CLASSIFICATION OF WHOLE-SLIDE HISTOPATHOLOGICAL IMAGES OF RADIOGENIC PAPILLARY THYROID CARCINOMA

A.A. Kosareva[1], V.A. Kovalev[1], M.V. Fridman[2]

The United Institute of Informatics Problems of the National Academy of Sciences of Belarus1, Minsk, Belarus
Belarusian State Medical University2, Minsk, Belarus
E-mail: kosarevaaleksandra4317@gmail.com

Aiming to analyse radiogenic papillary thyroid carcinoma features, different methods for preparing whole-slide histopathological images were investigated. The Bit-S R50x3 neural model was trained using four datasets: images with min-max rescaling, optical density of image values, dedicated hematoxylin and eosin channels. The best results were achieved using dataset with dedicated hematoxylin channel (*F1*-score = 0.9504).

Key words: whole-slide images; deep learning; neural networks; data preparation.

КЛАССИФИКАЦИЯ ГИСТОПАТОЛОГИЧЕСКИХ ПОЛНО-СЛАЙДОВЫХ ИЗОБРАЖЕНИЙ РАДИОГЕННОГО ПАПИЛЛЯРНОГО РАКА ЩИТОВИДНОЙ ЖЕЛЕЗЫ

А.А. Косарева[1], В.А. Ковалев[1], М.В. Фридман[2]

Объединенный институт проблем информатики Национальной академии наук Беларуси[1], Минск, Беларусь
Белорусский государственный медицинский университет[2], Минск, Беларусь
E-mail: kosarevaaleksandra4317@gmail.com

Обсуждены методы подготовки наборов гистопатологических изображений для решения задачи классификации фрагментов папиллярной карциномы щитовидной железы. Нейросетевая модель Bit-S R50x3 была обучена на четырёх наборах данных: с использованием мин-макс. нормализации, значений оптической плотности пикселов изображений, двух отдельно выделенных каналов изображений (соответствующих позитивной реакции с ядерным красителем гематоксилином и цитоплазматическим красителем эозином). Лучший результат при обучении показал набор данных с искусственно выделенным H- каналом (F1-score = 0.9504).

Ключевые слова: изображения в виде целого слайда; глубокое обучение; нейронные сети; подготовка данных.

INTRODUCTION

The vast majority of thyroid malignancies are diagnosed as papillary carcinoma [1]. As a rule, the disease does not threaten patient's life, however, clinical and histological variants of this tumour with an unfavorable prognosis are encountered [2,3]. Accordingly, to reveal specific biomarkers associated with relapse, progression or even deaths of patients are highly desirable.

“Reading” histological whole-slide images is time-consuming. A large number of characteristics can be missed during a routine investigation. On the other hand, using artificial intelligence methods in the field of whole-slide image analysis opens up new prospects for precise and reproductible detection of carcinoma features that are important for diagnosis and prognostication. Therefore, in this paper we aimed to compare several image datasets for establishing the best approach to deal with whole-slide images of papillary thyroid carcinoma. The following tasks were identified to reach objective:

1. Dataset labeling;
2. Preliminary preparation of tiles of whole-slide images;
3. Training a neural network on four datasets prepared in different ways;
4. Determining the best way to prepare images when solving the problem of binary classification of images of normal tissue and tissue with the presence of papillary carcinoma features.

This paper is organized in three sections:

1. The *Materials* section provides complete information about the prepared dataset for the development of a histopathological image analysis application. The current study used a dataset including images containing carcinoma features and images of normal tissue.
2. The *Method* section describes methods for preparing tiles of whole-slide histopathology images.
3. The *Results* section shows the results of training a neural network model on four datasets of images prepared in different ways. The metric for analysis is *F1-score*.

MATERIALS

The image dataset for the study was combined from whole-slide histopathological images of 129 patients. All these individuals were subdivided into two groups including patients born before the accident at the Chernobyl nuclear power plant (n=104, 35 men and 69 women), and their counterparts born after the disaster (n=25, 14 men and 11 women).

The materials for this study were represented by fragments of images of papillary thyroid carcinoma. The microscopic analysis was performed using routine hematoxylin and eosin staining. Histological slides containing samples of tumor and surrounding tissue were scanned with maximum resolution using the Aperio AT2 scanner system by Leica Biosystems, Germany.

Whole-slide images were labeled by an experienced histopathologist. The labeling resulted in the following characteristics: tumor size, type of histological structure, architectonics, type of tumor growth, exact localization in the thyroid, type and extension of mononuclear infiltrates, invasion of lymphatic and blood vessels,

presence of layered microcalcifications (psamoma bodies), fibrosis, and background pathology.

For this study labeled images were evenly distributed between two classes: normal tissue images and images with carcinoma features.

Finally, each image class included 1900 image tiles of whole-slide images of 256 x 256 pixels in size. The test dataset contained of around 500 image tiles in each class.

METHODS

The *Bit-S R50x3* model was used as a neural network model for the study. The choice of normalization method was based on the classification results of the trained neural network model. Classification was carried out into two classes: images with and without carcinoma features.

In the first step, the images were normalized using min-max scaling. Min-max normalization, rescaling is the simplest method and consists in rescaling the range of features to scale the range in [0, 1] (fig.1, a, fig.2, a).

The following normalization steps were considered:

1. Normalizing images using Optical Density (*OD*) values. The values were determined according to the described method [4] using formulas:

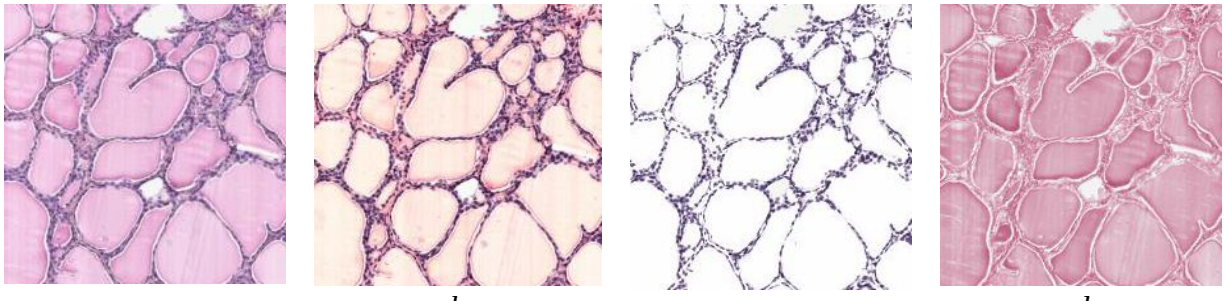
$$OD = -log_{10}(I), \quad (1)$$

$$S = V^{-1}OD, \quad (2)$$

where I – the value of the normalized color image vector in the range of [0;1], S and V - saturation vectors of two-color spots corresponding to different dyes during coloring (fig.1, b, fig.2, b).

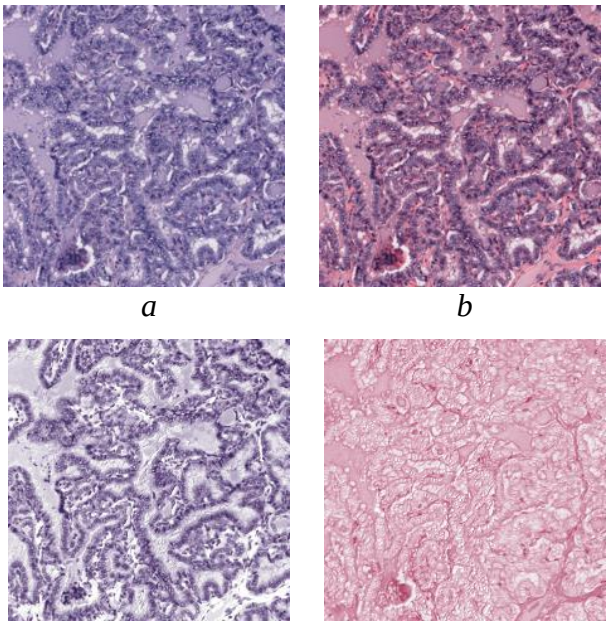
2. Using a dedicated channel corresponding to hematoxylin (*H*-channel) (fig.1, c, fig.2, c).

3. Using a dedicated channel corresponding to eosin (*E*-channel) (fig. 1, d, fig. 2, d).



a *b* *c* *d*

Fig. 1. Original image of normal tissue was analysed with:
min-max rescaling *(a)*, Optical Density *(b)*, dedicated channel corresponding to hematoxylin *(c)*, dedicated channel corresponding to eosin *(d)*



a *b*

c *d*

Fig. 2. Original image of papillary thyroid carcinoma was analysed with: min-max rescaling *(a)*, Optical Density *(b)*, dedicated channel corresponding to hematoxylin *(c)*, dedicated channel corresponding to eosin *(d)*

RESULTS

F1-score [5], calculated from the training results for four test datasets, is presented in Table 1.

Result of training a Bit-S neural network model on different datasets

Image dataset	*F1-score* on the test dataset
Original images with min-max rescaling	0.8951
OD-images	0.9454
E-channel images	0.9194
H-channel images	0.9504

Using images with selected H-channel, it was demonstrated the best results achieves *F1-score* = 0.9504. Image optical density values generally had a positive effect on classification. Besides, separate *E*-channel showed the worst result when classifying tiles with features of carcinoma. Therefore, it can be concluded that the trained *Bit-s R50x3* model has great capacity to detect nuclei and more precisely classify tiles-containing carcinoma features.

In the future, it is planned to use the selected image preparation method for the development an application for searching and quantitative description of tiles of histopathological whole-slide images of papillary thyroid carcinoma.

REFERENCES

1. *Scott A., Hundahl S.A., Fleming I.D., Fremgen A.M., Menck H.R.* A national cancer data base report on 53,856 cases of thyroid carcinoma treated in the U.S., 1985-1995. // Cancer. 1998. Vol. 83. P. 2638-22948.
2. *Bogolyubova A.V., Abrosimov A.Iu., Selivanova L.S., Belousov P.V.* Histopatological and molecular ge-netic characteristics of clinically aggressive variants of papillary thyroid carcinoma. // Arkhiv Patologii. 2019. Vol. 81. P. 46-51.
3. *Fridman M., Lam A.K., Krasko O., Schmid K.W., Branovan D.I., Demidchik Y.* Morphological and clinical presentation of papillary thyroid carcinoma in children and adolescents of Belarus: the influence of radiation exposure and the source of irradiation. // Exp. Mol. Pathol. 2015. Vol. 98. P. 527-531. DOI: https://doi.org/10.1016/j.yexmp.2015.03.039.
4. *Macenko M., Niethammer M., Marron J., Borland D., Woosley J., Guan X., Schmitt C., Thomas N.* A method for normalizing histology slides for quantitative analysis. Proceedings // 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009. P. 9. DOI: 1107-1110. 10.1109/ISBI.2009.5193250
5. The truth of the F-measure, 2007 [Electronic resource]. URL: https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf (date of access: 17.02.2023).