Toward automatic Segmentation and Quantification of Tumor and Stroma in Whole-Slide Images of H&E stained Rectal Carcinomas

Oscar G. F. Geessink[*abc], Alexi Baidoshvili[a], Gerard Freling[a],
Joost M. Klaase[b], Cornelis H. Slump[c], Ferdinand van der Heijden[c]
[a]Laboratory of Pathology East Netherlands, Hengelo, The Netherlands; [b]Department of Surgery, Medisch Spectrum Twente, Enschede, The Netherlands; [c]MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands

ABSTRACT

Visual estimation of tumor and stroma proportions in microscopy images yields a strong, Tumor-(lymph)Node-Metastasis (TNM) classification-independent predictor for patient survival in colorectal cancer. Therefore, it is also a potent (contra)indicator for adjuvant chemotherapy. However, quantification of tumor and stroma through visual estimation is highly subject to intra- and interobserver variability. The aim of this study is to develop and clinically validate a method for objective quantification of tumor and stroma in standard hematoxylin & eosin (H&E) stained microscopy slides of rectal carcinomas. A tissue segmentation algorithm, based on supervised machine learning and pixel classification, was developed, trained and validated using histological slides that were prepared from surgically excised rectal carcinomas in patients who had not received neoadjuvant chemotherapy and/or radiotherapy. Whole-slide scanning was performed at 20× magnification. A total of 40 images (4 million pixels each) were extracted from 20 whole-slide images at sites showing various relative proportions of tumor and stroma. Experienced pathologists provided detailed annotations for every extracted image. The performance of the algorithm was evaluated using cross-validation by testing on 1 image at a time while using the other 39 images for training. The total classification error of the algorithm was 9.4% (SD = 3.2%). Compared to visual estimation by pathologists, the algorithm was 7.3 times (P = 0.033) more accurate in quantifying tissues, also showing 60% less variability. Automatic tissue quantification was shown to be both reliable and practicable. We ultimately intend to facilitate refined prognostic stratification of (colo)rectal cancer patients and enable better personalized treatment.

Keywords: Segmentation, Quantification, Stroma, Hematoxylin and Eosin, Rectal cancer, Colorectal cancer, Pixel classification, Supervised machine learning

1. INTRODUCTION

Colorectal cancer (CRC) has become the third most common cancer in the world, with a global incidence of nearly 1.4 million cases in 2012 [1]. In the United States, the 5-year survival for CRC is 64.7%, based on follow-up data from 2004-2010 [2]. Currently, the two most important prognostic factors in CRC are depth of tumor invasion and the presence or absence of lymph node metastases. The prognostic value of these factors were first described in 1935 and they now form the core of the routinely used Tumor-(lymph)Node-Metastasis (TNM) classification and staging system [3]. For patients with early stage disease and those with advanced CRC, the TNM system provides strong prognostic information. However, for predicting individual patient outcome in stage II CRC, the TNM system proves to be less adequate [4]. Patients diagnosed at the same stage of disease often show significantly varying outcomes [5, 6]. Furthermore, adjuvant chemotherapy (fluorouracil combined with folinic acid) was found to provide only a small benefit (3.6%) for stage II CRC patients, compared to no adjuvant treatment [7]. This clearly demonstrates the need for advanced diagnostics that will allow a refined prognostic stratification of CRC patients and accurately identify those who may benefit from adjuvant therapy.

By now it is established that intra-tumoral stroma, often referred to as the tumor microenvironment, has an important role in tumor cell invasion and the ability to metastasize [8]. In CRC, patient prognosis was shown to be associated with the number of myofibroblasts in the stroma and the degree of stromal desmoplasia [9-12]. Furthermore, recent studies on

*Contact: o.geessink@labpon.nl

Medical Imaging 2015: Digital Pathology, edited by Metin N. Gurcan, Anant Madabhushi, Proc. of SPIE Vol. 9420, 94200F · © 2015 SPIE
CCC code: 1605-7422/15/$18 · doi: 10.1117/12.2081665

CRC prognostics revealed that quantification of relative tumor and stroma proportions in microscopy images, through visual estimation, yields a strong, TNM classification-independent predictor for patient survival [13-16]. A high percentage of stroma was related to a low disease-specific survival, whereas a low stroma percentage predicted a more favorable outcome. These results suggest that quantification of tumor and stroma should be used as a new prognostic marker for CRC and as a (contra)indicator for adjuvant chemotherapy.

However, tissue quantification through visual estimation is highly subject to intra- and interobserver variability. Although pathologists could employ a less subjective technique for estimating tissue proportions, like point counting, this would require the selection of a (very) small part of the whole-slide image in order for this method to remain practicable [14]. Since CRC histology often shows significant variations in local tumor and stroma proportions, the assessment of only a small part of the malignancy could render tissue quantities that are not representative for the malignancy as a whole. Therefore, in this study we aim to develop and clinically validate an automated method for objective quantification of tumor and stroma in standard hematoxylin & eosin (H&E) stained microscopy slides of rectal carcinomas. Through the application of our automated method on large malignant areas in whole-slide images, we ultimately intend to achieve refined prognostic stratification of (colo)rectal cancer patients and allow for better personalized treatment options.

2. METHODS

2.1 Images

Histological slides were prepared from surgically excised rectal carcinomas in patients who had not received neoadjuvant chemotherapy and/or radiotherapy. Slide preparation involved standard fixation of 5μm tissue sections and H&E-staining. Whole-slide scanning was performed at 20× magnification (~0.21μm^2/pixel) using a Hamamatsu NanoZoomer scanner (Herrsching, Germany). In accordance with Dutch law and national ethical guidelines ("Code for Proper Secondary Use of Human Tissue", Dutch Federation of Medical Scientific Societies), no specific patient consent was required for this study, provided that all material is anonymized.

For training and validation of the algorithm, 20 whole-slide images showing proper staining and fixation quality were selected (i.e. showing the full H&E color spectrum, red/pink to purple/blue and very few tears/cracks in the tissue). Since these whole-slide images contain various cells/tissues (e.g. fat, muscle, healthy epithelium, stroma, erythrocytes, tumor, necrotic debris, lymphocyte infiltrations, etc.) a Region Of Interest (ROI), which would include only tumor and (desmoplastic) stroma, was drawn in every selected whole-slide image. Large fields of necrosis were explicitly left out of the ROI's. From every selected whole-slide image, two smaller images of 2000-by-2000 pixels were extracted from within the ROI, while ensuring that various tumor-stroma proportion ratios were represented by the resulting 40 images.

In this study, we considered the tumor and stroma proportions to be complementary. Therefore, the 40 images were annotated using only tumor or stroma labels (Figure 4b). These annotations were provided by experienced pathologists using GIMP (GNU Image Manipulation Program, version 2.8.4). The annotated images were considered as ground truth.

2.2 Algorithm

The proposed algorithm for automatic segmentation and quantification of tumor and stroma was developed in MATLAB R2014a and the Image Processing Toolbox (The MathWorks, Natick, MA, USA). It is based on supervised machine learning and pixel classification. For training and classification, PRTools, a Pattern Recognition Toolbox for MATLAB, was used [17]. The algorithm can be described in the following steps (Figure 1):

Pre-classification: Each pixel in an input image is pre-classified as "white", "nucleus" or "cytoplasm" (Figure 2). White pixels, or optically empty pixels, are found by converting the Red-Green-Blue (RGB) input image to CIE L*a*b* color space followed by the application of an intensity threshold on the luminance channel L* (Figure 2b). The identified white pixels have hereby received their final classification label. The rest of the algorithm serves to find a final classification label for all non-white pixels. Since H&E-stained nuclei are blue/purple and cytoplasm is pink/red, the distinction between nucleus and cytoplasm pixels is achieved by converting the RGB image to the Hue-Saturation-Value (HSV) color space followed by local adaptive thresholding of the hue channel H with a disk-shaped kernel. Pre-classification results in a rough segmentation of nucleus and cytoplasm pixels which is required in the next step for the extraction of descriptive features (Figure 2c).

Feature extraction: In order to differentiate between tumor and stroma, a total of 6 features are extracted:

- Local density of nucleus pixels (Figure 3a). This feature was inspired by local variations in nucleus pixel density.
- Local density of cytoplasm pixels (Figure 3b). This feature was inspired by local variations in cytoplasm pixel density.
- Local density of small white objects (Figure 3c). This feature was inspired by local variations in tissue looseness.
- Luminance (Figure 3d). This feature was inspired by local variations in H&E-stained tissue brightness.
- Hue (Figure 3e). This feature was inspired by local variations in H&E-stained tissue color.
- Local direction magnitude (Figure 3f). This feature was inspired by local variations in the tissue's tendency to form unidirectional structures at a pixel level.

The first three features (local densities) are based on pre-classification output (Figure 2c) and are calculated using the Gaussian-weighted neighborhood of each pixel. The luminance and hue features are calculated during the pre-classification step and are used directly as features here. The local direction magnitude is based on the principal component of the gradient vectors that are derived from the Gaussian-weighted neighborhood of each pixel in the luminance image.

Training: A normal-density-based quadratic classifier was trained using the extracted features, the detailed annotations that were provided by the pathologists and PRTools (Figure 1). Only those feature values belonging to non-white pixels (Figure 2b) were considered during training.

Classification: The pre-trained quadratic classifier and the extracted features from a "new image" are used to classify every non-white pixel in the new image (Figure 1).

Correction: Since our method is pixel-based, some classification errors occur, often concerning small groups of pixels. From a contextual perspective, small groups of pixels isolated in a large field of pixels with a different class-label are very unlikely to be correct (e.g. 4 stroma pixels isolated in a large field of tumor pixels). During the correction step, classification labels are flipped for these small groups of "unlikely pixels". Furthermore, small "stroma islands" in large fields of white pixels are re-classified as tumor (for they are most likely to be tumor debris or bits of necrotic tumor cells). After the corrections have been applied, each pixel has received its final classification label (Figure 4c).

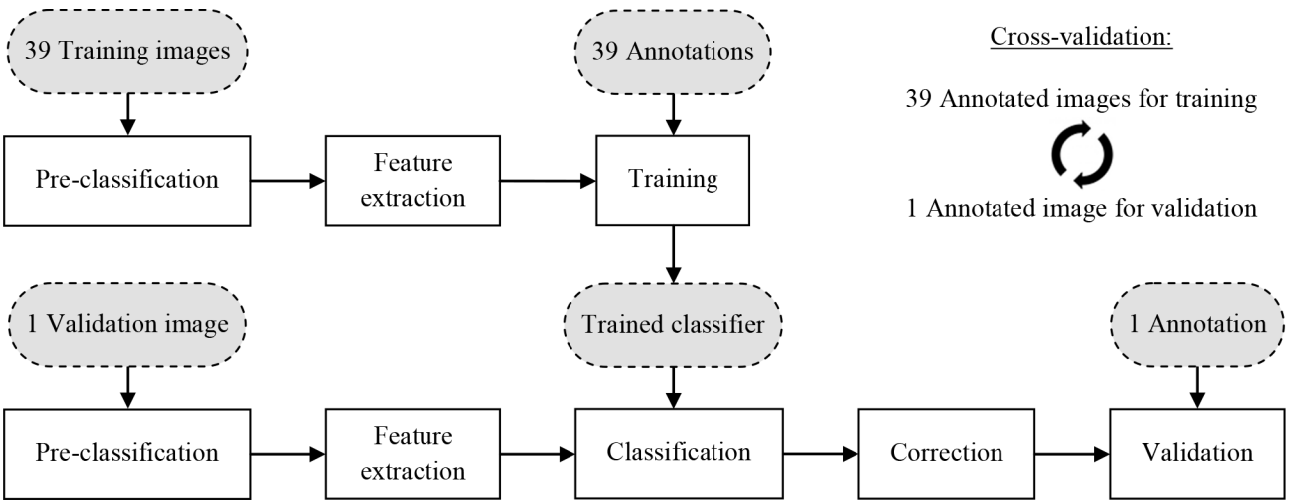

Figure 1. Block diagram of the proposed tissue segmentation and quantification algorithm. Through the application of cross-validation, the 40 annotated images are consecutively used for training and validation.

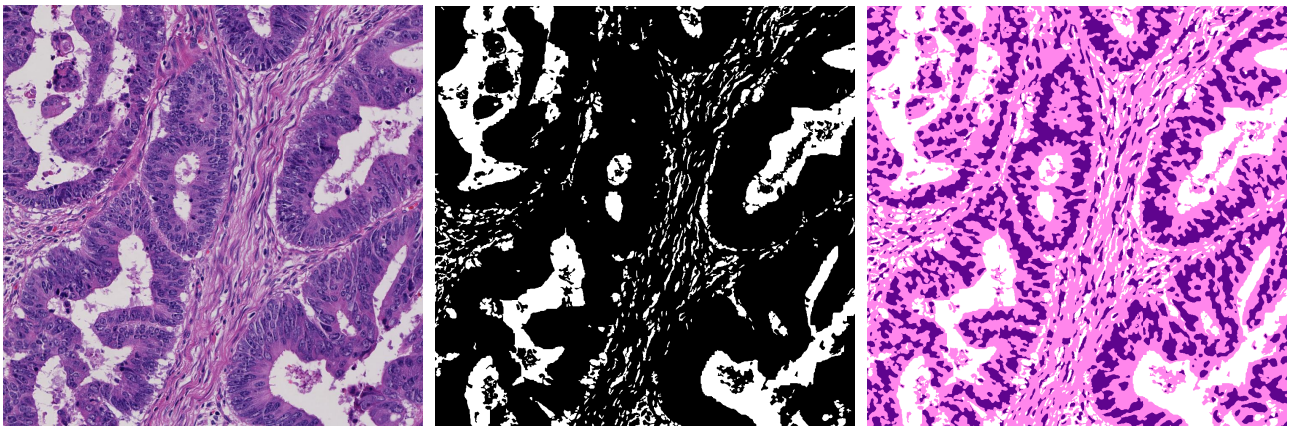

Figure 2. Pre-classification. (a) Input image (1 megapixel example). (b) Segmented white pixels, these pixels have received their final classification label. (c) Rough segmentation of nucleus (purple), cytoplasm (pink) and white pixels.

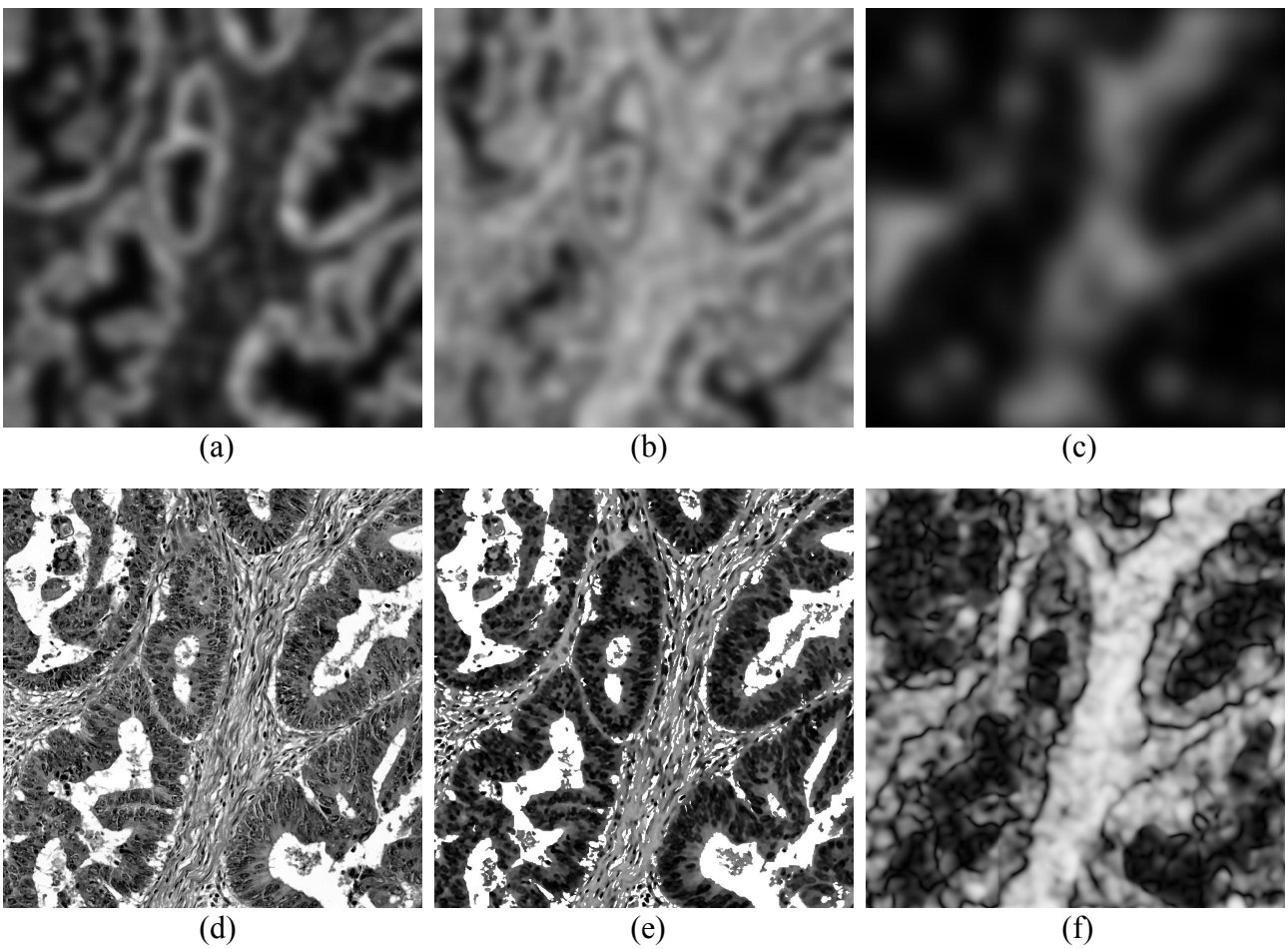

Figure 3. Features (extracted from the input image in Figure 2a). (a) Local density of nucleus pixels. (b) Local density of cytoplasm pixels. (c) Local density of small white objects. (d) Luminance. (e) Hue. (f) Local direction magnitude.

2.3 Validation

The performance of the algorithm was evaluated using cross-validation (Figure 1). One of the 40 annotated images was used for testing/validation while the other 39 images were used for training a classifier. In the next cycle, a different image is used for validation and the previous validation image is added to the training set. This process was repeated for each of the 40 images.

The tissue segmentation performance, i.e. classification performance, of the algorithm was evaluated separately from its tissue quantification performance. The classification error was found by calculating the average percentage of incorrectly classified non-white pixels, i.e. stroma pixels incorrectly classified as tumor, and vice versa (Figure 4). Apart from its classification performance, the tissue quantification performance of the algorithm was also evaluated by comparing with the (averaged) quantification capabilities of two experienced pathologists using visual estimation. The pathologists were asked to make an estimation of the proportion of tumor in relation to the total amount of tissue in all 40 images (i.e. without considering white, optically empty space). During visual estimation, 1 percent increments were used. The tumor quantification performance was calculated for both methods (algorithm and visual estimation) by comparing them individually to the actual (ground-truth-based) tumor proportions (i.e. by comparing to the tumor quantities that could be derived from the precise annotation images).

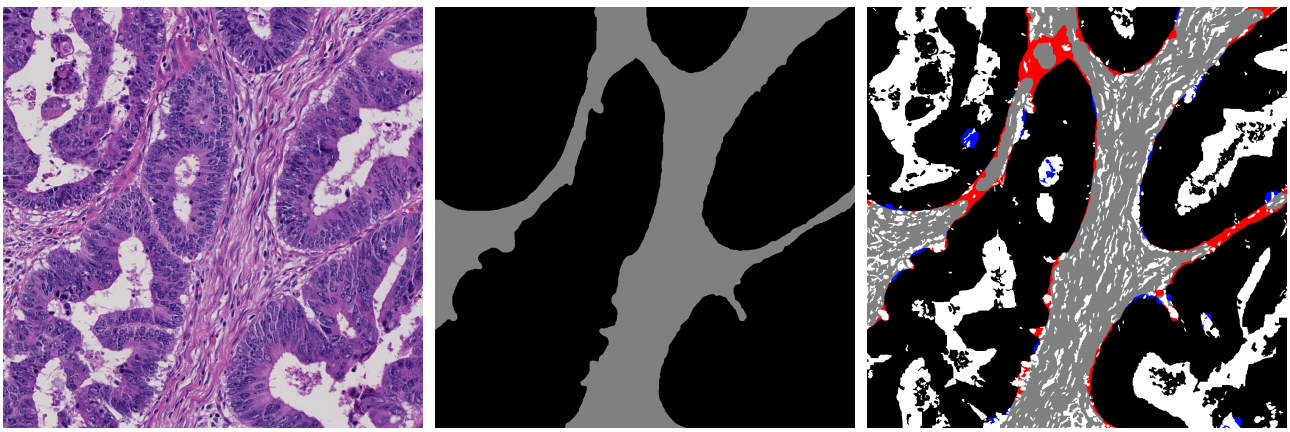

Figure 4. Pixel classification. (a) Image to be classified (1 megapixel example). (b) Pathologist annotations of tumor (black) and stroma (gray) . (c) Result after pixel classification and correction: Correctly classified tumor (black) and stroma (gray), as well as stroma pixels wrongly classified as tumor (red) and tumor pixels wrongly classified as stroma (blue).

3. RESULTS

Tumor and stroma were segmented in all 40 images through the application of pixel classification followed by a correction step. The total classification error of the algorithm was 9.4% (SD = 3.2%). On average, 4.4% (SD = 2.5%) of the non-white pixels was wrongly classified as tumor and 5.0% (SD = 3.2%) of the non-white pixels was wrongly classified as stroma. Classification errors mainly occurred at the transition between tumor and stroma (Figure 4c). Furthermore, some misclassifications were seen in dense lymphocyte infiltrations and patches of necrotic debris.

Based on the final segmentation/classification images, tumor and stroma proportions were calculated. Both tissue quantification methods (algorithm and visual estimation) showed an underestimation of the actual tumor proportions (Figure 5). The tumor quantification error was -0.6% (SD = 4.6%) for the algorithm and -4.4% (SD = 7.4%) for visual estimation (averaged for the two pathologists). On average, the algorithm was 7.3 times (P = 0.033) more accurate in quantifying tumor. However, this difference is likely due to the inability of the human eye/mind to ignore optically empty space when estimating tissue quantities. More importantly, quantification of tumor based on the algorithm showed a 60% smaller standard deviation compared to visual estimation by the two human experts.

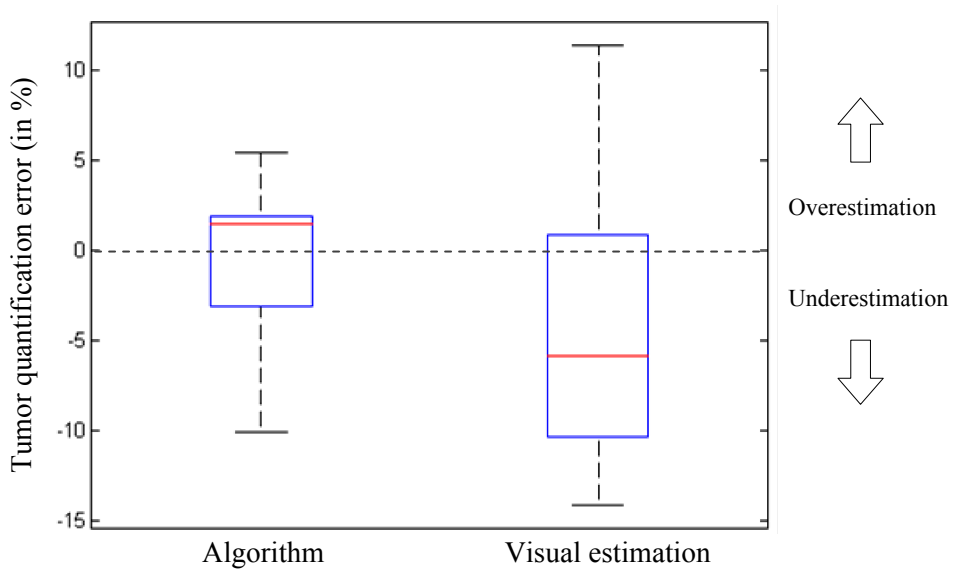

Figure 5. Tissue quantification performance: Algorithm versus visual estimation by two experienced pathologists (averaged). Algorithm results are based on cross-validation of 40 images (4 million pixels each).

4. CONCLUSION AND DISCUSSION

The work described in this paper shows that automatic segmentation of tumor and stroma is highly feasible through the application of a pixel classifier that has been trained on annotated images. During validation we found that the tissue quantification results based on automatic segmentation of tumor and stroma where more accurate and less disperse than the quantification results that relied on visual estimation by two pathologists. Although in this study, the pathologists also provided the ground truth segmentation for each image, and a nearly perfect quantification of tumor and stroma could therefore be derived directly from these annotations, such a method is highly impractical in everyday practice due to the time and effort is takes to accurately annotate large areas in histological slides. Implementation of a trainable algorithm is therefore the most practical way to achieve tissue quantification with high accuracy and low variability.

The proposed method uses only 6 descriptive features for training and classification. Although these features seem adequate for distinguishing tumor from stroma, additional features are probably needed for a reliable identification of other relevant tissue components, like necrotic debris and dense lymphocyte infiltrations. Our future research will be directed toward improvement of the algorithm and a more extensive validation. Through the application of our algorithm on large malignant areas in whole-slide images, we ultimately intend to facilitate refined prognostic stratification of (colo)rectal cancer patients and enable better personalized treatment.

REFERENCES

[1] World Cancer Research Fund International, "Colorectal Cancer Statistics," WCRF International, 1 January 2015, <http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/colorectal-cancer-statistics> (12 January 2015). http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/colorectal-cancer-statistics

[2] Surveillance Epidemiology and End Results Program, "Cancer of the Colon and Rectum - SEER Stat Fact Sheets," National Cancer Institute, 1 January 2015, <http://seer.cancer.gov/statfacts/html/colorect.html> (14 January 2015). http://seer.cancer.gov/statfacts/html/colorect.html

[3] Union for International Cancer Control, Sobin, L. H., Gospodarowicz, M. K., Wittekind, C., [TNM Classification of Malignant Tumours, 7th Edition], Wiley-Blackwell (2009).
[4] Schneider, N. I., Langner, C., "Prognostic stratification of colorectal cancer patients: current perspectives," Cancer Management and Research 6, 291-300 (2014).
[5] Lyall, M. S., Dundas, S. R., Curran, S., Murray, G. I., " Profiling markers of prognosis in colorectal cancer," Clinical Cancer Research 12(4), 1184-91 (2006).
[6] Compton, C. C., "Optimal pathologic staging: defining stage II disease," Clinical Cancer Research 13(22), 6862-70 (2007).
[7] QUASAR Collaborative Group, Gray, R., Barnwell, J., McConkey, C., Hills, R. K., Williams, N. S., Kerr, D. J., "Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study," Lancet 370(9604), 2020-9 (2007).
[8] De Wever, O., Mareel, M., "Role of tissue stroma in cancer cell invasion," Journal of Pathology 200(4), 429-47 (2003).
[9] Halvorsen, T. B., Seim, E., "Association between invasiveness, inflammatory reaction, desmoplasia and survival in colorectal cancer," Journal of Clinical Pathology 42(2), 162-6 (1989).
[10] Shepherd, N. A., Baxter, K. J., Love, S. B., "The prognostic importance of peritoneal involvement in colonic cancer: a prospective evaluation," Gastroenterology 112(4), 1096-102 (1997).
[11] Sis, B., Sarioglu, S., Sokmen, S., Sakar, M., Kupelioglu, A., Fuzun, M., "Desmoplasia measured by computer assisted image analysis: an independent prognostic marker in colorectal cancer," Journal of Clinical Pathology 58(1), 32-8 (2005).
[12] Tsujino, T., Seshimo, I., Yamamoto, H., Ngan, C. Y., Ezumi, K., Takemasa, I., Ikeda, M., Sekimoto, M., Matsuura, N., Monden, M., "Stromal myofibroblasts predict disease recurrence for colorectal cancer," Clinical Cancer Research 13(7), 2082-90 (2007).
[13] Huijbers, A., Tollenaar, R. A. E. M., Van Pelt, G. W., Zeestraten, E. C. M., Dutton, S., McConkey, C. C., Domingo, E., Smit, V. T. H. B. M., Midgley, R., Warren, B. F., Johnstone, E. C., Kerr, D. J., Mesker, W. E., "The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial," Annals of Oncology 24(1), 179-85 (2013).
[14] West, N. P., Dattani, M., Mcshane, P., Hutchins, G., Grabsch, J., Mueller, W., Treanor, D., Quirke, P., Grabsch, H., "The proportion of tumour cells is an independent predictor for survival in colorectal cancer patients," British Journal of Cancer 102(10), 1519-23 (2010).
[15] Mesker, W. E., Liefers, G. J., Junggeburt, J. M. C., Van Pelt, G. W., Alberici, P., Kuppen, P. J. K., Miranda, N. F., Van Leeuwen, K. A. M., Morreau, H., Szuhai, K., Tollenaar, R. A. E. M., Tanke, H. J., "Presence of a high amount of stroma and downregulation of SMAD4 predict for worse survival for stage I-II colon cancer patients," Cellular Oncology 31(3), 169-78 (2009).
[16] Mesker, W. E., Junggeburt, J. M. C., Szuhai, K., De Heer, P., Morreau, H., Tanke, H. J., Tollenaar, R. A. E. M., "The carcinoma-stromal ratio of colon carcinoma is an independent factor for survival compared to lymph node status and tumor stage," Cellular Oncology 29(5), 387-98 (2007).
[17] Duin, R. P. W., Juszczak, P., Paclik, P., Pekalska, E., De Ridder, D., Tax, D. M. J., Verzakov, S., "PRTools4.1, A Matlab Toolbox for Pattern Recognition," Delft University of Technology (2007).