and classification of celiac disease," *Bio-Medical Materials and Engineering*, 2014.

[7] R. Kwitt, S. Hegenbart, N. Rasiwasia, A. Vécsei, and A. Uhl, "Do we need annotation experts? a case study in celiac disease classification," in *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'14)*, September 2014, vol. 8674 of *Springer LNCS*, pp. 454–461.

[8] A. Vécsei, G. Amann, S. Hegenbart, M. Liedlgruber, and A. Uhl, "Automated marsh-like classification of celiac disease in children using an optimized local texture operator," *Computers in Biology and Medicine*, vol. 41, no. 6, pp. 313–325, June 2011.

[9] Michael Gadermayr, Andreas Uhl, and Andreas Vécsei, "Getting one step closer to fully automatized celiac disease diagnosis," in *Proceedings of the 4th IEEE International Conference on Image Processing Theory, Tools and Applications 2014 (IPTA'14)*, Oct. 2014, pp. 13–17.

[10] Michael Gadermayr, Andreas Uhl, and Andreas Vécsei, "Quality based information fusion in fully automatized celiac disease diagnosis," in *Proceedings of the German Conference on Pattern Recognition (GCPR'14)*, 2014, vol. 8753 of *Springer LNCS*, pp. 1–12.

[11] Antonio Gasbarrini, Veronica Ojetti, Lucio Cuoco, Giovanni Cammarota, Alessio Migneco, Alessandro Armuzzi, Paolo Pola, and Giovanni Gasbarrini, "Lack of endoscopic visualization of intestinal villi with the immersion technique in overt atrophic celiac disease," *Gastrointestinal endoscopy*, vol. 57, pp. 348–351, Mar. 2003.

[12] Sebastian Hegenbart, Andreas Uhl, Andreas Vécsei, and Georg Wimmer, "Scale invariant texture descriptors for classifying celiac disease," *Medical Image Analysis*, vol. 17, no. 4, pp. 458–474, 2013.

[13] Michael Gadermayr, Michael Liedlgruber, Andreas Uhl, and Andreas Vécsei, "Shape curvature histogram: A shape feature for celiac disease diagnosis," in *Medical Computer Vision. Large Data in Medical Imaging (Proceedings of the 3rd International MICCAI - MCV Workshop 2013)*, 2014, vol. 8331 of *Springer LNCS*, pp. 175–184.

[14] Michael Gadermayr, Andreas Uhl, and Andreas Vécsei, "Dealing with intra-class and intra-image variations in automatic celiac disease diagnosis," in *Proceedings of Bildverarbeitung für die Medizin 2015 (BVM'15)*, Mar. 2015, Informatik aktuell, pp. 461–466.

[15] Michael Gadermayr, Andreas Uhl, and Andreas Vécsei, "Degradation adaptive texture classification: A case study in celiac disease diagnosis brings new insight," in *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR'14)*, 2014, vol. 8815 of *Springer LNCS*, pp. 263–273.

[16] F. Emura, Y. Saito, and Ikematsu H., "Narrow-band imaging optical chromocolonoscopy: advantages and limitations," *World Journal of Gastroenterology*, vol. 14, no. 31, pp. 4867–4872, 2008.

[17] F. Valitutti, S. Oliva, D. Iorfida, M. Aloi, S. Gatti, C. M. Trovato, M. Montuori, A. Tiberti, S. Cucchiara, and G. Di Nardo, "Narrow band imaging combined with water immersion technique in the diagnosis of celiac disease," *Dig and Liver Dis*, vol. 46, no. 12, pp. 1099–1102, 2014.

[18] Sebastian Hegenbart, Andreas Uhl, and Andreas Vécsei, "Impact of histogram subset selection on classification using multiscale LBP," in *Proceedings of Bildverarbeitung für die Medizin 2011 (BVM'11)*, Mar. 2011, Informatik aktuell, pp. 359–363.

[19] Michael Gadermayr, Michael Liedlgruber, Andreas Uhl, and Andreas Vécsei, "Evaluation of different distortion correction methods and interpolation techniques for an automated classification of celiac disease," *Computer Methods and Programs in Biomedicine*, vol. 112, no. 3, pp. 694–712, Dec. 2013.

[20] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution Gray-Scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 24, no. 7, pp. 971–987, July 2002.

[21] Xiaoyang Tan and Bill Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Analysis and Modelling of Faces and Gestures*, Oct. 2007, vol. 4778 of *Springer LNCS*, pp. 168–182.

[22] Y. Xu, H. Ji, and C. Fermüller, "Viewpoint invariant texture description using fractal analysis," *International Journal of Computer Vision (IJCV*, vol. 83, no. 1, pp. 85–100, 2009.

[23] R. Kwitt and A. Uhl, "Modeling the marginal distributions of complex wavelet coefficient magnitudes for the classification of zoom-endoscopy images," in *Proceedings of the IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '07)*, Rio de Janeiro, Brasil, 2007, pp. 1–8.

[24] Jorge Sánchez, Florent Perronnin, Thomas Mensink, and Jakob J. Verbeek, "Image classification with the fisher vector: Theory and practice," *International Journal of Computer Vision (IJCV)*, vol. 105, no. 3, pp. 222–245, 2013.

[25] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, June 2007, pp. 1–8.

[26] Manik Varma and Andrew Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence," in *Proceedings of the 7th European Conference on Computer Vision (ECCV'02)*, 2002, Springer LNCS, pp. 255–271.

[27] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision (CVPR'99)*, 1999, vol. 2, pp. 1150 – 1157.

[28] Hervé Jégou, Florent Perronnin, Matthijs Douze, Jorge Sánchez, Patrick Pérez, and Cordelia Schmid, "Aggregating local image descriptors into compact codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 34, no. 9, pp. 1704–1716, 2012.

[29] Sebastian Hegenbart and Andreas Uhl, "A scale- and orientation-adaptive extension of local binary patterns for texture classification," *Pattern Recognition*, 2015, accepted.

[30] M. Cimpoi, S. Maji, I Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)*, 2014, pp. 3606–3613.

[31] Michael Gadermayr and Andreas Uhl, "Degradation adaptive texture classification," in *Proceedings of the IEEE International Conference on Image Processing 2014 (ICIP'14)*, Oct. 2014.