



Exploiting Multiple Color Representations to Improve Colon Cancer Detection in Whole Slide H&E Stains

Alex Skovsbo Jørgensen¹(✉), Jonas Emborg², Rasmus Røge^{3,4},
and Lasse Riis Østergaard¹

¹ Department of Health Science and Technology, Aalborg University,
Aalborg, Denmark
asj@hst.aau.dk

² Diagnostics and Genomics Group, Dako Denmark A/S, an Agilent
Technologies Company, Glostrup, Denmark

³ Institute of Pathology, Aalborg University Hospital, Aalborg, Denmark

⁴ Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

Abstract. Currently, colon cancer diagnosis is based on manual assessment of tissue samples stained with hematoxylin and eosin (H&E). This is a high volume, time consuming, and subjective task which could be aided by automatic cancer detection. We propose an algorithm for automatic cancer detection within WSI H&E stains using a multi class colon tissue classifier based on features extracted from 5 different color representations. Approx. 32000 tissue patches were extracted for the classifier from manual annotations of 9 representative colon tissue types from 74 WSI H&E stains. Colon tissue classifiers based on gray level or color features were trained using leave-one-out forward selection. The best colon tissue classifier was based on color texture features obtaining an average tissue precision-recall (PR) area under the curve (AUC) of 0.886 and a cancer PR-AUC of 0.950 on 20 validation WSI H&E stains.

Keywords: Classification · H&E stain · Colon · Cancer
Machine learning

1 Introduction

Colorectal cancer is the third most common form of cancer worldwide and the fourth most common cause of death from cancer. [1] Currently, the gold standard colon cancer diagnosis is manual histopathological assessment of tissue structures within biopsies or tissue samples. Whole slide image (WSI) hematoxylin and eosin (H&E) stains allow pathologists to assess the tissue structures to detect cancer tissue and either perform a diagnosis or determine if special stains has to be used for further diagnosis. This is a high volume and subjective task and subject to inter- and intra-observer variation in the diagnosis which may lead to suboptimal treatment of the patient. [2] Digitization of WSI H&E stains has

led to development of algorithms to aid pathologists in reducing workload and inter- and intra-observer variation in the diagnostic work flow. Automatic cancer detection within WSI H&E stains can be used as first step for an automated analysis of WSI H&E stains e.g. for prescreening of slides to detect regions with cancer tissue and discard slides with obvious benign tissue for further analysis. However, WSI H&E stains can contain many different benign and malignant tissue structures with different appearances, which can complicate development of automated cancer detection algorithms. Some benign tissue structures appear very dissimilar to cancer tissue and can be easy to classify while others more closely resembles cancer tissue e.g. mucosa tissue in colon. This may be illustrated by extracting features from manual tissue annotations of representative colon tissue structures and be used to take into consideration during algorithm development. Additionally, accurate tissue classification between representative tissue structures may be used e.g. to determine degree of tumor infiltration and tumor tissue composition within the WSI H&E stains as cancer tissue can consist of different sub-tissue structures e.g. necrosis.

Currently, patch based methods are popular to characterize the appearance of tissue structures and has been used to extract both gray level [6] and color based texture and intensity features for tissue classification. [3–9] Kather et al. [6] proposed a patch based multi class classification framework to discriminate between eight colon tissue types to determine tumor composition in colorectal cancer using texture features. The study showed promising results reporting a 87.4% classification accuracy between the eight tissue structures. However, no color information was exploited in the study and it was only applied on a limited data set of 10 independent H&E stains. However, many studies has proven that color features are useful for tissue classification in H&E stains. [3–9] However, the studies has only extracted features from one or two color representations even though many color transformations exists each which may provide unique information useful for tissue classification. Therefore, it would be interesting to assess classification performance when utilizing all the color information available in the images.

To the best of our knowledge, no previous studies has explored using features from a broad range of color representations for multi class colon tissue classification to detect colon cancer in WSI H&E stains. Therefore, the purpose of this study was to develop a patch based framework to detect colon cancer tissue within WSI H&E stains using a multi class colon tissue classifier trained using texture and intensity features from five color representations from patches extracted from manual tissue annotations of representative colon tissue structures.

2 Methods

A patch based multi class colon tissue classification framework was developed to detect cancer tissue within WSI H&E stains. Tissue patches were extracted from manual annotations of nine representative colon tissue structures to obtain color based texture and intensity features from the tissue structures. The features were used for training and validation of colon tissue classifiers designed

to classify extracted tissue patches within a WSI H&E stain and obtain tissue probabilities (P_{tissue}) of each colon tissue structure. Two multi class colon tissue classifiers were trained based on two initial feature sets (gray level vs. color based features) to assess feasibility in using features from many color representations for classification between multiple tissue structures.

2.1 Image Data

A total of 94 colon WSI H&E stains from independent subjects were available for the study (46 containing only non-neoplastic tissue and 48 adenocarcinoma). The tissue samples were fixed in formaline prior to embedding in paraffin. The tissue blocks were cut in 4 μm sections on a microtome at room temperature. Specimens were stained with H&E on a Dako CoverStainer with Dako Ready-to-Use reagents using the manufacturer’s validated protocol. The stained slides were scanned on Phillips IntelliSite Ultra-Fast Scanner (40x; 0.25 $\mu\text{m}^2/\text{pixel}$) and were processed on the 20x resolution level. The WSI H&E stains were randomly divided into a training set (38 adenocarcinoma and 36 non-neoplastic) and validation set (10 adenocarcinoma and 10 non-neoplastic). A pathologist annotated each specimen with ROI’s indicating gross regions containing cancer. Additionally, manual annotations of 9 representative colon tissue types (mucosa, muscle, fat, inflammation, red blood cells, necrosis, mucous, connective tissue, and cancer) were obtained. The manual annotations in the validation slides were all assessed and either approved or discarded by an experienced pathologist.

Patches with a size of 128×128 pixels were extracted from the manual annotations to obtain training and validation data from each colon tissue type. For the patch extraction, background pixels were first identified to discard patches without tissue information. Background pixels were defined as pixels with an intensity below 0.05 in the S-channel from the HSV color space, which was determined empirically. After the thresholding, background regions with an area smaller than a cell nuclei (700 pixels selected based on manual segmentations of representative cell nuclei within the data) was included as tissue pixels to ensure pixels within hypochromatic cells nuclei were included as tissue information. Finally, patches with more that 90% overlap with a manual tissue annotation and containing at least 30% tissue pixels were selected for training and validation data for each tissue structure. The criteria caused fat tissue to be excluded as a tissue class as patches overlapping with fat annotations consistently contained less than 30% tissue pixels. Each patch were labeled according to the tissue label of the overlapping manual segmentation. The final training set consisted of 4628 Mucosa, 4015 muscle, 5368 fat, 1435 inflammation, 4691 blood cells, 4396 necrosis, 609 mucous, 4300 connective tissue, and 4606 cancer patches and the validation set of 804 Mucosa, 722 muscle, 1189 fat, 305 inflammation, 282 blood cells, 932 necrosis, 6 mucous, 670 connective tissue, and 1006 cancer patches.

2.2 Feature Extraction

The feature extraction obtains features from the tissue patches for training and validation of the colon tissue classifier. Classic texture and intensity features were selected to assess contribution when using many color representations for the classification. Two initial feature sets were extracted for the study for comparative analysis: (1) Gray level texture features consisting of the 80 features proposed by Kather et al. [6] (2) Color texture features consisting of 7 intensity histogram and 18 gray level co-occurrence matrix (GLCM) texture features extracted from each color channel from RGB, HSV, CIELab, CMYK, and H&E color deconvolution as well as mean RGB and RGB gradient feature images (18 feature images in total). The algorithm proposed by Ruifrok et al. [10] was used for the color deconvolution. The intensity histogram features consisted of mean, standard deviation, coefficient of variation, skewness, kurtosis, 3rd moment, and entropy. The GLCM features were extracted using five distances (1, 3, 7, 15, and 20), four angles (0, 45, 90, and 135°), and 32 gray levels. For each distance the GLCM's were averaged to obtain rotation invariant features. The GLCM features proposed by Haralick et al. [11] were extracted: autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, variance, sum average, sum variance, sum entropy, difference entropy, information measure of correlation, inverse difference, and inverse difference moment. In total, the color feature set contained 1746 features for each tissue patch. The feature values were z-score normalized.

2.3 Colon Tissue Classifier

The extracted features were used to train 7 colon tissue classifiers using 7 initial feature sets: One only containing gray level texture, 5 only containing features from each separate color representation, and one containing all color based features. During a feature analysis it was observed that the feature distributions of each tissue class approx. could be modeled with multivariate Gaussian distributions. Therefore, for simplicity of the study, a Bayes classifier [12] was used for classification and to obtain P_{tissue} for each tissue structure. Each classifier was trained using a leave-one-out forward selection procedure for feature reduction to obtain the most compact and discriminative feature set for each classifier and prevent over-fitting. In each leave-one-out, all tissue patches extracted from one WSI H&E stain was used as validation data and all the other tissue patches extracted from the training data slides were used for training. Precision-recall area under the curve (PR-AUC) for each tissue class and average tissue PR-AUC (PR-AUC_{avg}) were used as performance metrics. PR-AUC for each tissue class were obtained by defining tissue patches from one tissue class as true positives (TP) class and all other tissue patches as true negatives (TN). In each iteration in the feature selection, the feature with the highest cancer PR-AUC (PR-AUC_{canc}) that also improved PR-AUC_{avg} was selected to improve cancer classification while also improving the overall tissue classification. The feature selection was stopped when the improvement in PR-AUC_{canc} or PR-AUC_{avg}

was less than 0.005 between iterations. For validation, the classification performance of the trained models were assessed on tissue patches extracted from the 20 validation WSI H&E stains using the same TP and TN definitions and performance metrics as in the model training. Additionally, one-vs-all PR-AUC's for each separate tissue type as well as one-vs-one PR-AUC between cancer and each separate tissue type were obtained for the best colon tissue classifier to highlight classification performance between different tissue types.

3 Results

3.1 Colon Tissue Classifier

PR-AUC_{canc} and PR-AUC_{avg} obtained on the training and validation data in each colon tissue classifier can be seen in Table 1 (PR-AUC's obtained during training are given in the brackets). The classifier obtained using all color features obtained superior PR-AUC_{canc} compared to the other classifiers with a PR-AUC_{cancer} of 0.930 and 0.950 in the training and validation data, respectively. Additionally, the all color feature classifier obtained high PR-AUC_{avg}'s of 0.886 and 0.836 on the training and validation data, respectively. Therefore, the all color feature classifier were selected as the best colon tissue classifier. PR-AUC's for each tissue type in the selected colon tissue classifier can be seen in Table 2. The tissue types with the worst classification performance were mucosa, inflammation, and mucous with PR-AUC's below 0.9. Additionally, results of the one-vs-one classification between patches from cancer and each separate tissue type can be seen in Table 3. A PR-AUC above 0.95 were obtained between cancer and each tissue types but the lowest classification performance were obtained between cancer and mucosa, inflammation, and necrosis patches, respectively.

Table 1. PR-AUC's obtained for each of the 7 colon tissue classifiers. PR-AUC's are given as validation PR-AUC [training data PR-AUC].

	Gray level texture	RGB	HSV	H& E color deconvolution	CIELab	CMYK	All color features
PR-AUC _{avg}	0.699 [0.677]	0.867 [0.855]	0.838 [0.823]	0.885 [0.846]	0.843 [0.874]	0.837 [0.895]	0.886 [0.836]
PR-AUC _{canc}	0.886 [0.836]	0.906 [0.861]	0.885 [0.916]	0.870 [0.877]	0.871 [0.836]	0.900 [0.875]	0.950 [0.930]

The selected features for the final colon tissue classifier consisted of two intensity features (coefficient of variation (Y-channel) and skewness (Eosin-channel)) and 7 texture features (information measure of correlation (C-channel, distance 3), information measure of correlation (V-channel, distance 3), correlation (L-channel, distance 3), information measure of correlation (RGB gradient, distance 1), correlation (Chromaticity A-channel, distance 15), difference entropy

(H-channel, distance 3), sum variance (S-channel, distance 3)). The selected features were obtained from 9 different color channels and channels from each of the five color representations were represented in the selected features. This indicate that using multiple color representations can provide additional information for the tissue classification. The selected texture features were mainly obtained with a GCLM distance of 1 or 3, indicating that local texture patters contained the most significant information in discriminating between tissue structures.

Table 2. PR-AUC’s obtained for each tissue class in the selected colon tissue classifier. PR-AUC’s are given as validation PR-AUC [training data PR-AUC].

	Cancer	Mucosa	Muscle	Blood cells	Inflammation	Necrosis	Mucous	Connective tissue
PR-AUC	0.950	0.881	0.942	0.900	0.891	0.924	0.626	0.974
	[0.930]	[0.865]	[0.936]	[0.835]	[0.951]	[0.805]	[0.410]	[0.957]

Application of the selected colon tissue classifier on three representative subimages of the H&E stains from the study can be seen in Fig. 1. First column show classification of benign colon tissue where only a few local patches are misclassified as cancer. It can also be seen that the colon tissue classifier can discriminate between different benign tissue structures such as red blood cells (green), muscle (cyan), connective tissue (brown), and mucousa (magenta). The two other columns shows subimages containing cancer tissue in the bottom to right corner in the images. Most of the patches located in cancer tissue are correctly classified as cancer (yellow patches) and most of the benign tissue are not classified as cancer. However, small misclassification problems between cancer tissue and mucosa can sometimes be observed (top of Fig. 1c). Additionally, the classifier could detect necrosis (white) within the tumor tissue in Fig. 1f.

4 Discussion

We have presented a framework for detecting cancer in WSI H&E stains of colon tissue using a multi class colon tissue classifier based on color intensity and texture features. The colon tissue classifier trained based on all color features obtained the best performance with a $PR-AUC_{avg}$ of 0.886 and $PR-AUC_{cancer}$ 0.950, respectively compared to the other classifiers based only on the gray level features proposed by Kather et al. [6] or features from separate color representations. The final feature set consisted of 9 intensity and texture features obtained from feature images from five different color representations. The study indicate that using information from multiple color representations can improve tissue classification within WSI H&E stains.

Table 3. One-vs-one PR-AUC's obtained on the validation data between cancer and each tissue class in the final colon tissue classifier based on all color features

	Mucosa	Muscle	Blood cells	Inflammation	Necrosis	Mucous	Connective tissue
PR-AUC	0.959	0.997	0.999	0.975	0.979	0.999	0.998

The worst one-vs-one PR-AUC's were obtained between cancer and mucosa, inflammation, and necrosis, respectively. This may be explained as tumor tissue may contain inflammatory cells and necrosis which may appear within the cancer annotations and the fact that colon adenocarcinoma originates from mucosa making cancer appearance more similar to mucosa than the other tissues. This confirm our hypothesis that some benign tissue structures makes accurate cancer classification more difficult within WSI H&E stains of colon tissue compared to

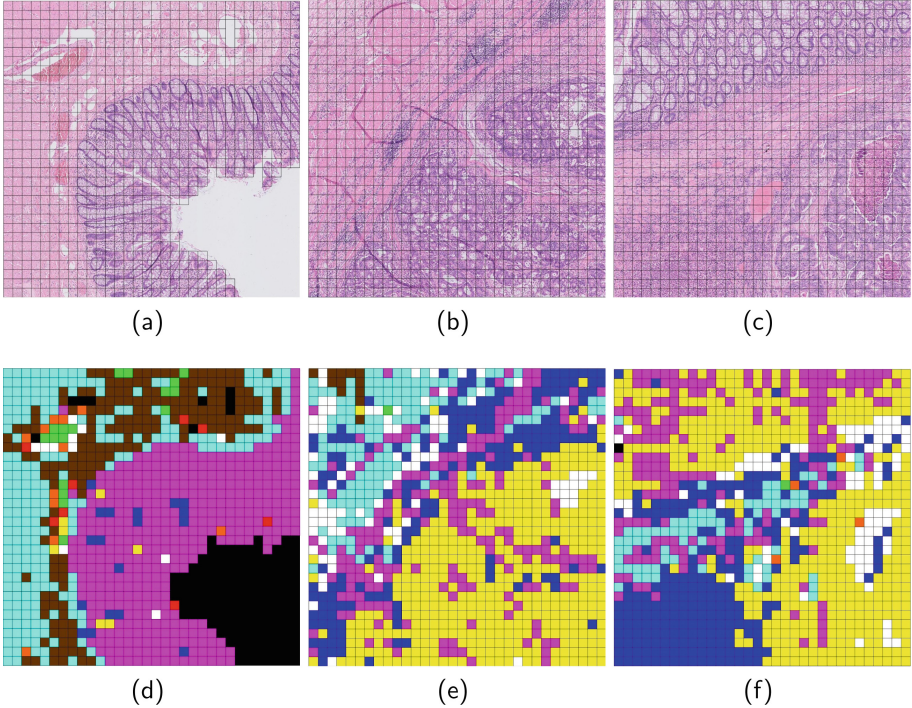


Fig. 1. Application of the colon tissue classifier in three representative H&E stain images (columns). First row show the location of the extracted patches in the original image. Second row show a color coded image of the tissue class with the highest P_{tissue} obtained within the patch (yellow = cancer, magenta = mucosa, cyan = muscle, green = blood cells, blue = inflammation, white = necrosis, red = mucous, and brown = connective tissue). (Color figure online)

others. This may have to be taken into consideration to improve classification accuracy when designing algorithms for automated cancer detection.

Stain normalization was not applied during this study as the framework was developed on WSI H&E stains stained with a standard staining protocol which minimized stain variation between slides. Therefore, the current framework is only expected to be robust to small variations in stain intensity. Stain normalization should therefore be applied as a preprocessing step to use the framework on WSI H&E stains stained with other staining protocols to ensure the generalized performance. Still, the results indicate that a good overall classification accuracy can be obtained between benign tissue and cancer using data obtained from manual tissue annotations of representative tissue structures. The framework can also be used to indicate the location of cancer tissue within WSI H&E stains which may be used for further analysis e.g. for cancer grading.

References

1. Ferlay, J., et al.: Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **386**(136), E359–E386 (2015)
2. Ismail, S.M., et al.: Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia. *BMJ* **298**(March), 707–710 (1989)
3. Doyle, S., Madabhushi, A., Feldman, M., Tomaszewski, J.: A boosting cascade for automated detection of prostate cancer from digitized histology. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) *MICCAI 2006*. LNCS, vol. 4191, pp. 504–511. Springer, Heidelberg (2006). https://doi.org/10.1007/11866763_62
4. Bahlmann, C., et al.: Automated detection of diagnostically relevant regions in H&E stained digital pathology slides. In: *Proceedings of SPIE, Medical Imaging* **8315**, 831504 (2012)
5. Peikari, M., Gangeh, M.J., Zubovits, J., Clarke, G., Martel, A.L.: Triaging diagnostically relevant regions from pathology whole slides of breast cancer: a texture based approach. *IEEE Trans. Med. Imaging* **35**(1), 307–315 (2016)
6. Kather, J.N., et al.: Multi-class texture analysis in colorectal cancer histology. *Sci. Rep.* **6**(1), 27988 (2016)
7. Tabesh, A., et al.: Automated prostate cancer diagnosis and Gleason grading of tissue microarrays. In: *Proceedings of the SPIE International Symposium on Medical Imaging*, vol. 5747, pp. 58–70 (2005)
8. Bejnordi, B.E., et al.: Automated detection of DCIS in whole-slide H & E stained breast histopathology images. *IEEE Trans. Med. Imaging* **35**, 1–10 (2016)
9. DiFranco, M.D., O’Hurley, G., Kay, E.W., Watson, R.W.G., Cunningham, P.: Ensemble based system for whole-slide prostate cancer probability mapping using color texture features. *Comput. Med. Imaging Graph.: Official J. Comput. Med. Imaging Soc.* **35**(7–8), 629–45 (2011)
10. Ruifrok, A., Johnston, D.: Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **23**, 291–299 (2001)
11. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification (1973)
12. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*, 2nd edn. Wiley, Hoboken (2012)