Computational Cell Classification Methodology for Hepatocellular Carcinoma

Chamidu Atupelage[*1], Hiroshi Nagahashi [*2], Fumikazu Kimura[#3],
Masahiro Yamaguchi [#4], Tokiya Abe[†5], Akinori Hashiguchi[†6], Michiie Sakamoto[†7]

[*]*Imaging Science and Engineering Laboratory,*
[#]*Global Scientific Information and Computing Center,*
Tokyo Institute of Technology, Japan.
[†]*Department of Pathology, School of Medicine, Keio University, Japan.*
[1]atupelage.c.aa@m.titech.ac.jp

Abstract— **Liver cancer is one of the frequent causes of death in the world. Hepatocellular carcinoma (HCC) is the most common histological type of primary liver cancer. HCC can be graded according to the malignancy of the tumors. Generally, a HCC grade is determined based on the characteristics of liver cell nuclei. This paper illustrates a methodology for classifying liver cell nuclei and grading HCC histological images quantitatively. The liver cell nuclei are classified in three consecutive tasks: nuclear segmentation, fibrous region detection, and nuclear classification. Each task utilizes the pixel-based textural features that are obtained through multifractal computation on digital images. First, the system segments every possible type of nuclei and excludes the nuclei within fibrous regions. Then, it classifies the rest of the nuclei to discriminate liver cell nuclei. For tumor grading, this method utilizes the following four categories of nuclear features: inner texture, geometry, spatial distribution, and surrounding texture. The proposed method was employed to classify a set of HCC histological images into five.**

Keywords— **Multifractal computation, Multifractal measures, Textural feature descriptor, Segmentation, Feature selection, Cancer grading, HCC histological images**

I. INTRODUCTION

Hepatocellular carcinoma (HCC) is the fifth most common cancer in the world and the most common histological type of primary liver cancer [1]. Histopathology is one of the popular HCC diagnostic techniques. In a histopathological examination, pathologists examine the microscopic images and analyse structural behaviour of cells or cellular components and determine the malignancy level of the tissues. Depending on the degree of malignancy, tumors are graded and treatments are determined. Edmondson and Steiner's grading is a commonly used standard for determining the grade of HCC [2]. It defines the four cancer grades: G1, G2, G3, and G4, where G1 is the lowest and G4 is the highest grade. In practice, cancerous regions may be extracted from non-neoplastic tissue regions. Therefore, we incorporated the images of non-neoplastic tissues and categorized them as G0.

Cancer grading is a great challenge for the pathologists, because of unevan growing of tumors and their indistinguishable structural appearance within the grades. In the last decade, a number of image classification-based computer aided diagnosis systems have been introduced to circumvent the limitations in manual histopathological diagnosing methods. Most of the studies proposed computational methods for prostate carcinoma, renal cell carcinoma, and brain tumor astocytomas. However, few studies have been reported on computer-assisted tumor grading of HCC [3][4].

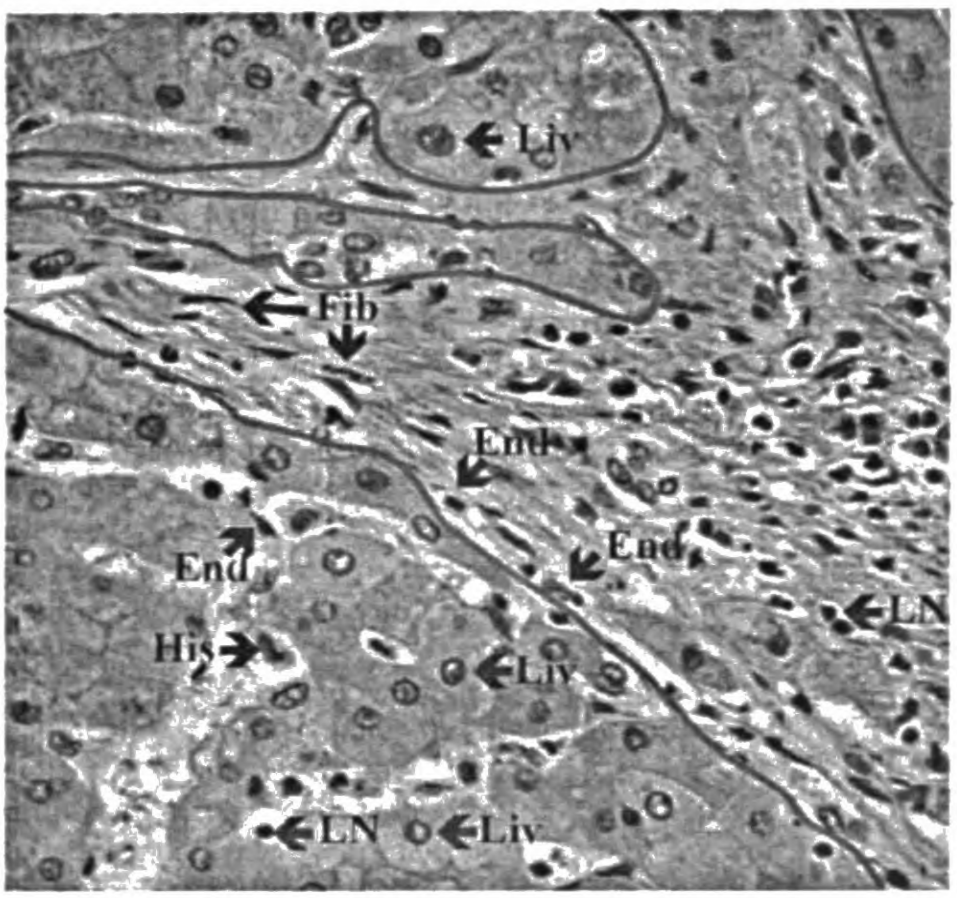

Fig. 1: Microscopic image of HE-stained hepatic liver biopsy specimen. **Liv**: liver cell nucleus, **Fib**: fibroblast nucleus, **LN**: lymphocyte, **End**: endothelial nucleus, **H**: histiocyte. The area bounded within blue lines indicates fibrous region.

A selected region of interest (ROI) in a HCC biopsy specimen may contain different types of cells, such as liver cells, fibroblasts, lymphocytes, endothelial cells, and histiocytes. Fig. 1 shows annotations of different types of cell nuclei in a ROI of hematoxylin and eosin (HE)-stained liver biopsy specimen. HCC grading is performed based on the characteristics of the liver cell nuclei. Therefore, for computational HCC grading, it is important to classify liver cell nuclei.

This paper illustrates a methodology for classifying the liver cell nuclei and investigates the impact of utilizing only the liver cell nuclei for HCC tumor grading. An Overview of the proposed method is illustrated in Fig. 2. We performed a pixel-based textural feature extraction using our previously proposed multifractal feature descriptor [4]. Using the multifractal feature of each pixel, the nuclear and fibroblast regions are segmented. Subsequently, the segmented nuclei located in the fibrous regions are excluded. The rest of the nuclei are classified into two classes: liver cell nuclei and others (endothelial cell nuclei, histiocytes, lymphocyte, etc.). Finally, tumor classification is performed using the textural, morphological and spatial distributions of the liver cell nuclei. In this approach, task 2 and task 3 subsequently exclude the

978-1-4799-1276-6/13/$31.00 ©2013 IEEE

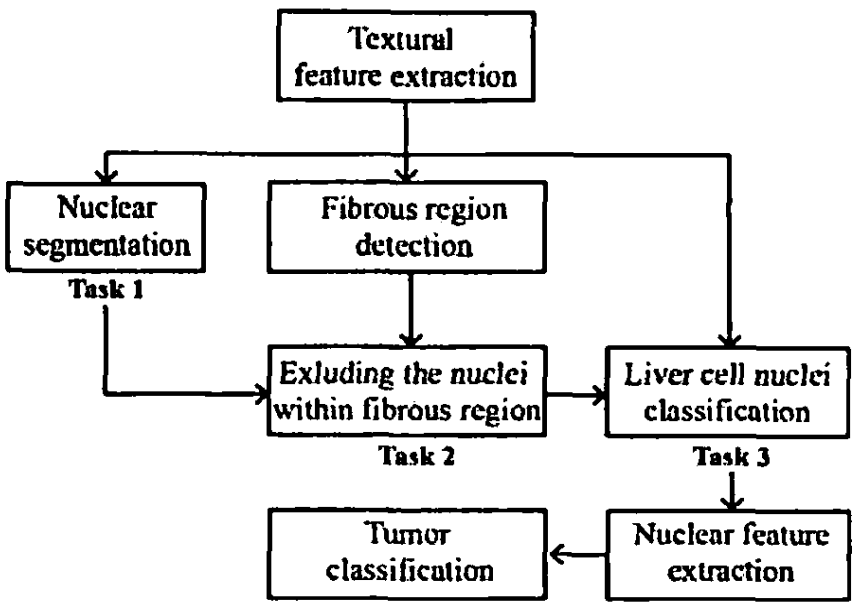

Fig. 2: Overview of the system

nuclei that are not explicitly taken into account in manual HCC grading. Therefore, we investigated the effectiveness of each task for HCC tumor classification. In particular, utilizing a set of HE-stained HCC biopsy image dataset, we analyse the effectiveness of each task by classifying non-neoplastic vs. tumor tissues. Experimental results showed that each task significantly contributes to HCC classification by increasing the correct classification rate (CCR) from task 1 to task 3 continuously. Finally, we utilized the characteristics of liver cell nuclei to grade HCC histological images into five classes. In practice, cells of the G4 grade are easily identified by pathologists because their nuclei are relatively larger and closely located or bounded. In addition, the cytoplasmic regions of these images are very narrow. Therefore, the entire classification was performed in two stages. First, the entire dataset was classified into two groups, G4 and G0—G3, using only the nuclear textural features. Subsequently, the images classified into G0—G3 were employed in multi-class classification using the four categories of nuclear features.

This paper is organized as follows. Sec. 2 describes the previous work. Sec. 3 illustrates the techniques of multifractal feature computation and bag-of-feature (BOF) classification model. Sec. 4 provides the experimental information and evaluation results. Sec. 5 concludes the entire work.

II. Previous Studies

Based on the morphological and textural features of the nuclei, Huang and Lai have proposed a HCC hitological image classification method [3]. They segmented the nuclear regions using a duel-morphological operation and estimated the contours using GVF (gradient vector flow) active contour method. Three categories of features were extracted from the segmented nuclei such as geometrical relationship between nucleus and cytoplasm, geometry of the nucleus, and texture of the nuclei. These features are represented as a 13-dimension feature vector. In the experiments, the authors divided a given ROI into 12 non-overlapping sub-images and used the majority of the prediction for classifying the ROI. Using SVM-based decision-graph classifier, they obtained 94.54% CCR.

Multifractal computation-based textural feature descriptor was proposed for grading the HCC images by Atupelage et al. [4]. They employed BOF-based classification model to compute 300 dimensional feature vector (histogram) for characterizing a given image. They selected 50 sample patches from a given ROI and thresholded the majority of the predictions (threshold-based majority voting rule) for classifying the ROI. They obtained 95.03% CCR for five class classification.

The proposed method in [3] utilized all segmented nuclei for grading the HCC images. Their results might be influenced by non-liver cell nuclei such as lymphocyte, histiocyte, etc. Also the method in [4] utilized the entire texture to characterize the tumors and this might extract non-informative textural regions such as muscles, fiber and fat.

III. Materials and Methods

A. Multifractal feature descriptor

Fractal dimension (FD) is a non-integer exponent used to describe the complexity of fractal structures. Multifractal analysis is a generalization of fractal analysis that can be used to characterize natural images as a spectrum of FDs, i.e., multifractal spectrum [5]. Multifractal computation for digital images is illustrated below.

We can find the local irregularity at a given point (x, y) in an image using a function μ called "multifractal measure", which is a non-integer exponent and described by the Hölder Exponent $h_\mu(x, y)$,

$$h_\mu(x,y) = \lim_{\varepsilon \to 1} \frac{\log(\mu(W_\varepsilon(x,y)))}{\log(\varepsilon)} \quad (1)$$

where, $W_\varepsilon(x, y)$ denotes a square window of side length ε centered at (x, y).

We plot $\log(\mu(W_\varepsilon(x, y)))$ against $\log(\varepsilon)$ for different sizes of neighbors and estimate $h_\mu(x, y)$ by computing the gradient of the linear regression line. This $h_\mu(x, y)$ is called α exponent. Accordingly, we compute α exponents for each pixel in the image and derive a matrix called α feature matrix.

Subsequently, we quantize the entire range of α (from minimum to maximum) into R discrete sub-ranges and derive a binary image for each sub-range. Let α_r be all α values quantized into r^{th} sub-range and I_{α_r} gives a binary value matrix as,

$$I_{\alpha_r}(x,y) = \begin{cases} 1, & \alpha_{rMin} \le \alpha(x,y) < \alpha_{rMax} \\ 0, & otherwise \end{cases} \quad (2)$$

where, α_{rMin} and α_{rMax} denote lower and upper limits of r^{th} sub-range and $\alpha(x, y)$ is the value at point (x, y) in α matrix. Then, we compute a FD for each image I_{α_r} using a method called box-counting algorithm. This method covers the entire image with grids of side length ε', and counts the number of non-empty boxes $N_{\varepsilon'}(I)$. When ε' tends to 1, the limiting value of $N_{\varepsilon'}(I)$ follows the power law $N_{\varepsilon'}(I) \sim \varepsilon'^{-d_B}$, where d_B is a constant, i.e., the FD of I.

$$d_B(I) = \lim_{\varepsilon' \to 1} \frac{\log(N_{\varepsilon'}(I))}{\log(\varepsilon'^{-1})} \quad (3)$$

We compute the FD of I by plotting $\log(N_{\varepsilon'}(I))$ against $\log(\varepsilon'^{-1})$, and estimate the gradient of the linear regression line, i.e., the FD of I. The FDs obtained for each sub-range α_r form a spectrum called $f(\alpha)$, i.e., multifractal spectrum. In addition, for each element in the α matrix has a corresponding FD in the $f(\alpha)$ and it yields a matrix called $f(\alpha)$ feature matrix.

The above description shows that α features observe the pixel's local behavior with respect to the neighborhood and $f(\alpha)$ features contain the pixel's spatial distribution characteristics.

B. BOF-based classification model

BOF-model is a promising approach for characterizing of the visual content of images (or regions). The key ideas behind the BOF classification model are the construction of a codebook and representation of the image by a simple frequency analysis of the codewords in the image. Generally, a codebook is a collection of distinct feature vectors that are called codewords, in which the most representative patterns are coded. These codewords are used to represent the pixel's characteristics. In particular, every pixel is assigned to a certain index of the codeword, where the codeword has a minimum Euclidean distance to each pixel's feature vector. This process may be called labeling. The frequency analysis of the codewords in an image or a region results a histogram, in which each bin represents a codeword in the codebook and its value indicates the frequency of the codeword (labels) contained in the entire image (or region). When different sizes of images or image regions are used, the areas of the histograms are different. Therefore, it is necessary to normalize the histograms. From a statistical point of view, this normalization turns the histogram into a probability distribution function, in which the sums of the total observations are equal to 1. Moreover, the codebook should contain an optimal number of discriminative codewords to obtain a higher classification performance in the classifier. Different codebook optimization (feature-selection) methods have been proposed by [6], [7], [8].

Our proposed method used different BOF models (different codebooks) for the detection of fibrous regions, liver cell nuclei classification, and nuclear feature extraction processes.

IV. Experiments and Results

A. Data acquisition

We obtained HE-stained liver biopsy specimens of 109 HCC patients and each specimen was scanned into a WSI using a scanner, Nano-Zoomer (Hamamatsu Photonics K.K.) with an objective lens of 20x magnifications. The size of the WSI is 33600×21000 pixels. Several experienced pathologists examined these WSIs, and annotated a set of ROIs with one of four categories, G0, G1, G2, G3, and G4. G0 is non-neoplastic tissues and the others are four grades of tumors. Consequently, we obtained a dataset; 93, 72, 67, 48, and 25 ROIs for G0, G1, G2, G3, and G4 categories, respectively. The size of a ROI is 2174×2174 pixels.

B. Textural feature extraction

The α and $f(\alpha)$ features are associated with the multifractal measure μ that was used in (1). In particular, different textural characteristics can be obtained for different multifractal measures. We utilize four multifractal measures to extract α and $f(\alpha)$ feature matrices.

The proposed feature descriptor utilizes four multifractal measures: Maximum:μ_{Max}, Minimum: μ_{Min}, Summation: μ_{Sum}, and Ndiff: μ_{NDiff} as defined in (4a), (4b), (4c), and (4d), respectively [5], [9]. These four measures observe the disparity of the intensities from four different viewpoints.

$$\mu_i(m,n) = \max_{(k,l)\in\Omega} g(k,l) \tag{4a}$$

$$\mu_i(m,n) = \max_{(k,l)\in\Omega^*} g(k,l) \tag{4b}$$

$$\mu_i(m,n) = \sum_{(k,l)\in\Omega} g(k,l) \tag{4c}$$

$$\mu_i(m,n) = \frac{\left(\max_{(k,l)\in\Omega} g(k,l) - \min_{(k,l)\in\Omega} g(k,l)\right)}{\varepsilon} \tag{4d}$$

where, $\mu_{(.)}(m,n)$ represents the measurement at point (m,n). Ω is the square window with side length ε centered at point (m,n). Ω^* represents all of the non-zero pixels of Ω. $g(k,l)$ is the image intensity at point (k,l).

In this study, we compute α and $f(\alpha)$ feature matrices for Red, Green, and Blue (RGB) color channels using the above mentioned four multifractal measures, which yield a 24 dimensional multifractal feature space. Subsequently, we combine the RBG color intensities with the multifractal features and obtain 27-dimensions feature space. More precisely, each pixel of an image can be characterized by a 27-dimension feature vector in the combined feature space.

C. Nuclear segmentation

We utilize a training dataset, in which nuclei and their background regions are manually annotated. We compute the feature vectors for each pixel in the annotated regions and train a random forest classifier for two classes: nuclei and background. For a given image, we perform a pixel-based classification and compute the prediction's posterior probabilities for nucleus class. Subsequently, we normalize these probabilities into range [0, 255] and obtain a gray-scale image, in which high intensities represent the regions of nucleus. The noise in the gray image is removed and boundaries of the nuclei are refined by morphological closing operator with three pixels radius of disk shape structural element. Subsequently, we employ level-set contour estimation method [10] on the resulted image and compute the boundaries of the nuclei. Furthermore, the results are refined by deleting the smaller region, of which area is less than 50 pixels. Fig. 2 gives an example of nuclear segmentation on selected regions of three ROIs.

D. Fibrous region detection

The fibrous regions are detected as block-based classification. All images in the training dataset are divided into 32×32 pixels of non-overlapped blocks and annotated into two classes: fiber and background blocks. Utilizing the proposed textural feature descriptor with a BOF model, we compute the histograms for each block and train a random forest classifier. The histograms computed for non-overlapped blocks are classified using the trained classifier and the noise (false positives) is further refined using non-linear filtering approach.

Let $\{C, C'\}$ be the outcomes of two class classification, $P(b_i|C)$ be posterior probability that b_i belongs to class C, and b_i^k be eight neighborhoods of b_i, where $k = 1, 2, 3, \ldots, 8$. If C be the classifier's prediction of block b_i, then the refined prediction of the block $R_{pred}(b_i)$ is defined as,

$$R_{pred}(b_i) = \begin{cases} C, & P(b_i|C) > mean_k(P(b_i^k|C')) \\ C', & otherwise \end{cases} \tag{5}$$

We annotated 115 ROIs, which include 14520 blocks of fibrous regions and 55383 background regions. The accuracy of the fibrous region detection was verified by a two-class classification of the annotated dataset with 10-fold cross validation. The results indicated that the entire classification obtained 80.72%, 74.10% and 86.20% for precision, recall and CCR, respectively. Fig. 3 shows an example of fibrous region segmentation (red shaded regions indicated the fibrous regions) on six ROIs.

978-1-4799-1276-6/13/$31.00 ©2013 IEEE

E. Liver cell nuclear classification

We exclude the segmented nuclei within the fiber regions and the rest are used for classifying the liver cell nuclei. The selected set of nuclei was annotated by the experts into two classes: liver cell nuclei and others (lymphocyte, histiocyte, endothelial nuclei, etc.). From each annotated nuclei, we extracted textural and morphological features (refer Sec. *F* for detail information) and trained a random forest classifier. To evaluate the accuracy of the liver cell nuclei classification approach, we performed two-class classification for the annotated dataset with 10-fold cross validation. The annotated dataset contained 2585 liver cell nuclei and 2027 other types of nuclei, which are drawn from 78 ROIs. The results indicated that the entire classification obtained 95.17%, 97.13% and 95.12% for precision, recall and CCR, respectively. Fig. 4 graphically shows the results of liver cell nuclei classification method.

F. Nuclear feature extraction for tumor grading

1) Textural features (f_1): We extract the texture of the segmented nucleus regions using the proposed textural feature descriptor and compute a normalized histogram using BOF model. Employing a codebook of 50 codewords, we obtain 50-dimensional feature vector that describes nuclear textural characteristics of a given image.

2) Morphological features (f_2): The morphological features are composed of the area, perimeter and shape of the nucleus. Area is the total number of pixels in the segmented region, perimeter is the length of the contour, and circularity is $perimeter^2/(4\pi \times area)$. A sample image (size: 1024×1014 pixels) contains a large number of nuclei. Therefore, we empirically select three percentiles such as 5%, 50%, and 95% from the average cumulative distributions of each morphological feature. Consequently, nine-dimension of feature vector is obtained to describe the nuclear morphological characteristics of the entire image.

3) Spatial distribution features (f_3): The spatial distribution characteristics of nuclei are extracted from the density and FD. To estimate the density, we randomly locate 50 sub-windows of size 256×256 pixels in the image and count the total number of nuclei within each sub-window. We select three percentiles: 5%, 50%, and 95% from the average cumulative distribution of the densities. We obtain the configuration pattern of the segmented nuclei as a binary image (nuclei: 1, background: 0) and compute its FD using the box-counting method. This feature provides the irregularity and the complexity of the nuclear configurations over the entire image. As a consequence, four-dimensional feature vector is generated to describe spatial distribution patterns of the nuclei.

4) Surrounding textural features (f_4): The segmented nuclei regions were eroded using a structural element of radius 10 and the eroded regions are taken as the surrounded texture. Textural characteristics of the surrounded regions are extracted using the similar manner as described in Sec F.1. Consequently, the surrounding textural features are also represented as a 50-dimensional feature vector.

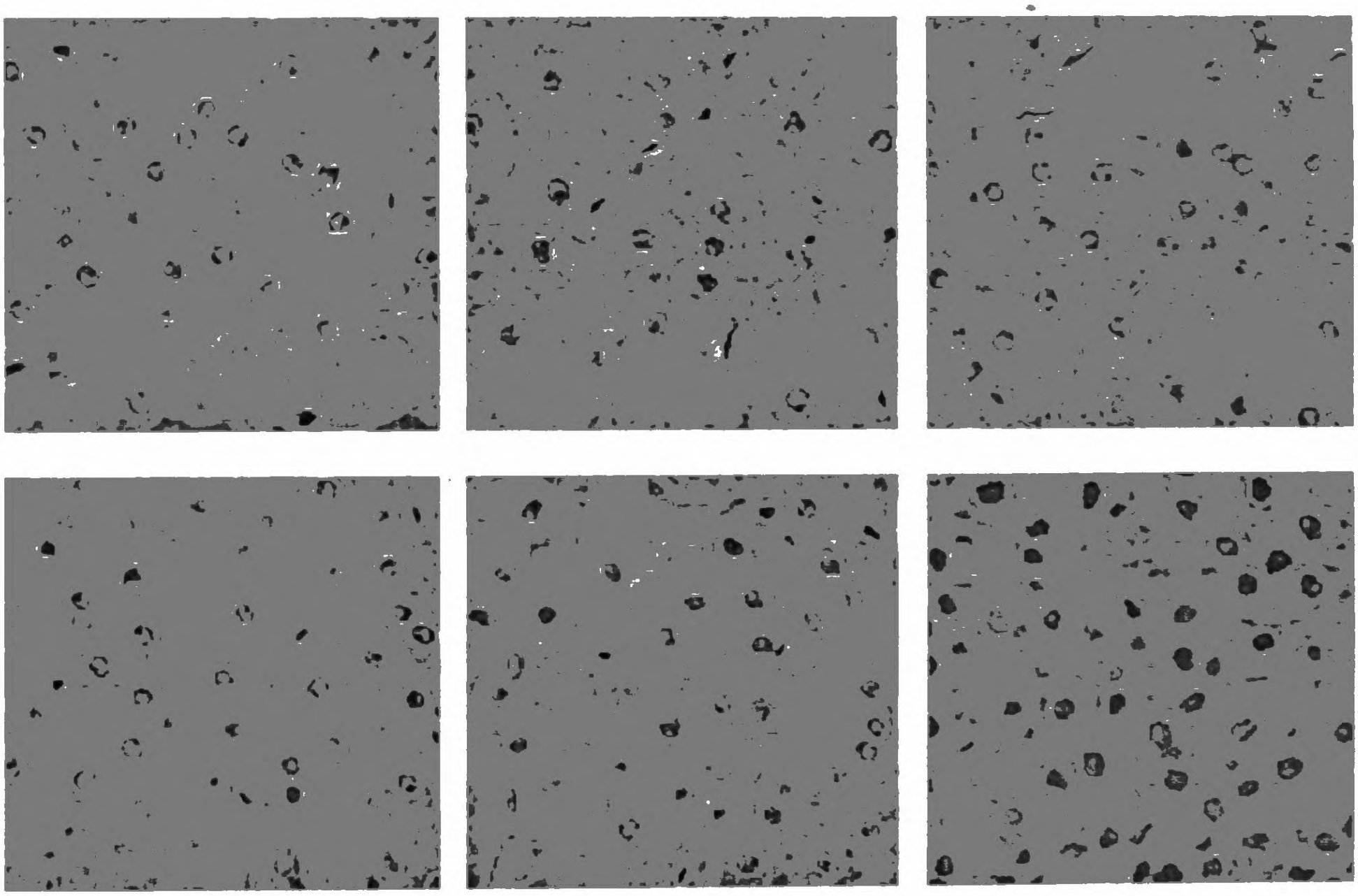

Fig. 2: Results of nuclear segmentation.

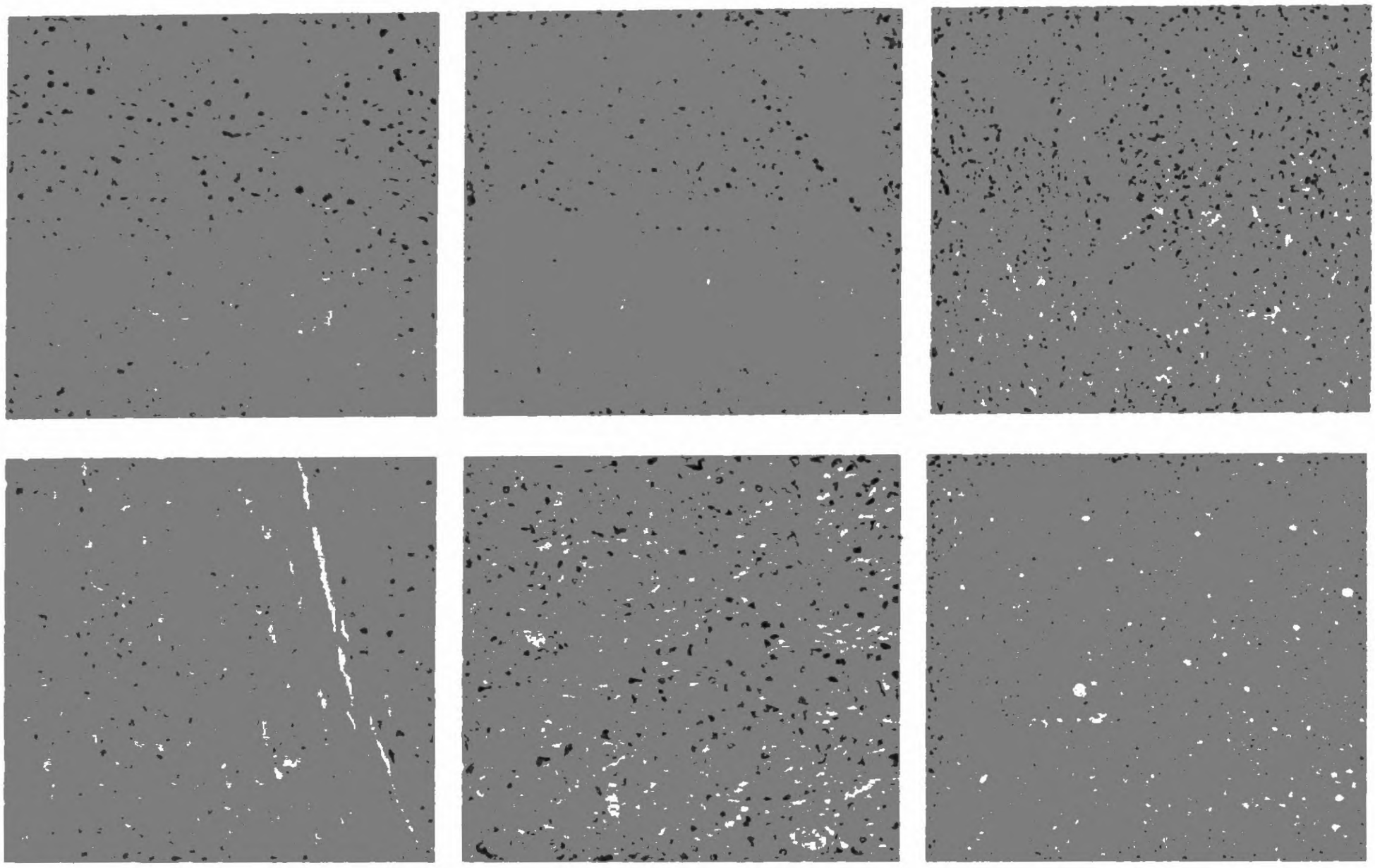

Fig. 3: Results of fibrous region segmentation

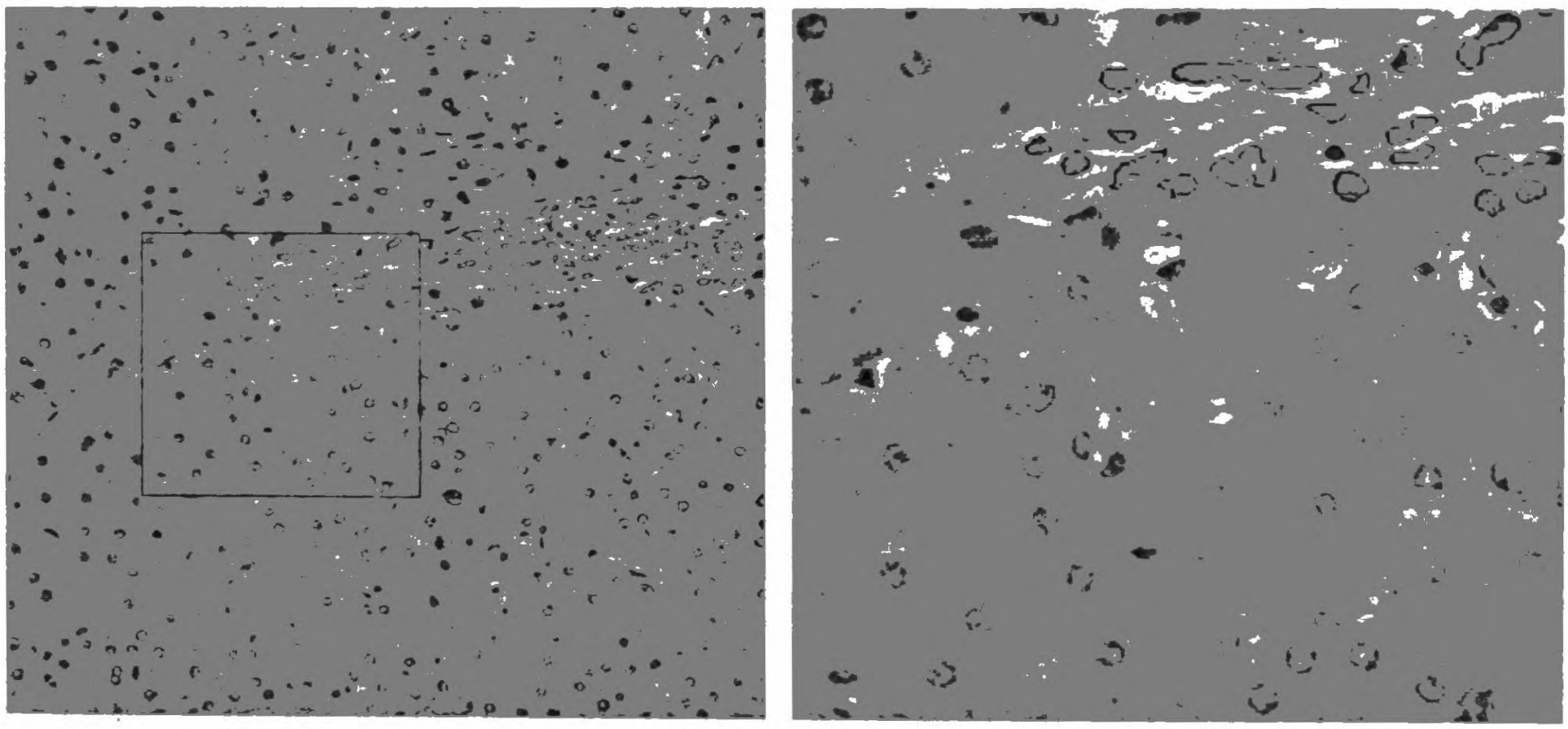

Fig. 4: Computational annotations of nuclei segmentation and classification. (a) HCC image of **1024 × 1024** pixels. (b) a high resolution image of the selected region in (a) (black square). Fibrous regions are shaded in red. Annotations of nuclear contours; fibroblast nuclei: black, liver cell nuclei: red, other nuclei: blue.

G. Evaluation of the method

The entire method contains the following three major tasks: Task 1: nuclear segmentation, task 2: excluding the nuclei within fibrous regions, and task 3: liver cell nuclear classification. Tasks 2 and 3 were used to exclude non-liver cell nuclei. Furthermore, this study extracted the following four categories of nuclear features; f_1:inner texture, f_2: geometry, f_3: spatial distribution, and f_4: surrounding texture. To evaluate the effectiveness of each task and each feature set, we conducted

978-1-4799-1276-6/13/$31.00 ©2013 IEEE

an experiment as a two-class classification. In particular, we extracted the four sets of nuclear features from the outcomes (segmented regions) of each task and utilized each feature set individually to classify non-neoplastic (G0) vs. neoplastic tissues (G1, G2, and G3). The results of the two-class classification for each task and for each feature set are shown in Fig. 5.

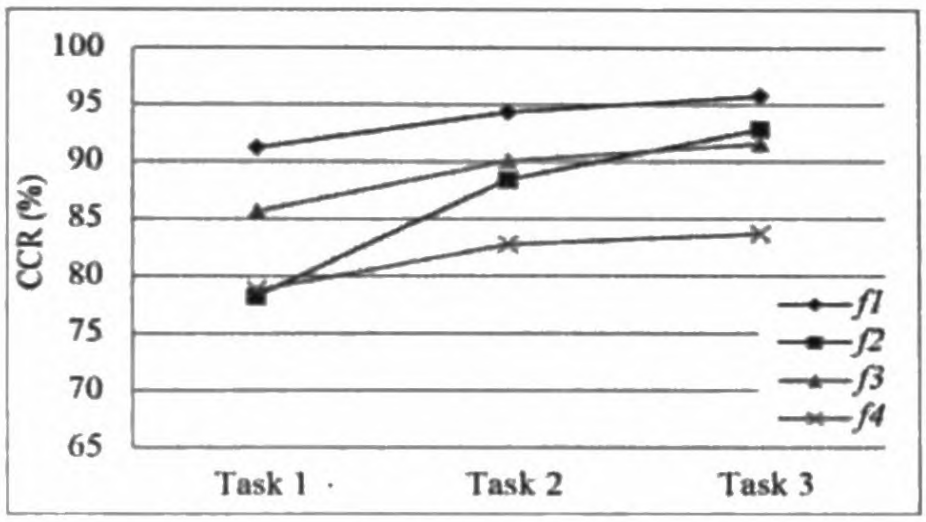

Fig. 5: Two class classification for each task for each feature set.

H. HCC classification

We utilized the proposed method for the grading of HCC histological image dataset into five classes. The experiment was performed as a two-stage classification. First, a two-class classification was performed using a 10-fold cross validation to differentiate G4 and others (G0—G3). Subsequently, the images that were classified as G0—G3 were further classified by multi-class classification.

Generally, tumor regions are arbitrary distributed in a WSI, and an ROI may, therefore, contain small regions of muscles, fiber, fat, and so on. For example, an ROI that is annotated as G1 may contain cells of G0 and G2, as well as smaller amounts of intermediate grade tumor cells that are between G0 and G1, and G1 and G2. Typically, pathologists examine the entire ROI and assign a label on the basis of majority of tumor tissues in the ROI. Accordingly, we drew 16 non-overlapped patches of size 256 × 256 to cover the ROI and 34 patches of the same size from randomly selected locations. Based on the prediction of each patch in a ROI, we estimate the final prediction (grade) of the ROI using a threshold-based majority voting rule.

Let t_i be the total number of predictions of class i out of 50 samples in an ROI. The percentage of the probability of that ROI being categorized into class i is,

$$P_i = \frac{t_i}{50} \times 100\% \quad (6)$$

For a given threshold T, if $maximum(P_i) > T$, then the ROI is labeled as class i, otherwise it is labeled as non.

The multiclass classification result for threshold 70% (in majority voting method) presented in Table 1 in the form of a confusion matrix. The second column shows the total number of images used in each class, and the spanned column "Predictions" shows the predictions of each sub-image. For example, the second row indicates that 72 ROIs were labeled as G1; there were 70 ROIs that were correctly classified, one ROI was classified as G0, and one ROI was classified as G3. Thus, the CCR that was obtained for G1 category was 97.22%.

V. Discussion and Conclusion

Fig. 5 shows that nuclear texture feature (f_1) has achieved the highest classification accuracy for each task. It is an evident for the discrimination capability of the proposed feature descriptor. Morphological features (f_2) has achieved the highest incremental rate. The nuclei in fiber regions are irregular, very close or overlapped, and difficult to compute isolated contours for each nucleus. Therefore, the geometrical properties of the nuclei in fiber regions significantly mislead the classifier. The textural, geometrical, and spatial distribution characteristics of the nuclei such as fiber, lymphocyte, endothelial, etc., may not significantly changed with respect to the grades of HCC. Avoiding the characteristics of these nuclei for computational grading should increase the performance of the classifier and it is obvious from the results.

Generally pathologists utilize only the characteristics of liver cell nuclei for HCC tumor grading. This study extracted texture of cytoplasm regions (nuclei surrounding textural feature (f_4)) using multifractal features and examined their discrimination capacity. In fact, it is difficult to observe the discriminant characteristics of cytoplasmic texture visually. The experimental results shown in Fig. 5 indicated that cytoplasmic texture substantially contributed to improve the accuracy of tumor classification.

Table 1 indicated the results of classifying 305 ROI in to five classes. Among the 305 ROIs only one ROI of G2 has been classified as *non*. As a high-throughput system, the images classified into *non* should be further assessed by the expert. In addition, Table 1 indicated the proposed method is significant to classify low grade cancers and non-neoplastic tissues by obtaining over 94% for G0, G1, and G2 categories. In overall, the proposed method showed significant classification accuracy as 95.97% for five grades of tumors.

This paper proposed a methodology to classify HCC by modelling manual HCC diagnostic routines into mathematical method. In particular, it segmented every possible nucleus and subsequently, classified liver cell nuclei. According to HCC categorization, only liver cell nuclei were used for classifying ROI of microscopic images. Significant of the method is evaluated by analysing the effectiveness of each intermediate process. Finally, entire methodology is assessed by using a large number of ROIs and obtained over 95% overall classification accuracy.

Table 1: HCC multiclass classification results.

Grade	#of ROI	Predictions						
		G0	G1	G2	G3	G4	*non*	CCR%
G0	93	90	2	0	0	0	0	96.77
G1	72	1	70	0	1	0	0	97.22
G2	67	0	1	63	2	0	1	94.03
G3	48	0	0	2	46	0	0	95.83
G4	25	0	0	0	1	24	0	96.00
Total	305							95.97

ACKNOWLEDGMENT

This research is supported by the New Energy and Industrial Technology Development Organization (NEDO) in Japan under the research and development project of Pathological Image Recognition.

REFERENCES

[1] K. Ishak, Z. Goodman, and T. Stochker, *Tumors of the Liver and Intrahepatic Bile Ducts (Atlas of Tumor Pathology)*. Silver Spring (American Registry of Pathology Press), 2001.

[2] H. A. Edmondson and P. E. Steiner, "Primary carcinoma of the liver. A study of 100 cases among 48,900 necropsies," *Cancer*, vol. 7, no. 3, pp. 462–503, 1954.

[3] P. W. Huang and Y. H. Lai, "Effective segmentation and classification for HCC biopsy images," *Pattern Recognit.*, vol. 43, no. 4, pp. 1550–1563, 2010.

[4] C. Atupelage, H. Nagahashi, M. Yamaguchi, A. Tokiya, A. Hashiguchi, and M. Sakamoto, "Computational grading of hepatocellular carcinoma using multifractal feature description," *J. Comput. Med. Imaging Graph.*, 2012.

[5] T. Stojić, I. Reljin, and B. Reljin, "Adaptation of multifractal analysis to segmentation of microcalcifications in digital mammograms," *Phys. Stat. Mech. Its Appl.*, vol. 367, pp. 494–508, 2006.

[6] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proceedings of the international workshop on Workshop on multimedia information retrieval*, New York, NY, USA, 2007, pp. 197–206.

[7] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Proceedings of the Fourteenth International Conference on Machine Learning (ICML 1997)*, Tennessee, USA,, 1997, pp. 412–420.

[8] A. Cruz-Roa, J. C. Caicedo, and F. A. González, "Visual pattern mining in histology image collections using bag of features," *Artif. Intell. Med.*, vol. 52, no. 2, pp. 91–106, 2011.

[9] A. Z. Górski, "Pseudofractals and the box counting algorithm," *J. Phys. Math. Gen.*, vol. 34, pp. 7933–7940, 2001.

[10] L. A. Vese and T. F. Chan, "A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model," *Int J Comput Vis.*, vol. 50, no. 3, pp. 271–293, Dec. 2002.

[11] C. Atupelage, H. Nagahashi, M. Yamaguchi, T. Abe, A. Hashiguchi, and M. Sakamoto, "Multifractal Analysis for HCC Biopsy Images," in *2012 IEICE General conference*, Okayama, Japan, 2012, p. 201.

978-1-4799-1276-6/13/$31.00 ©2013 IEEE