Method of Immunohistochemical Slide Analysis*

Serhii Potapov[1,†] and Vitaliy Gargin[1,2*,†]

[1] Kharkiv National Medical University, 4, Nayky, 61022, Kharkiv, Ukraine
[2] Kharkiv International Medical University, 38, Molochna, 61001, Kharkiv, Ukraine

Abstract

Immunohistochemical (IHC) studies allow obtaining additional information about the state of tissues, in particular the presence of a tumor process. Methods of computer analysis of color digital images make it possible to reduce or even completely eliminate the subjectivity of the study, as well as obtaining reliable quantitative data, which makes IHC studies more objective when solving, for example, diagnostic, prognostic and research tasks. In that connection, the goal of our study was to develop a method for objective analysis of IHC in order to improve the interpretation of the obtained results. The authors propose a two-stage, colorimetry-grounded pipeline for objective analysis of immunohistochemical slides. Images are acquired with a light microscope and processed in MATLAB; RGB values are converted to CIE XYZ and then to CIE Lab, after which K-means clustering is applied first to segment marker, background, nuclei, and membranes, and then to stratify expression by lightness into weak, medium, and strong levels. The approach argues for hardware-independent color description and uses ΔE as the clustering metric, yielding a scalar "S" intended to reduce observer subjectivity. Qualitative comparisons indicate that grayscale and raw RGB segmentations confound marked and unmarked tissues, whereas Lab-space segmentation isolates immunopositive regions and grades expression on testicular tumor exemplars. Suggested algorithm when first stage applies K-means clustering in CIE Lab space to split an image into four classes, second stage 2 re-segments the "marker" class by lightness L* into three expression bands using fixed L* thresholds, and reports the relative positive area S as the quantitative readout (implementation is in MATLAB and relies on an RGB-XYZ-Lab conversion and Euclidean ΔE in Lab as the distance metric) allows to obtain objective data about IHC microspecimen.

Keywords

pathology, image, analysis, immunohistochemistry, CIE Lab

1. Introduction

Morphological research has been a crucial part of medicine for many years [1]. Immunohistochemical studies (IHCS) allow obtaining additional information about the state of tissues, in particular the presence of a tumor process [2, 3]. Modern treatment protocols require objective interpretation of microscopic preparations, especially in relation to the genitals [4, 5].

Methods of computer analysis of color digital images make it possible to reduce or even completely eliminate the subjectivity of the study, as well as obtaining reliable quantitative data, which makes IHCS more objective when solving, for example, diagnostic, prognostic and research tasks [6, 7]. However, there are also problems of quantitative assessment of digital images, which are associated with the limitations of IHCS preparation techniques [8, 9]. This is the quality of the camera and microscope, and the different thickness of tissue sections, and the time of visualization reaction, which is always selected empirically, and the lack of standardized indicators and parameters for quantitative assessment of IHCS [10, 11].

The main differences between the images are as follows: the expression of markers can be determined in different cellular and tissue structures (nuclei, cytoplasm of cells, intercellular substance, membrane structures), and the Relative area (S) of marker expression, as well as its lightness (L), in different tumors can vary significantly. It is obvious that the immunopositive areas in the images have different sizes, the labeled areas are characterized by different structural and

* ProfIT AI'25: 5th International Workshop of IT-professionals on Artificial Intelligence, October 15–17, 2025, Liverpool, UK
1* Corresponding author.
† These authors contributed equally.
pathomorphologist@gmail.com (S. Potapov); vitgarg@ukr.net (V. Gargin)
0000-0002-5718-3341 (S. Potapov); 0000-0001-8194-4019 (V. Gargin)

color properties, and the only constant feature of such areas remains the “brown color” in the subjective perception.

It should be noted that a feature of digital color images of histological preparations is the significant instability of their color content. It is explained by the fact that in different preparations under the influence of the same markers, chemical reactions proceed differently and the properties of tissue structures in different preparations differ from each other, and these differences lead to slightly different results of tissue reactions to markers. Therefore, in different digital images, the color coordinates of the same types of tissues differ significantly. In addition, in the areas stained with the marker, it is necessary to distinguish different levels of expression - from the lightest to the darkest, and determine the level L of these areas. Automated methodic could be useful in that area as it is developed with other medical aspects [12-15].

Based on the above, the goal of our study was to develop a method for objective analysis of immunohistochemical micropreparations in order to improve the interpretation of the obtained results.

2. Methods

The proposed method allows to separate the areas of labeled and unlabeled tissues in the image by color features and to determine the levels of marker expression intensity. The color is encoded in the coordinates of the CIE Lab space, which is designed according to the principles of the human visual system. Therefore, the representation of the colors of labeled and unlabeled tissues in such coordinates is similar to human visual perception. Unlike the representation in grayscale or RGB values, the proposed method at the first stage of segmentation allows to separate the areas of labeled and unlabeled tissues in the image, and during repeated segmentation – to separate the labeled areas by the levels of marker expression precisely by the CIE Lab color coordinates. The use of the automatic mode of digital image segmentation at the first and second stages allows to avoid the subjective factor during the morphometric study, and at the same time accelerates the process of obtaining the result, and also allows to reduce the number of errors associated with the subjectivity of the researcher's perception, to develop clear morphometric criteria for evaluating the results of the expression of certain markers.

For segmentation of color digital images, the most widely used hierarchical method is the k-means clustering method [16, 17].

At its core, the k-means algorithm is based on minimizing an objective function that is equal to the sum of the squares of the distances from all points of a cluster to its center. The objective function, which is based on the sum of the least squares criterion, is defined as follows (1):

$$J = \sum_{i=1}^{K} \sum_{C \in S_i} w(C) \left\| C - \bar{C}_{S_i} \right\|^2 \quad (1)$$

where $\bar{C}_{Si}$ means the center of the cluster, K is the number of clusters, $w(C)$ weight of point c, $\|C - \bar{C}_{Si}\|^2$ expression of the quadratic norm for calculating the distance between points.

In classical segmentation methods [15], local features are smoothed and represented as vectors in a metric space, thus describing each image region by an averaged feature vector (center). The square of the (weighted) Euclidean distance is most often used as a measure of difference in this approach.

Given a dataset X, the K-means algorithm minimizes the objective function iteratively. This process consists of several steps:

Step 1. Select K initial centers $\bar{C}_{S1}, \bar{C}_{S2}, \ldots, \bar{C}_{Si}$.

Step 2. At the t-th iterative step, distribute the elements of the set X between K clusters taking into account the relation (2):

$$C \in S_i^t \; if \; \|C - \bar{C}_{Si}^t\|^2 < \|C - \bar{C}_{Sji}^t\|^2 \tag{2}$$

for all $j = 1,2,\ldots,K$, $j \neq i$, where S_i^t denotes the set of points for which $\bar{C}_{Si}$ is the center of the cluster. In other words, the cluster S_i^t is filled with points for which the condition (3)

$$S_i^t = \left\{ C \left| \left\| C - \bar{C}_{S_i}^t \right\|^2 < \left\| C - \bar{C}_{S_j}^t \right\|^2 \right. \right\} \tag{3}$$

for all $j = 1,2,\ldots,K$, $j \neq i$.

Step 3. Based on the results of step 2, new cluster centers are calculated $\bar{C}_{Si}^{t+1}$ such that the objective function decreases. New centers are formed according to the relation (4):

$$\bar{C}_{S_i}^{t+1} = \frac{\sum\limits_{C \in S_i^t} w(C) \times C}{\sum\limits_{C \in S_i^t} w(C)} \tag{4}$$

Step 4. If all cluster centers have not changed as the iteration step increases, the procedure stops. Go to step 2.

The behavior of this algorithm strongly depends on the value of K, the choice of cluster centers, and the geometric properties of the input data. Nevertheless, the simplicity of the method has provided it with wide application in pattern recognition, image processing, and machine vision problems. An overview of the capabilities of the method and algorithms for its implementation can be found, for example, in the work of S.M.A. Burney, H. Tariq [18].

One of the main criteria of homogeneity for a group of pixels in a cluster when performing image segmentation is color. Historically, the earliest approaches relied on the use of the RGB space to describe color data [19, 20]. However, this color space poorly describes the features of human color vision, so its application for segmentation problems is not always effective. Consequently, many alternative color spaces have been proposed and applied [21-24].

Among the color spaces used are HSV, Y'I'Q', XYZ, L*U*V*, and LAB. Due to their different features, they are all used in segmentation and further analysis of medical images. It should be noted that the color coordinates of the HSV, Y'I'Q' spaces are derived from the RGB values of the brightness of the image pixels [24]. RGB is a hardware-dependent space, that is, the values of the RGB color coordinates depend on the type of device that reproduces the color. Therefore, the use of the HSV, Y'I'Q' spaces reduces the accuracy of segmentation methods. The hardware-independent color spaces XYZ, L*U*V*, and LAB do not have this drawback – they are constructed on descriptions of the properties of a standard observer in the form of color matching functions and are associated only with the features of the human visual system.

XYZ values are calculated from spectrometric measurements of visual stimuli and, in turn, are the basis for calculating the color coordinates L*U*V* and LAB. For color specification, the most widespread is the LAB space, the coordinates of which are calculated using the following formulas (5-7) [25]:

$$L = 25\left(\frac{100Y}{Y_0}\right)^{1/3} - 16 \quad (5)$$

$$a = 500\left[\left(\frac{X}{X_0}\right)^{1/3} - \left(\frac{Y}{Y_0}\right)^{1/3}\right] \quad (6)$$

$$z = 200\left[\left(\frac{Y}{Y_0}\right)^{1/3} - \left(\frac{Z}{Z_0}\right)^{1/3}\right] \quad (7)$$

where X, Y, Z are the coordinates of the specified colors; X0, Y0, Z0 are the coordinates of the nominal white color stimulus of the standard lighting source.

In this space, the difference between color stimuli is calculated using the Euclidean metric (8) [25]:

$$\Delta E = \left[\Delta L^2 + \Delta a^2 + \Delta b^2\right]^{/2} \quad (8)$$

This formula has undergone multiple improvements and modifications since its introduction in 1976, but in its current form it remains the principal standard in modern colorimetric technology and is widely used to assess the accuracy of color reproduction. Therefore, the use of this color difference formula as a metric in the color segmentation algorithm is natural and justified.

3. Results

The method is carried out as follows. IHC-stained histological sections of the studied tissues are recorded using a microscope and a digital camera. The obtained images are processed in the Matlab software package using standard digital image processing tools. First, the auxiliary CIE XYZ color coordinates are calculated based on the brightness values of the RGB color channels in each pixel of the original image, and then the CIE Lab color coordinates. Thus, the original digital image corresponds to a three-dimensional array of CIE Lab color coordinates, one of which is L, the values of which can vary within 0–100. Next, the primary automatic segmentation is performed using the K-means method with the calculation of the values of the color differences between the image pixels and the selection of areas of marker expression, background, nuclei and membranes. After that, these areas are visually assessed to determine in which regions the target marker color is present. At the second stage of repeated automatic segmentation for the selected area using the K-means method, the values of the differences between pixels in the CIE Lab color space are calculated, due to which the marked area is divided into three levels of marker expression: weak, medium, and strong, with the determination of the marker expression value (S). For morphometric measurement of S, which is occupied by immunopositive structures, the ratio of the number of pixels of the digital image of the immunopositive reaction area to the total number of pixels in the image, expressed as a percentage, is automatically calculated in the selected area.

Each pixel of the image in each of the three color channels of red – R, green – G and blue – B has 256 possible intensity values, ranging from 0 (darkest) to 255 (lightest). To calculate the color characteristics of each pixel of the image in the CIE Lab color space, the following steps are performed:

1. Auxiliary CIE XYZ color coordinates are calculated using formulas (9-11):

$$X = 0.49R + 0.31G + 0.20B; \quad (9)$$

$$Y = 0.17697R + 0.81240G + 0.01063B; \quad (10)$$

$$Z = 0.00R + 0.01G + 0.99B; \quad (11)$$

2. CIE Lab color coordinates are calculated using formulas:

$$L = 25\left(\frac{100Y}{Y_0}\right)^{1/3} - 16; \quad (12)$$

$$a = 500\left[\left(\frac{X}{X_0}\right)^{1/3} - \left(\frac{Y}{Y_0}\right)^{1/3}\right]; \quad (13)$$

$$b = 200\left[\left(\frac{Y}{Y_0}\right)^{1/3} - \left(\frac{Z}{Z_0}\right)^{1/3}\right]; \quad (14)$$

Where the coordinates X_0, Y_0, Z_0 correspond to the reference white color and are given as: X_0=95.04; Y_0=100; Z_0=108.89;

3. Using the K-means method, automatic segmentation of the digital image into four areas is performed. Each area combines pixels that differ significantly in color, represented in CIE Lab coordinates. These areas correspond to the marker, background, nuclei, and membranes in the image. The areas are displayed on the screen for observation.

4. The observer visually evaluates these areas and determines which one contains the desired marker color; this group of pixels will then undergo further secondary segmentation and analysis;

5. At the second stage (secondary automatic segmentation), the identified image area with labeled tissues is further segmented using the K-means method based on lightness (L) into three expression levels: L=0–40 corresponds to a strong marker expression level; L=40–50 corresponds to a medium marker expression level; L=50–100 corresponds to a weak marker expression level;

6. The identified pixel groups are displayed on the screen as images of areas with different marker expression levels.

To illustrate the effectiveness of the proposed approach, sets of images (digital photographs) of various histological types of testicular tumors are presented. These images show the results of staining of histological sections with different IHC markers. Photography was performed using an Olympus BX-41TF microscope with the Olympus DP-Soft software (Version 3.1). Some images contained structures marked by the markers, and the S of these structures varied across different images. Examples of the original images are shown in Fig. 1.

As already noted, the main differences between the images are as follows: the expressed markers may be identified in different cellular and tissue structures; the S value of the marker, as well as its L value, may also vary significantly across different tumors.

It is evident that the objects in the images vary in size, while the marked areas are characterized by different structural and color properties. The only consistent feature of such areas remains the “brown color” in subjective perception. The proposed two-stage segmentation algorithm using K-means in the Lab space, as well as single-stage K-means segmentation with grayscale and RGB image representations, were applied to the analyzed images (Fig. 2–6).

The presented example shows that the use of grayscale representation does not allow the separation of marked tissues from unmarked ones (Fig. 2, 3). The RGB representation leads to segmented image components that, although differing in brightness (and consequently in L), still contain a mixture of marked and unmarked tissues (Fig. 4).

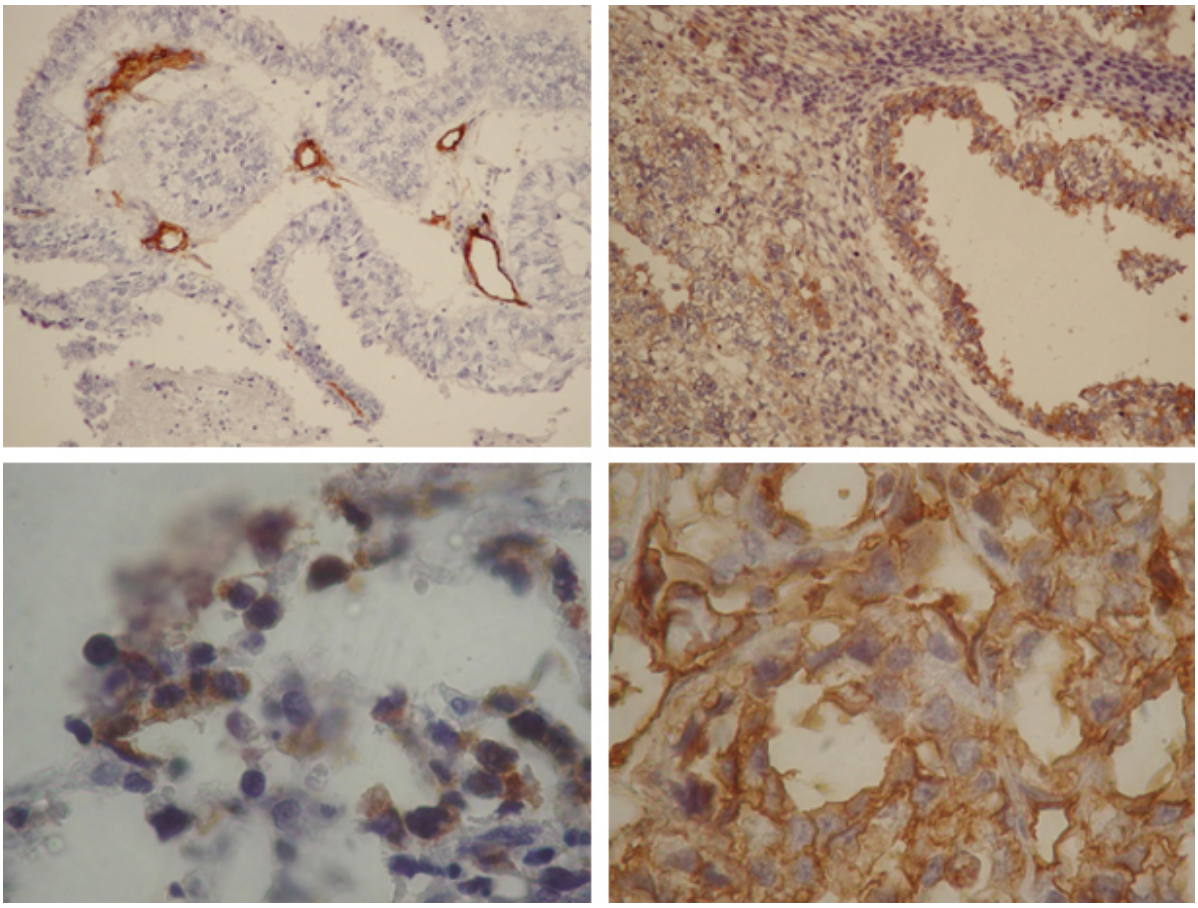

Figure 1: Examples of the original images.

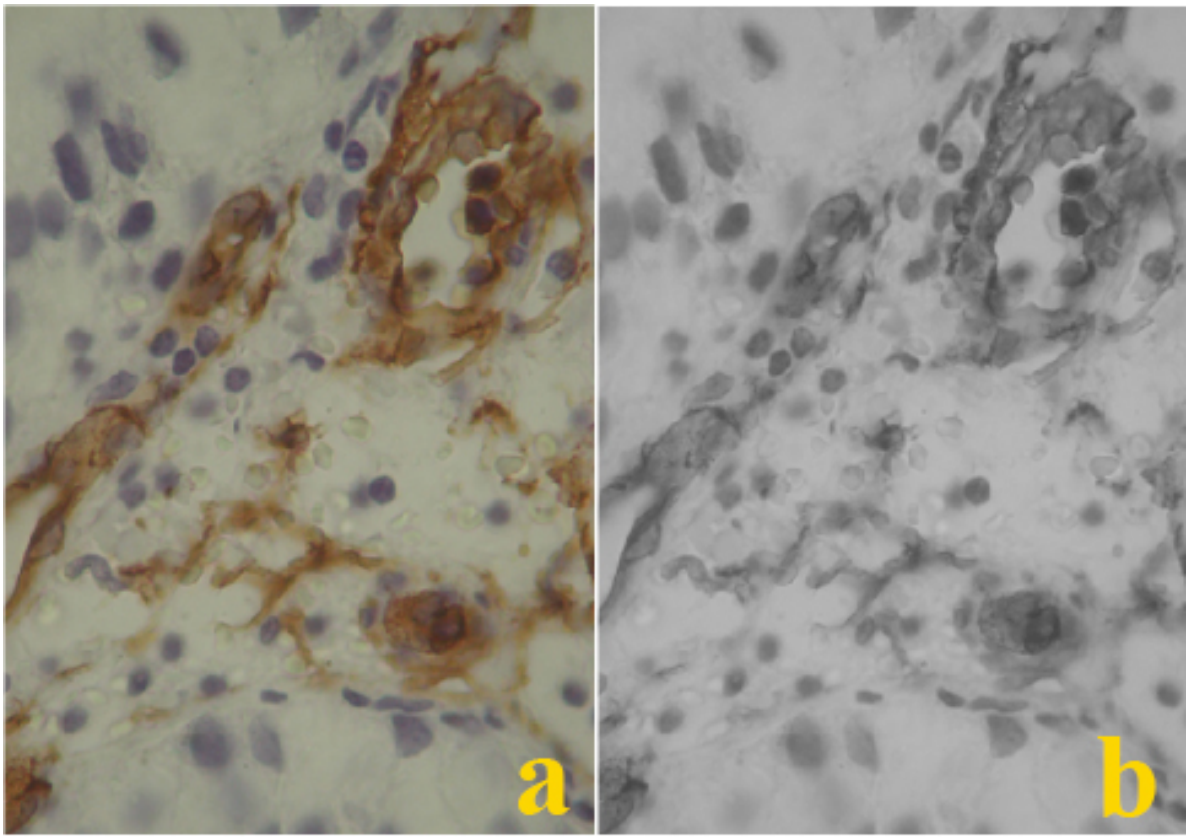

Figure 2: Examples of the original images.

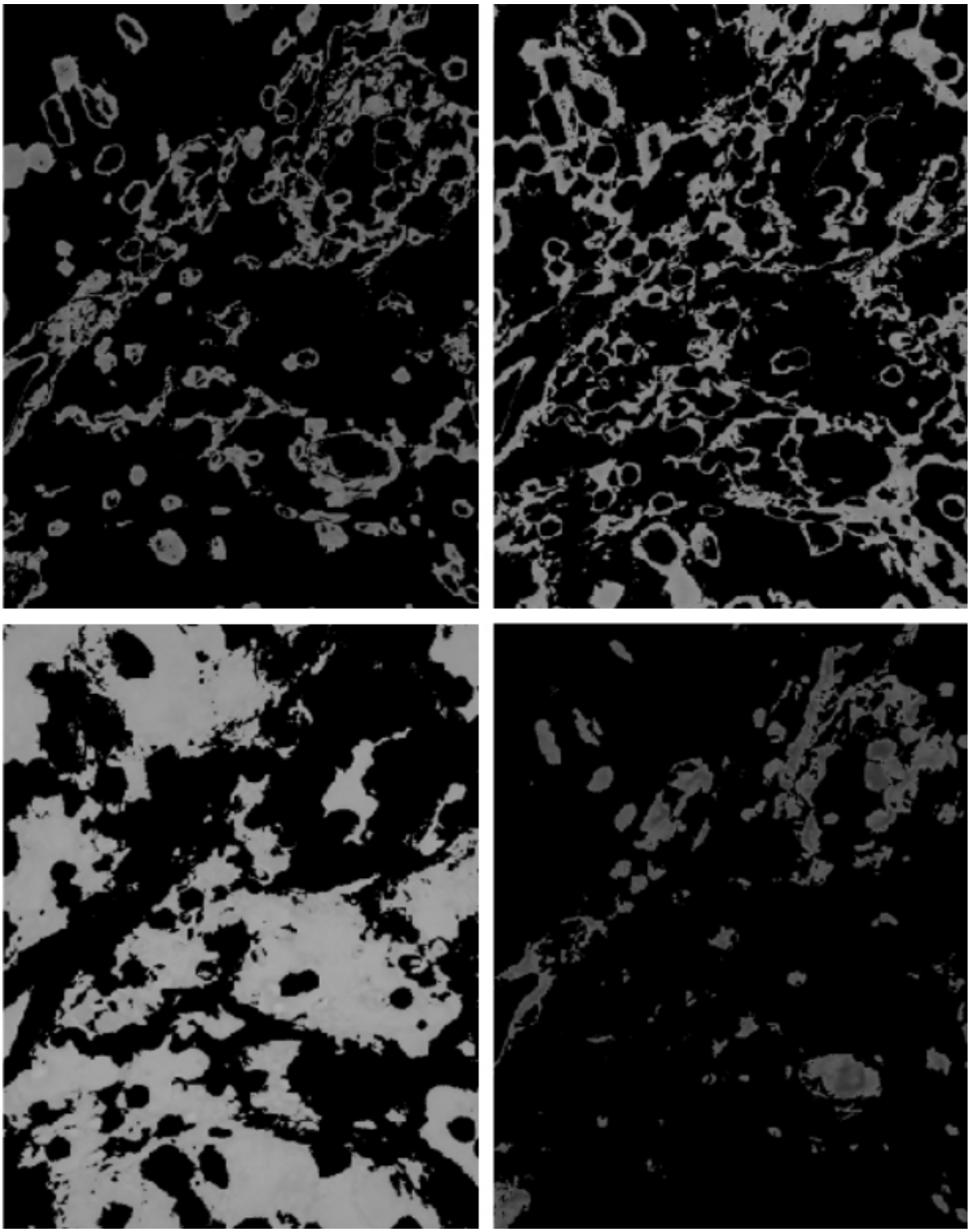

Figure 3: Results of one-step grayscale image segmentation (a – objects in cluster 1; b – objects in cluster 2; c – objects in cluster 3; d – objects in cluster 4).

In contrast, the proposed algorithm, at the first stage of segmentation, makes it possible to separate marked and unmarked tissue areas in the image (Fig. 5), while at the secondary segmentation stage, it allows the division of marked regions according to marker expression levels (Fig. 6). As a result, the determined S of the marker expression areas provides an objective assessment of the main biological properties of different tumors and other pathological processes [2-11].

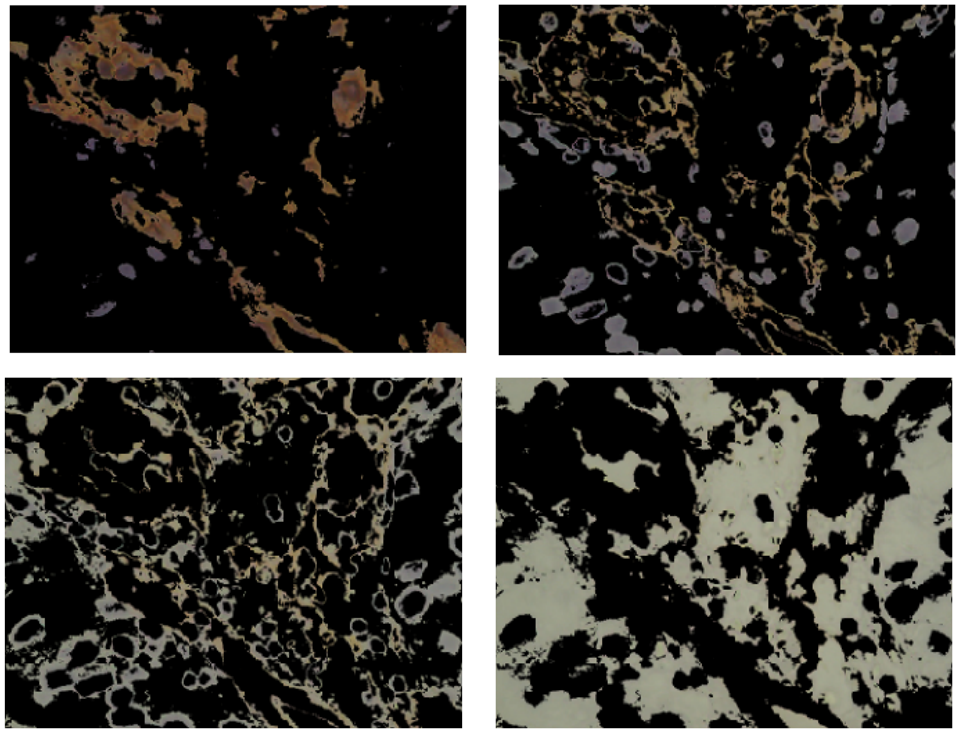

Figure 4: Results of image segmentation in RGB representation (a – objects in cluster 1; b – objects in cluster 2; c – objects in cluster 3; d – objects in cluster 4).

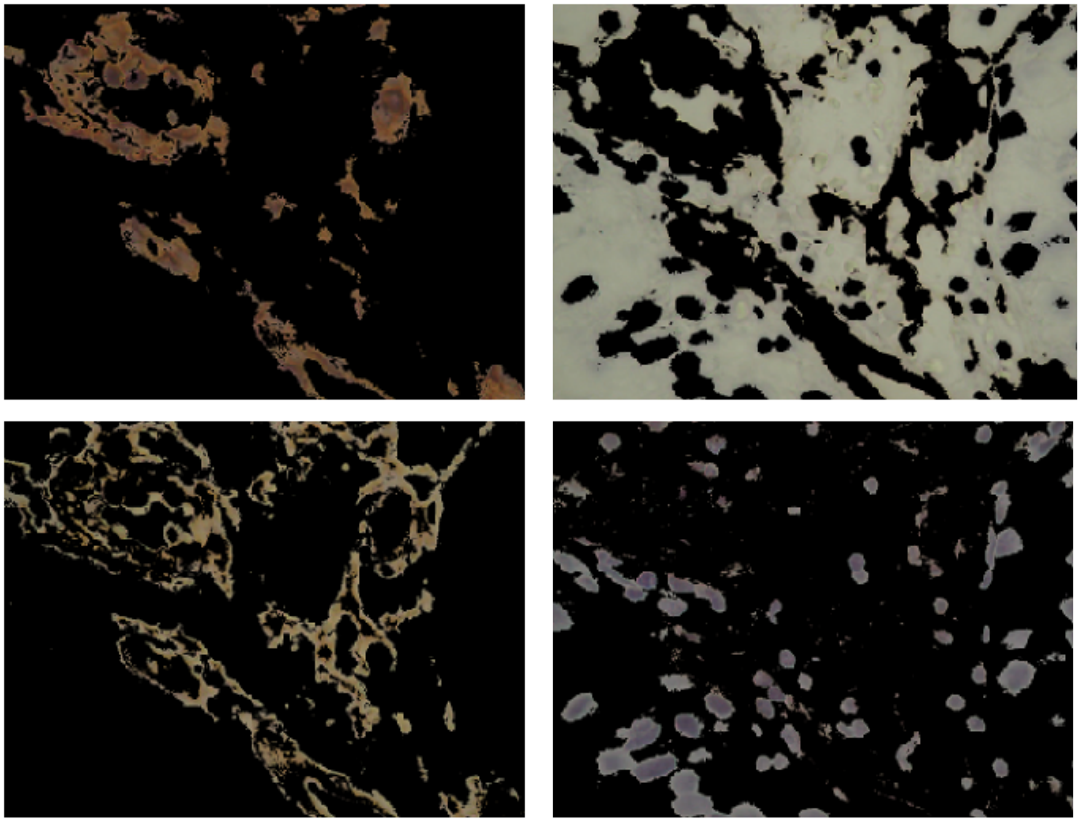

Figure 5: Results of primary image segmentation using K-means in the Lab color space (a – objects in cluster 1; b – objects in cluster 2; c – objects in cluster 3; d – objects in cluster 4).

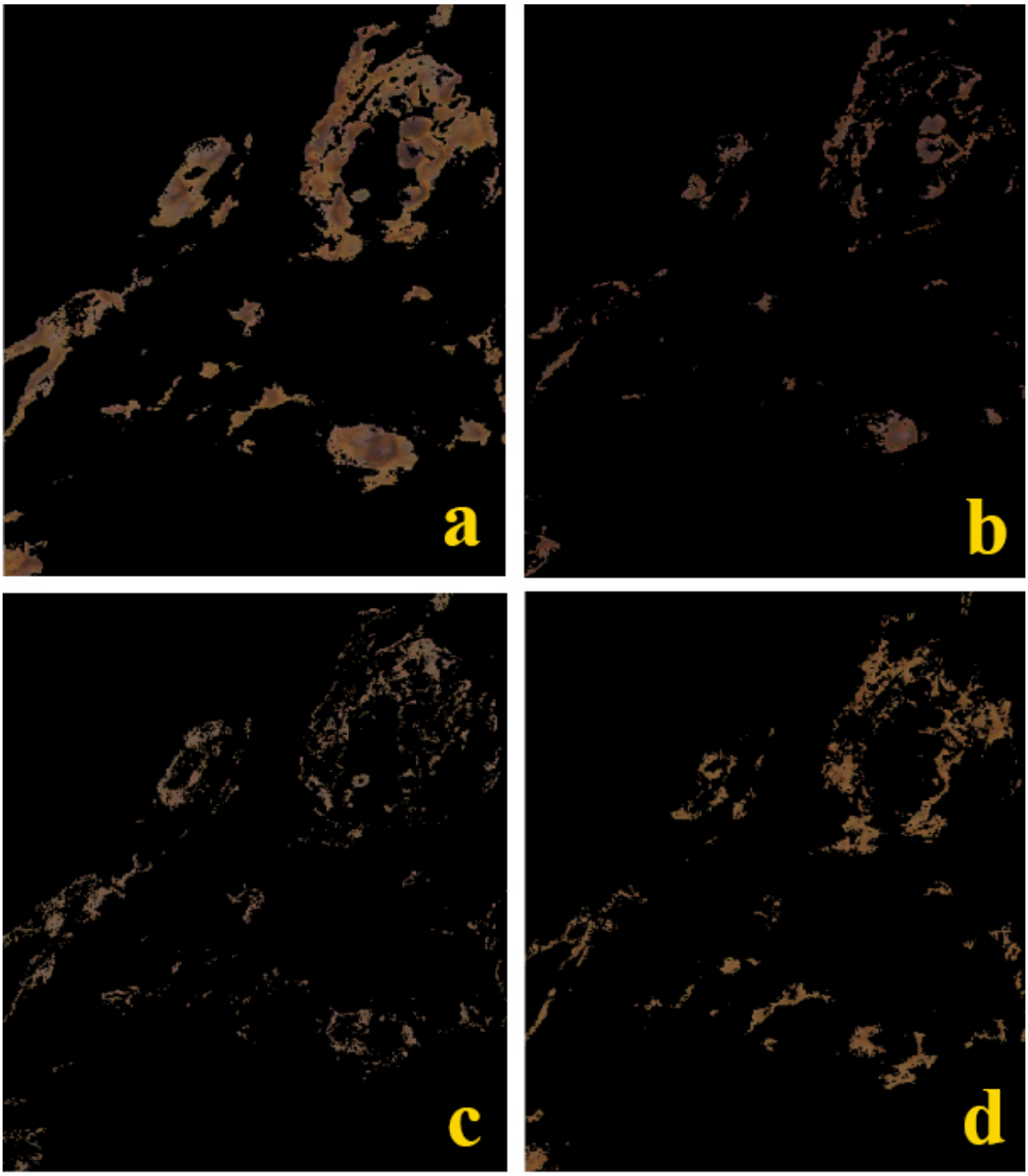

Figure 6: Initial image for secondary segmentation and its results with highlighted areas according to marker expression levels (a – image for secondary segmentation, S of the marker 11.29%; b – Lab coordinates of the marker 29.49; 12.64; 13.52, S of the marker 3.08%; c – Lab coordinates of the marker 38.85; 8.47; 16.22. S of the marker 3.60%; d – Lab coordinates of the marker 38.90; 11.84; 24.67. S of the marker 4.61%.

Practical examination of suggested algorithm with comparison and estimation obtained result by pathologists was realized in objective consensus about peculiarities of investigated microspecimen.

4. Conclusions

Suggested algorithm when first stage applies K-means clustering in CIE Lab space to split an image into four classes, second stage 2 re-segments the “marker” class by lightness L* into three

expression bands using fixed L* thresholds, and reports the relative positive area S as the quantitative readout (implementation is in MATLAB and relies on an RGB-XYZ-Lab conversion and Euclidean ΔE in Lab as the distance metric) allows to obtain objective data about IHC micropreparations.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] A. Schabadasch, "Intramurale nervengeflechte des darmrohrs," Z. Zellforsch., vol. 10, no. 2, pp. 320–385, 1930, doi: 10.1007/BF02450699.

[2] M. Lyndin, N. Hyriavenko, V. Sikora, Y. Lyndina, Y. Soroka, and A. Romaniuk, "Invasive breast carcinoma of no special type with medullary pattern: morphological and immunohistochemical features," Turk Patoloji Derg., vol. 38, no. 3, pp. 205–212, 2022, doi: 10.5146/tjpath.2021.01559.

[3] A. Romaniuk, M. Lyndin, V. Sikora, Y. Lyndina, and K. Panasovska, "Histological and immunohistochemical features of medullary breast cancer," Folia Medica Cracoviensia, vol. 2, pp. 41–48, 2015.

[4] I. O. Vynnychenko, A. O. Pryvalova, O. I. Vynnychenko, M. S. Lyndin, V. V. Sikora, and A. M. Romaniuk, "PIK3CA-mutant circulating tumor DNA in patients with breast cancer," Azerbaijan Medical Journal (ATJ), vol. 3, pp. 79–88, 2020, doi: 10.34921/amj.2020.3.010.

[5] A. Romaniuk, N. Gyryavenko, M. Lyndin, A. Piddubnyi, V. Sikora, and A. Korobchanska, "Primary cancer of the fallopian tubes: histological and immunohistochemical features," Folia Medica Cracoviensia, vol. 4, pp. 71–80, 2016.

[6] N. Hyriavenko et al., "Serous adenocarcinoma of fallopian tubes: histological and immunohistochemical aspects," J. Pathol. Transl. Med., vol. 53, pp. 236–243, 2019, doi: 10.4132/jptm.2019.03.21.

[7] M. Lyndin et al., "COX2 effects on endometrial carcinomas progression," Pathol. Res. Pract., vol. 238, p. 154082, 2022, doi: 10.1016/j.prp.2022.154082.

[8] N. Hyriavenko et al., "Neuroendocrine tumor of the fallopian tube and serous adenocarcinoma of the ovary: multicentric primary tumors," Turk Patoloji Derg., vol. 39, no. 2, pp. 161–166, 2023, doi: 10.5146/tjpath.2022.01589.

[9] O. I. Kravtsova et al., "The role of Hsp70 and Hsp90 in the endometrial carcinomas progression," Azerbaijan Medical Journal (ATJ), vol. 3, pp. 136–146, 2021, doi: 10.34921/amj.2021.3.019.

[10] O. Shevchenko and V. Shevchenko, "The role of inflammation in the development of dental diseases," Kharkiv Dental Journal, vol. 2, no. 1(32), pp. 78–91, 2025, doi: 10.26565/3083-5607-2025-3-08.

[11] D. O. Kovalchuk and N. M. Savielieva, "Matrix metalloproteases and proinflammatory proteins in serum as markers for the efficiency of temporomandibular joint disorders therapy," Kharkiv Dental Journal, vol. 1, no. 2, pp. 134–142, 2024, doi: 10.26565/3083-5607-2024-2-04.

[12] V. Alekseeva, A. Nechyporenko, M. Frohme et al., "Intelligent decision support system for differential diagnosis of chronic odontogenic rhinosinusitis based on U-Net segmentation," Electronics (Switzerland), vol. 12, no. 5, 2023, doi: 10.3390/electronics12051202.

[13] A. Nechyporenko et al., "Galvanic skin response and photoplethysmography for stress recognition using machine learning and wearable sensors," Appl. Sci., vol. 14, no. 24, p. 11997, 2024, doi: 10.3390/app142411997.

[14] A. Nechyporenko, V. Alekseeva, R. Nazaryan, and V. Gargin, "Biometric recognition of personality based on spiral computed tomography data," in Proc. IEEE 16th Int. Conf.

Experience of Designing and Application of CAD Systems (CADSM), Lviv, Ukraine, Feb. 2021, pp. 11–15, doi: 10.1109/CADSM52681.2021.9385267.
[15] V. V. Alekseeva, A. S. Nechiporenko, A. V. Lupyr, N. O. Yurevych, and V. V. Gargin, "A method of complex evaluation of morphological structure of ostiomeatal complex components, lower wall of maxillary and frontal sinuses," Wiad Lek., vol. 73, no. 12, pp. 2576–2580, 2020, doi: 10.36740/WLek202012104.
[16] S. P. Lloyd, "Least squares quantization in PCM," IEEE Trans. Inf. Theory, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
[17] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proc. 5th Berkeley Symp. Math. Stat. Probab., vol. 1, Berkeley, CA: Univ. California Press, 1967, pp. 281–297. [Online]. Available: http://projecteuclid.org/euclid.bsmsp/1200512992
[18] S. A. Burney and H. Tariq, "K-means cluster analysis for image segmentation," Int. J. Comput. Appl., vol. 96, no. 4, pp. 1–8, 2014.
[19] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: Advances and prospects," Pattern Recognit., vol. 34, no. 12, pp. 2259–2283, 2001.
[20] L. Lucchese and S. K. Mitra, "Color image segmentation: A state of the art survey," Proc. Indian Nat. Sci. Acad., vol. 67, no. 2, pp. 207–221, 2001.
[21] B. Azam, R. J. Qureshi, Z. Jan, and T. A. Khattak, "Color based segmentation of white blood cells in blood photomicrographs using image quantization," Res. J. Recent Sci., vol. 3, no. 4, pp. 34–39, 2014.
[22] P. J. Baldevbhai and R. S. Anand, "Color image segmentation for medical images using Lab* color space," IOSR J. Electron. Commun. Eng. (IOSR-JECE), vol. 1, no. 2, pp. 24–45, 2012.
[23] S. Mohapatra, D. Patra, and S. Satpathy, "Unsupervised blood microscopic image segmentation and leukemia detection using color based clustering," Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 4, pp. 477–485, 2012.
[24] S. M. Praveena and I. Vennila, "Optimization fusion approach for image segmentation using K-means algorithm," Int. J. Comput. Appl., vol. 2, no. 7, pp. 18–25, 2010.
[25] M. D. Fairchild, Color Appearance Models, 3rd ed. Chichester, U.K.: Wiley, 2013, p. 439.