Development of Cytology Support System using Machine Learning Methods

Hiroki Kiyose, Masakazu Morimoto
Graduate School of Engineering
University of Hyogo
Himeji, Japan

***Abstract*— In order to develop a cytology support system, we first apply U-net to separate the cytoplasm region, nucleus region, and background. After that, we extract 28 features from cell regions and the cancer level was predicted by regression analysis. As a result, we could predict cancer on Mean Squared Error (MSE) 0.0390 and Area Under the Curve (AUC) 0.987 with high accuracy for urothelial cells without nucleus contact.**

***Contribution*—We develop a cytology support system for urothelial cell to reduce the burden on pathologists.**

Keywords—Diagnosis support system; Cytology; Random Forest; Regression

I. INTRODUCTION

Pathologists make a diagnosis by observing glass samples made of affected tissue or cells with a microscope. Among them, cytology is a diagnosis that is performed to detect cancer, in which cells collected from a patient are stained and examined with a microscope. However, it takes a lot of effort to discover abnormal cells among many cells dispersed on the prepared slide without missing them. Currently, it is possible to scan the entire preparation and generate digital images, but it has not led to a reduction in the burden on the pathologist.

Current criteria for cytology involve subjective judgments, and the basis of the judgment may vary by pathologists. When two types of indicators show signs of cancer, the pathologist considers that it is cancer. However, if there is only one type of symptom, it is necessary to judge by combining other features. Therefore, the subjectivity of each pathologist is a problem. Furthermore, the number of pathologists in Japan is about 2,500 and a serious labor shortage. So, in this research, we are going to develop a support system that distinguishes bladder cancer from the microscopic image of urothelial cells and provides objective evidence to support pathologist's diagnosis. Now, there are several methods using deep learning [1] [2], but these methods cannot show the basis of judgement. In this paper, we apply several machine-learning methods to segment cell and nucleus regions and to predict normal and cancer cells by using extracted 28 features from the regions.

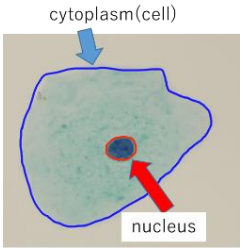

Figure 1. Structure of cell

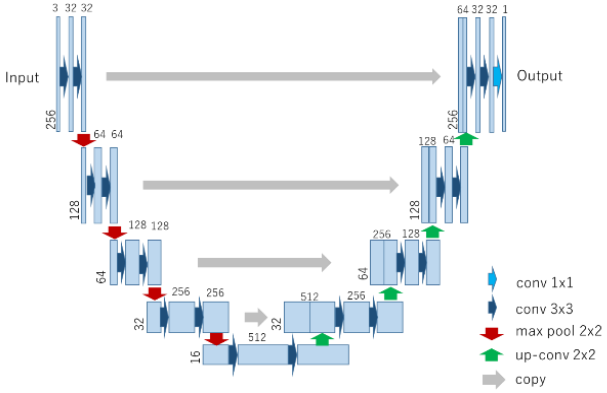

Figure 2. U-net model we used

II. CELL SEGMENTATION BY U-NET

First, we have to segment the cytoplasm (cell) region, nucleus region, and the background region. Here we use U-net [3] to segment the regions. The structure of the U-net used in our study is shown in Figure 2. The procedure for creating a mask is as follows.

1. Resize the input image to 256*256 pixels
2. Input to learned U-net for cytoplasm and nuclear extraction
3. Resize U-net output to original size
4. Binarization with 50% luminance as the threshold
5. Noise removal by opening process

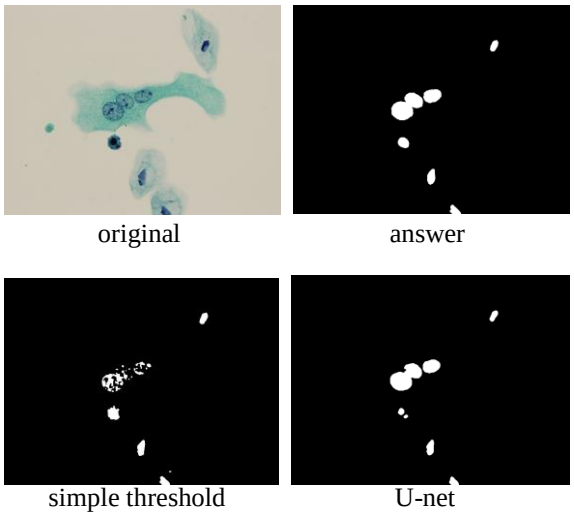

Figure 3 The nuclei mask generated by U-net

The mask created by the above method is shown in Figure 3. We used 53 cytoplasms and 48 nuclei answer masks for learning, and data augmentation was performed 4 times by rotation and inversion processing. The answer mask is cut out of a portion that seems to be the target area visually.

III. Feature Extraction

Feature extraction is performed on the image of the labeled area. The points at which a doctor diagnoses cancer is: enlargement nuclei, irregular-shaped nuclei, coarseness of chromatin pattern, hyperchromatic nuclei, and NC ratio that is area ratio of cytoplasm and nucleus. In order to explain these features, 5group of 28 features were prepared: size, shape, texture luminance, and NC ratio.

Features other than NC ratio are extracted from the nucleus, and the calculated features are assigned to each nucleus. The texture features are calculated from GLCM (Gray Level Co-occurrence Matrix) from the nucleus region, and it has 6 features. We used the GLCM distance which the separation between cancer and normal was highest.

IV. Regression of Cancer and Normal

We performed a regression analysis of cancer and normal. By performing regression, it is possible to show the abnormality of the cell. We used non-overlapping data in the nucleus region to calculate features. The data used were 94 cancer images and 52 normal images, but since the cancer images contained normal cells, we re-labeled 240 cancer samples and 196 normal samples by the nucleus unit. At the time of learning, three kinds of machine learning methods; Random forest, Support vector machine (SVM), and Logistic regression were tried. In logistic regression, it output the probability of cancer. Principal component analysis (PCA) was performed for each of the five groups. In addition, we performed dimension reduction by wrapper method. Evaluation performed by 10-fold cross-validation, and learning was performed with class 0 as normal and 1 as cancer. The result which is an average of 5 times validation is shown to TABLE 1. From the result, it can find out that the case of Random forest is the best MSE and AUC. Also, Figure 4 shows the ROC curve of the Random forest result. The correct answer rate when 99% of cancers were correctly judged was 0.832 for Random forest, 0.837 for SVM, and 0.617 for Logistic regression.

TABLE I. Result of Regression

Machine Learning Method	*MSE*	*AUC*
Random Forest	0.0390	0.987
SVM	0.0518	0.983
Logistic Regression	0.0528	0.980

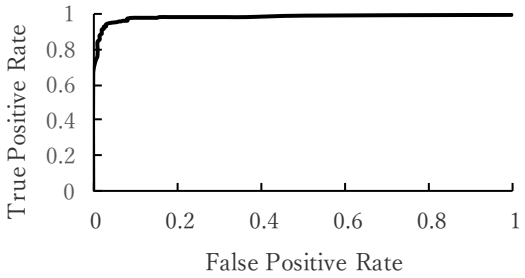

Figure 4. ROC curve of Random Forest

V. Conclusions

In this study, by using U-net for mask creation, we were able to create a mask of cell images with unclear borderlines. By using 28 features and Random forest, we could predict cancer and normal with MSE 0.0390 and AUC 0.987. In the future, we will out-of-focus cells and predict those with overlapping nuclei.

Acknowledgment

This paper is supported by Brain corporation and Dr. Yasunari Tsutihashi of Louis Pasteur Center for Medical Research. This research was partially supported by Ministry of Economy, Trade, and Industry as the Strategic Core Technology Advancement Program 2019.

References

[1] Ling Zhang, Le Lu, Isabella Nogues, Ronald M Summers, Shaoxiong Liu et al., "DeepPap: Deep Convolutional Networks for Cervical Cell Classification", IEEE Journal of Biomedical and Health Informatics, 2017.

[2] Meiquan Xu, Weixiu Zeng, Yanhua Sun, Hunhui Wu, Tingting Wu et al., "Cervical cytology intelligent diagnosis based on object detection technology", 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands, 2018.

[3] Ronneberger, O., Fischer, P. and Brox, T., "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.