Int. J. Appl. Math. Comput. Sci., 2021, Vol. 31, No. 1, 85–106
DOI: 10.34768/amcs-2021-0007

BREAST CANCER NUCLEI SEGMENTATION AND CLASSIFICATION BASED ON A DEEP LEARNING APPROACH

MAREK KOWAL [a,*], MARCIN SKOBEL [a], ARTUR GRAMACKI [a], JÓZEF KORBICZ [a]

[a] Institute of Control and Computation Engineering
University of Zielona Góra
ul. Szafrana 2, 65-516 Zielona Góra, Poland
e-mail: M.Kowal@issi.uz.zgora.pl

One of the most popular methods in the diagnosis of breast cancer is fine-needle biopsy without aspiration. Cell nuclei are the most important elements of cancer diagnostics based on cytological images. Therefore, the first step of successful classification of cytological images is effective automatic segmentation of cell nuclei. The aims of our study include (a) development of segmentation methods of cell nuclei based on deep learning techniques, (b) extraction of some morphometric, colorimetric and textural features of individual segmented nuclei, (c) based on the extracted features, construction of effective classifiers for detecting malignant or benign cases. The segmentation methods used in this paper are based on (a) fully convolutional neural networks and (b) the marker-controlled watershed algorithm. For the classification task, seven various classification methods are used. Cell nuclei segmentation achieves 90% accuracy for benign and 86% for malignant nuclei according to the F-score. The maximum accuracy of the classification reached 80.2% to 92.4%, depending on the type (malignant or benign) of cell nuclei. The classification of tumors based on cytological images is an extremely challenging task. However, the obtained results are promising, and it is possible to state that automatic diagnostic methods are competitive to manual ones.

Keywords: breast cancer, nuclei segmentation, classification, image processing.

1. Introduction

According to the world statistics, breast cancer was the most common type of cancer among women in 2018. Moreover, the estimated number of new cases of breast cancer reached 2 million while that of deaths was around 600,000 in 2018 (Bray *et al.*, 2018). There is a need to increase the effectiveness of classifying new cases of the disease. One of the most popular methods in the diagnosis of breast cancer is fine-needle biopsy without aspiration (FNB). Despite the controversy (Litherland, 2002), the FNB method is still used as an effective and minimally invasive technique of breast cancer diagnosis. The FNB-based biological material is analyzed by doctors for malignancy. To facilitate the task of cytological image diagnosis, the samples are stained using the hematoxylin and eosin (H&E) method, which is cheap and effective. Hematoxylin is responsible for staining basophilic objects (e.g., cell nuclei) in blue, while eosin is responsible for staining eosinophilic objects (e.g., cytoplasm) in red.

The development of medical imaging techniques opens new paths for computer-aided diagnosis research. Progress also involves the analysis of cytological images. Recently, the material obtained as a result of FNB can be scanned into a digital form and analysed by various computer techniques.

Cell nuclei are the most important elements of cancer diagnostics based on cytological images. Therefore, the first step to successful classification of cytological images is effective automatic segmentation of cell nuclei. In order to check how effective the given segmentation method is, a set of test images should be prepared with manually segmented cell nuclei. Unfortunately, the prepared set of test images contains samples of very different quality. There are images having well-separated nuclei, but there are also samples where the separation between adjacent nuclei is very poor. The variety of samples allows verifying the effectiveness of the proposed methods for both simple to difficult cases.

*Corresponding author

This paper presents the process of (a) segmentation, (b) feature extraction and (c) classification of individual cell nuclei of cytological images. The segmentation task is carried out using deep learning techniques. Extracted features may be divided into three characteristic groups, containing morphometric, colorimetric and textural features. Having the extracted features of individual nuclei, it is possible to provide effective classifiers for detecting malignant or benign cases. The main purpose of the presented experiment is to demonstrate whether automatic segmentation of cell nuclei based on an artificial convolutional neural network and the segmentation algorithm gives watershed classification results comparable to manual segmentation.

2. Related works

Computer-assisted cytology is usually composed of a set of methods of (a) semantic segmentation, (b) cell nuclei detection, (c) instance segmentation, (d) feature extraction and selection, and (e) classification.

Semantic segmentation of cytological and histopathological images usually begins with digital stain separation and color normalization. Stain separation approaches can be divided into supervised and unsupervised methods. The former are based on the color deconvolution algorithm (Ruifrok and Johnston, 2001). By contrast, the latter are based on independent component analysis (ICA) or non-negative matrix factorization (NMF) (Alsubaie *et al*., 2017; Macenko *et al*., 2009; Rabinovich *et al*., 2004). After color separation we usually need to normalize the resulting images, because cytological samples coming from different laboratories may differ in color.

Color variation can arise due to different staining protocols, different stain brands, the shelf life of stains, or due to using different microscopy scanners. To tackle this problem, various color normalization approaches have been proposed. They can be generally categorized into histogram matching methods, color transfer methods, and spectral matching methods (Piorkowski and Gertych, 2019; Santanu *et al*., 2018). After color separation and normalization, we obtain the image of dye concentration that deposits in cell nuclei. In the next step, we need to segment this image using either image thresholding methods or artificial neural networks (Hayakawa *et al*., 2019; Xing and Yang, 2016). Both approaches have their pros and cons; image thresholding is unsupervised and fast, but not very accurate. On the other hand, artificial neural networks require training data, but the results of semantic segmentation are far better (Cui *et al*., 2018; Kowal *et al*., 2018; Naylor *et al*., 2017; Sadanandan *et al*., 2017).

Having the image segmented into a background and a foreground, the next step is to detect and segment individual cell nuclei located in the foreground. The common approaches to this problem are based on active contours, mathematical morphology, region growing, or the watershed transform (Irshad *et al*., 2014). There are also other approaches such as tensor voting followed by nuclei boundary extraction (Paramanandam *et al*., 2016) or intelligent gravitational search (Mittal and Saraswat, 2019). However, the most promising one seems to be artificial neural networks because last years have brought enormous progress in the area of object detection and segmentation using convolutional neural networks (CNNs) (Kowal *et al*., 2018; Höfener *et al*., 2018; Wang *et al*., 2016)

A lot of papers that deal with computer-assisted cytology usually concentrate either on the cell nuclei segmentation problem or on their classification. Here, we develop and test the entire image processing pipeline, including image segmentation and cell nuclei classification. This allows us to verify the effectiveness of the detection and segmentation of cell nuclei in terms of the accuracy of their classification. Results of similar studies are presented by Dudzińska and Piórkowski (2020), Kowal *et al*. (2018), Kowal and Filipczuk (2014), Fondón *et al*. (2018) or Spanhol *et al*. (2016).

3. Image database

This article investigates digital cytological images of breast cancer. The material used in the research is an archival collection of samples taken from patients before 2014. Samples were taken using fine needle biopsy without aspiration (under ultrasonography support) using 0.4 or 0.5 millimeter needles. The biopsy procedure was carried out by pathologists from the University Clinical Hospital in Zielona Góra, Poland. Cellular material was used clinically for cancer diagnosis, so there is no doubt about the type of cancer found in the samples. Nowadays, materials collected under the core needle biopsy procedure are used more often for diagnostic tasks. Nevertheless, the Polish Society of Clinical Oncology allows the use of fine needle biopsy in the diagnosis of breast cancer according to the guidelines issued in 2018 and still valid (Jassem and Krzakowski, 2018). Core needle biopsy has not fully supplanted fine needle biopsy due to the benefits of its use (e.g., lower invasiveness and price). Besides, in the case of computer-assisted cytology, the specific type of biopsy is less important because the machine "sees" the digital image differently than the human eye.

The collected material was fixed with spray and dyed with hematoxylin (used to stain basophilic tissue structures) and eosin (used to stain acidophilic tissue structures). As a result of staining, the cell nuclei acquire a blue tint, while the red blood cells and cytoplasm are red. After staining and fixing, glass slides were digitized

with the help of the Olympus VS120 scanner. As a result, we obtained a set of 50 virtual slides (25 malignant and 25 benign cases) with the size of approximately 200,000 by 100,000 pixels.

To train our diagnostics system, we needed a certain number of manually segmented cell nuclei, both benign and malignant. Benign cell nuclei can be selected from virtual slides classified as benign, and malignant cell nuclei from malignant cases. It should, however, be borne in mind that, in the case of cytological preparations of malignant cases, there is a probability that a randomly selected region of interest (ROI) will contain malignant as well as benign cell nuclei. To avoid any mistakes during ROI selection, we involved a pathologist in this task. The doctor marked 11 small ROIs on each virtual slide. The main selection criterion was to ensure that (a) the ROI taken from malignant samples would contain only malignant cell nuclei, (b) the number of nuclei would be at least 10, (c) nuclei would be well visible. The result was a set of 550 ROIs for 50 patients (set A), where 275 ROIs represent benign cases and 275 ROIs stand malignant ones.

Unfortunately, the time costs associated with manual segmentation of 550 images were far too high for our capabilities. Therefore, we created a smaller set of images (set B) by selecting two ROIs for each patient, resulting in 50 ROIs for benign cases and 50 ROIs for malignant ones (in total, 100 ROIs for 50 patients). All ROIs from set B were manually segmented and then used to train and validate a U-Net neural network (see Fig. 1). For this purpose, set B, consisting of 100 manually segmented ROIs, was divided into two subsets: training (set B1) and validation (set B2). The training set (B1) included 50 ROIs from 12 randomly chosen benign patients and 13 randomly chosen malignant patients (two ROIs for each patient). The validation set (B2) included 50 ROIs from 13 benign patients and 12 malignant patients (two ROIs for each patient).

We decided not to create a test set because the number of manually segmented images was relatively small, so we did not want to deplete the already small training and validation sets. The validation set (B2) was used to determine an optimal point to stop the training of U-Net to avoid overfitting. The U-Net network test was carried out on ROIs selected from set A, which were not included in set B. In this case, the evaluation of the quality of segmentation could only be visual because the images from the test set were not manually segmented. The quantitative assessment of the segmentation quality could only be carried out on the validation set (B2). The subsequent stages of dividing a set of images are represented schematically in Fig. 2.

It should be emphasized that the experiment is not about classifying images but individual cell nuclei. Finally, about 1500–1700 benign and about 600–750

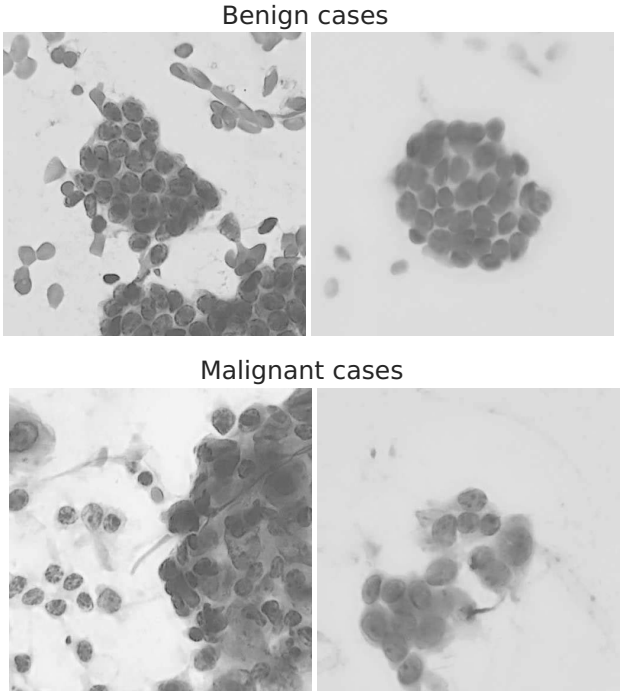

Fig. 1. Cell nuclei: fragments selected from virtual slides.

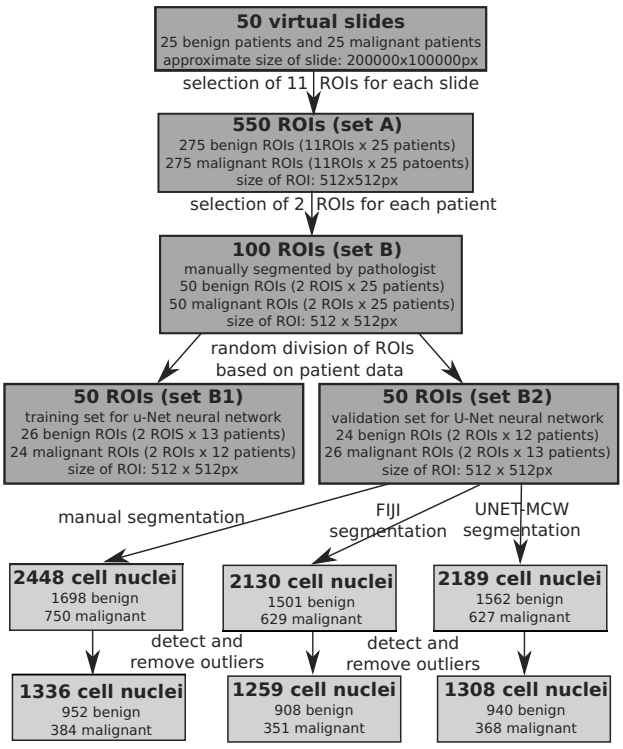

Fig. 2. Data distribution for various data sets.

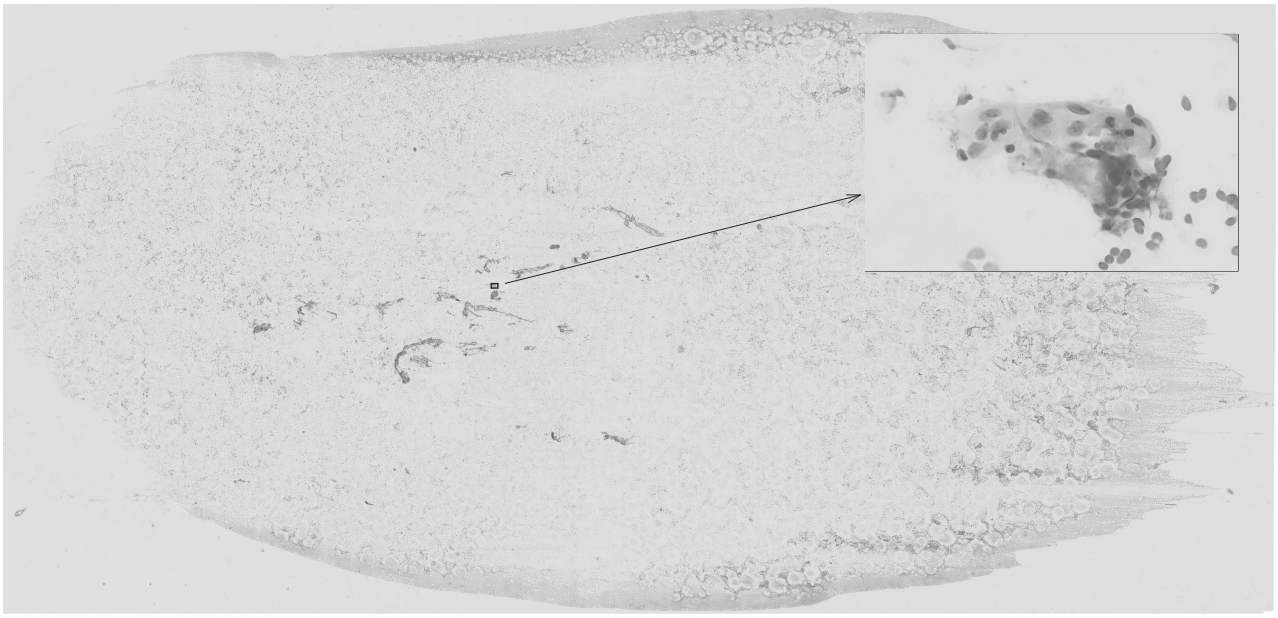

Fig. 3. Full view of a selected virtual slides digitized using the Olympus VS120 Virtual Microscopy System. A small fragment, marked with a small black rectangle and an arrow, is selected and then shown enlarged.

malignant cell nuclei were obtained based on 50 ROIs from set B2.

Manually selected ROIs are required only to train our system. After that, the system works in inference mode and classifies cell nuclei from new virtual slides. It can even process the entire cytological specimen, so no ROI selection is required. But still, it is more likely that the pathologist will choose the sample he or she is interested in to shorten the processing time.

4. Methods

4.1. Overview. The system aims to classify cellular material coming from an examination of breast cancer as benign or malignant. The system's input is the image of size 512 by 512 pixels, which represents a very small fragment of the cytological specimen (see Fig. 3). A pathologist can select which fragment of the specimen should be examined. The chosen fragment of the microscopic image is processed, and then cell nuclei detected in the specimen are classified as malignant or benign.

The system works in two modes: training and prediction. In training mode, the system learns to detect and segment cell nuclei, afterwards it also learns to classify cell nuclei based on their morphology, color and texture. In prediction mode, the system takes the input image and segments it, detects nuclei, extracts features of detected cell nuclei and finally predicts the class (benign or malignant) of each nucleus.

The system has the form of an image processing pipeline which consists of six steps:

1. image preprocessing,
2. semantic segmentation,
3. cell nuclei detection and instance segmentation,
4. feature extraction,
5. feature selection,
6. classification.

Each step has an associated image or feature processing algorithm (see Table 1). In the first step, the input image is subjected to color separation (to extract cell nuclei, which are stained with hematoxylin) and image normalization. During the second step pixels are classified into three categories: nuclei interiors, nuclei edges and background. This step is carried out by the U-Net neural network (Ronneberger *et al.*, 2015). In the third step, the system detects centers of cell nuclei. This task is realized using another U-Net neural network. In the fourth step, the marker-controlled watershed algorithm is applied to instance segmentation of cell nuclei. Having precise shapes of cell nuclei, we extract features for each nucleus (morphometric, colorimetric and textural features). Finally, each nucleus is classified as benign or malignant based on its features. This is done using a set of classifiers: LDA, QDA, SVM, NB, RF, KNN, RPART (see Section 4.7).

Table 1. Scheme of the system.

Step	Step name	Objective	Algorithm(s)
1	Image preprocessing	Color separation (extracting cell nuclei which are stainedusing hematoxylin) and image normalization	Colour deconvolution
2	Semantic segmentation	Segment the image into three categories (regions): nuclei interiors, nuclei edges, background	U-Net neural network
3	Cell nuclei detection and instance segmentation	Detect the center of each cell nuclei and identify the precise shape of each detected nuclei	U-Net neural network + marker-controlled watershed (UNET-MCW), or alternatively ultimate eroded points + watershed (FIJI)
4	Feature extraction of cell nuclei	Extract features for each detected nucleus	A set of morphometric colorimetric and textual features are computed for every detected cell nucleus
5	Feature selection of cell nuclei	Reduction of the number of features	LASSO
6	Classification of cell nuclei	Classify each nucleus as benign or malignant	LDA, QDA, SVM,NB, RF, KNN, RPART(see Section 4.7)

4.2. Image preprocessing. Cytological preparations used in our experiment are stained with hematoxylin and eosin. Cell nuclei mainly react with hematoxylin staining (blue), whereas cytoplasm and red blood cells mainly react with eosin staining (red). Our aim is to separate cell nuclei from the other structures using the information about stain concentration. Unfortunately, absorption spectra of hematoxylin and eosin (H&E) overlap in RGB space.

To overcome this problem we can use the stain separation method proposed by Ruifrok and Johnston (2001). It uses the Beer–Lambert equation to associate the attenuation of the light transmitted through a cytological sample with the stain concentration (see Fig. 4). Based on this law, we can extract stain concentrations at each pixel from RGB intensities using the color deconvolution algorithm. The method requires providing a color deconvolution matrix tuned for a specific set of stains. It can be determined empirically by measuring the relative absorption for the red, green and blue channels on samples stained with a single stain. The color deconvolution matrix for H&E is generally well known, so in our experiments we used the matrix provided by the colour deconvolution plugin for FIJI software (Landini *et al*., 2004). Before further processing, the image representing hematoxylin concentration is normalized. First, a mean value of the image intensity is subtracted from each pixel value and then each pixel value is divided by the standard deviation of the image intensity.

4.3. Semantic segmentation. Semantic segmentation of an image is in fact a kind of classification of each pixel into a predefined category. In our case, pixels can be categorized as (a) the cell nuclei interior, (b) the cell nuclei edge or (c) the background. Each pixel must have associated the label of one of these categories. Unfortunately, this process is quite difficult because cell nuclei are heterogeneous and form complex structures that overlap and create clumps. Creating a segmentation program that would be based only on domain knowledge is practically impossible. It seems that such a program must absolutely use some form of machine learning techniques.

However, the problem is that these techniques require large amounts of data. In our case, we need to provide a lot of images with precisely marked shapes of nuclei. This is very time consuming and requires huge human effort. Nonetheless, it was shown by Ronneberger *et al*. (2015) that it is possible to train the neural network using a relatively small dataset (30 images) and not overfit to these data. This was possible due to using the U-Net architecture of the neural network and multiplicating training samples with a data augmentation algorithm.

This approach fits well to our scenario, where we have 50 training images and 50 validation images with manually marked nuclei shapes (see Fig. 5). Therefore, we developed a slightly modified U-Net structure that takes as the input a cytological image preprocessed by colour deconvolution and produces as the output a semantic map with labels that indicate pixel classes (i.e., nuclei interiors, nuclei borders and background).

The structure consists of a downsampling path (left) and upsampling path (right), see Fig. 6. Between the two, there are direct connections that allow propagation

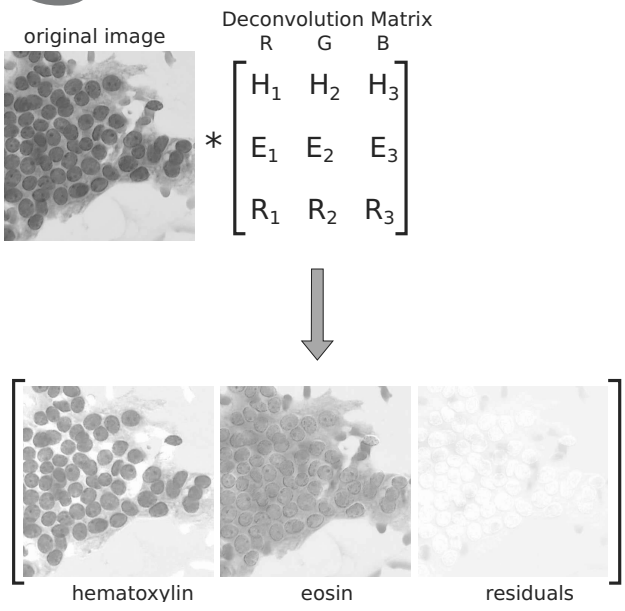

Fig. 4. H&E colour deconvolution.

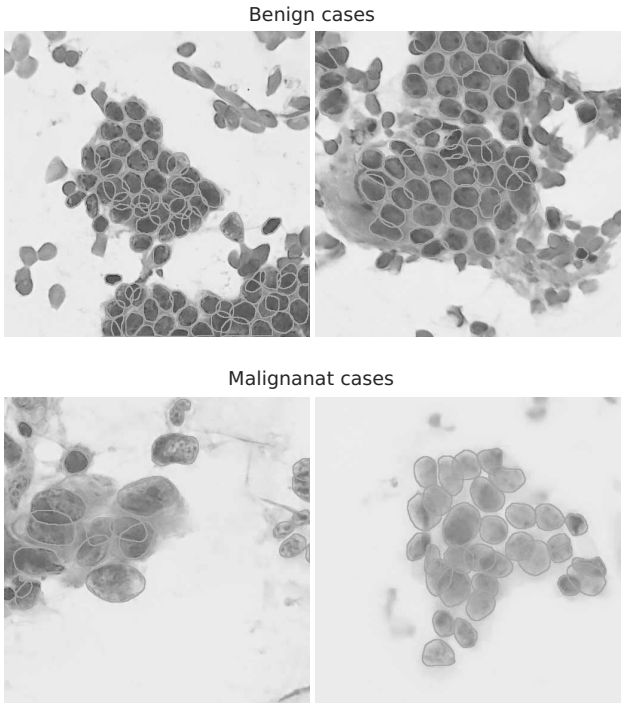

Fig. 5. Manual segmentation results.

of feature maps generated in the downsampling part to the corresponding layers in the upsampling part. The former path is assembled using repeated applications of the convolutional block, each followed by paths maxpooling operation with stride 2. At each downsampling step, the spatial size of the feature map is reduced by a factor of 0.5 and the number of feature channels is doubled. The convolutional block consists of a convolution layer (kernel size 3×3, padded with zeros) with ReLU (rectified linear unit) activations followed by the batch normalization and dropout layers, which in turn are followed by the second convolutional layer with ReLU activations (kernel size 3×3, padded with zeros) and the second batch normalization layer. The upsampling path is responsible for successive doubling of the spatial size of the feature map and halving of the number of feature channels. It consists of convolutional blocks followed by the deconvolutional operation (transposed convolution). At the top of the network, there is a convolutional layer with ReLU activations (kernel size 3×3, padded with zeros) followed by the softmax layer. They both generate the output of the network in the form of three feature maps (channels).

These maps have the same spatial size as the input image, and they define for each pixel a class probability distribution (Fig. 7). They must be transformed into a semantic map to find the regions of the image where there are nuclei interiors, nuclei edges, and background. Each pixel receives the label of the class for which it gained the highest probability. Compared with the original U-Net structure, ours is equipped with the batch normalization and dropout layers, it has three downsampling (maxpooling) and upsampling layers instead of four, the size of the output map is the same as the input image and the number of filters in convolutional layers is reduced by half.

Our training data set is relatively small. Therefore, we implemented an image augmentation procedure to enrich its diversity. Input images and corresponding output maps undergo synchronized random transformations (image scaling, random rotation, vertical and horizontal flipping). The procedure takes place online during the learning process. Each image that goes into learning has a 0.5 chance of being transformed. Thanks to this, the data fed into the neural network in subsequent epochs are slightly different each time. As a result, the overfitting problem can be eliminated more effectively.

4.4. Cell nuclei detection and instance segmentation. To extract the features of individual cell nuclei, it is necessary to (a) detect (localize) them on the picture and (b) determine their instance segmentation (shapes). The simplest way to do this is to take the binary map of nuclei interiors and to find and label all connected components. Such components can be seen as clusters of pixels with the same value, which are connected to each other through 8-pixel connectivity. Unfortunately, it has been observed that this approach tends to create objects consisting of groups of clumped nuclei (see Fig. 8). This distorts the features of cell nuclei and makes their classification less

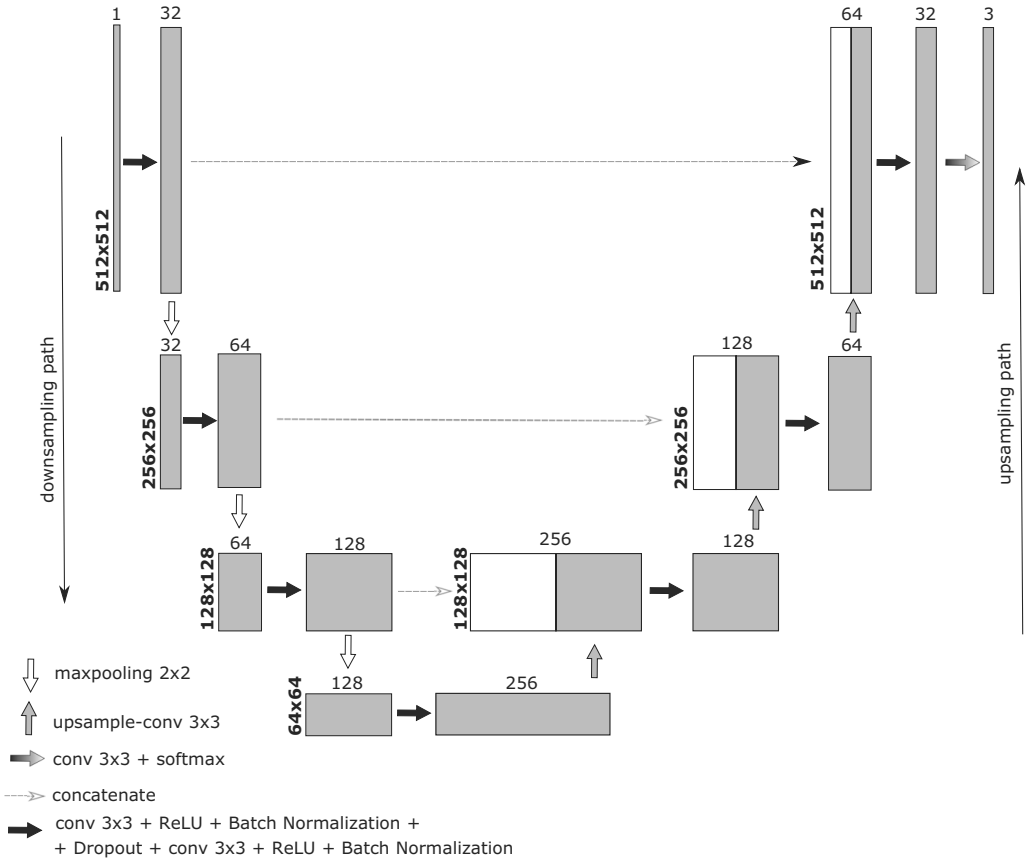

Fig. 6. Scheme of the U-Net neural network.

accurate.

To overcome this shortcoming, we developed an alternative method (UNET-MCW) of cell nuclei instance segmentation that is able to split touching and overlapping objects given a binary map of nuclei interiors.

The method is based on the marker-controlled watershed algorithm (Yang *et al.*, 2006; Koyuncu *et al.*, 2016; Skobel *et al.*, 2019). It is schematically presented in Fig. 9. First, the binary map of nuclei interiors is transformed into a virtual topographical surface by a distance transform. Next, we must detect nuclei centers because they will play the role of cell nuclei markers. Further, markers are imposed into the topographic surface using the pointwise maximum operation and morphological reconstruction. Finally, the watershed algorithm is carried out on the complement of the topographic surface. It floods subsequent basins (local minima) on the topographic surface and separates adjacent basins by barriers when different water sources meet.

However, before we can apply marker-controlled watershed to our data, we have to detect cell nuclei centers to provide markers for the watershed algorithm. We employed for this task the U-Net neural network. It has the same architecture as the network used for semantic segmentation (see Fig. 6). The only difference is in the last convolutional block of the network (placed at the top of the network), which is now composed of a convolutional layer (kernel 3×3, padded with zeros) with a sigmoid activation function. The input of the neural network is the image of size 512 px$\times$512 px preprocessed with color deconvolution and the image normalization procedure. As the output, we get a single feature map. It assigns each pixel a probability of being a center of the cell nuclei. The feature map is thresholded to get a binary map that indicates pixels belonging to cell nuclei centers. The neural network training process is based on 100 images for which cell nucleus centers have been manually marked. Images were divided into training (50 images) and validation images (50 images). Because the training data set is relatively small, training images were artificially augmented using randomized transformations: scaling by a factor from 0.8 to 1.2, random rotation, vertical and horizontal flip. The probability of occurrence of each augmentation was 0.5.

To verify the effectiveness of our approach for cell nuclei detection and instance segmentation, we compared it with an alternative method that is based on ultimate eroded points (UEP) with the watershed

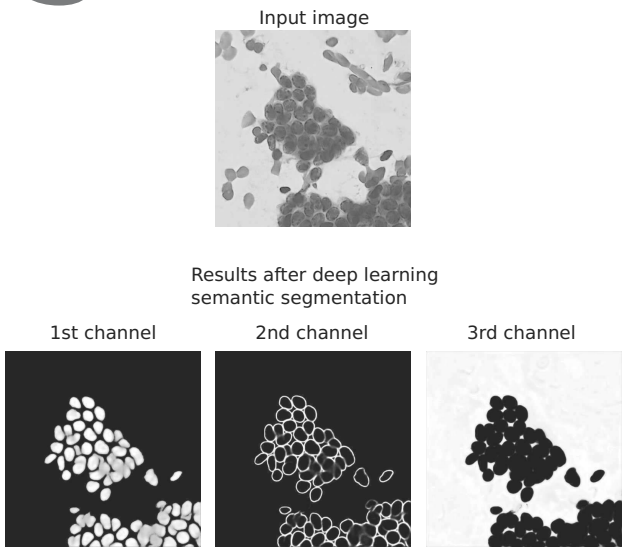

Fig. 7. Result of U-Net based semantic segmentation. Pixel classes: 1st channel—probability of the nuclei interiors class, 2nd channel—probability of the nuclei edges class, 3rd channel—probability of the background class.

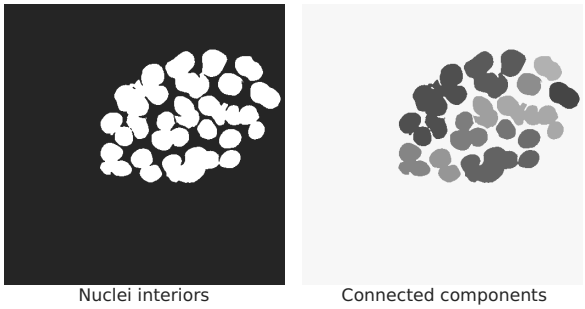

Fig. 8. Instance segmentation: groups of clumped nuclei.

algorithm (Vincent and Soille, 1991) (FIJI). Such an approach is available in the *Fiji* image processing software (Schindelin *et al*., 2012). We processed with this method the same images that were processed by the U-Net cell nuclei detector and marker-controller watershed (UNET-MCW). Sample results can be seen in Fig. 11. The results of a comprehensive comparison of both methods are presented in Sections 5.2 and 5.3.

4.5. Feature extraction. To perform nuclei classification one must first do nuclei segmentation, as was described in Section 4.3. In the next step, based on the segmented nuclei, classification of individual items is possible. The classification workflow of cytological images starts from the extraction of some characteristic features (such as the area, perimeter, etc.; see below). In our research we work with a set of 48 features (see Table 2) that can characterize every single cell nucleus. All the features can be further divided into three groups:

1. *morphometric* and location features,
2. *colorimetric* features based on (a) red, green and blue color channels, (b) grayscale image representation and (c) image after the H&E deconvolution,
3. features based on the analysis of cell nucleus *textures*. This group can be further divided into two subgroups of features: (a) based on the gray-level co-occurrence matrix (GLCM), (b) based on the gray-level run length matrix (GLRLM).

The GLCM stores four elements (Guan *et al*., 2020):

- contrast,
- correlation,
- energy,
- homogeneity.

They were calculated for images both in grayscale and after H&E deconvolution, which gives eight features in total. Based on the GLRLM, seven elements were considered:

- short-run emphasis,
- long-run emphasis,
- low gray-level run emphasis,
- high gray-level run emphasis,
- gray-level non-uniformity,
- run length non-uniformity,
- run percentage.

They were calculated also for two types of images: in grayscale and after H&E deconvolution, which gives 14 features in total. Hence, the total number of textural-based features gives 22 different items. For exact mathematical formulas to calculate the above mentioned features, see the work of Kowal and Filipczuk (2014). All the above mentioned features are also summarized in Table 2.

4.6. Feature selection. The above mentioned 48 variables (see Table 2) should not be used directly in this form. First, there are too many of them and, second, they are very strongly correlated with each other. In addition, they are not standardized, which can cause some numerical problems. Therefore, before they can be used to build classifiers, this set of 48 variables should be modified accordingly. To this end, the following preparatory assignments will be made.

Table 2. List of extracted nuclei features. The items printed in italics are highly correlated with some other features and they were excluded from further analysis

Group of features	Feature names	
Morphometric and location	(1) area, (2) perimeter, (3) major axis length, (4) minor axis length, (5) extent, (6) eccentricity, (7) equivalent diameter, (8) circularity, (9) aspect ratio, (10) distance to nearest nuclei centroid, (11) ellipse factor, (12) skeleton size, (13) number of pixel sharing with neighbour nuclei , (14) bending energy	
Colorimetric	(15) mean value in red channel, (16) mean value in green channel, (17) mean value in blue channel, (18) mean value on deconvolution image, (19) variance in red channel, (20) variance in green channel, (21) variance in blue channel, (22) variance on deconvolution image, (23) mean value on grayscale, (24) variance on grayscale, (25) entropy on grayscale, (26) entropy on deconvolution image	
Textural based on gray-level co-occurrence matrix	in grayscale	after deconvolution
	27) contrast, (28) correlation, (29) energy, (30) homogeneity	(31) contrast, (32) correlation, (33) energy, (34) homogeneity
Textural based on grey-level run length matrix	in grayscale	after deconvolution
	(35) short-run emphasis, (36) long-run emphasis, (37) gray-level non-uniformity, (38) run percentage, (39) run length non-uniformity, (40) low gray-level run emphasis, (41) high gray-level run emphasis	(42) short-run emphasis, (43) long-run emphasis, (44) gray-level non-uniformity, (45) run percentage, (46) run length non-uniformity, (47) low gray-level run emphasis, (48) high gray-level run emphasis

Step 1: The first step in preparing data for further analysis should always be data standardization. Its main goal is adjusting values measured on different scales to a common scale because non-standardized coefficients are not directly comparable. Typically the standard scores are used (also called z-scores) defined as

$$Y = \frac{X - \mu}{\sigma}. \tag{1}$$

Data standardization can in principle be considered mandatory in almost all types of analyses.

Step 2: In the next step, it is worth looking at the data in terms of the occurrence of outliers. They can greatly distort the final results, so it is good to exclude them from the data set. Outliers are easily detected using classical boxplots. Any data points which lie beyond the extremes of the whiskers are simply treated as outliers.

Step 3: Among the 48 designated features of cell nuclei, many are strongly correlated with each other, directly or indirectly. Therefore, it is worth choosing only the most important features. This procedure is commonly known as variable (or feature) selection. For this purpose, the LASSO algorithm was used, which is known for its very good quality (Tibshirani, 1996; Hastie *et al.*, 2009). It is a type of regularization method that penalizes with L1-norm. This step is the most time consuming because LASSO must be executed many times to find an optimal set of features. The procedure uses the k-fold cross-validation technique and its results are random, since the folds are selected at random. Users can reduce this randomness by running LASSO many times, and averaging the results.

Step 4: After feature selection it is worth looking at the correlation matrix to see if there are still some highly correlated features. If so, we arbitrarily decide to exclude from the model one of the two features where their correlation factor r is greater than 0.95.

Step 5: After completing the steps described above, we get a modified dataset with a reduced number of instances (result of removing outliers, Step 2) and reduced dimensionality (result of using LASSO, Step 3). Moreover, there are no strongly correlated features in the data (Step 4). Such modified datasets are ready to start building and testing classifiers.

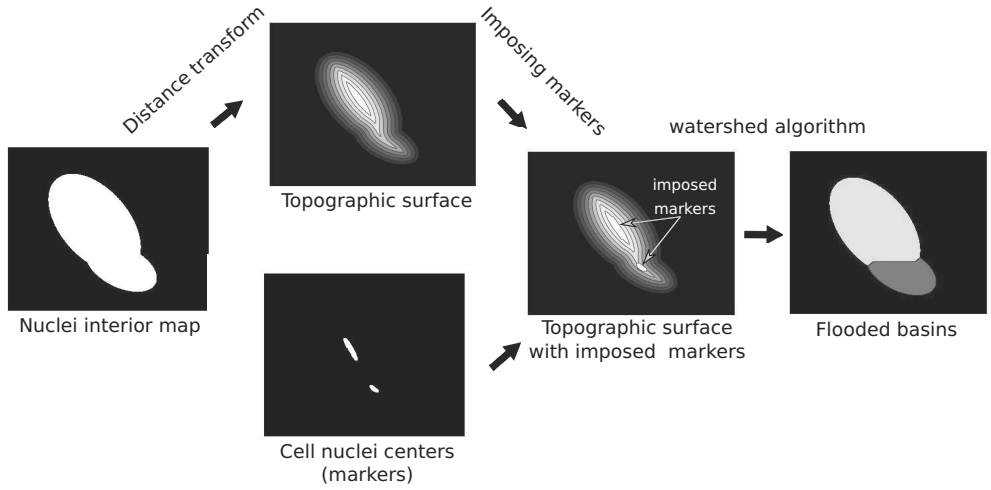

Fig. 9. Marker-controlled watershed.

4.7. Classification. Many different classification methods can be used to classify the analyzed objects. In this work 7 classical *parametric classifiers* are used. These include:

- LDA (*linear discriminant analysis*),
- QDA (*quadratic discriminant analysis*),
- SVM (*support vector machine*),
- RF (*random forests*),
- NB (*naive Bayes*),
- KNN (*k-nearest neighbors*),
- RPART (*recursive partitioning and regression trees*).

These methods are widely known and therefore we do not discuss them in detail here. For more information, the reader is referred to, e.g., Hastie *et al*. (2009) and James *et al*. (2013). Various authors also propose new methods or modify existing ones (see, e.g., Kantavat *et al*., 2018).

4.8. Methods of assessing the quality of classifiers. A natural step after building a classifier is to evaluate its performance. A large number of measures have been developed and, typically, the training dataset is used for this task. Four approaches are the most common (Hastie *et al*., 2009):

1. *Reclassification* method. After building a classifier using the training data set, the same data set is used for evaluating its performance. In a sense, these results can be considered less binding, because it can be regarded as controversial to use exactly the same full training dataset both for building and assessment of the resulted classified.
2. *Holdout* method. This is the most typical type of validation, in which the training data set is divided randomly into independent sets: the training one and the test one. Typically, the test set is smaller than 1/3 of the training set. Such a procedure is carried out repeatedly hundreds of times and at the end the average rate of correct classifications is calculated.
3. K*-fold cross validation* method (K-fold CV). The original dataset is randomly divided into K equal sized subsets. Out of these, a single subset is retained as the validation data for testing the model, and the remaining $K-1$ subsets are used as training data. The cross-validation process is then repeated K times (the folds), with each of the K subsets being used exactly once as the validation data. The K results from the folds can then be averaged to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once. A 10-fold cross-validation is commonly used but, in general, K remains a design parameter.
4. *Leave-one-out cross validation* method (LOOCV). This is a variation of the K-fold approach when the N-element data set is divided into N subsets, containing one element. The method involves using one observation as the test data set and the remaining $N-1$ observations as the training data set. This method is often used for small data sets.

All the above mentioned approaches were used to assess the quality of the constructed classifiers.

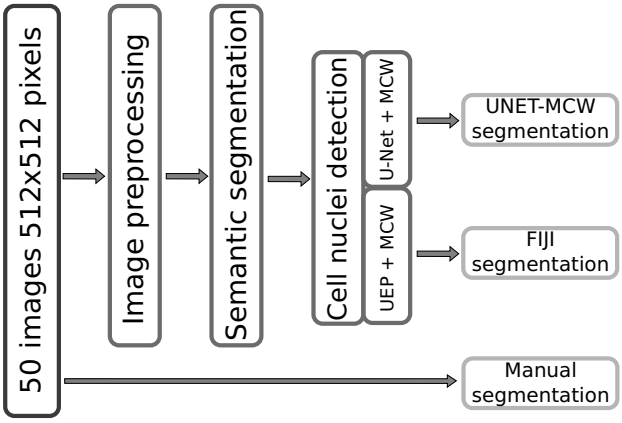

Fig. 10. Segmentation datasets: MCW—marker controlled watershed, UEP—ultimate eroded points.

5. Results

5.1. Cell nuclei segmentation. The contours of the cell nuclei are marked in each picture from test data. This process, called manual segmentation, is the ground truth for automatic segmentation methods (FIJI, UNET-MCW). In summary, the evaluation of automatic segmentation methods consists in comparing the location of cell nuclei detected using these methods to the location of nuclei detected during manual segmentation. A simplified scheme of performing data sets for individual segmentation methods is shown in Fig. 10.

In addition, both manual and automatic segmentation methods are involved in the classification process.

5.2. Evaluation of cell segmentation results. Demonstration of the segmentation accuracy of the proposed methods requires quantitative research. For this purpose a *manual* segmentation of cell nuclei was performed on the basis of a reference set of cytological images; see Fig. 5. The knowledge of the shape and position of the reference cell nuclei allows one to determine the number of true positive (TP) objects. Detection of TP nuclei requires a method for measuring their similarity. The most intuitive and one of the most commonly used method for testing the similarity of two sets is Jaccard's index:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}. \tag{2}$$

In this approach, we treat cell nuclei as binary sets of pixels. Mutual comparison of the shape and position between the reference and detected object determines the value of the Jaccard index. This index can take values in the range from 0 to 1, where 1 means full similarity of the examined objects and 0 means no similarity. The approach applied also requires setting a threshold for which the detected cell nucleus will be classified as TP. In our experiments, the minimum threshold value was set to 0.5. The set of objects detected by the automatic segmentation method, which do not have their equivalent in the set of manually detected objects, reach the false positive (FP) category. Cell nuclei that were not detected fall into the false negative (FN) category.

Table 3. Automatic segmentation results: 'B' stands for benign, 'M' stands for malignant.

		FIJI	UNET-MCW
B	Recall	0,85	**0.88**
	Precision	**0.93**	0.92
	F-score	0.88	**0.90**
M	Recall	**0.82**	0.81
	Precision	**0.92**	**0.92**
	F-score	**0.86**	**0.86**

Three classical coefficients were used to assess the quality of segmentation: *recall*, *precision* and the *F-score*. Recall is defined as a quotient of the number of correctly detected nuclei (TP) and the actual number of nuclei in the image:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

Precision is the ratio of TP to all automatically detected nuclei:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{4}$$

Finally, the F-score is calculated based on the following formula:

$$\text{F-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \tag{5}$$

All the above mentioned coefficients are well known and are commonly used in the evaluation of cytological image segmentation results as well as in many other areas. Moreover, the F-score is the harmonic mean of precision and recall, and therefore it is the most important indicator of segmentation quality in our experiments.

5.3. Cell nuclei segmentation results. We start the presentation of results obtained by showing a qualitative assessment. The results for sample images are presented in Fig. 11. In this figure, it can be seen that the FIJI segmentation method perfectly separates the area of cell nuclei from the background, but it has problems with the proper separation of individual objects. The application of marker-controlled watershed segmentation methods (FIJI, UNET-MCW) solves the problem of separation of individual nuclei.

The paper proposes two different approaches for detecting markers that are necessary to improve the watershed segmentation process. The overall

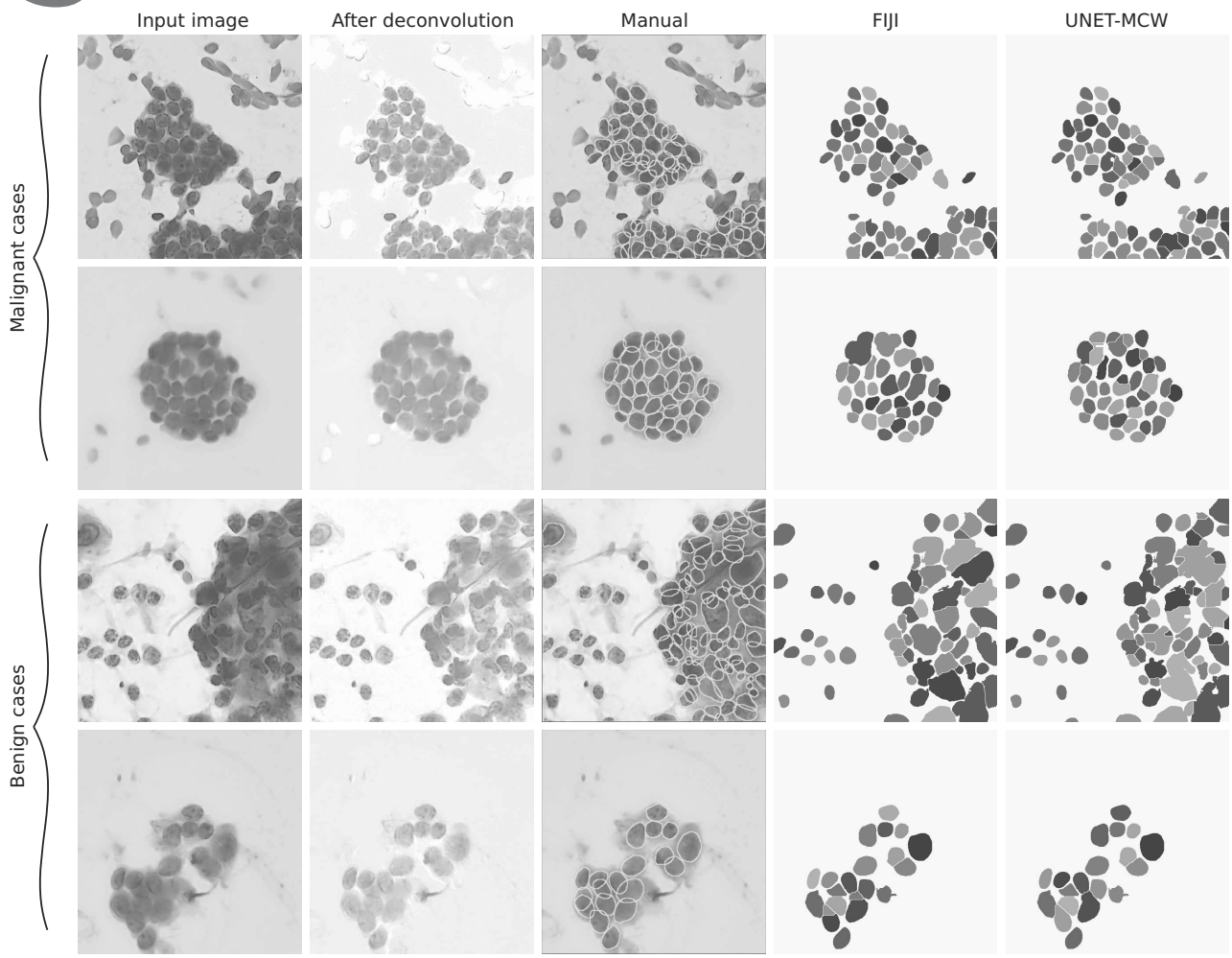

Fig. 11. Examples of semantic segmentation results.

quantitative results obtained from both approaches based on marker-controlled watershed segmentation are almost identical; see Table 3 for a comparative study.

5.4. Feature selection results. The first step of feature selection should be data standardization, as already mentioned in Section 4.6, Step 1. This is a very simple transformation of data, and we will not deal with it here.

In the next step (Section 4.6, Step 2) all obvious outliers should be removed. Figure 12 shows boxplots of sample data from the *manual* collection before and after removing outliers. Table 4 shows a summary of the number of observations in individual datasets before and after removing outliers. It can be seen that approximately half of the observations were removed. However, this is not a serious problem, as the observations that remain are sufficient to train classifiers.

Feature selections (Section 4.6, Step 3) were made using the LASSO algorithm (Tibshirani, 1996),(Hastie *et al*., 2009). The implementation available in the R system (R Core Team, 2019) was used here. We employed two functions from the glmnet package: *predict.cv.glmnet* and *cv.glmnet*. The functions can be parameterized and therefore we get slightly different final results. In order to make the results reproducible, in Table 5 we give the values of the parameters used. Note that the *cv.glmnet* function does k-fold cross-validation for the *glmnet* function and, as a consequence, the results are somewhat random. Therefore, this function is run multiple times (it was decided 100 times) and variables that have been selected at least 90 times are included in the final set of features.

In the last step (Section 4.6, Step 4), inspecting the correlation matrix, some additional features can be removed. When the correlation coefficient r between two variables is greater than 0.95, we remove one of these variables.

In Table 6 we show which features were finally selected for each of the three analyzed datasets (being the result of the FIJI and UNET-MCV segmentation methods

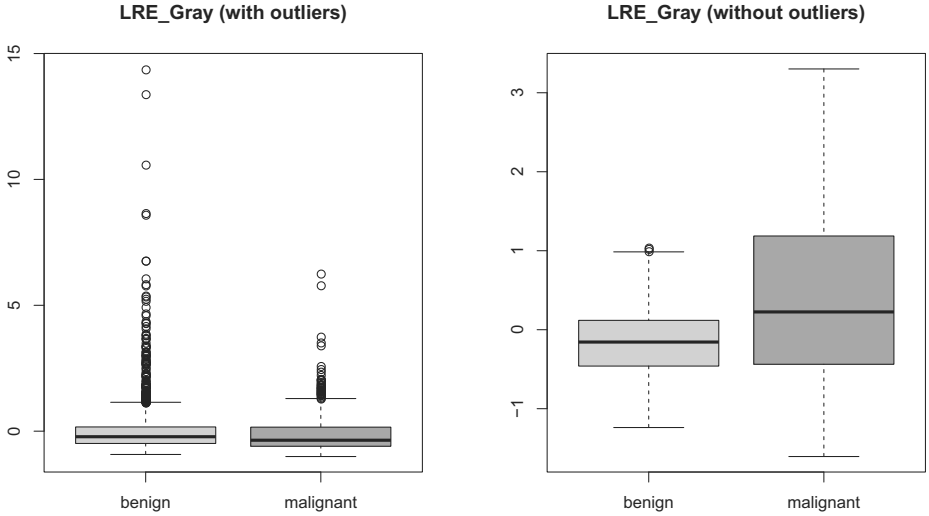

Fig. 12. Sample boxplots showing standardized data with the outliers present (left panel) and data after automatic outliers deletion (right panel). The picture was prepared for the *manual* dataset for the feature number 36 (see Table 2).

Table 4. Overall number of detected nuclei.

Dataset	Number of observations					
	Before outlier detection			After outlier detection		
	B	M	Total	B	M	Total
Manual	1698	750	2448	952	384	1336
FIJI	1501	629	2130	908	351	1259
UNET-MCW	1562	627	2189	940	368	1308

as well as manual cell nuclei selection).

5.5. Quality of the employed classifiers. Evaluation of the quality of constructed classifiers is not subject to discussion. Omitting this step may lead to a situation where we do not really know whether the obtained classifiers have any practical value. Methods suggested for this task were briefly discussed in Section 4.8.

It is also worth remembering that classifier quality assessment is made based on the data set from a single medical center. A consequence of this fact is that we cannot precisely determine how our classifier will work when classifying a completely different dataset coming from another source than the one used to build the classifier.

Nevertheless, the obtained results certainly tell us something about the quality of the obtained classifiers. The results (see Table 7) definitely prove that the quality of our classifiers is quite good. In most cases the SVM method gives the best results. Manual segmentation is treated as the reference one. It is difficult to point out a definite winner here. However, we can see that the classification results are in fact comparable, which somewhat proves that the methods of automatic cell classification we use are effective, and this is a very positive conclusion.

5.6. Classification results. One of the most natural ways to demonstrate the quality of any classifier is to show its results in the form of the so-called confusion matrix. Each row of the matrix represents instances in a predicted class while each column represents instances in an actual class (or vice versa). Tables 8–11 show the results for classification of the manual dataset (see Section 4.3 and Fig. 5) with the LDA, QDA, SVM, NB, RF, KNN and RPART classification technique. The best result was obtained for the SVM method and the worst one for the RPART method. This result is not surprising, because the SVM method is widely regarded as very robust and reliable one.

Finally, it is also proper to mention that the proposed procedure does not always work perfectly well. Analysing Tables 8–11 it is not difficult to notice that the classification of 'M' observations is clearly much worse than that of 'B' observations. Unfortunately, it is not entirely clear why this happens, and further research and experiments are required here. Some explanation for this behavior is that, for correct identification of most of the cell nuclei, the images must be sufficiently 'good and clear'. The two worst cases are presented in Figs. 13(a)

Table 5. Details on the R functions and packages used.

R function (package)	Notes
lda (MASS)	LDA classifier
qda (MASS)	QDA classifier
svm (e1071)	SVM classifier
randomForest (randomForest)	RF classifier
NaiveBayes (claR)	NB classifier
knn (class)	KNN classifier. The number of neighbours set according the popular rule-of-thumb: $N = \sqrt{n}/2$, n is the number of examples.
rpart (rpart)	RPART classifier. The complexity parameter cp was set according the 1SE rule.
cv.glmnet (glmnet)	LASSO feature selection. The following tuning parameters were set: type.measure = "class", alpha = 1, family = "binomial"
predict.cv.glmnet (glmnet)	LASSO feature selection. The following tuning parameters were set: s = "lambda.min", type = "nonzero"

and (c). The problem here is that the individual nuclei are too close to each other and the proposed algorithms, in their current version, cannot cope with correct separation of such objects. What connects these images is the blurring of the image and the lack of clear textures inside the nuclei. In turn, Figs. 13(b) and(d) are examples of very good cases, here the classification rate is very high. Example 13(c) is one of three cases where the SVM classifier scored accuracy below 60% and the only one where it did not score at least 50% accuracy.

5.7. Some notes on the software used. Color separation (Landini *et al*., 2004) and manual segmentation were carried out using the Fiji software (Schindelin *et al*., 2012). Semantic segmentation was carried out using the TensorFlow library for the Python language. The UEP based segmentation was carried out with the help of Fiji software (Schindelin *et al*., 2012). Matlab was used to implement the UNET-MCW algorithm.

All classifier calculations for the needs of this paper were done with the R software, Version 3.6.1 (R Core Team, 2019). Table 5 shows detailed information about the most important functions used in the calculation.

6. Discussion

During our experiments two strategies of cell nuclei segmentation were tested. Both approaches were based on semantic segmentation realized with the help of the U-Net network.

Unfortunately, the network was not able to extract all cell nuclei. Especially, it had a problem separating clumped cell nuclei. Therefore additional processing was necessary. We used two alternative approaches to tackle this issue. In the first approach, we trained another U-Net network to detect centers of cell nuclei, and then we applied the marker controlled watershed algorithm (markers were defined by detected cell nuclei centers) to separate clumped cell nuclei. The second (reference) approach was based on two algorithms implemented in the Fiji software. The UEP algorithm was used to detect cell nuclei centers, and then Fiji's implementation of the watershed algorithm was used to separate clumped cell nuclei. You can compare the obtained results in Fig. 11.

The evaluation of the results showed that the proposed approach achieved 90% segmentation accuracy for benign nuclei and 88% for malignant nuclei, according to the F-score. We point out here that various segmentation results can be found in the literature. For example, the accuracy of 85%–90% (depending on the test set used) was achieved by Veta *et al*. (2013). According

Table 6. List of features selected by LASSO for each of the three analyzed datasets. Feature numbers (not names) were used and they correspond to the numbers given in Table 2.

Data set	Features selected	No. of features
Manual	4 5 8 10 11 13 14 15 19 20 21 22 25 26 27 28 29 31 32 33 36 40 43 47	24
FIJI	5 6 8 9 10 12 13 14 15 16 17 19 20 22 25 26 27 28 32 33 34 40 47	23
UNET-MCW	1 5 8 10 11 13 19 22 28 32 36	11

Table 7. Accuracy of different classifiers for the manually segmented cell nuclei. Four classical tests were used, as described in Section 4.8.

No.	Method	Reclassification	Holdout	K-fold CV	LOOCV
Manual					
1	LDA	85.0	84.5	84.2	84.1
2	QDA	88.9	86.1	86.2	86.2
3	SVM	**92.4**	**87.6**	**88.1**	**87.9**
4	NB	80.7	79.3	79.8	80.2
5	RF	86.2	85.6	86.3	85.8
6	KNN	83.2	82.6	81.9	82.3
7	RPART	81.5	80.6	81.2	80.5
FIJI					
1	LDA	85.6	84.7	85.1	85.1
2	QDA	87.0	85.3	85.1	85.3
3	SVM	**90.9**	85.9	86.4	86.1
4	NB	81.9	81.2	81.6	81.5
5	RF	87.3	**86.6**	**86.5**	**86.7**
6	KNN	84.7	82.9	84.0	83.9
7	RPART	86.1	85.1	84.3	83.1
UNET-MCW					
1	LDA	86.0	84.7	85.1	85.6
2	QDA	87.2	86.1	85.9	85.6
3	SVM	**90.7**	**87.3**	**87.8**	**88.2**
4	NB	85.7	84.7	85.4	85.5
5	RF	87.5	86.5	87.4	87.1
6	KNN	85.5	83.7	84.7	85.0
7	RPART	87.7	85.8	86.3	86.9

to the F-score, Husham *et al*. (2016) reports the accuracy of 92%. The results obtained by us are more or less at the same level. It is also worth emphasizing here that it is difficult to compare the segmentation results if they are obtained from various test sets.

The results of the segmentation evaluation indicate that both methods based on the watershed algorithm (i.e. FIJI and UNET-MCW) are characterized by similar quality. Differences in the quality of segmentation obtained by individual methods can be seen only at the stage of classification.

Classification results show that automated methods are able to indicate benign cells nuclei with accuracy exceeding 90%. Unfortunately, malignant nuclei are classified with a much lower accuracy of 65–70%. It seems that the results obtained are worse than they might have been due to the poor obtained results for six out of 25 images (three of them are depicted in Fig. 13). So it seems that in order to improve the classification of malignant cases some additional discussions with

Table 8. Confusion matrices for classification of the segmented cell nuclei for the *reclassification* method: 'B' stands for benign, 'M' stands for malignant.

Manual					FIJI					UNET-MCW				
LDA	B	M	Total	Correct[%]	LDA	B	M	Total	Correct[%]	LDA	B	M	Total	Correct[%]
B	907	45	952	95.27	B	865	43	908	95.26	B	912	28	940	97.02
M	155	229	384	59.64	M	138	213	351	60.68	M	155	213	368	57.88
Total	1062	274	1336	85.03	Total	1003	256	1259	85.62	Total	1067	241	1308	86.01
QDA	B	M	Total	Correct[%]	QDA	B	M	Total	Correct[%]	QDA	B	M	Total	Correct[%]
B	897	55	952	94.22	B	834	74	908	91.85	B	885	55	940	94.15
M	93	291	384	75.78	M	90	261	351	74.36	M	112	256	368	69.57
Total	990	346	1336	88.92	Total	924	335	1259	86.97	Total	997	311	1308	87.23
SVM	B	M	Total	Correct[%]	SVM	B	M	Total	Correct[%]	SVM	B	M	Total	Correct[%]
B	937	15	952	98.42	B	881	27	908	97.03	B	919	21	940	97.77
M	86	298	384	77.60	M	87	264	351	75.21	M	100	268	368	72.83
Total	1023	313	1336	92.44	Total	968	291	1259	90.95	Total	1019	289	1308	90.75
NB	B	M	Total	Correct[%]	NB	B	M	Total	Correct[%]	NB	B	M	Total	Correct[%]
B	826	126	952	86.76	B	787	121	908	86.67	B	872	68	940	92.77
M	132	252	384	65.62	M	107	244	351	69.52	M	119	249	368	67.66
Total	958	378	1336	80.69	Total	894	365	1259	81.89	Total	991	317	1308	85.70
RF	B	M	Total	Correct[%]	RF	B	M	Total	Correct[%]	RF	B	M	Total	Correct[%]
B	925	27	952	97.16	B	875	33	908	96.37	B	909	31	940	96.70
M	158	226	384	58.85	M	127	224	351	63.82	M	133	235	368	63.86
Total	1083	253	1336	86.15	Total	1002	257	1259	87.29	Total	1042	266	1308	87.46
KNN	B	M	Total	Correct[%]	KNN	B	M	Total	Correct[%]	KNN	B	M	Total	Correct[%]
B	945	7	952	99.26	B	892	16	908	98.24	B	933	7	940	99.26
M	217	167	384	43.49	M	177	174	351	49.57	M	183	185	368	50.27
Total	1162	174	1336	83.23	Total	1069	190	1259	84.67	Total	1116	192	1308	85.47
RPART	B	M	Total	Correct[%]	RPART	B	M	Total	Correct[%]	RPART	B	M	Total	Correct[%]
B	912	40	952	95.80	B	868	40	908	95.59	B	920	20	940	97.87
M	207	177	384	46.09	M	135	216	351	61.54	M	141	227	368	61.68
Total	1119	217	1336	81.51	Total	1003	256	1259	86.10	Total	1061	247	1308	87.69

Table 9. Confusion matrices for classification of the segmented cell nuclei using the *holdout* method: 'B' stands for benign, 'M' stands for malignant.

Manual					FIJI					UNET-MCW				
LDA	B	M	Total	Correct[%]	LDA	B	M	Total	Correct[%]	LDA	B	M	Total	Correct[%]
B	138	9	147	93.88	B	128	8	136	94.12	B	136	5	141	96.45
M	22	32	54	59.26	M	21	32	53	60.38	M	25	31	56	55.36
Total	160	41	201	84.58	Total	149	40	189	84.66	Total	161	36	197	84.77
QDA	B	M	Total	Correct[%]	QDA	B	M	Total	Correct[%]	QDA	B	M	Total	Correct[%]
B	134	12	146	91.78	B	123	13	136	90.44	B	132	9	141	93.62
M	15	39	54	72.22	M	14	38	52	73.08	M	18	38	56	67.86
Total	149	51	200	86.50	Total	137	51	188	85.64	Total	150	47	197	86.29
SVM	B	M	Total	Correct[%]	SVM	B	M	Total	Correct[%]	SVM	B	M	Total	Correct[%]
B	139	8	147	94.56	B	129	7	136	94.85	B	136	5	141	96.45
M	17	37	54	68.52	M	20	33	53	62.26	M	20	36	56	64.29
Total	156	45	201	87.56	Total	149	40	189	85.71	Total	156	41	197	87.31
NB	B	M	Total	Correct[%]	NB	B	M	Total	Correct[%]	NB	B	M	Total	Correct[%]
B	124	22	146	84.93	B	117	19	136	86.03	B	131	10	141	92.91
M	19	35	54	64.81	M	16	36	52	69.23	M	20	36	56	64.29
Total	143	57	200	79.50	Total	133	55	188	81.38	Total	151	46	197	84.77
RF	B	M	Total	Correct[%]	RF	B	M	Total	Correct[%]	RF	B	M	Total	Correct[%]
B	141	6	147	95.92	B	131	6	137	95.62	B	137	5	142	96.48
M	23	31	54	57.41	M	20	33	53	62.26	M	22	34	56	60.71
Total	164	37	201	85.57	Total	151	39	190	86.32	Total	159	39	198	86.36
KNN	B	M	Total	Correct[%]	KNN	B	M	Total	Correct[%]	KNN	B	M	Total	Correct[%]
B	145	33	178	81.46	B	132	28	160	82.50	B	141	31	172	81.98
M	2	21	23	91.30	M	4	24	28	85.71	M	1	24	25	96.00
Total	147	54	201	82.59	Total	136	52	188	82.98	Total	142	55	197	83.76
RPART	B	M	Total	Correct[%]	RPART	B	M	Total	Correct[%]	RPART	B	M	Total	Correct[%]
B	139	8	147	94.56	B	131	5	136	96.32	B	139	2	141	98.58
M	31	24	55	43.64	M	23	30	53	56.60	M	26	30	56	53.57
Total	170	32	202	80.69	Total	154	35	189	85.19	Total	165	32	197	85.79

Table 10. Confusion matrices for classification of the segmented cell nuclei using *k-fold cross validation*: 'B' stands for benign, 'M' stands for malignant.

Manual					FIJI					UNET-MCW				
LDA	B	M	Total	Correct[%]	LDA	B	M	Total	Correct[%]	LDA	B	M	Total	Correct[%]
B	902	50	952	94.75	B	859	49	908	94.60	B	906	34	940	96.38
M	161	223	384	58.07	M	138	213	351	60.68	M	161	207	368	56.25
Total	1063	273	1336	84.21	Total	997	262	1259	85.15	Total	1067	241	1308	85.09
QDA	B	M	Total	Correct[%]	QDA	B	M	Total	Correct[%]	QDA	B	M	Total	Correct[%]
B	883	69	952	92.75	B	819	89	908	90.20	B	874	66	940	92.98
M	115	269	384	70.05	M	99	252	351	71.79	M	118	250	368	67.93
Total	998	338	1336	86.23	Total	918	341	1259	85.07	Total	992	316	1308	85.93
SVM	B	M	Total	Correct[%]	SVM	B	M	Total	Correct[%]	SVM	B	M	Total	Correct[%]
B	911	41	952	95.69	B	860	48	908	94.71	B	905	35	940	96.28
M	118	266	384	69.27	M	123	228	351	64.96	M	125	243	368	66.03
Total	1029	307	1336	88.10	Total	983	276	1259	86.42	Total	1030	278	1308	87.77
NB	B	M	Total	Correct[%]	NB	B	M	Total	Correct[%]	NB	B	M	Total	Correct[%]
B	818	134	952	85.92	B	786	122	908	86.56	B	871	69	940	92.66
M	136	248	384	64.58	M	110	241	351	68.66	M	122	246	368	66.85
Total	954	382	1336	79.79	Total	896	363	1259	81.57	Total	993	315	1308	85.40
RF	B	M	Total	Correct[%]	RF	B	M	Total	Correct[%]	RF	B	M	Total	Correct[%]
B	925	27	952	97.16	B	869	39	908	95.70	B	911	29	940	96.91
M	156	228	384	59.38	M	131	220	351	62.68	M	136	232	368	63.04
Total	1081	255	1336	86.30	Total	1000	259	1259	86.50	Total	1047	261	1308	87.39
KNN	B	M	Total	Correct[%]	KNN	B	M	Total	Correct[%]	KNN	B	M	Total	Correct[%]
B	943	9	952	99.05	B	888	20	908	97.80	B	933	7	940	99.26
M	233	151	384	39.32	M	181	170	351	48.43	M	193	175	368	47.55
Total	1176	160	1336	81.89	Total	1069	190	1259	84.03	Total	1126	182	1308	84.71
RPART	B	M	Total	Correct[%]	RPART	B	M	Total	Correct[%]	RPART	B	M	Total	Correct[%]
B	906	46	952	95.17	B	864	44	908	95.15	B	921	19	940	97.98
M	205	179	384	46.61	M	154	197	351	56.13	M	160	208	368	56.52
Total	1111	225	1336	81.21	Total	1018	241	1259	84.27	Total	1081	227	1308	86.31

Table 11. Confusion matrices for classification of the segmented cell nuclei using *leave-one-out cross validation*: 'B' stands for benign, 'M' stands for malignant.

Manual					FIJI					UNET-MCW				
LDA	B	M	Total	Correct[%]	LDA	B	M	Total	Correct[%]	LDA	B	M	Total	Correct[%]
B	901	51	952	94.64	B	860	48	908	94.71	B	907	33	940	96.49
M	161	223	384	58.07	M	140	211	351	60.11	M	156	212	368	57.61
Total	1062	274	1336	84.13	Total	1000	259	1259	85.07	Total	1063	245	1308	85.55
QDA	B	M	Total	Correct[%]	QDA	B	M	Total	Correct[%]	QDA	B	M	Total	Correct[%]
B	881	71	952	92.54	B	823	85	908	90.64	B	870	70	940	92.55
M	113	271	384	70.57	M	100	251	351	71.51	M	118	250	368	67.93
Total	994	342	1336	86.23	Total	923	336	1259	85.31	Total	988	320	1308	85.63
SVM	B	M	Total	Correct[%]	SVM	B	M	Total	Correct[%]	SVM	B	M	Total	Correct[%]
B	906	46	952	95.17	B	858	50	908	94.49	B	907	33	940	96.49
M	116	268	384	69.79	M	125	226	351	64.39	M	121	247	368	67.12
Total	1022	314	1336	87.87	Total	983	276	1259	86.10	Total	1028	280	1308	88.23
NB	B	M	Total	Correct[%]	NB	B	M	Total	Correct[%]	NB	B	M	Total	Correct[%]
B	822	130	952	86.34	B	785	123	908	86.45	B	870	70	940	92.55
M	134	250	384	65.10	M	110	241	351	68.66	M	120	248	368	67.39
Total	956	380	1336	80.24	Total	895	364	1259	81.49	Total	990	318	1308	85.47
RF	B	M	Total	Correct[%]	RF	B	M	Total	Correct[%]	RF	B	M	Total	Correct[%]
B	920	32	952	96.64	B	871	37	908	95.93	B	906	34	940	96.38
M	158	226	384	58.85	M	131	220	351	62.68	M	135	233	368	63.32
Total	1078	258	1336	85.78	Total	1002	257	1259	86.66	Total	1041	267	1308	87.08
KNN	B	M	Total	Correct[%]	KNN	B	M	Total	Correct[%]	KNN	B	M	Total	Correct[%]
B	938	14	952	98.53	B	885	23	908	97.47	B	929	11	940	98.83
M	222	162	384	42.19	M	180	171	351	48.72	M	185	183	368	49.73
Total	1160	176	1336	82.34	Total	1065	194	1259	83.88	Total	1114	194	1308	85.02
RPART	B	M	Total	Correct[%]	RPART	B	M	Total	Correct[%]	RPART	B	M	Total	Correct[%]
B	897	55	952	94.22	B	857	51	908	94.38	B	926	14	940	98.51
M	206	178	384	46.35	M	162	189	351	53.85	M	158	210	368	57.07
Total	1103	233	1336	80.46	Total	1019	240	1259	83.08	Total	1084	224	1308	86.85

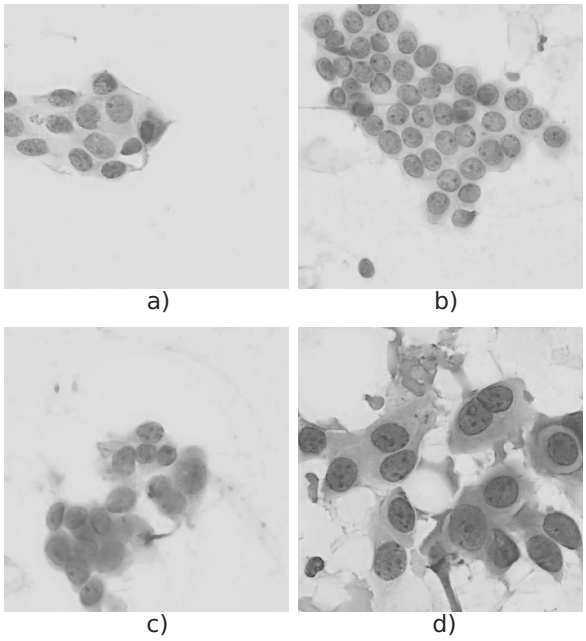

Fig. 13. Four example images with classification results for manual segmentation according to the SVM classification method: benign case with 83% accuracy (the worst) (a), benign case with 100% accuracy (b), malignant case with 44% accuracy (the worst) (c), malignant case with 100% accuracy (d).

physicians required to detect the characteristics of cell nuclei that have been classified as malignant. We also suspect that worse results for malignant cases may arise due to the fact that the collection of malignant nuclei is significantly smaller than that of benign nuclei. Moreover, malignant cell nuclei are much more diverse than benign nuclei. As a result, in a small population of malignant cell nuclei, it is difficult to find those that are similar to each other. The class imbalance is due to the fact that the ROIs that represent benign samples contain more cellular material than the ROIs of malignant samples. Unfortunately, this fact only became obvious to us after we performed manual segmentation of selected ROIs.

Considering all the above facts, we cannot solve our problem by simply reducing the benign data set (by resampling). We need more malignant cells in the training collection. Our present experimental results show that the classification accuracy for manually segmented cell nuclei and automatically segmented cell nuclei are practically the same. We are going to utilize this fact to prepare a semi-automatic tool that will help pathologists segment and label cell nuclei. This will significantly accelerate the process of collecting training data in future. Thanks to this, we will be able to collect a much larger and diverse base of cell nuclei for more patients.

7. Conclusions

The article presented the problem of automatic segmentation and classification of cytological images. Segmentation based on UNET-MCW, in which the markers are detected using convolutional neural networks, achieved the best results. In the case of images classified by the doctor as benign, their automatic classification is very effective. The classification of malignant nuclei is unfortunately worse. It seems that this is due to the inferior quality of images which are more blurred and devoid of important details. It can be assumed, therefore, that after obtaining better images, it will be possible to improve the results. The aim of the conducted experiments was to show whether or not the methods of automatic segmentation worsen the classification results compared to manual segmentation. Fortunately, the classification results for automatic segmentation were certainly no worse than the results of the classification for manual segmentation. This is a very positive conclusion.

An important open problem is automatic selection of good quality fragments from virtual slides. It is a necessary step to finally get good quality segmentation. It would be best if the cell nuclei were not too smudged, distorted or overlapped. For instance, the sample shown in Fig. 13(d) can be considered as very good, while the one shown in Fig. 13(c) it is difficult to carry out its automatic segmentation. However, virtual slides contain a large number of potential fragments. Therefore, a certain challenge is automatic selection of good quality samples that will allow good quality segmentation.

Another important problem is the right choice of features for the classification task. In our experiments, we used the LASSO algorithm, although of course there are other techniques. Further testing would be required to select the method most appropriate for our data.

Acknowledgment

The authors express sincere thanks and appreciation to Dr. Roman Monczak, University Hospital in Zielona Góra, Poland, for preparing test images, and to Dr. Michał Żejmo, University of Zielona Góra, Poland, for helping us to prepare results of semantic segmentation.

References

Alsubaie, N., Trahearn, N., Raza, S., Snead, D. and Rajpoot, N. (2017). Stain deconvolution using statistical analysis of multi-resolution stain colour representation, *PLOS ONE* **12**(1): e0169875.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018). Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: A Cancer Journal for Clinicians* **68**(6): 394–424.

Cui, Y., Zhang, G., Liu, Z., Xiong, Z. and Hu, J. (2018). A deep learning algorithm for one-step contour aware nuclei segmentation of histopathological images, *arXiv*: 1803.02786.

Dudzińska, D. and Piórkowski, A. (2020). Tissue differentiation based on classification of morphometric features of nuclei, *in* H. Florez and S. Misra (Eds), *Applied Informatics*, Springer, Cham, pp. 420–432.

Fondón, I., Sarmiento, A., García, A.I., Silvestre, M., Eloy, C., Polónia, A. and Aguiar, P. (2018). Automatic classification of tissue malignancy for breast carcinoma diagnosis, *Computers in Biology and Medicine* **96**: 41–51.

Guan, H., Zhang, Y., Cheng, H.-D. and Tang, X. (2020). Bounded-abstaining classification for breast tumors in imbalanced ultrasound images, *International Journal of Applied Mathematics and Computer Science* **30**(2): 325–336, DOI: 10.34768/amcs-2020-0025.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Second Edition*, Springer Series in Statistics, Springer, New York.

Hayakawa, T., Prasath, S., Kawanaka, H., Aronow, B. and Tsuruoka, S. (2019). Computational nuclei segmentation methods in digital pathology: A survey, *Archives of Computational Methods in Engineering* **28**: 1–13.

Höfener, H., Homeyer, A., Weiss, N., Molin, J., Lundström, C.F. and Hahn, H.K. (2018). Deep learning nuclei detection: A simple approach can deliver state-of-the-art results, *Computerized Medical Imaging and Graphics* **70**: 43–52.

Husham, A., Hazim Alkawaz, M., Saba, T., Rehman, A. and Saleh Alghamdi, J. (2016). Automated nuclei segmentation of malignant using level sets, *Microscopy Research and Technique* **79**(10): 993–997.

Irshad, H., Veillard, A., Roux, L. and Racoceanu, D. (2014). Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential, *IEEE Reviews in Biomedical Engineering* **7**: 97–114.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, Springer Series in Statistics, Springer, New York.

Jassem, J. and Krzakowski, M. (2018). Breast cancer, *Oncology in Clinical Practice* **14**(4): 171–215.

Kantavat, P., Kijsirikul, B., Songsiri, P., Fukui, K.-I. and Numao, M. (2018). Efficient decision trees for multi-class support vector machines using entropy and generalization error estimation, *International Journal of Applied Mathematics and Computer Science* **28**(4): 705–717, DOI: 10.2478/amcs-2018-0054.

Kowal, M. and Filipczuk, P. (2014). Nuclei segmentation for computer-aided diagnosis of breast cancer, *International Journal of Applied Mathematics and Computer Science* **24**(1): 19–31, DOI: 10.2478/amcs-2014-0002.

Kowal, M., Skobel, M. and Nowicki, N. (2018). The feature selection problem in computer-assisted cytology, *International Journal of Applied Mathematics and Computer Science* **28**(4): 759–770, DOI: 10.2478/amcs-2018-0058.

Koyuncu, C.F., Akhan, E., Ersahin, T., Cetin-Atalay, R. and Gunduz-Demir, C. (2016). Iterative h-minima-based marker-controlled watershed for cell nucleus segmentation, *Cytometry Part A* **89**(4): 338–349.

Landini, G., Rueden, C., Schindelin, J., Hiner, M. and Pavie, B. (2004). Image colour deconvolution, `https://imagej.net/Colour_Deconvolution`.

Litherland, J.C. (2002). Should fine needle aspiration cytology in breast assessment be abandoned?, *Clinical Radiology* **57**(2): 81–84.

Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C. and Thomas, N.E. (2009). A method for normalizing histology slides for quantitative analysis, *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), Boston, USA*, pp. 1107–1110.

Mittal, H. and Saraswat, M. (2019). An automatic nuclei segmentation method using intelligent gravitational search algorithm based superpixel clustering, *Swarm and Evolutionary Computation* **45**: 15–32.

Naylor, P., Laé, M., Reyal, F. and Walter, T. (2017). Nuclei segmentation in histopathology images using deep neural networks, *IEEE 14th International Symposium on Biomedical Imaging, Melbourne, Australia*, pp. 933–936.

Paramanandam, M., O'Byrne, M., Ghosh, B., Mammen, J.J., Manipadam, M.T., Thamburaj, R. and Pakrashi, V. (2016). Automated segmentation of nuclei in breast cancer histopathology images, *PLOS ONE* **11**(9): 1–15.

Piorkowski, A. and Gertych, A. (2019). Color normalization approach to adjust nuclei segmentation in images of hematoxylin and eosin stained tissue, *in* E. Piętka *et al.* (Eds), *Information Technology in Biomedicine*, Springer, Cham, pp. 393–406.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, `https://www.R-project.org`.

Rabinovich, A., Agarwal, S., Laris, C., Price, J. and Belongie, S. (2004). Unsupervised color decomposition of histologically stained tissue samples, *in* S. Thrun *et al.* (Eds), *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge, pp. 667–674.

Ronneberger, O., P.Fischer and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation, *in* N. Navab *et al.* (Eds), *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, Cham, pp. 234–241.

Ruifrok, A.C. and Johnston, D.A. (2001). Quantification of histochemical staining by color deconvolution, *Analytical & Quantitative Cytology & Histology* **23**(4): 291–299.

Sadanandan, S.K., Ranefall, P., Le Guyader, S. and Wahlby, C. (2017). Automated training of deep convolutional neural networks for cell segmentation, *Scientific Reports* **7**(7860): 1–7.

Santanu, R., Alok, J., Shyam, L. and Jyoti, K. (2018). A study about color normalization methods for histopathology images, *Micron* **114**: 42–61.

Schindelin, J., Arganda-Carreras, I. Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Albert A. (2012). FIJI: An open-source platform for biological-image analysis, *Nature Methods* **9**: 676–682.

Skobel, M., Kowal, M., Korbicz, J. and Obuchowicz, A. (2019). Cell nuclei segmentation using marker-controlled watershed and Bayesian object recognition, *in* E. Piętka *et al.* (Eds), *International Conference on Information Technologies in Biomedicine*, Springer International Publishing, Cham, pp. 407–418.

Spanhol, F.A., Oliveira, L.S., Petitjean, C. and Heutte, L. (2016). Breast cancer histopathological image classification using convolutional neural networks, *International Joint Conference on Neural Networks, Vancouver, Canada*, pp. 2560–2567.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1): 267–288.

Veta, M., van Diest, P.J., Kornegoor, R., Huisman, A., Viergever, M.A. and Pluim, J.P.W. (2013). Automatic nuclei segmentation in H&E stained breast cancer histopathology images, *PLOS ONE* **8**(7):e70221.

Vincent, L. and Soille, P. (1991). Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(6): 583–598.

Wang, D., Khosla, A., Gargeya, R., Irshad, H. and Beck, A.H. (2016). Deep learning for identifying metastatic breast cancer, *arXiv*: 1606.05718.

Xing, F. and Yang, L. (2016). Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review, *IEEE Reviews in Biomedical Engineering* **9**: 234–263.

Yang, X., Li, H. and Zhou, X. (2006). Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy, *IEEE Transactions on Circuits and Systems I: Regular Papers* **53**(11): 2405–2414.

Marek Kowal received his PhD degree in electrical engineering from the University of Zielona Góra, Poland, in 2004, and his DSc degree in computer science from the Częstochowa University of Technology in 2020. Currently, he is an assistant professor at the Institute of Control and Computation Engineering of the University of Zielona Góra. He has published about 60 journal and conference papers. His current interests include image analysis, pattern recognition, stochastic geometry and machine learning.

Marcin Skobel holds an MSc degree in geodesy, surveying and cartography (2013) from the AGH University of Science and Technology in Kraków and an MSc degree in computer science (2018) from the University of Zielona Góra. Currently, he is a PhD candidate at the Faculty of Computer, Electrical and Control Engineering of the University of Zielona Góra. His main research interests are focused on image processing and machine learning in computer vision.

Artur Gramacki received his PhD degree in electrical engineering from the Technical University of Zielona Góra, Poland, in 2000 and his DSc degree in computer science from the Częstochowa University of Technology in 2018. Currently, he is an associate professor in the Institute of Control and Computation Engineering, University of Zielona Góra. His research focuses on database systems, data mining, exploratory data analysis and statistical data analysis.

Józef Korbicz has been a full-rank professor of automatic control at the University of Zielona Góra, Poland, since 1994. In 2007 he was elected a corresponding member and in 2020 an ordinary member of the Polish Academy of Sciences (PAS). His current research interests include fault detection and isolation, control theory and computational intelligence. He has published more than 370 scientific publications, authored or coauthored 8 books and coedited 7 books. He served as the IPC chairman of the *IFAC SAFEPROCESS* Symposium held in Beijing, China (2006), and as the chairman of the NOC for *SAFEPROCESS* held in Warsaw, Poland (2018). Together with Prof. J.M. Kościelny he founded the Polish conferences on *Diagnostics of Processes and Systems, DPS*, organized every two years. He is currently the chair of the Committee on Automatic Control and Robotics of the PAS, a senior member of the IEEE and a member of the IFAC SAFEPROCESS TC. More at `http://www.uz.zgora.pl/~jkorbicz/`.

Received: 2 March 2020
Revised: 3 November 2020
Re-revised: 29 November 2020
Accepted: 7 December 2020