

Symposium – International Academy of Digital Pathology (IADP)

Enhancing automatic classification of hepatocellular carcinoma images through image masking, tissue changes and trabecular features

Maulana Abdul Aziz¹, Hiroshi Kanazawa², Yuri Murakami², Fumikazu Kimura², Masahiro Yamaguchi^{1,2}, Tomoharu Kiyuna³, Yoshiko Yamashita^{1,3}, Akira Saito³, Masahiro Ishikawa⁴, Naoki Kobayashi⁴, Tokiya Abe⁵, Akinori Hashiguchi⁵, Michiie Sakamoto⁵

¹Department of Information Processing, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259-S1-17 Nagatsuta, Midori, Yokohama, Kanagawa 226-8503, ²Global Scientific Information and Computing Center, Tokyo Institute of Technology, 2-12-1-17-6 Ookayama, Meguro-ku, Tokyo 152-8550, ³Medical Solutions Division, NEC Corporation, 5-7-1 Shiba Minato-ku, Tokyo 108-8001, ⁴Faculty of Health and Medical Care, Saitama Medical University, 1397-1 Yamane, Hidaka-shi, Saitama 350-1241, ⁵Department of Pathology, School of Medicine, Keio University, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan

E-mail: *Mr. Maulana Abdul Aziz - abdulaziz.m.aa@m.titech.ac.jp

*Corresponding author

Received: 4 March 15

Accepted: 24 April 15

Published: 03 June 2015

This article may be cited as:

Aziz MA, Kanazawa H, Murakami Y, Kimura F, Yamaguchi M, Kiyuna T, et al. Enhancing automatic classification of hepatocellular carcinoma images through image masking, tissue changes and trabecular features. *J Pathol Inform* 2015;6:26.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2015/6/1/26/158044>

Copyright: © 2015 Abdul Aziz M. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Background: Recent breakthroughs in computer vision and digital microscopy have prompted the application of such technologies in cancer diagnosis, especially in histopathological image analysis. Earlier, an attempt to classify hepatocellular carcinoma images based on nuclear and structural features has been carried out on a set of surgical resected samples. Here, we proposed methods to enhance the process and improve the classification performance. **Methods:** First, we segmented the histological components of the liver tissues and generated several masked images. By utilizing the masked images, some set of new features were introduced, producing three sets of features consisting nuclei, trabecular and tissue changes features. Furthermore, we extended the classification process by using biopsy resected samples in addition to the surgical samples. **Results:** Experiments by using support vector machine (SVM) classifier with combinations of features and sample types showed that the proposed methods improve the classification rate in HCC detection for about 1-3%. Moreover, detection rate of low-grades cancer increased when the new features were appended in the classification process, although the rate was worsen in the case of undifferentiated tumors. **Conclusions:** The masking process increased the reliability of extracted nuclei features. The additional of new features improved the system especially for early HCC detection. Likewise, the combination of surgical and biopsy samples as training data could also improve the classification rates. Therefore, the methods will extend the support for pathologists in the HCC diagnosis.

Key words: Hepatocellular carcinoma, histopathological image classification, image masking, liver tissue, liver trabecular, support vector machine

Access this article online

Website:
www.jpathinformatics.org

DOI: 10.4103/2153-3539.158044

Quick Response Code:

BACKGROUND

Hepatocellular carcinoma (HCC) is a malignant tumor with hepatocellular differentiation and one of the most common cancers in the world. This type of cancer is often diagnosed when the survival time is measured in months which leads to high death rates.^[1] In the meantime, recent advances in computer vision and digital microscopy have expanded cancer diagnosis by automatic classification on histopathological images in various type of cancers, including lung cancer,^[2] breast cancer,^[3] prostate cancer,^[4] and thyroid cancer.^[5] For the purpose of supporting histopathology diagnosis of HCC, NEC Corporation has developed an anatomical pathology diagnosis assisting system for automatic classification of HCC images.^[6] We then tried to enhance the system by developing an experimental system of “Feature measurement software for liver biopsy.”^[7] The system provides pathologists the quantitative measurement of tissue morphology using a digital slide of hematoxylin and eosin (H&E) stained liver tissue specimen, as well as the HCC detection based on those measurement results.

Liver tissues are primarily constructed by liver cell called hepatocyte, stroma and sinusoid (liver’s capillary). The morphological features of hepatocyte nuclei are mostly used as histopathological features in conventional HCC diagnosis.^[6,8] Nevertheless, the other regions of tissue liver also has potential to provide crucial information for the analysis, such as morphological features based on the characteristics of hepatic trabecular, sinusoid, stroma structures, cytoplasm, and fat droplets for diagnosing HCC images.^[8,9]

Hepatocellular carcinoma images itself can be extracted from two kind of liver tissues; surgically resected samples and biopsy resected samples. In surgical sample, the tissue

is removed from the patient through operation by cutting the affected area. Meanwhile, biopsy sample is removed using needles and often utilize computed tomography scan or ultrasound to help guide the surgeon to the right area. Due to the extracting method, the area of extracted tissue in the surgical sample is large and clear while it is thin, torn and a lot of glass area in the biopsy samples. Figure 1 shows the difference between liver tissues from surgical and biopsy samples.

In this study, we are focusing on the classification process of HCC images in the system. Earlier, an automatic classification of HCC images has been introduced by Kiyuna *et al.* based on 13 types of nuclear and structural features, where each feature consists of 6 statistical distributions.^[6] The experiment was carried out on a set of histopathological images extracted from a collection of surgically resected samples. In order to improve the classification performance, we have developed methods to segment the liver tissue and quantify additional tissue features such as trabecular morphology and the changes of the tissue morphology.^[8] We also employed both surgical and biopsy resected samples in the classification process. This is important since the characteristics of both tissue types may give different influence on the feature values extracted from the images. Thus, by combining surgical and biopsy samples in the process, the system will be able to handle and classify HCC images with higher extent by facilitating the characteristics of biopsy samples in addition to the increment of the amount of training data. Some experiments are set up to evaluate the proposed methods using support vector machine (SVM) classifier. The evaluation results on the impact of the segmentation and the additional features and type of data in the HCC detection performance are reported in this paper.

METHODS

Segmenting Histological Components of Liver Tissue

Liver tissues are constructed by several components, such as sinusoids, lymphocytes, red blood cells, and hepatic trabecular.^[8,9] With regards to HCC diagnosis, features from certain components can be more crucial compared to the other. Hepatocyte nuclei, major components of trabecular other than the cytoplasm, have been utilized as one of the main sources of features used by pathologist in HCC diagnosis. However, liver tissue does not consist of hepatocyte only, thus there are nuclei that belong to the other cells in the tissue. In addition, existed some components that are relatively looks similar with hepatocyte nuclei such as lymphocytes. Both nuclei and lymphocytes have round shape and dark color compared to its surrounding in H&E stained image, although lymphocytes have smaller size. When employing image processing to analyze HCC, it is necessary to identify nuclei that belong

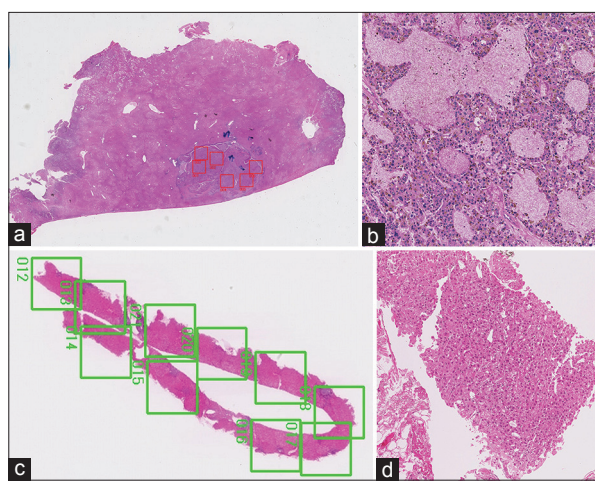


Figure 1: Liver tissue types; (a), (b), (c), and (d) shows the examples of whole slide image (WSI) of surgical sample, region-of-interest (ROI) of surgical sample, WSI of biopsy sample, and ROI of biopsy sample respectively

to hepatocyte only in order to get a reliable hepatocyte nuclei features.

To overcome this problem, we have developed several methods to segment the histological components, including sinusoids,^[8] fat droplets, stroma, hepatic cell cords (trabecular), and nuclei.^[6] Segmenting liver structures can provide important information that will facilitate recognition on each components. For example, by segmenting and extracting sinusoid regions, we can recognize hepatic cell cords and stroma regions. Additionally, some features can only be derived when the histological components have been segmented first, such as the thickness of trabecular.^[8] From the developed methods, 6 masked images were generated for each region-of-interest (ROI) images that is, (1) fat mask, (2) fiber mask, (3) glass mask, (4) nuclei mask, (5) red blood cell mask, and (6) sinusoid mask. Figure 2 shows the example of a masked ROI image along with its original image.

Feature Extractions

Nuclei Features

Kiyuna *et al.* have introduced 13 nuclei and structural features where 6 statistical distributions were calculated for each features, which consisted of, 10, 30, 50, 70, 90 percentiles and standard deviation.^[6] The features are (1) area, (2) peripheral, (3) circularity, (4) length of long axis of elliptic fit, (5) length of short axis of elliptic fit, (6) ratio of short and long axis lengths, (7) texture angular second moment of grey level co-occurrence matrix (GLCM), (8) texture contrast of GLCM, (9) texture homogeneity of GLCM, (10) texture entropy of GLCM, (11) contour complexity of cell, (12) nuclei density, and (13) trabecular thickness. This study uses the same list of features for nuclei features. However, we also extract the 13 features on HCC images that have been masked by the stroma segmentation first, such that those features are derived from hepatocytes only. Therefore, we have two groups of nuclei features; 78 nuclei features that are calculated from unmasked tissues and another 78 nuclei features that are calculated from masked tissues.

Trabecular Features

In the previous study, Kiyuna *et al.* introduced two features related to liver structure, which are nuclei density and trabecular thickness.^[6] To include some other characteristics of hepatocyte structure that are also used by pathologists in analyzing HCC image, we developed several techniques to calculate some features. Komagata *et al.* and Ishikawa *et al.* have developed techniques to quantify some morphological features for hepatic trabecular analysis related to HCC cases.^[6,10-13] For example, hepatocyte trabecular will get thicker and may become multilayer in the case of cancer. Overall, new 10 trabecular features were extracted. They are (1) nuclei-cytoplasm ratio, (2) nuclei density, (3) average number of layers, (4), standard deviation of number of layers,

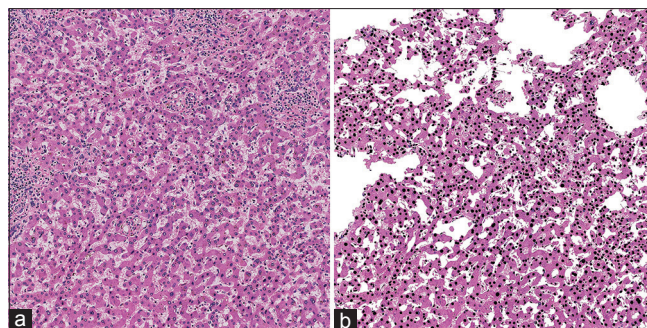


Figure 2: Unmasked (a) and masked (b) liver tissue. Here, all area other than trabecular are masked in white color, hepatocyte nuclei in black color while cytoplasm are kept using the original color

(5) average of nuclei eccentricity, (6) standard deviation of nuclei eccentricity, (7) graph-based nuclei distance, (8) graph-based nuclei alignment, (9) sinusoid weighted sum of mutual angle distribution histogram (WSMA), and (10) trabecular WSMA.

Tissue Changes Features

In addition to trabecular features, we also extended the features by including 11 features of tissue changes. These are features related to the state of fat droplets, cytoplasm, and stroma. This group of features describes the characteristics of the tissue and not directly related to the cancer. Nevertheless, the features will provide additional information to the nuclei and trabecular features for the classification. The features are (1) area ratio of fatty degeneration, (2) number of fat droplets, (3) average value of fat droplet area, (4) standard deviation of fat droplet area, (5) red color of cytoplasm, (6) green color of cytoplasm, (7) blue color of cytoplasm, (8) average value of clearness index, (9) standard deviation of clearness index, (10) area ratio of stroma, and (11) structural feature of stroma. More details on these features can be referred to the works by Ishikawa *et al.*^[6,14] and Murakami *et al.*^[15]

Experiments

Hepatocellular Carcinoma Image Collection

The experiments were performed on a collection of histopathology images supplied by the Department of Pathology at Keio University, Japan. The images consisted of 147 and 116 H&E stained whole slide images (WSI) of surgical and biopsy resected samples respectively. The surgical samples were collected from 111 patients (varying from 1 to 16 samples per patient) while the biopsy samples were collected from 108 patients (1 or 2 samples per patient). From the surgical samples, 1076 ROI images were extracted as 2048×2048 pixels of TIFF images (2-11 images per slide), where 543 of the images were HCC positive. As for the biopsy samples, 1054 ROI images having the same size and image type were extracted (1-27 images per slide), where 550 images were HCC positive. All slide images were captured using the NanoZoomer 2.0HT slide scanner (Hamamatsu Photonics K.K., Hamamatsu, Japan) at $\times 20$ (equivalent to $0.46 \mu\text{m}/\text{pixel}$) as NanoZoomer

Digital Pathology Image (before being converted to TIFF images), and had been labeled by experienced pathologists at the department. Additionally, the pathologist had also graded the HCC positive cases from the surgical samples based on Edmondson and Steiner grading system (EGS) that is, well differentiated (G1), moderately differentiated (G2), poorly differentiated (G3), and undifferentiated (G4). To reduce the color variation due to the staining process, color correction processes introduced by Murakami *et al.*^[15] were applied to each WSIs.

Hepatocellular Carcinoma Image Classification

We used SVM for the classification with radial basis function as the kernel type and LibSVM as the library. Five-fold cross validation was used in order to evaluate the classification performance. When partitioning the ROI images, we kept ROIs from the same slide on the same group in order to avoid ROIs from the same slide being divided into both training and test data. In the process, we made some combinations on the sets of training data from both biopsy and surgery samples. To evaluate the new features, we also did the experiments on several sets of features. As described in section 2.2, we have 4 groups of features as follow.

- 78 unmasked nuclei features
- 78 masked nuclei features
- 10 trabecular features
- 11 tissue changes features.

To see the effect of masking, we did experiments on the 78 unmasked nuclei features and 78 masked nuclei features so that we could compare the rates of each classification. We then combined the masked nuclei features with 10 trabecular and 11 tissue changes features to evaluate the effect of the new features. Note that since the 10 trabecular features are more comprehensive in representing the liver structural characteristics, we removed the “trabecular thickness” feature from the nuclei features, thus generating in total 93 features.

RESULTS

The results of HCC image classifications are summarized in Table 1. It shows that combinations of the new features

with the nuclei and structural features can improve the accuracy for about 1–3% depending on the type of training and test data. For example, in biopsy samples, the sensitivity is improved from 84.7% to 88.2% while the specificity is unchanged (91.9%). Figure 3(a) shows the receiver operating characteristics (ROC) curves when biopsy samples having different set of features are used as the test data along with their sensitivity values when specificity is 90%. The area under curves (AUC) values are 0.941, 0.940, 0.940, and 0.960 in the case of B0, B1, B2, and B3 respectively. On the other hand, Figure 3(b) shows the ROC curves in the case of surgical samples, having AUC values 0.957, 0.955, 0.951, and 0.968 in S0, S1, S2, and S3 respectively.

Table 2 shows the comparison between the true positive rates on HCC grades between the method by Kiyuna *et al.*^[6] with the proposed method in surgical samples. Here, we remove ROI images having 2 or more grades detected in an image for the comparison and only use images that are labeled to one of the grades in EGS. From the table, we can see that the detection rate for the well-differentiated tumors has the highest improvement that is, from 65.0% to 77.5%, by the addition of new features. This is because the nuclei shapes in the case of well-differentiated tumors are similar with normal liver, yet there are deformations in the arrangement of cell pattern. Therefore, appending the new features to the classification process can expand the difference

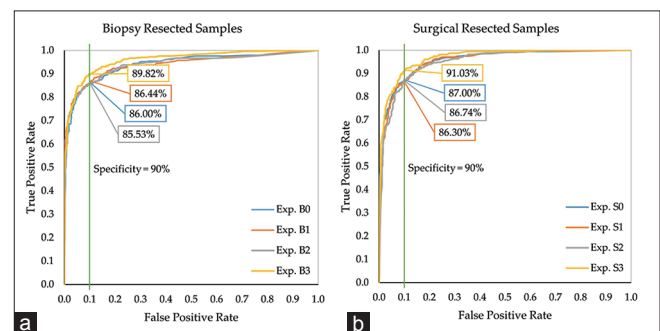


Figure 3: Receiver operating characteristics curves for hepatocellular carcinoma classification on (a) experiment B1, B2, B3, and (b) experiment S1, S2, S3

Table 1: Experiment results

Training data	Test data	ID	Set of features	Sensitivity (%)	Specificity (%)
Biopsy sample	Biopsy sample	B0	78 nuclei (unmasked)	86.36	88.29
Surgery sample	Surgery sample	S0	78 nuclei (unmasked)	88.21	87.99
Combination of biopsy and surgery sample	Biopsy sample	B1	78 nuclei (unmasked)	84.73	91.87
		B2	78 nuclei (masked)	85.27	90.67
		B3	72 nuclei (masked) +21 new features	88.18	91.87
	Surgery sample	S1	78 nuclei (unmasked)	88.95	87.62
		S2	78 nuclei (masked)	89.50	87.62
		S3	72 nuclei (masked) +21 new features	91.34	89.68

Table 2: Comparison on true positive rates between Kiyuna et al. method and our proposed method

Grade	Previous method %	Proposed method %
1.0	65.00 (52/80)	77.50 (62/80)
2.0	91.29 (262/287)	93.03 (267/287)
3.0	96.88 (62/64)	100.00 (64/64)
4.0	95.45 (21/22)	86.36 (19/22)
Overall	87.64 (397/453)	90.95 (412/453)

between the normal tissues and the tissues having low-grade cancer. Even so, the rate of undifferentiated tumors are worsen. As the state of cancer is getting worst, the sinusoid areas become disappear and thus the sinusoid segmentations for undifferentiated tumors are sometimes inaccurate. Moreover, the lack of samples in this grade may also contribute to the misclassification. Further works are required to investigate the grade using more training data and to improve the sinusoid segmentation.

CONCLUSIONS

The combination of nuclear, trabecular and other tissue features enables improved classification rate in HCC detection using SVM. Furthermore, the masking process on the nuclei features facilitates the reliability of the nuclei features since falsely detected nuclei are removed from the quantification. The proposed method improves the chance of early HCC detection, which is very important and has been one of the main challenges in the area, and thus it is recommended by the pathologists. The HCC classification scheme introduced in this paper is implemented in the prototype “feature measurement software for liver biopsy,” and the probability of HCC is visualized for every ROI in the WSI. It will support pathologists in the HCC diagnosis along with the quantitative measurements of tissue morphology.

ACKNOWLEDGMENTS

This work is supported by New Energy and Industrial Technology Development Organization of Japan. Mr. Abdul Aziz would also like to thank the Indonesia Endowment Fund for Education for their support on his study. The authors would like to thank

the reviewers for their valuable comments and suggestions to improve the manuscript.

REFERENCES

- Bialecki ES, Di Bisceglie AM. Diagnosis of hepatocellular carcinoma. *HPB (Oxford)* 2005;7:26-34.
- Wang CW, Yu CP. Automatic morphological classification of lung cancer subtypes with boosting algorithms for optimizing therapy. In: Suzuki K, Wang F, Shen D, Yan P, editors. *Machine Learning in Medical Imaging*. Berlin, Heidelberg: Springer; 2011. p. 217-24.
- Ojansivu V, Linder N, Rahtu E, Pietikäinen M, Lundin M, Joensuu H, et al. Automated classification of breast cancer morphology in histopathological images. *Diagn Pathol* 2013;8:S29.
- Doyle S, Feldman MD, Shih N, Tomaszewski J, Madabhushi A. Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. *BMC Bioinformatics* 2012;13:282.
- Wang W, Ozolek JA, Rohde GK. Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images. *Cytometry A* 2010;77:485-94.
- Gurcan MN, Madabhushi A, editors. *Medical Imaging 2013: Digital Pathology*. Proc SPIE 8676. Florida, USA, Washington: SPIE; 2013.
- Yamashita K, Kiyuna T, Sakamoto M, Hashiguchi A, Ishikawa M, Murakami Y, et al. Development of a prototype for hepatocellular carcinoma classification based on morphological features automatically measured in whole slide images. *Anal Cell Pathol* 2014;2014:2. [Abstract].
- Ishikawa M, Ahi ST, Kimura F, Yamaguchi M, Nagahashi H, Hashiguchi A, et al. Segmentation of sinusoids in hematoxylin and eosin stained liver specimens using an orientation-selective filter. *Open J Med Imaging* 2013;3:144-55.
- Bosman FT, Carneiro F, Hruban RH, Theise ND. *WHO Classification of Tumours of the Digestive System*. 4th ed. Lyon: IARC Press; 2010.
- Ishikawa M, Fujita Y, Aziz MA, Murakami Y, Yamaguchi M, Kobayashi N, et al. Examination of improve the accuracy of HCC classification by quantification of tissue structure from HE-Stained liver specimens. *Proc. of JAMIT Annual Meeting* 2014;33:1-9 (ID OP3-7).
- Ishikawa M, Fukui T, Fujita Y, Murakami Y, Yamaguchi M, Abe T, et al. A method to disarray quantification of trabeculae from HE stained liver specimens. *Proc. of JAMIT Annual Meeting* 2013;32:1-7 (ID OP7-5).
- Ishikawa M, Fukui T, Murakami Y, Yamaguchi M, Abe T, Hashiguchi A, et al. An automatic measurement of morphological feature for analysis of the hepatic trabecula. *IEICE Tech Rep* 2013;112:255-60.
- Komagata H, Kobayashi N, Ishikawa M, Shinoda K, Yamaguchi M, Abe T, et al. High speed algorithm to calculate feature quantities of nuclei graph structures in hepatic histological images. *IEICE Tech Rep* 2014;113:239-44.
- Ishikawa M, Kobayashi N, Komagata H, Yamaguchi M, Abe T, Hashiguchi A, et al. A method to extract fat drops from liver biopsy image using machine learning algorithm. *IEICE Tech Rep* 2014;113:103-8.
- Murakami Y, Abe T, Yamashita Y, Yamaguchi M, Ishikawa M, Hashiguchi A, et al. Color Processing in Pathology Image Analysis System for Liver Biopsy. *Color and Imaging Conference, 22nd Color and Imaging Conference Final Program and Proceedings, Society for Imaging Science and Technology; 2014. p. 184-8.*