# Computer-Aided Image Analysis in Onco-Pathology

WLODZIMIERZ KLONOWSKI, ANNA KORZYNSKA, RYSZARD GOMOLKA, PAWEL STEPIEN
Nalecz Institute of Biocybernetics and Biomedical Engineering
Polish Academy of Sciences
Ks. Trojdena 4, 02-109 Warsaw
POLAND
wklon@ibib.waw.pl

*Abstract:* - ***Introduction:*** We have proposed new quantitative methods for assessment and classification of selected tumors to assist oncopathological diagnostics. These methods are based on computer-aided analysis of histopathological images. ***Materials and Methods*:** In this paper we propose a simple new methods of quantitative image-based assessment of color tissue histopathology slides. The method is based on **c**olor **f**iltration **p**ixel-by-**p**ixel of the whole virtual slides will be called *CFPP method.* Information contained in such slides that results from staining tissues from biopsies with different dyes helps to reveal details that might otherwise not be apparent. ***Results*:** We demonstrated that with appropriate image preprocessing our method helps out in diagnosis of Diffuse Large B-Cell Lymphoma (DLBCL or DLBL). ***Conclusions*:** In collaboration with pathologists and oncologists our computer-aided method may be adapted and adjusted to assist in diagnosis of other problems in digital pathology.

*Key-Words:* - digital pathology, tumor grading, proliferation index, biopsy, lymphoma, DLBCL, CFPP method

## 1 Introduction

When classifying and grading tumors pathologists often spend a lot of their time sieving through benign tissues [1]. Results of tumor grading often show considerable interobserver and even intraobserver variation. To help out with oncological diagnosis we have proposed new quantitative methods of assessment and classification of selected tumors, methods based on computer-aided analysis of histopathological images.

In particular we applied our own method called *Image Landscapes' Fractal Dimension method (ILF)* that consists in constructing from the analyzed 2-D image two 1-D sequences that we call landscapes, which are subsequently analyzed using Higuchi fractal dimension method (cf. [2]). For example, our studies demonstrated significant correlation between tumor grading made by trained pathologists and results obtained with ILF method in grading Anal Intraepithelial Neoplasia (AIN) [3]. ILF method may be applied in diagnosis of tumors like lymphomas, since it does not require preliminary marking of the tumor's contours that is rather complicated task even for massive tumors.

In this paper we present a very simple computer-aided method of grading *Diffuse large B-cell lymphomas* (DLBCL or DLBL). DLBCL is a neoplasm of large B lymphocytes, a fast-growing, aggressive form of Non-Hodgkin lymphoma (NHL), in fact the most common type of NHL [4]. If left untreated DLBCL is fatal, but with timely and appropriate treatment approximately 70 percent of all patients can be cured. That is why quick diagnosis of DLBCL is so important.

## 2 Materials and methods

We obtained 9 digital slides (below referred to as *specimens*) of tissue samples from patients diagnosed with DLBCL in high grade. We had no specimens of healthy subjects nor of patients with low grade DLBCL. The specimens had been prepared by Pathology Department of the Military Institute of Medicine, Warsaw, stained with hematoxilin (H) and 3,3'-diaminobenzidine (DAB), scanned by 3DHistech slide scanner and saved in .mrxs file format. Each specimen was evaluated and its proliferation index was estimated by the experienced pathologist. The original .mrxs images were of approximate sizes ranged from 0.5 GB to 3.5 GB, but if converted to .tiff format their sizes would rise to 7 - 40 GB. Moreover, empty white background bearing no information made up from 73% up to even 95% of the original images area. So, for further analysis the *representative fragments*, free of artefacts and of endothelial or adipocyte cells' images from blood vessels or fatty tissue, were manually chosen using the Panoramic Viewer 1.15.4.43.061 software (3DHistech). Subsequently, for each of the original specimens its representative fragments were stacked together to form a single .tiff image, below referred as the *whole slide* (Fig. 1). The size of so created whole slides varied from about 1.5 GB up to 15 GB.

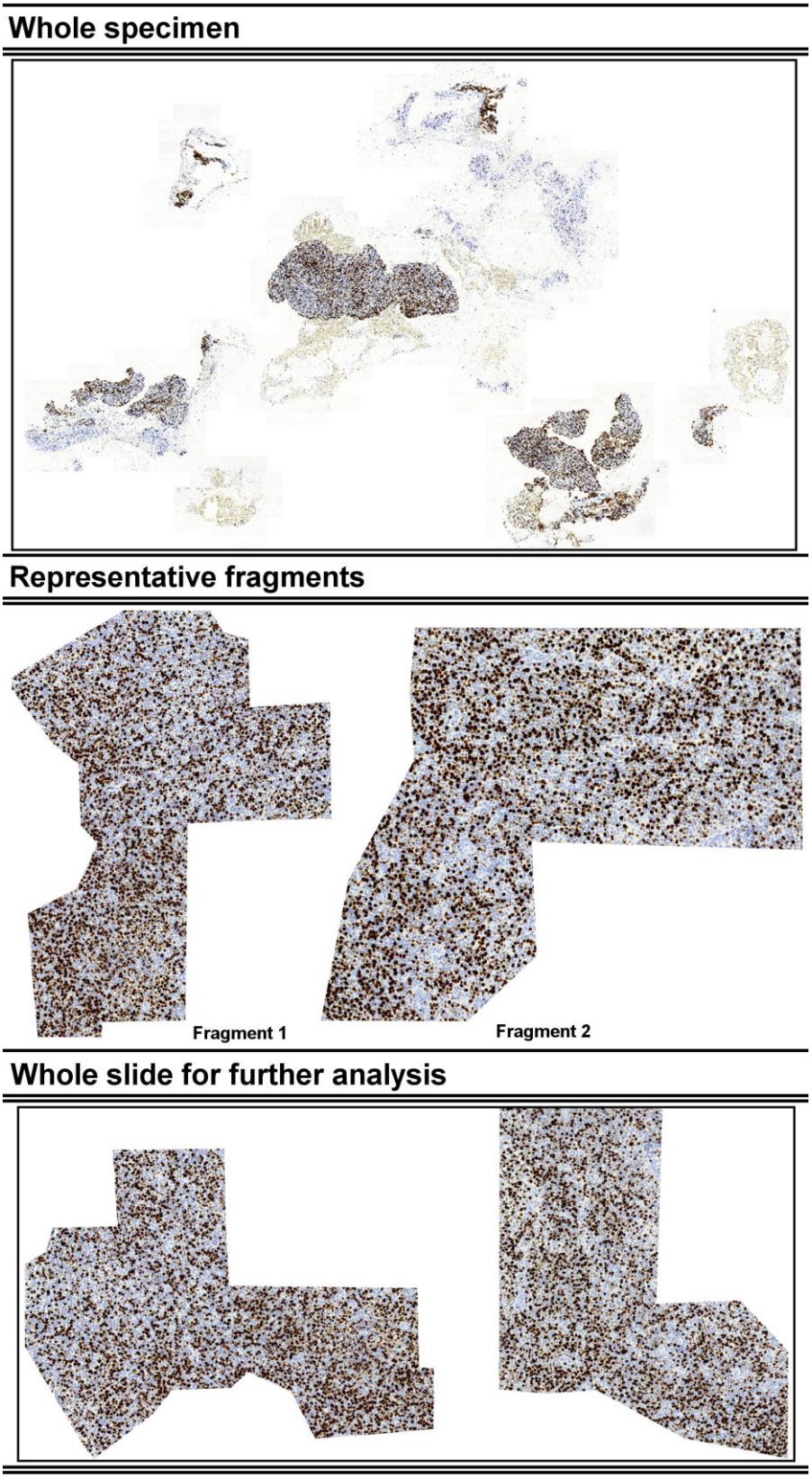


Fig. 1. Example of the whole specimen image, of its representative fragments, and of the whole slide used for further analysis with CFPP method; for this case $PI_p = 0.60$, $PI_c = 0.43$ (cf. Fig. 2. Specimen no. 1.)

## 2.1 Standard method of specimen assessment

When grading DLBCL pathologist concentrates his/her attention on some fragments of the whole slide. First, several *regions of interests (ROIs)* where there are many *hot-spots* i.e. the small boxes containing many DAB-stained nuclei, are manually chosen from the representative fragments. Then for each ROI one counts the number $L_H$ of the *objects* (cells' nuclei) that are H-stained and so turned blue, and the number $L_D$ of the *objects* (cells' nuclei) that are DAB-stained and so turned brown. Then *proliferation index*, $PI_P$ , for each ROI is calculated as

$$PI_P = L_D / (L_H + L_D) \quad (1)$$

and these values are then averaged to calculate the proliferation index, $PI_P$ , for each representative fragment and finally for the whole specimen.

Proliferation index is supposed to serve as a quantitative measure of the level of aggressivity and malignancy of a tumor. The definition (1) of proliferation index is widely used for DLBCLs but in literature there exist different definitions of proliferation indices so it often leads to a confusion.

Computerized DLBCL staging based on the proliferation index (1) is time consuming and expensive since it requires several steps, in particular choosing ROIs and counting of H-stained and DAB-stained objects (cf. [5]-[6]).

## 2.2 CFPP method

Our objective here is to develop a new method which enables quick assessment of tumor proliferation index for dyes-stained images of DLBCLs directly from the whole slides. For this purpose we propose to define proliferation index differently; this proliferation index will then be denoted $PI_C$ (cf. (2)). Instead of choosing manually artefact-free representative fragments and then trying to detect hot-spots to count numbers $L_H$ and $L_D$ of the *objects* we propose to built up the whole slide from each of the whole specimens (cf. Fig. 1.) and calculate in each whole slide the numbers of *pixels* belonging to the nuclei that are H-stained, $N_H$ , and that of *pixels* belonging to the nuclei that are DAB-stained, $N_D$. The new proliferation index, $PI_C$ , is defined as

$$PI_C = N_D / (N_H + N_D) \quad (2)$$

This method that is based on **c**olor **f**iltration **p**ixel-by-**p**ixel of the whole virtual slides will be called *CFPP method.*

Each whole slide is appropriately color filtered pixel by pixel using MATLAB. Color-coding of images most often uses RGB system. The color of a pixel is expressed as a triplet, *(r,g,b)* , each component of which can vary from *0* for the darkest one to *255* for the brightest one.

We open each whole slide (.tiff file) in MATLAB. To compute $N_H$ (number of pixels belonging to the nuclei that are H-stained) computer counts pixels with *(r,g,b)* components greater than minimum values, further denoted with index *Hm* , but smaller than maximum values, further denoted with index *HM*, that is pixels fulfilling the condition

$$(r_{Hm}<r<r_{HM}) \; AND \; (g_{Hm}<g<g_{HM}) \; AND \; (b_{Hm}<b<b_{HM}) \quad (3)$$

Similarly, to compute $N_D$ (number of pixels belonging to the nuclei that are DAB-stained) computer counts pixels with *(r,g,b)* components greater than minimum values, further denoted with index *Dm* , but smaller than maximum values, further denoted with index *DM*, that is pixels fulfilling the condition

$$(r_{Dm}<r<r_{DM}) \; AND \; (g_{Dm}<g<g_{DM}) \; AND \; (b_{Dm}<b<b_{DM}) \quad (4)$$

It is important that for at least one component, *r* or *g* or *b*, the intervals of values for pixels belonging to H-stained (3) and to DAB-stained (4) cells should be completely disjoint. Such color filtration also filters out the remainings of the empty background and majority of artefacts.

When any fragment of a slide is open in MATLAB then clicking on a pixel shows *(r,g,b)* components of this pixel. To choose *Hm* values one clicks subsequently on several dark-blue pixels belonging to a hot-spot, one writes down the shown *(r,g,b)* values and for calculation of $N_H$ one takes the average values of the corresponding component as the limits in (3); similarly, to choose *HM* values one clicks subsequently on several light-blue pixels belonging to a hot-spot. To choose *Dm* values one clicks subsequently on some dark-brown pixels, and to choose *DM* values on some light-brown pixels, one writes down the shown *(r,g,b)* values and for calculation of $N_D$ one takes the average values of the corresponding component as the limits in (4).

In the presented case to count the number of pixels belonging to H-stained nuclei, $N_H$ , we choose the following *(r,g,b)* intervals:

$$(<45 , 180> , <50 , 185> , <160 , 215>) \quad (5)$$

To count the number of pixels belonging to DAB-stained nuclei, $N_D$ , we choose the following *(r,g,b)* intervals:

$$(<40 , 115> , <6 , 80> , <10 , 75>) \quad (6)$$

To calculate $N_H$ and $N_D$ the whole slide may be subdivided into any number of disjoint parts or even read into RAM pixel by pixel. It is important since it makes possible to analyze specimens even on a PC with 16 GB RAM. $N_H$ and $N_D$ are then used to calculate proliferation index $PI_C$ (2). For one specimen time necessary for calculations varied from about 50 seconds to about 7 minutes depending on the size of the whole specimen.

## 3 Results

We have calculated proliferation index $PI_C$ for 9 whole slides of DLBCLs and we juxtapose the results with proliferation index $PI_P$ evaluated for the same slides by a trained pathologist (Fig. 2.).

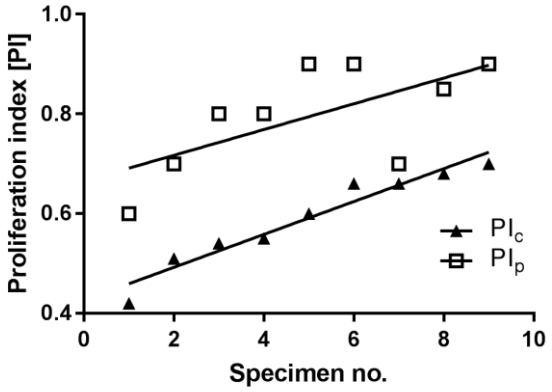


Fig. 2. Proliferation indices $PI_c$ and $PI_p$ for 9 specimens.

The ranges of $PI_C$ for whole slides and of $PI_P$ evaluated for these slides by pathologist are comparable - *0.28 vs. 0.30* - but absolute values of $PI_C$ are about *0.20* lower than corresponding values of $PI_P$. This difference is obvious since pathologist while visually evaluating a specimen tries to find hot-spots and $PI_P$ takes its highest values locally just for the hot-spots. The calculated $PI_C$ and the $PI_P$ evaluated by pathologist change concurrently i.e. if for a given case $PI_P$ is equal or greater than for another case then also $PI_C$ computed for the given slide is equal or greater than that for the other case. Statistics shows that correlation coefficient of $PI_C$ *vs* $PI_P$ is equal *0.71*. Only one specimens behaves not in agreement with this rule. The most important is that calculated $PI_C$ still shows that these are high grade DLBCLs.

## 4 Conclusions

CFPP method of calculating proliferation index $PI_C$ for grading DLBCL neoplasms from stained histological slices is quick and simple (cf. [7]). It gives the results that well correlate with proliferation index $PI_P$ estimated by a trained pathologist. The method is much simpler and much less time consuming than the methods that try to emulate pathologist's way of thinking when estimating $PI_P$. Of course, medical decision still remains to be made by a pathologist, but such a quick computer-assisted grading of DLBCLs may be quite helpful in diagnosis. By appropriate changing of color filtration thresholds CFPP method may be adapted to analysis of other problems in digital pathology.

**Acknowledgements**

This work was partially supported by Nalecz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, through statutory activities.

***References:***

[1] R.K.Kumar et al. Virtual microscopy for learning and assessment in pathology, *The Journal of Pathology*, 304:613-618, 2004.

[2] W. Klonowski, Applications of Chaos Theory Methods in Clinical Digital Pathology, in: *Handbook of Applications of Chaos Theory*, Ch.H.Skiadas & C.Skiadas, Eds. CRC Press, Boca Raton, New York, 2016, pp.681-690.

[3] W. Klonowski et al. Application of Higuchi's Fractal Dimension in Analysis of Images of Anal Intraepithelial Neoplasia, *Chaos, Solitons and Fractals*, 48:54-60, 2013.

[4] S.H.Swerdlow et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* 127:2361-2364,

[5] T.Markiewicz, A.Korzynska et al. MIAP - web-based platform for the computer analysis of microscopic images to support the pathological diagnosis. *Biocybernetics & Biomedical Engineering* 36(4):597-609, 2016,

[6] Z.Swiderska, A.Korzynska et al. Comparison of the manual, semiautomatic, and automatic selection and leveling of hot spots in whole slide images for Ki-67 quantification in meningiomas. *Anal Cell Pathol* 2015:1–15.

[7] E.M.Brey, Automated selection of DAB-labeled tissue for immunohistochemical quantification, *The Journal of Histochemistry & Cytochemistry*, 51:575–584, 2003.