Classification System for BilIN-1 and BilIN-2 based on Normal Nuclei-Array Model

Jae-Won Song[1], Ju-Hong Lee[2], Joon-Hyuk Choi[3]

[1,2] Department of Computer Science & Information Engineering
Inha University, Incheon, Korea
[3]Department of Pathology
YeungNam University College of Medicine, Daegu, Korea

sjw@datamining.inha.ac.kr, juhong@inha.ac.kr, joonhyukchoi@yun.ac.kr

Abstract. Biliary intraepithelial neoplasm with intraductal papillary neoplasm of bile duct is known as a precursor lesion for intrahepatic cholangiocarcinoma. Therefore, to prevent the intrahepatic cholangiocarcinoma, it is important to diagnose the biliary intraepithelial neoplasm. The biliary intraepithelial is classified into BilIN-1, BilIN-2, and BilIN-3 by the atypia of cell and nuclei, loss of nuclear polarity, and disorder of the arrangement of nuclei. This paper classifies BilIN-1 and BilIN-2 using the proposed features based on the Normal Nuclei-Array Model with existing features for nuclei. For showing how the proposed features to improve the accuracy of classification, the experiment is performed by using the proposed features with the existing features of nuclei. The Classifier is used multilayer neural networks with back-propagation algorithm. The experiment results show that the accuracy of classification with the existing nuclei features and proposed features is improved about 13% than the classification with only the existing nuclei features.

Keywords: BilIN, ICC, Normal Nuclei-Array Model, Linear Regression

1 Introduction

Biliary Intraepithelial Neoplasia(BilIN) and Intraductal Papillary Neoplasm of bile duct(IPN) are known as precursors of ICC. So, the detection of these precursors and the classification of stages for them are very important for preventing progression to ICC and providing appropriate treatment to a patient. The BilIN is classified into BilIN-1, BilIN-2, and BilIN-3 by the atypia of cell and nuclei, loss of nuclear polarity, and disorder of the arrangement of nuclei [1]. The grading between BilIN-1 and BilIN-2 of three stages for BilIN is difficult. BilIN-3 just does not occur invasive, but the stage is significantly advanced. So, it is clearly distinguished from BilIN-1 and BilIN-2.

Pathological diagnosis using the tissue of a patient is the most reliable diagnostic method. But, because the pathological diagnosis is performed through screening of tissue in high magnification, this process takes a lot of time and effort [5]. Also, the

pathological diagnosis is very subjective, because it depends on experiences and knowledge of each individual doctor. The pathological diagnosis that requires much time and work has made recent rapid advances with the advent of the digital pathology equipment [12]. For tissue slides that have been converted to digital image by digital pathology equipment, the quantitative analysis and the rapid diagnosis are possible by using the digital image processing techniques. This paper proposed new features based on the Normal Nuclei-Array Model for distinguishing the stages of BilINs. The classification accuracy for the BilIN-1 and BilIN2 that are difficult to distinguish is improved by using the proposed features with the existing features.

The rest of this paper is organized as follows. Section 2 describes the pathological characteristics for distinguishing stages of BilINs. Section 3 talks about the proposed features for classifying the stages. Section 4 compares the experiment results for the existing features and the proposed features. Finally, section 5 presents our conclusions

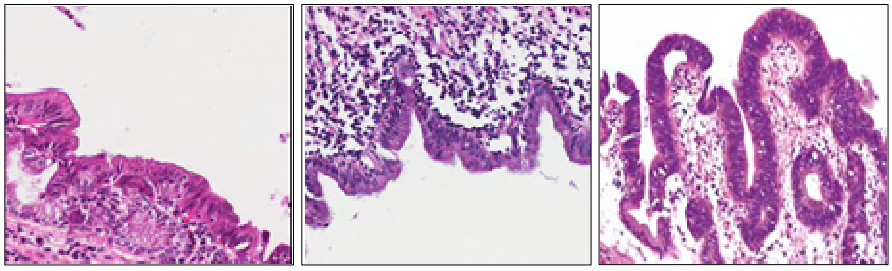
2 Pathological Characteristics of BilIN

BilINs are characterized by atypical epithelial cells with multilayering of nuclei and micropapillary projections into the duct lumen. The atypical cells have an increased nucleus-to-cytoplasm ratio, partial loss of nuclear polarity, and nuclear hyperchromasia. BilINs are divisible into BilIN-1, BilIN-2, and BilIN-3 according to degree of atypia [8]. Table 1 describes the histopathological characteristics for classifying stages of BilINs.

Table 1. Pathological Characteristics for Classifying stages of BilIN [1]

BilIN-1 (biliary intraepithelial neoplasia-1)   These lesions show flat or micropapillary architecture. Nuclei are basally located. Some lesions show focal nuclear pseudostratification; however, the nuclei remain within the lower two thirds of the epithelium. Cytologically, mild nuclear abnormalities, such as subtle irregularities of nuclear membrane, high nuclear/cytoplasmic ratios and nuclear elongation are seen. Nuclear sizes and shapes are relatively uniform, and the presence of large nuclei suggest a diagnosis of BilIN-2 or BilIN-3.
BlIN-2 (biliary intraepithelial neoplasia-2)   These lesions show flat, pseudopapillary or micropapillary architecture. Loss of cellular polarity is easily found, but it is not a diffuse feature. Nuclear pseudostratification reaching the luminal surface is common. Cytologically, dysplastic nuclear changes, which include enlargement, hyperchromasia and irregular nuclear membrane, are evident. Some variations in nuclear sizes and shapes are seen. Peribiliary glands are sometimes involved (glandular involvement). Mitoses are rare.
BilIN-3 (biliary intraepithelial neoplasia-3)   These lesions usually show pseudopapillary or micropapillary architecture, and are only rarely flat. They cytologically resemble carcinoma, but invasion through the basement membrane is absent. Cellular polarity is diffusely and severely distorted with nuclei reaching and piling on the luminal surface. 'Budding off' of small clusters of epithelial cells into the lumen and cribriforming can be seen. Cytologically malignant features with severe nuclear membrane irregularities, hyperchromasia or abnormally large nuclei are typically noted. Mitoses can be observed. Peribiliary gland involvement is sometimes found.

Fig. 1 shows tissue images for BilIN-1, BilIN-2, and BilIN-3.



(a) (b) (c)

Figure 1. (a) Histological picture of BilIN-1, (b) Histological picture of BilIN-2, (c) Histological picture of BilIN-3

In conclusion, the stages for BilINs are distinguished by the atypia of cell nuclei and the disorder of arrangement of nuclei. There are many existing studies for morphological features of nuclei [9, 10, 11]. But, the features for measuring the disorder of arrangement of nuclei have not been studied. In this paper, we propose the features for representing the disorder of arrangement of nuclei for distinguishing stages of BilINs. This paper is focused on the classification of BilIN-1 and BilIN-2. In BilIN-3, the nuclei and cell is seen a very distinct change and the arrangement of nuclei is disorderly maintained. Therefore, we are not concerned with BilIN-3 because it is easily distinguished from other stages.

3 The Proposed Method

3.1 Normal Nuclei-Array Model

The arrangement of nuclei for epithelial cells according to progress stages of BilINs are more disordered. That is, in normal case, the nuclei of epithelial cells are linearly arranged. But, in BilINs, the arrangement of nuclei of epithelial cells becomes more and more disordered and seems papillary form according to progress the stages. From this perspective, we define the Normal Nuclei-Array Model for representing the normal arrangement of nuclei for epithelial cells. The disorder of the arrangement of nuclei for epithelial cells according to the stages of BilINs caused the change of lumen boundary. So, we can measure the disorder of nuclei for epithelium through the lumen boundary.

Now, we suggest the Normal Nuclei-Array Model for representing the arrangement of normal nuclei for epithelial cells using the lumen boundary for the given tissue image. First, for generating the Normal Nuclei-Array Model, the lumen boundary must be identified. The Lumen having white color can be easily identified by the

region growing algorithm because it is adjacent to the epithelial cells. The identified lumen boundary, B_L, by region growing method is defined as follow.

$$B_L = \{p(x, y) \mid \text{boundary points that consist the lumen}\} \quad (1)$$

Now, the Normal Nuclei-Array Model is generated by using the identified lumen boundary B_L. If the given tissue image is normal, the lumen boundary is shown as a straight line. That is, the lumen boundary of normal tissue will appear the linear relationship at x-y coordinate system for the given tissue image. We first assume that the coordinates, x and y, of points that make up the lumen boundary has linear relationship. Therefore, the Normal Nuclei-Array Model is represented as a simple linear equation form.

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2)$$

Where, β_0 and β_1 as unknown parameter for determining the linear equation for the lumen boundary are the intercept and slope respectively. ε_i is the model error. Then, let us assume that the ε_i are random variables with mean zero and constant variance σ^2. Now, the relationship between the coordinates, x and y, of the lumen boundary is assumed as $Y = \beta_0 + \beta_1 x$ that called by the population Normal Nuclei-Array Model. Then, the population parameters, β_0 and β_1, for determining the Normal Nuclei-Array Model are estimated by the points, (x_1, Y_1), …, (x_n, Y_n) for lumen boundary. The Least Square Estimation (LSE) method is used for estimating the population parameters, β_0 and β_1, for the model. The LSE is the method for finding the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1, respectively, that the residual sum of squares is minimized. Then, the estimated Normal Nuclei-Array Model is defined as follows.

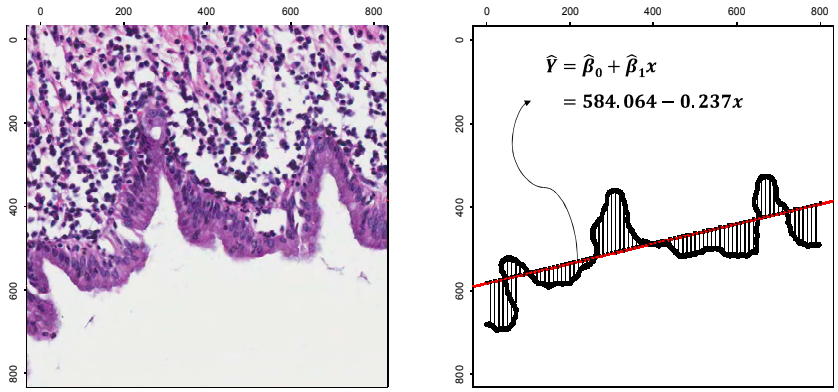
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (3)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (4)$$

$$\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{x} \quad (5)$$

.

Fig. 2 (b) shows the estimated Normal Nuclei-Array Model (red line) for Fig. 1 (a). In Fig. 2, the estimated population parameters, $\hat{\beta}_0$ and $\hat{\beta}_1$, are 584.064 and -0.237 respectively.



(a) (b)

Figure 2. (a) BilIN-2 with abnormal nuclei array, (b) The estimated normal nuclei array model for (a).

3.2 The proposed features

The proposed Normal Nuclei-Array Model is based on the linear regression model in statistics. Therefore, the proposed model can use the statistics for the linear regression model. We propose the features for measuring the disorder for the arrangement of epithelial nuclei using the proposed model. In regression analysis, the coefficient of determination R^2 is used for measuring how well the estimated model (Normal Nuclei-Array Model) performs as a predictors of observation Y. R^2 is computed as SSR(regression sum of squares) divided by SST(Total sum of square).

$$R^2 = SSR / SST \tag{6}$$

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = S_{yy} - SSE \tag{7}$$

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}} \tag{8}$$

$$SST = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = S_{yy} \tag{9}$$

Therefore, for measuring the disorder for the arrangement of epithelial nuclei, we first obtain R^2 from the estimated Normal Nuclei-Array Model. Now, the disorder for the arrangement of epithelial nuclei that denoted by $NormalArrayAbnormality$ is computed as follow equation by using the obtained coefficient of determination R^2.

$$NucleiArrayAbnormality = 1 - R^2 \tag{6}$$

The second feature is to measure the average of residual that means the size of dislocation from the Normal Nuclei-Array for representing the normal arrangement of epithelial nuclei. This feature is defined as *MSE* in the linear regression.

$$MSE = \frac{SSE}{n-2} \tag{7}$$

Where, n-2 is a degree of freedom for *SSE*. These features, *NucleiArrayAbnormality* and *MSE*, based on the Normal Nuclei-Array Model are used as features for distinguishing the stages of BilINs.

4 Experimental Results

Tissue Samples used in experiments are stained by Hematoxylin & Eosin. We received 3 tissue slides from 3 patients with the help of YeungNam University's Department of Pathology. The tissue slides are scanned into a digital image of 200x optical magnification using aperio's slide scanner equipment, "ScanScope CS System (www.aperio.com)". We generate each 30 images including epithelium as a 800x800, 24bit tiff file for BilIN-1 and BilIN-2. Table. 2 shows the number of slides and images that used in experiments.

Table 2. Experiment Data

Type	Number of slides	Number of images
BilIN-1	2	30
BilIN-2	1	30

The experiments are designed to classify between BilIN-1 and BilIN-2 using the existing nuclei features and the proposed features based on the Normal Nuclei-Array Model. The existing features for nuclei used in the experiment as follows: Area, perimeter, width, height, major axis, minor axis, angle, circularity, skewness, ferret angle, feret's diameter, aspect ratio, roundness, and solidity. Refer to [9, 10] for the description of these features. Features for epithelial nuclei and non-epithelial nuclei are extracted individually because epithelial cell and non-epithelial cell are pathologically divided. So, the categories of features used in experiments are epithelial nucleus, non-epithelial nucleus, and the proposed disorder features. For classifying between BilIN-1 and BilIN-2 using the extracted features, we used the multilayer neural network as classifier.

The ratio of training set and test set for learning the neural network is 50 to 50. Fig. 3 shows classification results for the 3 categories for feature and the combined features. The experiment with non-epithelial nuclei features shows the features are not suitable for classifying the stages of BilINs. In experiment using features for epithelial nuclei, classification performance is improved as 73% than non-epithelial nuclei

features. The proposed features based on the Normal Nuclei-Array Model showed classification accuracy, 80%. Finally, the combination of epithelial nuclei and the proposed features has been improved the accuracy of classification by up to 86%. The experiment with the epithelial nuclei and the proposed features shows that the features of these categories are complementary to each other for improving the classification performance.

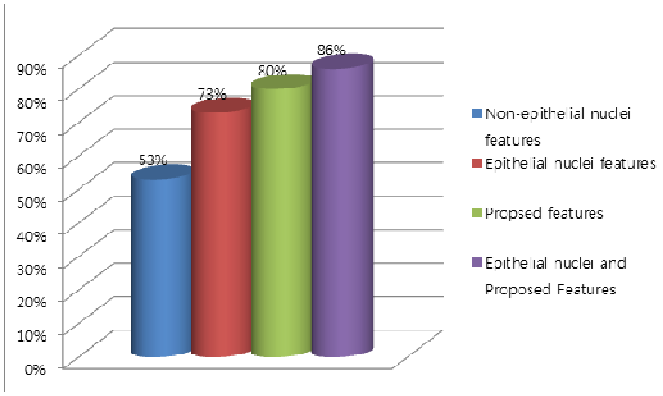


Figure 3. Classification Accuracy for each feature categories

5 Conclusion

The stages of BilINs are divided by the atypia of epithelial nuclei and the state of arrangement for epithelial nuclei. In this paper, we first proposed the Normal Nuclei-Array Model based on the regression model for measuring the disorder of arrangement for epithelial nuclei. The Normal Nuclei-Array Model estimated the normal arrangement of epithelial nuclei using the lumen boundary. We proposed the features for measuring the disorder of arrangement of epithelial nuclei based on the proposed model. The *NormalArrayAbnormality* feature is to measure the disorder of the arrangement of epithelial nuclei as measuring how well the lumen boundary unfitted to the proposed model. Also, we measured MSE for measuring the average size of dislocation from the proposed model to the epithelial nuclei. The experiment shows that the proposed features based on the Normal Nuclei-Array Model contribute to classify the stages of BilINs.

Acknowledgments.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0024216).

References

1. Zen Y, Adsay NV, Bardadin K, Colombari R, Ferrell L, Haga H, et al. Biliary intraepithelial neoplasia: an international interobserver agreement study and proposal for diagnostic criteria. Mod Pathol 2007; 20: 701–709.
2. Jong Ryul Eun, Byung Ik Jang, Jun Young Lee, Kyung Ok Kim, Si Hyung Lee, Tae Nyeun Kim, Heon Ju Lee, Clinical Characteristics of Intrahepatic Cholangiocarcinoma and Prognostic Factors in Patients Who Received Non-surgical Treatment, Korean J Gastroenterol 2009;54:227-234
3. Lee WJ, Park JW. Review: analysis of survival rate and prognostic factors of intrahepatic cholangiocarcinoma: 318 cases in single institue. Korean J Hepatol 2007;13:125-128.
4. Chen MF, Jan YY, Jeng LB, et al. Experience of cholangiocarcinoma in Taiwan. J Hep Bil Pancr Surg 1999;6:136-141.
5. M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot, B. Yener, "Histopathological Image Analysis: A Review," IEEE REIVEWS IN BIOMEDICAL ENGINEERING, vol. 2, pp. 147-171, 2009.
6. A. Andrion, C. Magnani, P.G. Betta, A. Donna, F. Mollo, M. Scelsi, P. Bernardi, M. Botta, B. Terracini, Malignant mesothelioma of the pleura: inter-observer variability, J. Clin. Pathol. 48(1995) 856-860.
7. S.M. Ismail, A.B. Colclough, J.S. Dinnen, D. Eakins, D.M. Evans, E. Gradwell, J.P. O'Sullivan, J.M. Summerell, R.G. Newcombe, "Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia," BMJ, vol. 298, no. 6675, pp. 707-710, 1989.
8. Bosman, F.T., Carneiro, F., Hruban, R.H., Theise, N.D. (Eds.): WHO Classification of Tumours of the Digestive System. IARC, Lyon, pp.222-223 (2010)
9. W.N. Street, W.H. Wolberg, O.L. Mangasarian, "Nuclear feature extraction for breast tumor diagnosis," IS&T/SPIE International Symposium on Electronic Imaging: Science and Technology, vol. 1905, pp. 861-870, 1993
10. T. Allen, "Particle Size Measurement. Powder Sampling and Particle Size Measurement," 5th ed., vol. 1, Chapman & Hall, London, 1997.
11. D. Cigdem, and Y. Bülent, "Automated cancer diagnosis based on histopathological images: a systematic survey," Tech. Rep. TR-05-09, Rensselaer Polytechnic Institute, Department of Computer Science, 2005
12. J.P. Steven, "Digital pathology in drug discovery and development: multisite integration," Drug Discovery Today, vol. 14, no. 19-20, pp.935-941, October 2009.