Correspondence author
O. Yo. Pitsun
oleh.pitsun@gmail.com

O. Yo. Pitsun

West Ukrainian National University, Ternopil, Ukraine

MULTI-THREAD PARALLELIZING OF CELL CHARACTERISTICS OF BIOMEDICAL IMAGES

An approach to the parallelization of the process of calculating the quantitative characteristics of cell nuclei on biomedical images (cytological, histological, immunohistochemical) is proposed, which will speed up the process of making a diagnosis. The relevance of this task lies in the fact that there are a large number of micro-objects in the image that need to be investigated, and optimization of time and rational distribution of resources will speed up the stage of calculating the area of cell nuclei and their average brightness level. In the future, these data are stored in the database for further use as a dataset for the tasks of classification, clustering, and intellectual analysis. Modern means of data classification and intellectual analysis are used to make a diagnosis. When using convolutional neural networks, the input data to the classifier are images in the format .jpg, .png, .bmp, etc. Alternative algorithms and data processing tools in most cases require quantitative characteristics. In the case of using biomedical images, the quantitative characteristics are the area, perimeter, circumference, length, and major and lateral axes of the cell nucleus. The area and other characteristics of cell nuclei characterize the normal state or the presence of pathologies.

Calculating quantitative characteristics on immunohistochemical and cytological images is time-consuming because the number of cell nuclei in one image can be in the range of 10-20 units. To create a dataset of quantitative characteristics of cell nuclei, it is necessary to perform calculations on a large number of images, which in turn requires significant resources, at a particular time. The parallelization of calculating the biomedical image characteristics is implemented on the basis of computer vision algorithms to select the necessary objects and means of software parallelization of tasks at the thread level to speed up the process of calculating the cell nucleus characteristics. It was established that the existing systems of automated microscopy and diagnostic systems based on images do not have the presence of a large number of characteristics of cell nuclei and do not have mechanisms for parallelizing the process of their calculation. The proposed approach makes it possible to speed up the process of calculating the quantitative characteristics of cell nuclei by 25 %. The relevance of the problem of parallelization is due to the need to process a large amount of data for their further reduction and classification. Thread-level parallelization improves image processing speed and does not require specialized hardware.

Keywords: immunohistochemistry; histology; threads; cell nucleus.

Introduction / Вступ

The use of results of immunohistochemical and histological research is actively applied in the process of diagnosing pathological conditions of breast cancer. The main criteria used by diagnosticians are the area of cell nuclei and the average level of their brightness. With the help of various biomarkers, the doctor visually assesses the coverage area on the image and its intensity. Given the significant progress in the development of computer vision algorithms and artificial intelligence systems, the task of automating diagnosis based on immunohistochemical images arises.

Calculating quantitative characteristics on immunohistochemical and cytological images is time-consuming because the number of cell nuclei in one image can be in the range of 10–20 units. To create a dataset of quantitative characteristics of cell nuclei, it is necessary to perform calculations on a large number of images, which in turn requires significant resources, at a particular time.

With the development and widespread implementation of software and hardware approaches to the parallelization of processes and flows, the task of parallelizing the calculation of quantitative characteristics of cell nuclei on biomedical images becomes urgent.

The object of the research – processes of immunohistochemical and histological image processing for breast cancer diagnosis.

The subject of research – methods of image processing parallelization at the threads level to speed up the calculation of quantitative characteristics of cell nuclei on immunohistochemical and cytological images.

The purpose of the research – is to develop an algorithm for parallelizing the process of calculating the micro-objects quantitative characteristics on biomedical images.

To achieve this *purpose*, the following *main research objectives* are identified:
- to determine the list of quantitative characteristics of cell nuclei that must be calculated;
- to develop an algorithm for parallelizing the image processing process for calculating quantitative characteristics;
- to perform a comparative analysis of the obtained results.

Analysis of recent research and publications. Scientists in the field of diagnosing breast cancer pay considerable attention to modern tools of computer vision and artificial intelligence for image processing. Kamyar Nazeri in his work [20] pays considerable attention to the problem of classification of biomedical images of breast cancer. The authors achieved a classification accuracy of 95 % using the BreAst Cancer Histology (BACH) dataset. Approaches to the classification of histological images are considered in works [3], [17], [18], [22].

To calculate the quantitative characteristics of cell nuclei, it is necessary first to segment the input image aimed at obtaining a new image of the "object – background" type. There are many approaches to image segmentation. In [7], the authors present an adaptive method of segmentation ba-

sed on the use of metrics to assess the quality of segmentation. Segmentation of biomedical images is a complex process since the original microscopic images are characterized by a high level of noise and the presence of unclear contours of the studied objects. Therefore, the task of researchers is to maximally separate cell nuclei in the image. The authors in [2], [19], [23] present the results of using modern approaches to biomedical segmentation, in particular, in the analysis of images used in the diagnosis of breast cancer. Analyzing the above articles, we can conclude that the use of convolutional networks for image segmentation is widely used.

The use of quantitative characteristics in the analysis of cell nuclei is considered in [12]. The application of quantitative characteristics for data clustering is presented in [9]. Analysis of cytological images based on quantitative characteristics is given in [6]. The utilization of numerical data allows the classification, clustering, and regression of biomedical images without the need to directly use images, the processing of which requires the exploitation of a large number of resources. Images in *.png, *.jpg, *.bmp and other formats are suitable for using convolutional neural networks, but for other types of algorithms, it is necessary to convert the information in the image into a numerical format.

The study of the approach to the segmentation of biomedical images based on fuzzy logic using tools for parallelization is presented in [1]. Image segmentation using parallel processing is discussed in [10]. The authors note the acceleration of the processing process by 3 times. Means of parallelization of the biomedical image processing process are also considered in works [14], [15], [16], [21]. The productivity of using parallelism for image processing tasks using Tensorflow is given in [11]. The resulting acceleration ranged from 3.6 to 15 times. In [4], a new data distribution strategy aimed at CPU-GPU hybrid machines is presented. Gancheva V. in [13] provides an approach to the parallelization of image processing algorithms at a low level of computer vision, in particular at the level of filtering.

Such a large number of studies in the direction of dispersing the process of processing biomedical images demonstrates the relevance and necessity in the development of this task.

Materials and methods of research. A generalized algorithm for parallel calculation of biomedical images' quantitative characteristics.

The result of the microscopic examination is a digital image in *.jpg, *.png, *.bmp format. Parameters of cell nuclei on immunohistochemical, histological, and cytological images characterize the presence or absence of pathologies. In most cases, diagnosticians visually evaluate each image when analyzing images. The implementation of automation in this diagnostic process requires the need for quantitative assessment of input parameters, in this case – cell nuclei. Evaluation of the characteristics of cell nuclei is based on the following criteria:

- area;
- perimeter;
- the length of the main axis;
- the length of the lateral axis;
- Feret's diameter;
- circumference;
- the area of the bounding rectangle;
- the angle of inclination to the OX axis.

In general, the generalized algorithm for parallel calculation of quantitative characteristics of biomedical images consists of the following steps:

Step 1: Download the image to the computer memory.

Step 2: Noise removal. At this stage, Gaussian and median filters are used to remove Gaussian and impulse noise, respectively. The rules for selecting input filter parameters (filter window size) are discussed in [8].

Step 3: Preprocessing. This stage involves the use of algorithms for adaptive alignment of the image histogram to avoid the presence of darkened or too-bright areas.

Step 4: Segmentation. One of the most important stages, since the quality of calculations of quantitative characteristics of cell nuclei, will depend on the quality of segmentation. This work uses adaptive segmentation based on elements of fuzzy logic. The rules of fuzzy logic for this problem were selected by the author based on experimental studies.

Step 5: Highlighting each kernel on the segmented image – Parallel execution.

Step 6: Calculation of quantitative characteristics – Parallel execution. The stage of calculating the quantitative characteristics of micro-objects consists of two parts: the selection of contours, and key points, and the use of known formulas for calculating the area, perimeter, circumference, and other characteristics of cells. The process of calculating cell characteristics is described in detail in [5].

Step 7: Output of calculation results to the screen and storage in the database – Parallel execution.
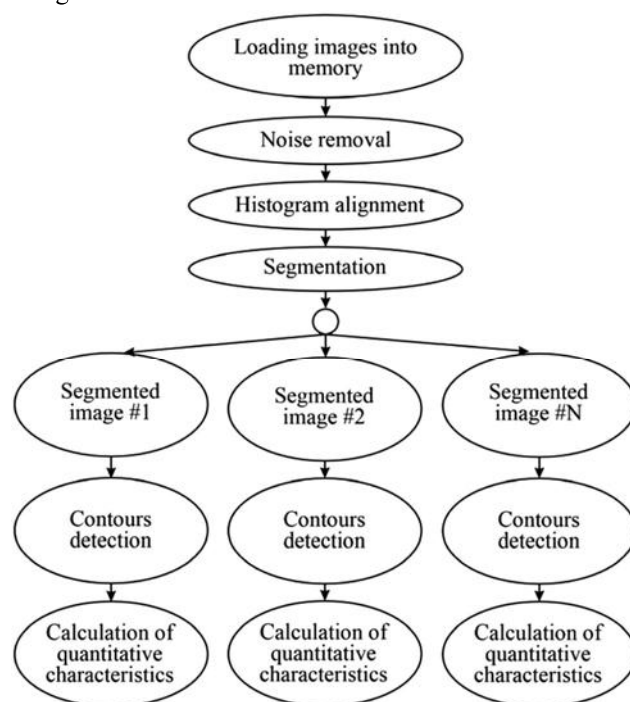
Figure 1. Tier-parallel form of representation of parallelization of the process of calculating quantitative characteristics / Ярусно-паралельна форма представлення процесу розпаралелення обчислення кількісних характеристик

Step 8: To ensure the correct operation of the program and to avoid unexpected termination of the program, a thread synchronization mechanism is provided. That is, the completion of all threads is expected and only after that the following components of the program are executed. Software implementation in the Java programming language occurs using the executor .awaitTermination method. The syntax for calling this method is as follows:

boolean awaitTermination(*long timeout, TimeUnit unit*), where *long timeout* – the maximum time to wait. *TimeUnit unit* – the unit of time of the timeout argument.

The tier-parallel form of representation of the parallelization of the process of calculating quantitative characteristics is shown in Figure 1.

The generated calculation results are stored in a list of type ObservableList<Nuclei>, where "Nuclei" is a designed entity containing information about quantitative characteristics such as area, perimeter, coordinates, etc. This list is global, that is, in the future, you can get a list of all records of the collection.

Research results and their discussion / Результати дослідження та їх обговорення

The processing of immunohistochemical and histological images is reduced to the presentation of images in the form of "white background, black objects".
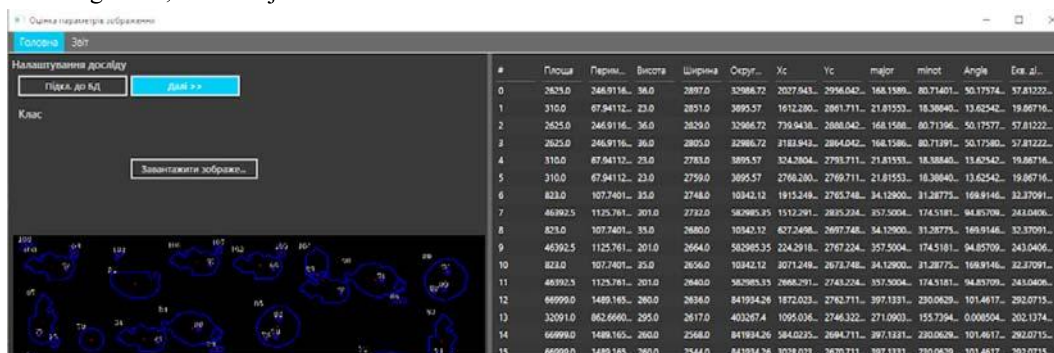
An example of a cell nucleus and the process of calculating its characteristics are shown in Figure 2.

Figure 2. Visualization of the process of calculating quantitative characteristics / Візуалізація процесу обчислення кількісних характеристик: a – original image / оригінальне зображення; b – area / площа; c – main axis / головна вісь; d – area of bounding rectangle / площа обмежувального прямокутника

The graphic interface of the developed program is shown in Figure 3.

Figure 3. The graphical interface of the developed program for calculating quantitative characteristics of cell nuclei / Графічний інтерфейс розробленої програми для розрахунку кількісних характеристик ядер клітин

The software module consists of the following components: image upload unit and database connection; processed image output unit; a unit for outputting quantitative characteristics in tabular form.

The provision of thread-based parallelization is implemented using the Executor framework, which is a component of the Java programming language. The working principle of the Executor framework is shown in Figure 4.
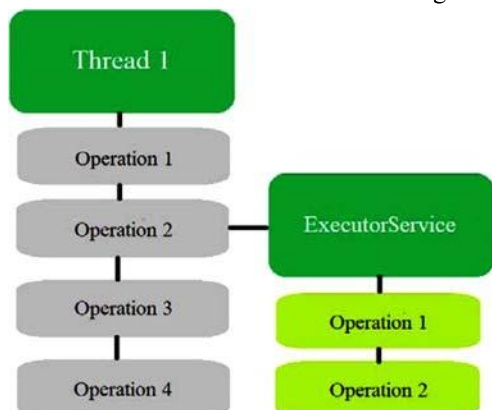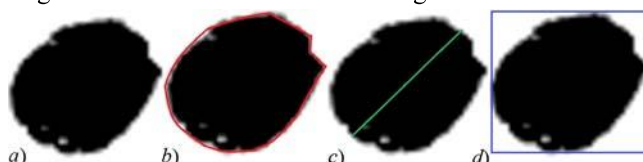

Figure 4. The working principle of the Executor framework / Принцип роботи фреймворку Executor

Executor Service allows you to use a convenient programming interface to manage the execution of operations at the thread level. Table 1 shows the hardware parameters of the computer on which the experiments were performed.

Table 1. Parameters of the hardware part of the computer / Параметри апаратного забезпечення

Parameter	Value
Processor	Intel (R) Core (TM) i5-7200U CPU @ 2.50GHz Intel(R) Core(TM) i5-8400 CPU @ 2.80GHz
RAM	8,00 Gb

The results of the research are shown in Figure 5. The average result of the execution time of the experiments using parallelism and without parallelism is shown in Figure 6.
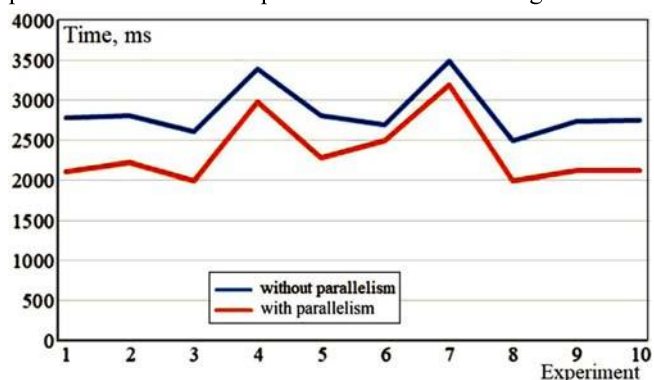

Figure 5. Comparative analysis of conducted experiments / Порівняльний аналіз проведених експериментів (Intel (R) Core (TM) i5-7200U)
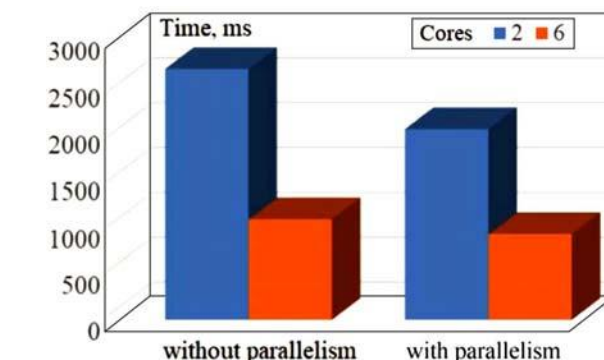

Figure 6. The average result of the execution time of experiments using parallelism and without parallelism / Середня тривалість виконання завдання з використанням паралелізму та без нього

Discussion of research results. From the results of the experiments, it can be concluded that the increase in the time of the calculation of quantitative characteristics of micro-objects is an average of 25 %. In the developed software module, the number of threads is determined automatically depending on the number of segments in the image.

So, based on the results of the work performed, it is possible to formulate the following scientific novelty and practical significance of the research results.

The scientific novelty of the obtained research results – the development of a tiered-parallel form of the algorithm for calculating the characteristics of biomedical images, which allows to speed up the image processing process.

The practical significance of the research results – is the software implementation of the module for calculating a stack of segmented biomedical images on JAVA programming language for the purpose of calculating the quantitative characteristics of cell nuclei.

Conclusions / Висновки

In this research:

1. A tiered-parallel form of the task of calculating the quantitative characteristics of cell nuclei on a stack of images was developed, which made it possible to speed up the process of executing the task using the Executor Service framework;

2. Approaches to the classification of biomedical images (immunohistochemical, histological) were determined, which made it possible to identify the necessary parameters of cell nuclei;

3. A comparative analysis of the obtained results demonstrates a 25 % acceleration of the image processing process compared to the case without the use of parallelism. For 6 cores, the increase is about 15%.

References

[1] AlZubi, Shadi, Shehab, Mohammed, Al-Ayyoub, Mahmoud, Jararweh, Yaser, Gupta, Brij. (2020). Parallel implementation for 3D medical volume fuzzy segmentation. *Pattern Recognition Letters*, 130, 312–318. https://doi.org/10.1016/j.patrec.2018.07.026

[2] Amgad, M., Elfandy, H., Hussein, H. (2019). Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics, 35*(18), 3461-3467. https://doi.org/10.1093/bioinformatics/btz083

[3] Araújo, T, Aresta, G, Castro, E, Rouco, J, Aguiar, P, Eloy, C, Polónia, A., & Campilho, A.. (2017). Classification of breast cancer histology images using Convolutional Neural Networks. *PLoS ONE, 12*(6), e0177544. https://doi.org/10.1371/journal.pone.0177544

[4] Barreiros, W., C.M.A. Melo, A., Kong, J., Ferreira, R., M. Kurc, T., H. Saltz, J., Teodoro, G. (2022). Efficient microscopy image analysis on CPU-GPU systems with cost-aware irregular data partitioning. *Journal of Parallel and Distributed Computing, 164*, 40–54. https://doi.org/10.1016/j.jpdc.2022.02.004

[5] Berezky, O. M. (Ed.), Batko, Yu. M., Berezka, K. M., Verbovy, S. O., Datsko, T. V., Dubchak, L. O., Ignatev, I V., Melnik, G. M., Nikoluk, V. D., & Pitsun, O. Y. (2017). Methods, algorithms and software tools for the processing of biomedical images: monograph. Ternopil: TNEU "Economic Thought", 330 p.

[6] Berezsky, O. M., Pitsun, O. Yo., Melnyk, G. M., & Datsko, T. V. (2021). Application of linear regression method for analysis of cytological images quantitative characteristics. *Ukrainian Journal of Information Technology, 3*(1), 73–77. https://doi.org/10.23939/ujit2021.03.073

[7] Berezsky, O., Pitsun, O., Batryn, N., Berezska, K., Savka, N., & Dolynyuk, T. (2018). Image Segmentation Metric-Based Adaptive Method. IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 554–557. https://ieeexplore.ieee.org/document/8478579

[8] Berezsky, O., Pitsun, O., Derish, B., Berezska, K., Melnyk, G., & Batko, Y. (2020). *Adaptive Immunohistochemical Image Pre-processing Method*. 10th International Conference on Advanced Computer Information Technologies (ACIT), 820–823. https://ieeexplore.ieee.org/document/9208920

[9] Berezsky, O., Pitsun, O., Dubchak, L., Berezka, K., Dolynyuk, T., & Derish, B. (2020). Cytological Images Clustering of Breast Pathologies. 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), 62–65. https://ieeexplore.ieee.org/document/9321867

[10] Cao, Jianfang, Chen, Lichao, Wang, Min, & Tian, Yun (2018). Implementing a Parallel Image Edge Detection Algorithm Based on the Otsu-Canny Operator on the Hadoop Platform. *Computational Intelligence and Neuroscience, 2018*, Article ID 3598284. https://doi.org/10.1155/2018/3598284

[11] Demirović, D., Skejić, E., & Šerifović – Trbalić, A. (2018). *Performance of Some Image Processing Algorithms in Tensorflow*. 25th International Conference on Systems, Signals and Image Processing (IWSSIP), 1–4. https://ieeexplore.ieee.org/document/8439714/

[12] Fujima, N., Hirata, K., Shiga, T., Li, R. (2018). Integrating quantitative morphological and intratumoural textural characteristics in FDG-PET for the prediction of prognosis in pharynx squamous cell carcinoma patients. *Clinical Radiology, 73*(12). https://doi.org/10.1016/j.crad.2018.08.011

[13] Gancheva, V. (2021). *Parallel Multithreaded Medical Images Filtering*. 2021 International Conference on Computational Science and Computational Intelligence (CSCI), 1788–1793. https://ieeexplore.ieee.org/document/9799119

[14] Haiyan, Gu, Han, Yanshun, Yang, Yi, Li, Haitao, Liu, Zhengjun, Soergel, Uwe, Blaschke, Thomas, & Cui, Shiyong (2018). An Efficient Parallel Multi-Scale Segmentation Method for Remote Sensing Imagery. *Remote Sensing, 10*(4), 590. https://doi.org/10.3390/rs10040590

[15] Hangün, B., Eyecioğlu, Ö. (2017). Performance Comparison Between OpenCV Built-i CPU and GPU Functions on Image Processing Operations. *International Journal of Engineering Science and Application, 1*, 34-41.

[16] Huang, B., Xue, J., Lu, K., Tan, Y., & Zhao, Y. (2021). MPNet: Multi-scale Parallel Codec Net for Medical Image Segmentation. In: Fang, L., Chen, Y., Zhai, G., Wang, J., Wang, R., Dong, W. (Eds). Artificial Intelligence. CICAI 2021. *Lecture Notes in Computer Science, 13069*. Springer, Cham. https://doi.org/10.1007/978-3-030-93046-2_42

[17] Li, Y., Wu, J., & Wu, Q. (2019). Classification of Breast Cancer Histology Images Using Multi-Size and Discriminative Patches Based on Deep Learning. *IEEE Access, 7*, 21400–21408. https://ieeexplore.ieee.org/document/8636921

[18] Peizhen, Xie, Ke, Zuo, Jie, Liu, Mingliang, Chen, & Shuang, Zhao (2021). Interpretable Diagnosis for Whole-Slide Melanoma Histology Images Using Convolutional Neural Network. Journal of Healthcare Engineering, 2021, Article ID 8396438. https://doi.org/10.1155/2021/8396438

[19] Qaiser, T., Tsang, Y.-W., Taniyama, D., Sakamoto, N., Nakane, K., Epstein, D., & Rajpoot, N. (2019). Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Medical Image Analysis, 55*, 1–14. https://doi.org/10.1016/j.media.2019.03.014

[20] Ramirez-Quintana, J., Acosta-Lara, I., Ramirez-Alonso, G., Chacon-Murguia, M., & Corral-Saenz, A. (2022). A Lightweight Convolutional Neural Network for Breast Cancer Diag-

nosis with Histology Images. Pattern Recognition. MCPR 2022. Lecture Notes in Computer Science, 13264. Springer, Cham. https://doi.org/10.1007/978-3-031-07750-0_30

[21] Roels, J., De Vylder, J., Saeys, Y., Goossens, B., & Philips, W. (2016). Decreasing Time Consumption of Microscopy Image Segmentation Through Parallel Processing on the GPU. In: Blanc-Talon, J., Distante, C., Philips, W., Popescu, D., Scheunders, P. (Eds). Advanced Concepts for Intelligent Vision Systems. ACIVS 2016. *Lecture Notes in Computer Science, 10016*. Springer, Cham. https://doi.org/10.1007/978-3-319-48680-2_14

[22] Vang, S., Chen, Y. Z., & Xie, X. (2018). Deep Learning Framework for Multi-class Breast Cancer Histology Image Classification. Computer Vision and Pattern Recognition, Computer Science, Arxiv. https://doi.org/10.48550/ar-Xiv.1802.00931

[23] Xu, H., Liu, L., Lei, X., Mandal, M., Lu, C. (2021). An unsupervised method for histological image segmentation based on tissue cluster level graph cut. *Computerized Medical Imaging and Graphics, 93*. https://doi.org/10.1016/j.compmedimag.2021.101974

О. Й. Піцун

Західноукраїнський національний університет, м. Тернопіль, Україна

БАГАТОПОТОКОВЕ РОЗПАРАЛЕЛЕННЯ ПРОЦЕСУ ОБЧИСЛЕННЯ ХАРАКТЕРИСТИК КЛІТИН БІОМЕДИЧНИХ ЗОБРАЖЕНЬ

Запропоновано підхід до розпаралелення процесу обчислення кількісних характеристик ядер клітин на біомедичних зображеннях (цитологічних, гістологічних, імуногістохімічних), що дасть змогу пришвидшити процес постановки діагнозу. Для постановки діагнозу використовують сучасні методи і засоби інтелектуального аналізу даних, складовою частиною якого є класифікація даних. При використанні згорткових нейронних мереж вхідними даними для їх класифікації є зображення у форматі *.jpg, *.png, *.bmp та ін. Альтернативні алгоритми та засоби оброблення даних здебільшого вимагають наявності кількісних характеристик. У випадку використання біомедичних зображень кількісними характеристиками є площа, периметр, окружність, довжина головної та бічної осі ядра клітин. Площа та інші характеристики ядер клітин характеризують нормальний стан або наявність патології.

Розпаралелення процесу обчислення характеристик біомедичних зображень реалізовано на підставі алгоритмів комп'ютерного зору для виділення необхідних об'єктів і засобів програмного розпаралелення задач на рівні потоків для пришвидшення процесу обчислення характеристик ядра клітин. Підхід полягає у програмному розпаралеленні на рівні потоків незалежних задач обчислення кількісних характеристик ядер клітин з використанням Executor framework. Встановлено, що наявні системи автоматизованої мікроскопії та системи діагностування на підставі зображень не володіють наявністю великої кількості характеристик ядер клітин та не мають механізмів до розпаралелення процесу їх обчислення. Запропонований підхід дає змогу пришвидшити процес обчислення кількісних характеристик ядер клітин на 25 %. Актуальність задачі розпаралелення обумовлена потребою опрацювання великого обсягу даних для подальшої їх редукції та класифікації. Розпаралелення на рівні потоків дає змогу збільшити швидкість опрацювання зображень та не вимагає наявності спеціалізованого апаратного забезпечення.

Ключові слова: імуногістохімія; гістологія; потоки; ядро клітини.

_____