

Abstract

<sup>1</sup>Kovalev V.,  
<sup>2</sup>Diachenko Y.,  
<sup>1</sup>Malyshev V.,  
<sup>3</sup>Rjabceva S.,  
<sup>4</sup>Kolomiets O.,  
<sup>4</sup>Lyndin M.,  
<sup>4</sup>Moskalenko R.,  
<sup>2</sup>Dovbysh A.,  
<sup>4</sup>Romaniuk A.,

<sup>1</sup>*Biomedical Image Analysis Department of United Institute of Informatics Problem Belarus National Academy of Sciences, Minsk, Belarus;*

<sup>2</sup>*Laboratory "Center of electron and light microscopy" Institute of Physiology of the National Academy of Sciences of Belarus, Minsk, Belarus;*

<sup>3</sup>*Department of Computer Sciences of Sumy State University, Sumy, Ukraine;*

<sup>4</sup>*Department of Pathology of Sumy State University, Sumy, Ukraine*

**COMPARATIVE FEATURES OF OPEN SOURCE SOFTWARE PRODUCTS FOR THE DEVELOPMENT OF AN AUTOMATED BREAST CANCER DIAGNOSTIC PROGRAM**

Breast cancer is one of the most common cancer diseases in the world among women. The reliability of histological verification of breast cancer depends on pathologist's experience, knowledge, his willingness to self-improve and study specialized literature. Digital pathology is also widely used for educational purposes, in telepathology, teleconsultation and research projects. Recently developed Whole Slide Image (WSI) system opens great opportunities in the histopathological diagnosis quality improvement. Digital whole-slide images provide the effective use of morphometry and various imaging techniques to assist pathologists in quantitative and qualitative evaluation of histopathological preparations. The development of software for morphological diagnosis is important for improving the quality of histological verification of diagnosis in oncopathology. The purpose of this work is to find and benchmark existing open-source software for the whole-slide histological images processing. Choosing an open source program is an important step in developing an automated breast cancer diagnosis program.

The result is a detailed study of open-source software: ASAP, Orbit, Cytomine and QuPath. Their features and methods of image processing were analyzed. QuPath software has the best characteristics for extending it with an automated module for the cancer diagnosis. QuPath combines a user-friendly, easy-to-use interface, customizable functionality, and moderate computing power requirements. Besides, QuPath works with whole-slide images with immunohistochemical markers; features implemented in this software allow making a morphometric analysis.

QuPath saves time for a graphical user interface development and provides a scalable system to add new key features. QuPath supports third-party MATLAB and Python extensions.

**Keywords:** breast cancer, digital pathology, whole-slide image, histopathological diagnosis, automated diagnostic program.

Corresponding author: [pathomorph@gmail.com](mailto:pathomorph@gmail.com)

**Резюме**<sup>1</sup>Ковальов В. О.,<sup>2</sup>Дяченко Є. В.,<sup>1</sup>Малишев В. Д.,<sup>3</sup>Рябцева С. М.,<sup>4</sup>Коломієць О. О.,<sup>4</sup>Линдін М. С.,<sup>4</sup>Москаленко Р. А.,<sup>2</sup>Довбиш А. С.,<sup>4</sup>Романюк А. М.,<sup>1</sup>Відділення біомедичних зображень, Об'єднаний інститут проблем інформатики НАН Білорусі, м. Мінськ, Білорусь;<sup>2</sup>Лабораторія "Центр електронної і світлової мікроскопії", Інститут Фізіології НАН Білорусі, м. Мінськ, Білорусь;<sup>3</sup>Кафедра комп'ютерних наук Сумського державного університету, м. Суми, Україна;<sup>4</sup>Кафедра патологічної анатомії Сумського державного університету, м. Суми, Україна**ПОРІВНЯЛЬНА ХАРАКТЕРИСТИКА МОЖЛИВОСТЕЙ ПРОГРАМНИХ ПРОДУКТІВ З ВІДКРИТИМ КОДОМ ДОСТУПУ ДЛЯ РОЗРОБКИ АВТОМАТИЗОВАНОЇ ПРОГРАМИ ДІАГНОСТИКИ РАКУ МОЛОЧНОЇ ЗАЛОЗИ**

Рак молочної залози (РМЗ) – одне з найчастіших онкологічних захворювань у світі серед жінок з високою захворюваністю та смертністю. Надійність гістологічної верифікації раку молочної залози залежить від досвіду, знань патолога, його готовності до самовдосконалення та вивчення спеціалізованої літератури. Цифрова патологія також широко використовується в навчальних цілях, в телепатології, телеконсультаціях та науково-дослідних проектах. Нещодавно розроблена система Whole Slide Image (WSI) відкриває великі можливості в покращенні якості гістопатологічної діагностики. Розробка різноманітного програмного забезпечення для морфологічної діагностики - важливий напрямок для підвищення якості гістологічної верифікації діагнозу в онкопатології. Метою роботи є пошук та порівняльний аналіз існуючих програмних засобів з відкритим вихідним кодом для обробки повно-слайдових гістологічних зображень.

У результаті здійснено детальне дослідження програмних продуктів з відкритим вихідним кодом ASAP, Orbit, Cytomine та QuPath. Проаналізовано їх особливості та методи обробки зображень, на основі яких було визначено, що програмне забезпечення QuPath має найкращі характеристики для використання у розробці автоматизованої програми діагностики онкологічної патології. QuPath поєднує в собі зручний простий інтерфейс, розширюваність функціоналу користувацькими модулями та відсутність високих вимог до обчислювальної потужності. Крім цього, QuPath направлений на роботу з повно-слайдовими зображеннями з імуногістохімічними маркерами; функції, реалізовані в даному програмному забезпеченні, дозволяють проводити морфометричний аналіз.

Використання QuPath допоможе зберегти час на розробку графічного інтерфейсу користувача і надасть систему, яку можна розширювати - додавати нові ключові функції, визначені метою. QuPath підтримує розширення, розроблені за допомогою різних мов програмування, включаючи MATLAB та Python.

**Ключові слова:** рак молочної залози, цифрова патологія, повно-слайдове зображення, гістопатологічна діагностика.

**Автор, відповідальний за листування:** [pathomorph@gmail.com](mailto:pathomorph@gmail.com)

**Introduction**

Breast cancer is one of the most common cancer diseases in the world among women [1]. It amounts 23–27% of malignant neoplasms found among women and develops twice as often as cancer of some other localization [1]. The reliability of histological verification of breast cancer depends on morphologist's experience, knowledge, his willingness to self-improve and study specialized literature. Due to the facts above, the issue of high-tech (using advanced information technologies) expert system development is relevant not only for

pathologists but also for every woman waiting for reliable morphological verification of the malignant process of the mammary gland with an accurate assessment of receptor status.

Digital pathology is also widely used for educational purposes, in telepathology, teleconsultation, and research projects [2]. Digital pathology allows you to share slides and annotate them much easier. Downloading and uploading the annotated sets of data creates new opportunities for e-learning and knowledge sharing in the sphere of pathological anatomy.

Recently developed Whole Slide Image (WSI) system opens great opportunities in the histopathological diagnosis quality improvement [3]. This innovation is the complete digitization of histological preparation of any pathology in high resolution. Digital whole-slide images provide the effective use of morphometry and various imaging techniques to assist pathologists in quantitative and qualitative evaluation of histopathological preparations [4].

Development of morphological diagnostics software is important for improving the quality of histological verification of the diagnosis in oncopathology [5].

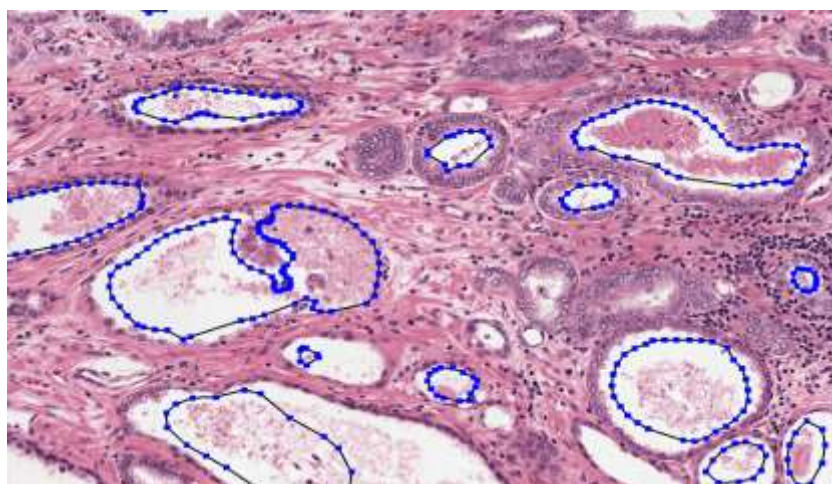
The purpose of this work is to find and benchmark existing open-source software for the whole-slide histological image processing.

**Results.** Among existing open-source software for the whole-slide histological image processing, we are to find the most suitable one with extensible program interface to supplement it with a custom automated module using Deep Learning with the methods of information-extreme intellectual technology of comparative pattern recognition. The analysis of existing packages and programs before the algorithm development phase allows reducing the time for the program and graphical interfaces implementation and focusing on the set goal. The analysis was performed on software with extensible program interface which is designed to work with whole-slide images.

We have studied ASAP, Orbit, Cytomine and QuPath open-source software products in details.

**ASAP** Software (Automated Slide Analysis Platform) is an open-source platform for visualization, annotation and analysis of multi-resolution histopathology whole-slide images. The developers of the platform note that “ASAP consists of several key-components (slide input/output, image processing, viewer) which can be used separately. It is built on open-source packages like OpenSlide, Qt and OpenCV but also tries to extend them in several meaningful ways” [7].

The feature of this software is its support for the third-party vendors’ whole-slide image formats via OpenSlide. ASAP can read not only scanned images from Olympus, Aperio, Hamamatsu, Ventana machines, but also fluorescence images in Leica Image File format (LIF). In the ASAP program environment, one can write “generic multi-resolution tiled TIFF files for ARGB, RGB, Indexed and monochrome images (including support for different data types)” [7]. ASAP supports Patches - image primitives, that can be used as a part of image processing filters. OpenCV connection and the viewer based on Qt allow visualizing whole-slide images responsively and fast. The platform provides various annotation tools (fig. 1) with spline, polygonal and point markers. ASAP stores the annotation info in XML-format to make it easier to use for some other purposes.



**Figure 1 – Flexible annotation tools for whole-slide images in ASAP**

One of the benefits of this program is on-slide visualization of image analysis and machine learning results such as segmentation masks with customizable lookup-tables. Libraries for reading and viewing the whole-slide images can be

extended with custom plugins in order to add new functionality to ASAP. One should use ‘file-formats’, ‘tools’, ‘extensions’ and ‘filters’ interfaces for that. Another feature, as specified in the official ASAP documentation “is the integration

of on-the-fly image processing while viewing (current examples include colour deconvolution and nuclei detection)” [7].

Currently one can install ASAP using only Linux or Windows x64 Operating Systems. ASAP uses non-specific for other platforms libraries, so the installation and compilation should not cause any problems. Configuring ASAP one should take into account that it depends upon the following open-source third-party libraries:

- Boost (<http://www.boost.org/>)
- DCMTK (<http://dicom.offis.de/dcmtdk.php.en>)
- JasPer (<http://www.ece.uvic.ca/~frodo/jasper/>)
- libjpeg (<http://libjpeg.sourceforge.net/>)
- libtiff (<http://www.libtiff.org/>)
- OpenCV (<http://www.opencv.org/>)
- OpenSlide (<http://openslide.org/>)
- PugiXML (<http://pugixml.org/>)
- Qt (<http://www.qt.io/>)
- SWIG (<http://www.swig.org/>) (only for Python wrapping of the IO library)
- unittest++ (<https://github.com/unittest-cpp/unittest-cpp>)
- zlib (<http://www.zlib.net/>)

ASAP components can be also built separately, so you can configure the system for specific purposes using custom libraries. The ‘Package-on-install’ option allows building a binary setup-package.

The open-source software for the whole-slide analysis **Orbit Image Analysis** was developed at Actelion Pharmaceuticals Ltd, now Idorsia Pharmaceuticals Ltd by Manuel Stritt in Java. Orbit

Image Analysis software has sophisticated image analysis algorithms: “tissue quantification using machine learning techniques, object/cell segmentation, and object classification are the basic ones. Region of interest (ROI) can be defined by manual annotations or via a trainable exclusion map” [8]. Orbit’s algorithms are aimed at processing the whole-slide scanned images that can be of gigapixel size. Integration with other software like Omero and Spark is available with support for using image servers. One can also run Orbit offline using NDPI-, SVS- and SVS-formatted whole-slide images. To reduce the load on the computing power of the machine, one can use Spark software as a scale-out infrastructure. The advantage of Orbit for developers is that it provides an extensible API to add custom modules.

On the Orbit software official web-page is noted: “machine learning-based tissue quantification allows the domain expert to train the system-specific (e.g stained) tissue classes and to quantify them (compute the ratio of different tissue classes, e.g. percentage of collagen in a tissue)” [8].

Orbit Image Analysis function allows to segment object (e.g. such objects like nerve cells etc). This feature is “based on trainable foreground/background classes” [8]. Image Analysis output such as intensity, shape etc can be used later for object classification.

Based on machine learning, object classification (fig. 2) helps to divide the output data after object segmentation into classes. One can specify these classes also by using examples. As well as ASAP, Orbit provides the ROI annotation feature: regions of interest can be highlighted manually or automatically through the trainable exclusion map.

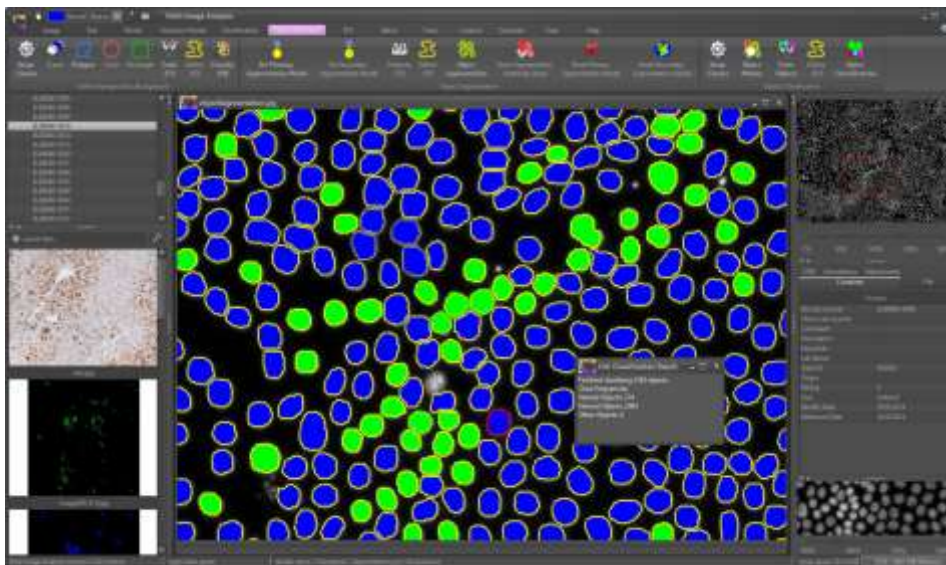


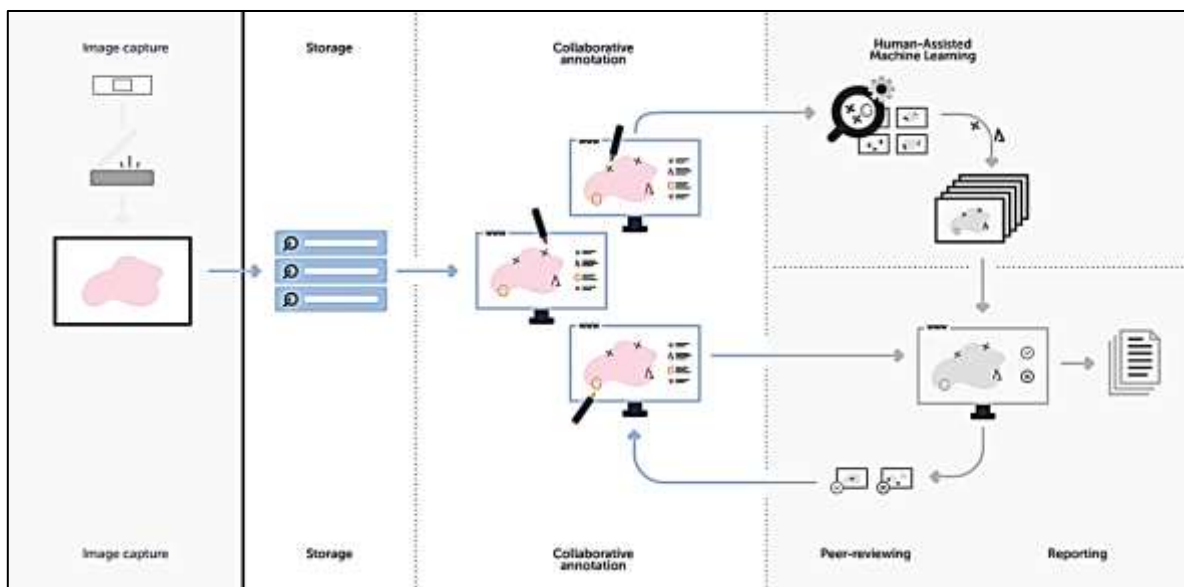
Figure 2 – ROI Object Classification in Orbit

“Orbit’s context-based structure classification is based on the so-called structure-size, a surrounding area for each pixel, which is used to compute several features on multiple image resolutions. These features describe the structure of the underlying tissue or other biological sample and are used as an input for a Support Vector Machine (SVM) to discriminate regions within the image” [9]. With this approach, one can create a model in a short amount of time specifying just a few training regions.

**Cytomine** is open-source software written in Java for the whole-slide image analysis using machine learning algorithms. The software was developed by The Cytomine Company. Cytomine allows to upload and work with 3DHistech MRXS-, Aperio SVS-, generic-tiled TIFF-, Hamamatsu VMS-, Leica SCN-, NDPI-, Philips TIFF- and OME-TIFF-formatted whole-slide images. Cloud-Based image storage and sharing services are provided (but paid).

Cytomine supports image annotations as well as ASAP and Orbit. ROI or regions of interest can be highlighted with polygon, rectangle, ellipse or custom annotation instruments. Unlike previous products, Cytomine allows describing ROI structure elements manually or using structured vocabularies (ontologies). Consolidation of image atlases is also possible: used for analysis custom images can be used as input to the machine learning algorithm.

Image analysis in the Cytomine software environment is performed using “machine learning algorithms (fig. 3) implemented on modern, multi-core architectures and computing clusters. Algorithms for the recognition of tissue substructures, cell types, and landmarks can be trained by experts to speed-up and refine the detection and quantification of the most relevant biological objects in your images.” [10] Extensible API allows changing or complementing algorithms in order to make final results more accurate.



**Figure 3 – Cytomine Whole-slide Image Processing Algorithm**

The benefits of Cytomine include its extensibility: its functions can be extended with custom algorithms, plug-ins and web applications. It is possible to interact with Cytomine from third-party software owing to RESTful API. Cytomine Python and Java clients are already available, but other programming languages can be used as well.

Another feature of the software is the existence of web-application; unlike the previous counterparts (ASAP and Orbit), Cytomine lets you work with the web application directly in the web

browser. The Cytomine software UI and UX is constantly updated.

Cytomine official web-page [10] gives full access to the Open Image Collection with the whole-slide images of the tissue morphology. One can use them for the program algorithm testing or some other purposes. Cytomine Official Documentation [11] provides detailed instructions for installing and operating this software.

**QuPath** is a cross-platform open-source software written in Java for digital pathology whole-slide image analysis. The software was

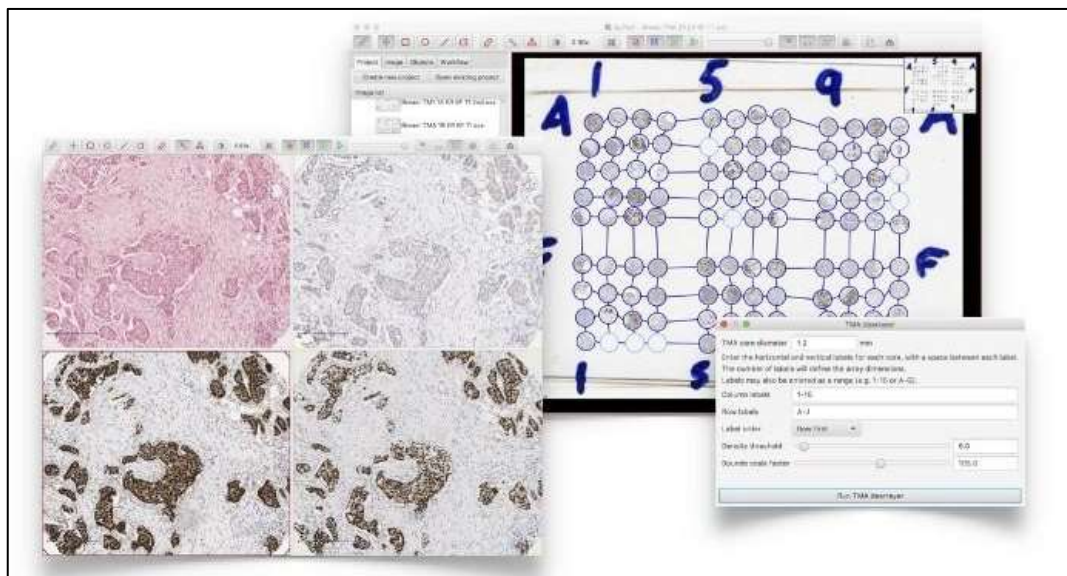
created at the Centre for Cancer Research & Cell Biology at Queen’s University Belfast, as part of research projects funded by Invest Northern Ireland and Cancer Research UK. Now the University of Edinburgh maintains the QuPath software [12].

QuPath software has extensible annotation and visualization tools, novel algorithms for common tasks, e.g. cell segmentation and tissue microarray dearraying. It also has interactive machine learning on the board, e.g. for cell and texture classification, an object-based hierarchical data model, with scripting support, extensibility with new features and different image sources support. QuPath provides integration with other tools, e.g. MATLAB and ImageJ.

Overall, QuPath aims to provide researchers with a new set of tools to help with bio-image analysis in a way that is both user- and developer-friendly.

QuPath software provides a flexible and fast whole-slide image (usually more than 40 GB) viewer. Such functionality as “biomarker quantification (nuclear, cytoplasmic & membranous biomarkers) can all be quantified quickly using automated segmentation algorithms combined with trainable cell classification” [12]. QuPath’s software benefits include support for tissue micro-arrays: automated dearraying of Tissue Microarrays and

the ability to view related cores side-by-side (fig. 4). Expert tumour detection is also available: tumour identification algorithms can be applied directly to regions of interest without the need to highlight separate tumour areas. QuPath provides swift whole-slide image analysis: huge areas are split into smaller ones and are analyzed with effective algorithms which do not require high computing power requirements. Another advantage is the flexible object classification: object classification can be done using default classifiers, such as Random Forest, or one can create customized algorithms changing classification parameters. QuPath features include interactive tools: “extensive tools for slide navigation, annotating areas, exporting image regions or counting cells - either manually, or using automated cell detection” [12]. The last but not the least is that one can perform a detailed analysis of the script using powerful debugging tools that allow one to log and compare performance results. Important benefits of the QuPath software platform are the ability to use custom scripts and share data using ImageJ or other open-source tools, image imports from cloud storages etc. QuPath data structures allow complementing original scripts to perform specific customized analysis.



**Figure 4 – QuPath Tissue Microarray Support**

Other features of QuPath software are analytics and data export (fig. 5). They allow creating “interactive results tables, histograms, scatterplots

& survival curves directly within QuPath, or exporting results in standard formats to import into other software if required” [12].

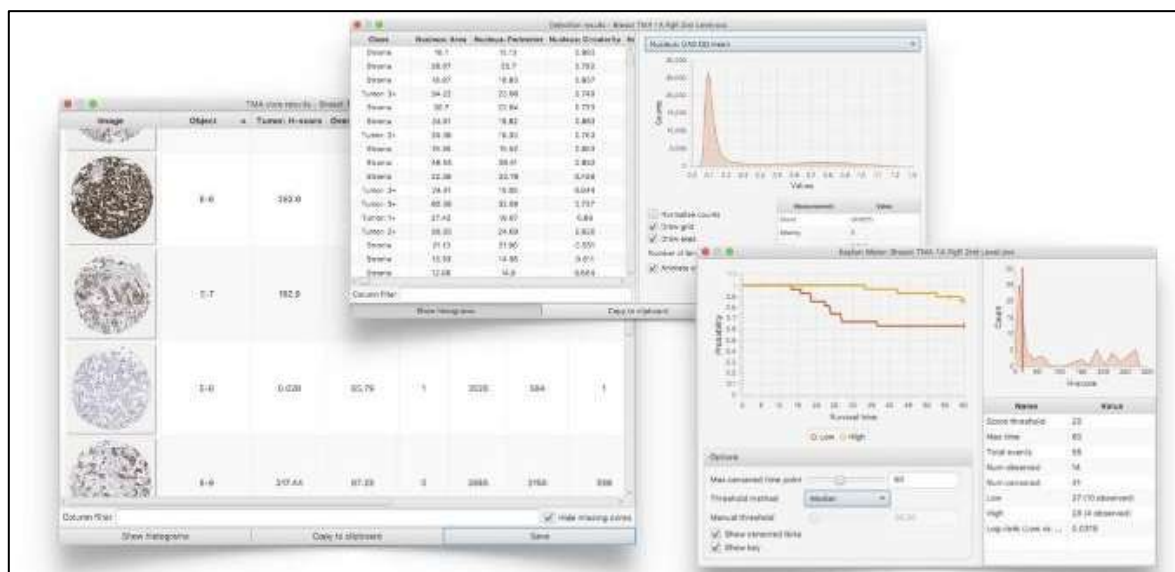


Figure 5 – QuPath Analytics and Export Tools

QuPath visualization codes the objects with specific colours corresponding to their features.

One of the main advantages of QuPath is the ability to work with a wide variety of OpenSlide whole-slide image formats: Aperio (.svs, .tif),

Hamamatsu (.vms, .vmu, .ndpi), Leica (.scn), MIRAX (.mrxs), Philips (.tiff), Sakura (.svslide), Trestle (.tif), Ventana (.bif, .tif), Generic tiled TIFF (.tif), ImageJ TIFF, JPEG, PNG.

Table 1 – Features of open-source software for the whole-slide histological images processing

Features	Software			
	ASAP	Orbit	Cytomine	QuPath
Reading scanned whole-slide images from different vendors	+	+	+	+
Output data in different formats	+	+	+	+
Annotation tools	+	+	+	+
Analysis result visualization	+	-	+	+
Extensibility	+	+	+	+
Cross-platform	-	-	+	+
Computing power dependence	+	+	+	-
Third-party services integration	-	+	+	+
Consolidation of image atlases	-	-	+	+
Web-app	-	-	+	-
Paid services	-	-	+	-
Data exchange	-	-	-	+

### Conclusions

In the course of the study, the open-source software for the analysis of whole-slide images of tissue histology was analyzed; they are ASAP, Orbit, Cytomine and QuPath. The features and methods of image processing software were identified. QuPath software has the best characteristics for extending it with an automated module for the cancer diagnosis. QuPath combines

a user-friendly, easy-to-use interface, customizable functionality, and moderate computing requirements. QuPath works with the whole slide images with immunohistochemical markers; this software features implemented allow doing morphometric analysis.

QuPath saves GUI development time and provides the extensible API. QuPath maintains custom MATLAB and Python extensions.

**Conflict of interest**

The authors declare no conflict of interest.

**Acknowledgements**

This research has been performed with the financial support of joint UKRAINE-BELARUS R&D project «Development of an automated

program for differential diagnosis of breast tumours with a morphometric evaluation of the receptor status of cancer cells» in 2019–2020.

**Відомості про авторів/Information about the authors**

Ковальов Василь Олексійович – к.т.н, завідувач відділення біомедичних зображень, Об'єднаний інститут проблем інформатики НАН Білорусі, Мінськ, Білорусь, vassili.kovalev@gmail.com;

Дяченко Єгор Вадимович – магістрант кафедри комп'ютерних наук Сумського державного університету, Суми, Україна, diachenko.yehor@gmail.com;

Малишев Валерій Дмитрович – інженер-програміст відділення біомедичних зображень, Об'єднаний інститут проблем інформатики НАН Білорусі, Мінськ, Білорусь, malyshevalery@gmail.com;

Рябцева Світлана Миколаївна – к.мед.н, завідувачка лабораторії “Центр електронної і світлової мікроскопії”, Інститут Фізіології НАН Білорусі, Мінськ, Білорусь, sveta.rjabceva@tut.by;

Коломієць Олена Олегівна – аспірантка кафедри патологічної анатомії Сумського державного університету, Суми, Україна, o.kolomiets@med.sumdu.edu.ua;

Линдін Микола Сергійович – к.мед.н, доцент кафедри патологічної анатомії Сумського державного університету, Суми, Україна, n.lyndin@med.sumdu.edu.ua;

Москаленко Роман Андрійович – д.мед.н, доцент кафедри патологічної анатомії Сумського державного університету, Суми, Україна, r.moskalenko@med.sumdu.edu.ua;

Довбиш Анатолій Степанович – д.т.н, професор, завідувач кафедри комп'ютерних наук Сумського державного університету, Суми, Україна, a.dovbysh@cs.sumdu.edu.ua;

Романюк Анатолій Миколайович – д.мед.н, професор, завідувач кафедри патологічної анатомії Сумського державного університету, Суми, Україна, pathomorph@gmail.com

**References**

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359- E386. doi: 10.1002/ijc.29210
2. Hamilton PW, Wang Y, McCullough SJ. Virtual microscopy and digital pathology in training and education. *APMIS* 2012. 120:305–15. doi: 10.1111/j.1600-0463.2011.02869.x
3. Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological image analysis: a review. *IEEE Rev Biomed Eng*. 2009;2:147-171. doi:10.1109/RBME.2009.2034865
4. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, Hulsbergen-van de Kaa C, Bult P, van Ginneken B, van der Laak J. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. rep.* 2016;6:26286. doi: 10.1038/srep26286
5. Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, Cathro HP, Cheng L, Cooper K, Dickey GE, Gill RM, Heaton RP Jr, Kerstens R, Lindberg GM, Malhotra RK, Mandell JW, Manlucu ED, Mills AM, Mills SE, Moskaluk CA, Nelis M, Patil DT, Przybycin CG, Reynolds JP, Rubin BP, Saboorian MH, Salicru M, Samols MA, Sturgis CD, Turner KO, Wick MR, Yoon JY, Zhao P, Taylor CR. Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study). *Am J Surg Pathol*. 2018;42:39–52. doi: 10.1097/PAS.0000000000000948
6. Computation Pathology Group. ASAP - Automated Slide Analysis Platform. *Geert Litjens*. Retrieved from:



- <https://computationalpathologygroup.github.io/ASAP/>
7. Computation Pathology Group. ASAP - Automated Slide Analysis Platform GitHub repository. [Geert Litjens.]. Retrieved from: <https://github.com/computationalpathologygroup/ASAP>
  8. Actelion Pharmaceuticals Ltd. Orbit Image Analysis. Manuel Stritt. Retrieved from: <https://www.orbit.bio/>
  9. Stritt M, Anna K. Stalder, Vezzali E. *Orbit Image Analysis: an open-source whole slide image analysis tool*. Retrieved from: <https://www.biorxiv.org/content/10.1101/731000v1>
  10. Vincke G, Hoyoux R. *Cytomine. The Cytomine Company*. Retrieved from: <https://cytomine.coop/>
  11. Rollus L, Marée R. *Cytomine Technical Documentation. The Cytomine Company*. Retrieved from: <https://doc.cytomine.be/>
  12. Bankhead P. Documentation Q. *Centre For Cancer Research & Cell Biology*. Retrieved from: <https://qupath.github.io/>
  13. Bankhead P, Loughrey MB, Fernández JA. *Open source software for digital pathology image analysis*. Retrieved from: <https://www.nature.com/articles/s41598-017-17204-5>

(received 17.11.2019, published online 29.12.2019)

(одержано 17.11.2019, опубліковано 29.12.2019)