Computerized Medical Imaging and Graphics 72 (2019) 13–21

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Deep learning for cell image segmentation and ranking

Flávio H.D. Araújo[a,b,*], Romuere R.V. Silva[a,b], Daniela M. Ushizima[c,d], Mariana T. Rezende[e], Cláudia M. Carneiro[e], Andrea G. Campos Bianchi[e], Fátima N.S. Medeiros[b]

[a] Federal University of Piauí, Brazil
[b] Federal University of Ceará, Brazil
[c] University of California, Berkeley, USA
[d] Lawrence Berkeley National Laboratory, USA
[e] Federal University of Ouro Preto, Brazil

ARTICLE INFO

Article history:
Received 16 February 2018
Received in revised form 3 December 2018
Accepted 15 January 2019

Keywords:
Convolutional neural network
Cervical cells
Quantitative microscopy
Segmentation

ABSTRACT

Ninety years after its invention, the Pap test continues to be the most used method for the early identification of cervical precancerous lesions. In this test, the cytopathologists look for microscopic abnormalities in and around the cells, which is a time-consuming and prone to human error task. This paper introduces computational tools for cytological analysis that incorporate cell segmentation deep learning techniques. These techniques are capable of processing both free-lying and clumps of abnormal cells with a high overlapping rate from digitized images of conventional Pap smears. Our methodology employs a pre-processing step that discards images with a low probability of containing abnormal cells without prior segmentation and, therefore, performs faster when compared with the existing methods. Also, it ranks outputs based on the likelihood of the images to contain abnormal cells. We evaluate our methodology on an image database of conventional Pap smears from real scenarios, with 108 fields-of-view containing at least one abnormal cell and 86 containing only normal cells, corresponding to millions of cells. Our results show that the proposed approach achieves accurate results (MAP = 0.936), runs faster than existing methods, and it is robust to the presence of white blood cells, and other contaminants.

1. Introduction

Cervical cancer is the fourth most frequent type of cancer among women. There are approximately 570,000 new cases every year, from which 80% of the cases occur in under-developed countries (WHO, 2018). While methods for early detection of precancerous lesions using cytology (e.g., Pap smear) have been effective in developed countries, they are still imprecise in impoverished regions due to their precarious health care systems and to the lack of trained people to manually inspect the samples. The Pap smear test relies upon a microscope and a cytologist to analyze cervical cell samples deposited on a slide (Lu et al., 2015). In this test, the cytologist appraises the collected sample and then searches for abnormalities driven by cell features, such as shape and texture of the nucleus and cytoplasm. Consequently, the visual inspection of Pap smear test results is highly repetitive and prone to human error. However, previous attempts to automate conventional Pap smear cell segmentation and analysis have lacked the quality required for a reliable medical application.

Segmentation algorithms that handle multiples cells per image may achieve promising results on cytoplasm and nuclei of normal cells (Ushizima et al., 2014; Lu et al., 2016). However, due to the differences in shape, size, and chromatin among normal and abnormal cells, most automated algorithms perform poorly in nucleus segmentation tasks when both normal and abnormal cells coexist. Cell segmentation methods designed for liquid-based cytology perform reasonably well due to the absence of the artifacts and cell overlapping that make segmentation difficult. Although most of the exams in the United States are liquid-based which are still widely used in developing countries. Moreover, the expectation is that liquid-based Pap smears will not be made available in the public health systems of developing countries within the next 10 years.

Cytology in liquid basis costs at least five times more than conventional cytology, which makes its use practically unfeasible in developing countries. The main difference between the two methods is that in conventional cytology the sample is collected from the cervix (ectocervix and endocervix) and transferred directly to the

* Corresponding author at: Federal University of Piauí, Brazil.
E-mail addresses: flavio86@ufpi.edu.br (F.H.D. Araújo), romuere@ufpi.edu.br (R.R.V. Silva), dushizima@lbl.gov (D.M. Ushizima), trevisanrezende@gmail.com (M.T. Rezende), carneirocm@gmail.com (C.M. Carneiro), andrea@ufop.edu.br (A.G. Campos Bianchi), fsombra@ufc.br (F.N.S. Medeiros).

https://doi.org/10.1016/j.compmedimag.2019.01.003
0895-6111/

glass slide using Ayre spatula and endocervical brush. For cytology in liquid basis, the sample is collected with a brush and transferred to a vial containing a fixative liquid. In the laboratory, the sample is treated so as to reduce cell overlap and remove contaminants that make it more challenging to search for abnormal cells.

As part of the Pap test, the cytopathologist inspects several fields of a slide, scanning for cervical cells at the microscope and often using magnification at 40×. This process typically involves checking thousands of cells (Gonzalez et al., 2016), including a large quantity of slide fields per analysis, which here corresponds to a 1383 × 1036 image. If a field contains a clump of abnormal cells, then the analysis stops instead of going through the whole slide (Nayar and Wilbur, 2015).

This paper introduces a deep learning methodology for abnormal cell segmentation which comes from digitized conventional Pap smear and it ranks images according to the probability of that image field to contain abnormal cells. Our main objective with this ranking list of images is to improve the cytopathologist's performance, prompting abnormal patterns that may support cervical cell screening and diagnosis, potentially reducing the amount of image fields to be examined. We have also compared the performance of our method with the algorithm introduced by Bora et al. (2017) and with a supervised classification tool for image segmentation, namely Trainable Weka Segmentation (TWS) (Arganda-Carreras et al., 2017). The literature reports that TWS has achieved relevant image segmentation results on different imaging modalities (Bredfeldt et al., 2014; Staniewicz and Midgley, 2015; Vyas et al., 2016).

The main contributions of the proposed methodology are: (a) A fast pre-processing routine removes poor quality images, i.e., images that contain only background or regions of mucus of the harvested material, based on the intensity variation cut-off technique introduced by Yen et al. (1995). This technique removed about 25% of poor quality images in 0.07 seconds per image: (b) Segmentation of abnormal cervical cell images using a CNN trained with more than a million patch-images. (c) Inclusion of techniques to avoid overfitting in the training process, such as down-sample of the majority class. Any examples of the majority class were replaced, and the minority classes were flipped and blurred after each k epoch. (d) A post-processing step to remove the false positives and a parameter estimation step using a training dataset. (e) A low computational cost algorithm for cervical cell analysis with about 4.75 s per image. (f) The use of a new image database containing 194 samples representing different slide fields, summing up thousands of cervical cells that are manually curated by at least three pathologists. (g) We publicise the database, the trained CNN to segment abnormal cells, and all of the algorithms developed in this paper to improve reproducibility and to support the benchmarking of new applications.

2. Related works

The nuclei segmentation of cervical cells can be divided into two main groups: only one nucleus per image and multiple nuclei per image (Zhang et al., 2014). Some single nucleus per image detection methods use: parametric fitting (Wu et al., 1998); difference maximization (Tsai et al., 2008); clustering (Chankong et al., 2014); or shape, color, and contour information combined with active contour models (Li et al., 2012). However, although these methods consider that the input image contains only a single nucleus, multiple nuclei per image are quite common.

Several techniques to detect multiple nuclei per image have been developed based on thresholding (Harandi et al., 2010), Hough transform (Bergmeir et al., 2012), morphology (watershed) (Gençtav et al., 2012; Plissiti and Nikou, 2012) and level set (Lu et al., 2015). Song et al. (2015) proposed a multiscale convolutional network (MSCN) and a graph-partitioning-based method for the segmentation of cervical cytoplasm and nuclei. More specifically, deep learning via MSCN is explored to extract features and then segment regions centered at each pixel. The coarse segmentation is refined by an automated graph partitioning method that is based on the pretrained features. The texture, shape, and contextual information of the target objects are learned to locate the appearance of distinctive boundary, which is also explored to generate markers to split the touching nuclei. It is worth mentioning that this method aims to segment all of the nuclei in an image and it does not differentiate nuclei from normal or abnormal patterns. A detailed review of nuclei segmentation and classification can be found in Irshad et al. (2014).

Although these methods may achieve promising results on the cytoplasm and nuclei of normal cell segmentation, they perform poorly in nucleus segmentation tasks when both normal and abnormal cells coexist due to the differences in shape, size, and chromatin of these cells. These methods also perform poorly in images from conventional Pap smears due to the presence of artifacts and cell overlapping, which make the segmentation difficult. Thus, inspired by Litjens et al. (2017); Sharma et al. (2017); Wong et al. (2017), who applied CNN to other types of medical images, in this work we investigate the use of CNNs for cervical cell image segmentation.

It is worth mentioning that most classical methods which identify abnormal cells segment candidate regions containing dark background, noise, and both normal and abnormal cells. Hence, they perform a classification step to eliminate false positive regions by using shape and texture features for each region. In our methodology, we use more than a million patch images to train a CNN that segments only abnormal cells without extracting high computational features from each region. These aspects differentiate our work from others in the literature and they contribute to its low computational cost.

3. Proposed methodology

The goal of this work is to develop a methodology that is able of segment both free-lying and clumps of abnormal cells with high overlapping from digitized images of conventional Pap smears. Therefore, our approach must be robust to the presence of neutrophils, noise, and artifacts, which are all very common in images of conventional Pap test. To achieve this performance, we propose the methodology illustrated in Fig. 1.

The input images are filtered and any images containing only background or poor information are eliminated without prior segmentation. The remaining cervical cell images are then preprocessed to crop the sub-images for the training process. The segmentation method that is based on CNN detects abnormal cells and then a post-processing step is applied to improve the abnormal cell segmentation. Finally, we used the average area of segmented regions to rank the images. This ranking process sorts the images according to the probability that they contain abnormal cells. Images with more abnormal cells are ranked in the first positions. Therefore, our goal is able to aid the cytopathologist in the diagnosis of premalignant and malignant lesions of the cervix. The following subsections describe the steps of the proposed methodology. It is worth emphasizing at this point that we did not find a public database of cervical cells that contains normal and abnormal cells, artifacts, and cell overlapping that make segmentation difficult and which are very common in real images from conventional Pap smears. Consequently, we create a database[1] from real scenarios of a conventional Pap test.

[1] This database will be made available upon acceptance to support the benchmarking of new applications: https://sites.google.com/view/centercric.

3.1. Database

We used a Carl Zeiss microscope and a Zeiss AxioCam MRc camera with magnification of 40× to digitize 194 glass slides from Pap smears (108 with at least one abnormal cell and 86 with only normal cells). Each image has 0.255 mm/pixel, size 1392 × 1040 and it represents a field of a slide containing different number and type of cells. It is worth to mention that the abnormal images have cells with 5 different types of abnormalities (Carcinoma, HSIL, LSIL, ASCUS and ASCH). However, the ground truth of the database currently lacks differentiation of abnormality subtypes.

Fig. 2 displays the image samples from the Brazilian Health System (BHS). This database has several desirable characteristics, such as: BHS samples come from a broad racial diversity, which is common in the Brazilian population; the BHS holds ground truth images for abnormal cells that are manually segmented by one or more cytopathologists; and, its collection keeps a record of cervical cells from routine conventional Pap smears, including overlapping cells.

In this paper, we randomly created two different datasets using these images. The first one was named the training dataset and it contains 26 images (24 with abnormal cells and two with only normal cells). We used this dataset to train the CNN and to estimate the parameters that are used in the proposed methodology. The second dataset was named the test dataset and it contains 168 images (84 with abnormal cells and 84 with only normal cells). We used this dataset to test the proposed methodology and to evaluate its performance.

3.2. Data reduction to highlight essential images

The scanning process of a Pap test slide generates a great amount of cell images which mostly present only background and poor information and, therefore, a low probability of abnormal cell presence. Consequently, the proposed methodology applies a refining process to remove these images and, therefore, reduce the number of input images for the segmentation step. In this process, Yens's (1995) thresholding-based technique removes the image background. Then, if at least one region has an area greater than 5000 pixels, the related image is selected for the next step. We chose to use this number of pixels because an image with more than 5000 pixels has a high probability of having a clump of abnormal cells in the image. If more than five regions have an area greater than 2000 pixels, then this image will be input to the segmentation step. In fact, abnormal cells may be split into more regions after the thresholding step. Both values (i.e., 5000 and 2000) were calculated using the ground truth images of the abnormal cells from the training dataset. Our methodology adopts these fixed values because the images were acquired at the same magnification, i.e., 40× and have the same size. However, for images with different size and resolution, the thresholds must be recalculated proportionally to the size and resolution of the new image. The computational time of this refining process is about 0.07 s per image. The images that are eliminated in this step are not segmented. Fig. 3 illustrates some of the images that are removed in the refining process.

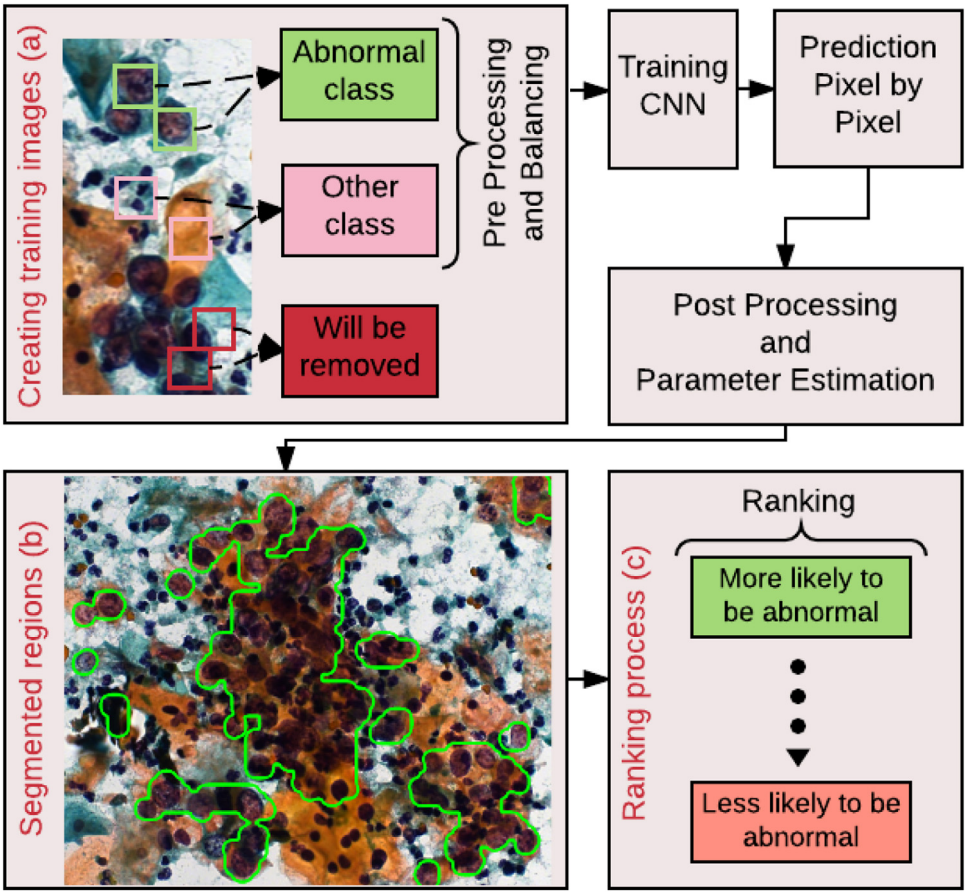

Fig. 1. Flowchart of the proposed methodology: (a) to create the training set of images, (b) to segment regions, and (c) to rank regions likely to be abnormal patterns.

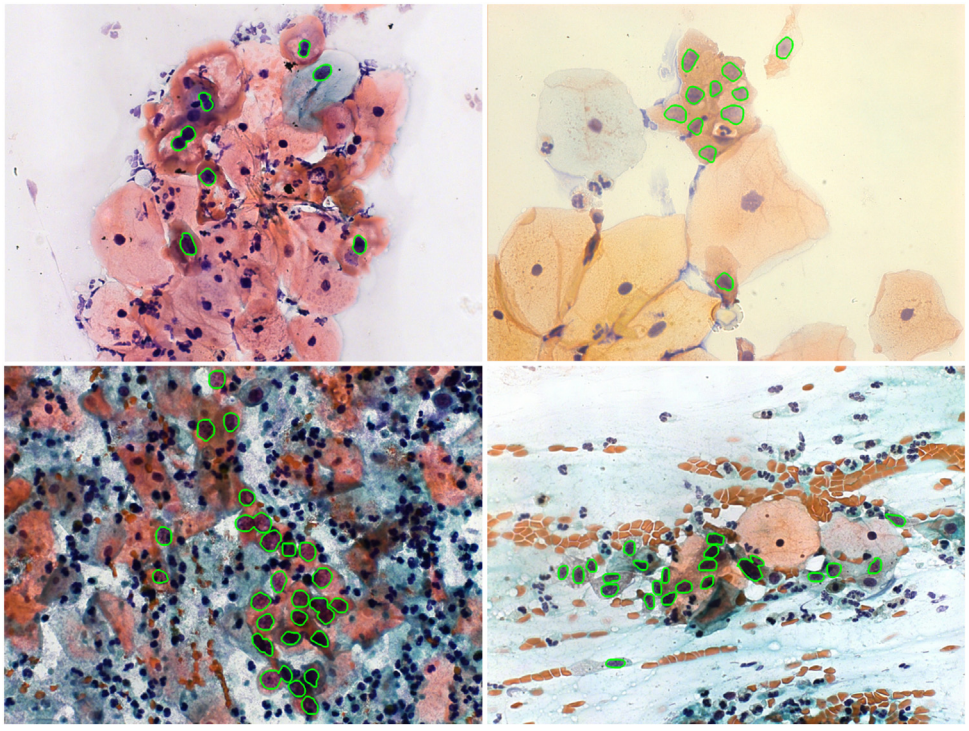

Fig. 2. Images from the BHS database which illustrate the diversity of the cervical cells. The green edges represent the ground truth of abnormal cells. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

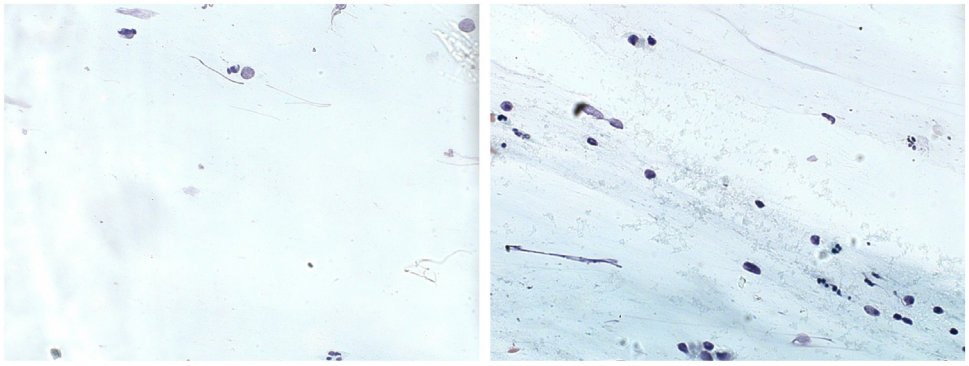

Fig. 3. Image samples selected in the refining process: (a) this sample presents only background and noise, and (b) this sample contains poor information.

3.3. Training a CNN for cytology

The Pap test slide images that are preserved in the refining process are then pre-processed and cropped to generate a set of images for the training process. This pre-processing step applies the median filter to reduce noise and artifacts. The algorithm then applies a sliding window to crop these pre-processed images. Here, the window size (52×52) corresponds to the average size of the regions which contain abnormal cells on the ground truth images of the training set. It is worth to mention that this value only changes if the magnification ($40\times$) of the microscope changes. These cropped images are then input to the training process and they may be classified into *abnormal class* or *other class*.

It is worth mentioning here that images from the conventional Pap test contain a lot of contaminants, which make the segmentation difficult. To improve the robustness of the method, we include images of neutrophil, artifact, and other contaminants in the samples of the *other class*. We then classify a cropped cell image as abnormal if it presents more than 50% of pixels belonging to an abnormal cell pattern. Otherwise, we remove it from the training set because this incomplete information may lead to convergence problems with the CNN. When an image does not include pixels that belong to abnormal cell patterns, it is classified as *other class*. This process is illustrated in Fig. 1(a).

The number of images in the *other class* (1,971,570) far exceeds the *abnormal cell* class (121,038). Therefore, we apply the random down-sample (Chawla, 2005) of the majority class to balance the training set. To avoid CNN overfitting, we augment the image collection by applying three operations after each five epochs: we flip the images horizontally, adjust image brightness and contrast by a random factor, and the linear scales of the image have zero mean and unit norm. In addition, we changed the examples of the *other class* after each five epochs. We trained the CNN for 50 epochs and the time of the training process was 6626 seconds.

Fig. 4 shows how the CNN error decreases during the training. The error decay in Fig. 4 demonstrates that our strategy avoids

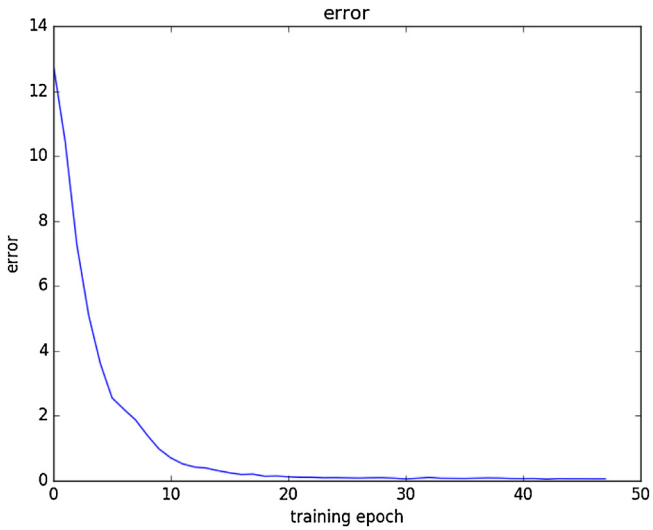

Fig. 4. The classification error during the CNN training for image segmentation.

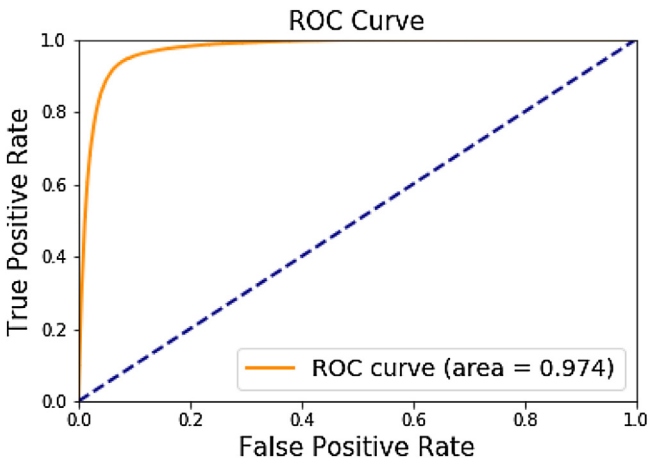

Fig. 5. ROC curve and AUC of the proposed methodology.

overfitting because the error does not increase and the error decay remains stable with changes in the training patterns.

We trained the CNN using raw pixel values (the RGB color space) from the training set images. Here, we employed the LeNet architecture (Lecun et al., 1998), which consists of two convolutional layers, with 64 filters of 5×5 and 3×3 sizes, with strides 1, no zero padding and Rectified Linear Units (ReLU) function. Two layers of max pooling immediately follow the two convolutional layers, with 2×2 size and stride 2. After the convolutional layers, there are two fully-connected layers with 382 and 192 neurons. The last layer is the softmax function, which calculates the output of the network. Notice that this scheme can be adapted to other microscope magnifications because the variation in the image size and resolution does not affect the CNN structure. Independently, the algorithm slides a window with fixed size (52×52) through the entire image, therefore, if the image size increases/decreases, the algorithm requires more/less iterations with the sliding window for image processing, leading to increased/decreased processing time.

3.4. Segmentation and ranking

The segmentation step consists of classifying the small cropped images from the new set of test images using a pixel-wise sliding window as input to the CNN. For each test image, the CNN model with a sliding window technique creates a probability map, where each pixel is assigned to a probability of being an abnormal cell or not. The total number of cropped images classified in the test phase was 8,477,960 normal and 57,390 abnormal crops. Fig. 5 shows the ROC Curve and Area Under ROC Curve (AUC = 0.974) obtained by the proposed methodology.

After this process, we perform a post-processing step to remove any small regions that are likely to be noise, neutrophil and small artifacts segmented as abnormal cells. In the post-processing step, we apply a morphological opening operator with a structuring element disk. We then fill the holes and remove all of those regions whose areas are smaller than a minimum threshold. If the structuring element radius and the minimum area threshold are smaller than a clump of neutrophils, then this region will not be removed. However, if the structuring element radius and the minimum area threshold are larger than regions of abnormal cells, then these regions will be removed. Consequently, we have performed a parameter estimation process to find the values of these two parameters (minimum area threshold and structuring element radius).

After the segmentation process, we calculate the average area of the segmented regions of each image. Then, the ranking list consists of images in the decreasing order of the average area, where those images with the highest values are tagged with a high probability to have abnormal cells.

4. Results

Our algorithm was implemented in Python and we used scikit-image (van der Walt et al., 2014) and tensorflow (Abadi et al., 2015). All of the experiments ran on a machine with six core Intel Xeon E5-2643 @ 3.40 GHz processors, four graphics processors (GeForce GTX Titan-X), and 251GB memory.

4.1. Metrics for performance evaluation

The assessment methodology applies the True Positive (TP), False Negative (FN) and False Positive (FP) rates to quantitatively evaluate the segmentation results. TP is the number of abnormal cells that are correctly segmented, FN corresponds to the number of abnormal cells that are not segmented, and FP is the number of regions that are segmented incorrectly as abnormal cells. To reckon the overall detection of the abnormal cells (TP), we considered that the algorithm correctly detected more than 60% of the pixels (Lu et al., 2015; Song et al., 2015). Based on these values, we calculated the F-Score (FS), Precision (P) and Recall (R) measures, which are given by

$$FS = 2\frac{P * R}{P + R}, \tag{1}$$

$$P = \frac{TP}{TP + FN}, \tag{2}$$

$$R = \frac{TP}{TP + FP}. \tag{3}$$

We have also used the Mean Average Precision (MAP) (Wang et al., 2015) to evaluate the ranking quality of the images. The MAP score is often used in Content Based Image Retrieval (CBIR) systems to evaluate retrieval results and it is computed by averaging the Average Precision score (AP) over all of the images in the ranking list. Considering each image Q in the ranking list, the $AP(Q)$ score is defined as

$$AP(Q) = \frac{\sum_{k=1}^{M}(P(k) * f(k)}{N}, \tag{4}$$

where $P(k)$ is the precision at cut-off k in the rank, and $f(k)$ is equal to 1 if the image at rank k is abnormal and is 0 otherwise. M is

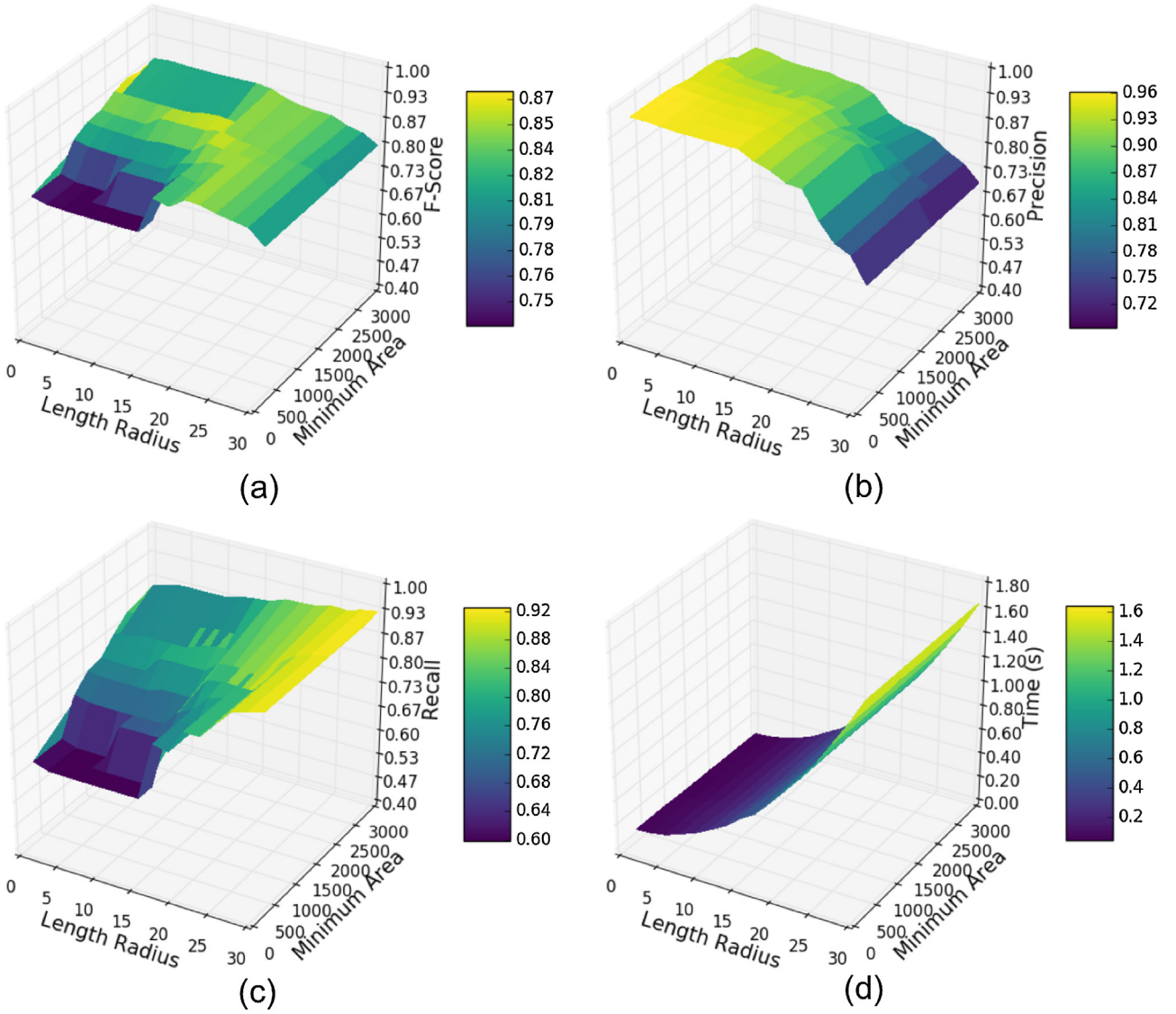

Fig. 6. Results obtained by varying the structuring element radius and minimum area threshold parameters. Values of (a) *F*-Score, (b) Precision, (c) Recall, and (d) Time per image in seconds.

the number of images in the rank and N is the number of abnormal images. The performance of the retrieval experiment is better when the MAP score is higher.

4.2. Parameter estimation

Some parameters (such as number of layers and size of filters) affect the learning efficiency and effectiveness of the proposed CNN. However, there is no scientific theory to select these parameters. Therefore, the current parameters of the CNN models are selected based on previous works (Dou et al., 2016; Yu et al., 2016). We selected the structuring element radius and minimum area threshold parameters in the post-processing step by applying a parameter estimation approach to the training set images. The tests were carried out with the structuring element radius in the interval [0, 30], steps of 2 and the minimum area threshold in the interval [0, 3000] with steps of 300. Fig. 6(a)–(c) display the *F*-Score, Precision, and Recall values, respectively, by varying these two parameters. Fig. 6(d) shows the impact on the processing time by applying the proposed approach.

Fig. 6(c) shows that when the structuring element radius is higher, the Recall value is larger. This is due to the high number of FN regions that have to be removed. In contrast, Fig. 6(b) demonstrates that the Precision measure decreases sharply for radius values greater than 20. This was due to the average radius of abnormal cells, which are between 20 and 25 pixels. Therefore, when the structuring element radius was greater than 20, regions of abnormal cells can be removed and, hence, this decreased the Precision value. Fig. 6(d) shows that the computational time increases remarkably with the increase of the structuring element radius.

The proposed methodology also removed regions with an area smaller than a threshold and, therefore, the increase of the threshold value led to an increase of the number of regions removed. The Precision value decreased for threshold values greater than 2000 pixels due to the average of abnormal cell regions, which was between 2000 and 5000 pixels. In contrast, the variation of the threshold value did not affect the computational cost. Based on these results, we set the structuring element radius and minimum area threshold values as 20 and 2000.

4.3. Segmentation results and quantitative evaluation

Several algorithms for cervical cell segmentation (Lu et al., 2015; Song et al., 2015; Bora et al., 2017) were designed for liquid-based

Table 1
Values of *F*-Score (FS), Precision (P), Recall (R), and mean time obtained by using the proposed methodology, the algorithm of Bora et al. (2017), and different classifiers of the TWS tool.

	TP	FN	FP	FS	P	R	Time(s)
TWS-NB	**938**	**11**	1249	0.66	**0.98**	0.43	193
TWS-RF	519	430	440	0.55	0.55	0.54	293
TWS-KNN	495	454	620	0.47	0.52	0.45	438
Bora et al. (2017)	860	89	2,584	0.40	0.91	0.25	9.96
Proposed	686	263	**375**	**0.69**	0.73	**0.65**	**4.75**

Bold values correspond to the best results.

cytology and they perform reasonably well for images without artifacts and cell overlapping that make segmentation difficult. Therefore, in addition to the method proposed by Bora et al. (2017), we also compared our approach with a machine learning tool for Microscopy Image Segmentation, which is named Trainable Weka Segmentation (TWS) (Arganda-Carreras et al., 2017). The TWS tool was used to segment different types of medical images (Ozcelikkale et al., 2017), including cell images (Kalinin et al., 2017). This tool performed a pixel by pixel classification using different classifiers and features grouped into the following types: edge detectors, texture filters, noise reduction filters, and membrane detectors.

Table 1 shows the *F*-Score, Precision, Recall, and mean time values obtained by using the proposed methodology, the algorithm introduced by Bora et al. (2017) and TWS with the following classifiers: Naive Bayes (NB) (John and Langley, 1995), Random Forest (RF) (Breiman, 2001) and K-Nearest Neighbor (KNN) (Aha and Kibler, 1991). The training set that was used to train these classifiers and the Bora et al. (2017) algorithm was the same that we used in our study. In addition, we used the random down-sample of the majority class to balance the number of samples of each class. TWS performed a pixel by pixel classification as well as our approach, and thus we used the same post-processing step of the proposed methodology to reduce the number of false positives.

A total of 39 images (about 23%) were removed in the refining process from the set of 168 test images and they contained only normal cells. Table 1 demonstrates that TWS with NB accomplished the highest Precision value, whereas Recall reached a lower value due to the high number of False Positives. The algorithm introduced by Bora et al. (2017) also resulted in a high Precision value, however showed also the lowest value of Recall because many clumps of neutrophils and dark regions were mis-classified as abnormal cells. The proposed approach achieved balanced values of Precision and Recall and, therefore, it attained the highest *F*-Score and Recall values. Our approach obtained a Precision value equal to 0.73. This means that 73% of the abnormal cells were segmented correctly. A great advantage of the proposed methodology is the low computational cost, which is about 4.75 seconds (0.07 for filtering, 3.95 for segmentation and 0.73 for post-processing) per image. Indeed, the low computational cost of our methodology is due to the CNN that was trained to segment only abnormal cells; therefore, it is not necessary to segment all of the cells of the image and then extract high computational features from each region.

Fig. 7 shows examples of success (green edges) and failure (red and blue edges) of the proposed methodology and the algorithms evaluated in Table 1. In Fig. 7(a), (c), (e) and (i) the regions in green edges were correctly segmented as abnormal cells while the regions in blue are abnormal regions that were not segmented. In the proposed methodology, these cells in blue were probably removed in the post-processing step due to its small size. In Fig. 7(a), (c), (e) and (i) the regions in red edges are clumps of neutrophil that were segmented incorrectly as abnormal cells. Fig. 7(b), (d), (f) and (j) show other examples of failures that are caused by the presence of dark regions formed by the overlapping of cells. Fig. 7(g) and (h) show that NB did not perform well and it segmented all of the dark regions of the images as abnormal regions. Fig. 7 confirmed that most *FP* regions were caused by the presence of clumps of neutrophil or were dark regions formed by overlapping cells.

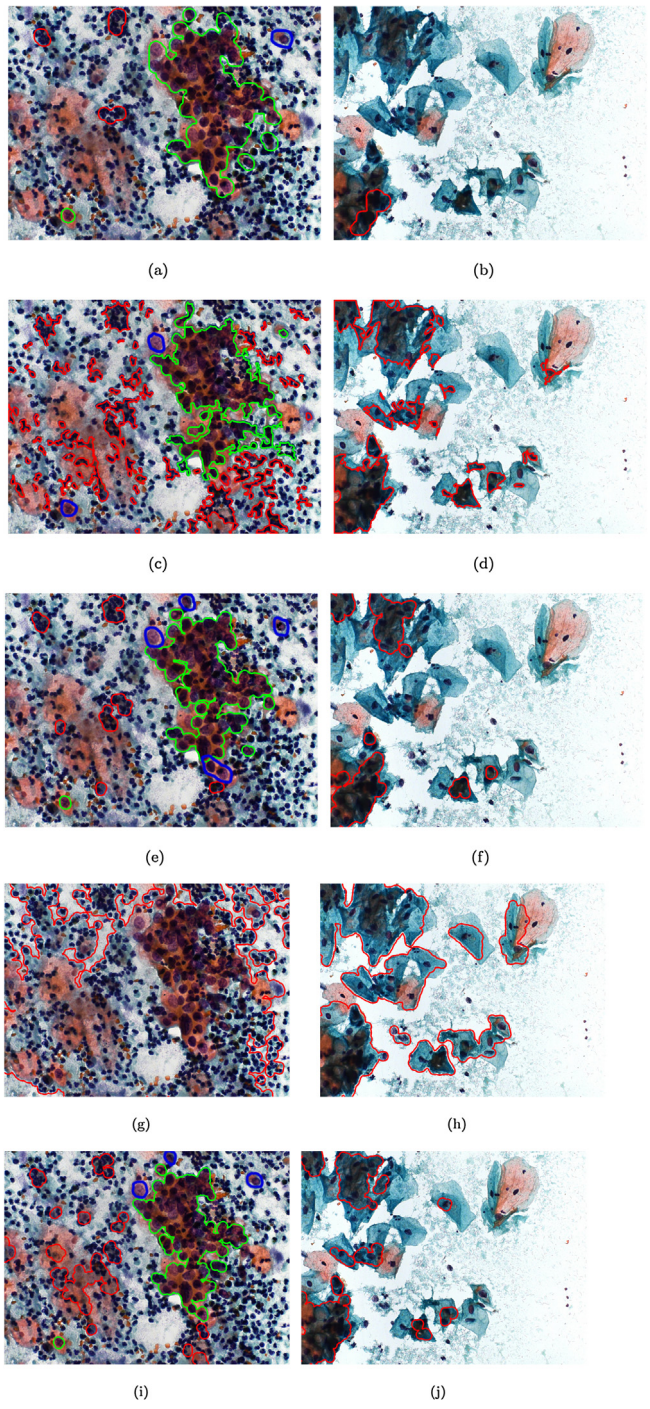

Fig. 7. Examples of success (green edges) and failure (red edges are FP regions and blue edges are FN regions) of the proposed methodology and the algorithms evaluated in Table 1. Segmentation obtained by the: (a) and (b) proposed methodology; (c) and (d) algorithm of Bora et al. (2017); (e) and (f) TWS with Random Forest; (g) and (h) TWS with Naive Bayes; (i) and (j) TWS with K-Nearest Neighbor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.4. Ranking evaluation results

We calculated the average area of the regions segmented by the proposed methodology and we obtained 20,718 (due to the clumps

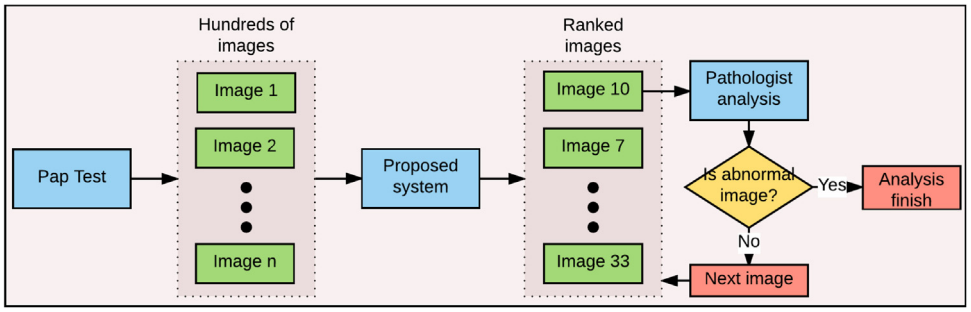

Fig. 8. The overall pipeline of the proposed methodology to assist cytopathologist in the Pap smear exam.

Table 2
MAP values for abnormal cell ranking: proposed methodology, the algorithm of Bora et al. (2017) and different classifiers of the TWS tool.

	TWS-NB	TWS-RF	TWS-KNN	Bora et al. (2017)	Proposed
MAP	0.695	0.821	0.71	0.746	**0.936**

of abnormal cells) for abnormal cells and 4984 for normal regions. These values demonstrated that the ranking process, performed in terms of average area of the segmented regions, was capable of sorting abnormal images accordingly. Table 2 shows the MAP values achieved by using the proposed methodology, the algorithm of Bora et al. (2017) and different classifiers of the TWS tool.

The proposed methodology achieved the MAP value equal to 0.936 by ranking the set of 168 test images (84 with abnormalities and 84 normal). An abnormal image has several abnormal cells and, therefore, our ranking approach succeeded, even when only 73% of the abnormal cells were correctly segmented. Although 98% of nuclei were correctly segmented, TWS using the NB reached the lowest MAP value due to the large number of FP regions that were mis-segmented.

5. Discussion

As part of the Pap test, the cytopathologist inspects several fields of a slide containing cervical cells. If a slide field contains two or more abnormal cells of the same type, then the analysis often ceases instead of going through the whole slide. Thus, Fig. 8 illustrates an alternative to reduce the slide analysis time by applying the proposed methodology. The slide containing the collected cells is automatically scanned and thousands of images with each field are generated. Nevertheless, most parts of these images consist of background or they present poor information (Fig 3). These images are discarded by the refining process without prior segmentation. The images that are preserved are then segmented and ranked by using the proposed methodology. These processed images are presented to the cytopathologist and sorted according to the probability to have abnormal cells. Although the cytopathologist provides the final diagnosis, the ranking list may potentially help to reduce the amount of image fields to be examined.

6. Conclusion

In this work, we introduced a methodology that segments abnormal cells from digitized images of conventional Pap smears by using the CNN. After segmenting abnormal cells, the average area of the segmented regions is used to rank the images according to the probability of that image field to contain abnormal cells. The proposed methodology segments both free-lying and clumps of abnormal cells with high overlapping. Moreover, it is robust to the presence of the neutrophils, noise, and artifacts that are very common in conventional pap smear images because our methodology applies a CNN and a post-processing step to eliminate these regions.

We carried out experiments on images from a cervical cell dataset that contains abnormal and normal cell patterns from real scenarios of conventional pap smear. The fast pre-processing routine was introduced to discard the poor quality images and it removed about 23% of these images in about 0.07 s per image. The segmentation failures of our methodology were caused by the presence of clumps of neutrophil or dark regions formed by overlapping cells. Overall, the proposed approach outperformed the other algorithms in terms of the MAP measure and time consumption because the experiments confirmed that it was more accurate (MAP = 0.936) and faster (with about 4.75 s per image). Most classical methods that identify abnormal cells present a high computational cost because there is a prior segmentation step to detect candidate regions containing dark background, noise, and both normal and abnormal cells. These methods usually perform a classification step to eliminate FP regions by using the shape and texture features for each region. In contrast, our methodology has a low computational cost because the CNN was trained to segment only abnormal cells and it did not extract high computational features from each region.

Future work includes extending tests to larger image collections, including classification among different abnormal stages in order to design and test more complex deep learning schemes, such as U-net, fully-connected neural networks, and mixed-scale dense CNNs.

Conflicts of interest

None.

Acknowledgments

This work was supported by Capes/CNPq-PVE (401442/2014-4), CNPq (306600/2016-1), PPSUS (APQ-03740-17), and the Moore-Sloan Foundation. Partial work on the development of machine learning algorithms was supported by the Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We are grateful to the cytologist Alessandra Tobias who helped to classify the cervical cells manually for this paper, to the Center for Recognition and Inspection of Cells (CRIC), and to the Berkeley Institute for Data Science (BIDS) data scientists and staff, specially Stefan van der Walt for incentivizing exploration of python packages in deploying open-source tools.

References

Abadi, M., et al., 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. http://tensorflow.org/.
Aha, D., Kibler, D., 1991. Instance-based learning algorithms. Mach. Learn. 6, 37–66.
Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K.W., Schindelin, J., Cardona, A., Seung, H.S., 2017. Trainable Weka Segmentation: A Machine Learning Tool for Microscopy Pixel Classification. Bioinformatics, Oxford, England, pp. 2424–2426, http://dx.doi.org/10.1093/bioinformatics/btx180.

Bergmeir, C., Silvente, M.G., Benítez, J.M., 2012. Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework. Comput. Methods Progr. Biomed. 107, 497–512.
Bora, K., Chowdhury, M., Mahanta, L.B., Kundu, M.K., Das, A.K., 2017. Automated classification of pap smear images to detect cervical dysplasia. Comput. Methods Progr. Biomed. 138, 31–47.
Bredfeldt, J.S., Liu, Y., Conklin, M.W., Keely, P.J., Mackie, T.R., Eliceiri, K., 2014. Automated quantification of aligned collagen for human breast carcinoma prognosis. J. Pathol. Inform. 5, http://dx.doi.org/10.4103/2153-3539.139707.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32, http://dx.doi.org/10.1023/A:1010933404324.
Chankong, T., Theera-Umpon, N., Auephanwiriyakul, S., 2014. Automatic cervical cell segmentation and classification in pap smears. Comput. Methods Progr. Biomed. 113, 539–556.
Chawla, N.V., 2005. Data mining for imbalanced datasets: an overview. In: Maimon, O., Rokach, L. (Eds.), The Data Mining and Knowledge Discovery Handbook. Springer, pp. 853–867.
Dou, Q., Chen, H., Yu, L., Qin, J., Heng, P.A., 2016. Multi-level contextual 3d cnns for false positive reduction in pulmonary nodule detection. IEEE Trans. Biomed. Eng., 1558–1567, http://dx.doi.org/10.1109/TBME.2016.2613502.
Gençtav, A., Aksoy, S., Önder, S., 2012. Unsupervised segmentation and classification of cervical cell images. Pattern Recognit. 45, 4151–4168 http://www.sciencedirect.com/science/article/pii/S0031320312002191.
Gonzalez, G.D., Silvente, M.G., Aguirre, E., 2016. A multiscale algorithm for nuclei extraction in pap smear images. Expert Syst. Appl. 64, 512–522.
Harandi, N., Sadri, S., Moghaddam, N., Amirfattahi, R., 2010. An automated method for segmentation of epithelial cervical cells in images of thinprep. J. Med. Syst. 34, 1043–1058, http://dx.doi.org/10.1007/s10916-009-9323-4.
Irshad, H., Veillard, A., Roux, L., Racoceanu, D., 2014. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential. IEEE Rev. Biomed. Eng. 7, 97–114, http://dx.doi.org/10.1109/RBME.2013.2295804.
John, G.H., Langley, P., 1995. Estimating continuous distributions in bayesian classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Mateo, pp. 338–345.
Kalinin, A.A., Allyn-Feuer, A., Ade, A., Fon, G.V., Meixner, W., Dilworth, D., de Wet, J.R., Higgins, G.A., Zheng, G., Creekmore, A., Wiley, J.W., Verdone, J.E., Veltri, R.W., Pienta, K.J., Coffey, D.S., Athey, B.D., Dinov, I.D., 2017. 3d Cell Nuclear Morphology: Microscopy Imaging Dataset and Voxel-Based Morphometry Classification Results., http://dx.doi.org/10.1101/208207, bioRxiv https://www.biorxiv.org/content/biorxiv/early/2018/04/22/208207.full.pdf.
Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 2278–2324.
Li, K., Lu, Z., Liu, W., Yin, J., 2012. Cytoplasm and nucleus segmentation in cervical smear images using radiating gvf snake. Pattern Recognit. 45, 1255–1264 http://www.sciencedirect.com/science/article/pii/S0031320311003979.
Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.
Lu, Z., Carneiro, G., Bradley, A., 2015. An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells. IEEE Trans. Image Process. 24, 1261–1272, http://dx.doi.org/10.1109/TIP.2015.2389619.
Lu, Z., Carneiro, G., Bradley, A., Ushizima, D., Nosrati, M.S., Bianchi, A., Carneiro, C., Hamarneh, G., 2016. Evaluation of three algorithms for the segmentation of overlapping cervical cells. IEEE J. Biomed. Health Inform. PP, 1, http://dx.doi.org/10.1109/JBHI.2016.2519686.
Nayar, R., Wilbur, D.C., 2015. Definitions, criteria, and explanatory notes. In: The Bethesda System for Reporting Cervical Cytology, Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-11074-5.
Ozcelikkale, A., Shin, K., Noe-Kim, V., Elzey, B.D., Dong, Z., Zhang, J.T., Kim, K., Kwon, I.C., Park, K., Han, B., 2017. Differential response to doxorubicin in breast cancer subtypes simulated by a microfluidic tumor model. J. Control. Release 266, 129–139.
Plissiti, M.E., Nikou, C., 2012. Overlapping cell nuclei segmentation using a spatially adaptive active physical model. IEEE Trans. Image Process. 21, 4568–4580.
Sharma, H., Zerbe, N., Klempert, I., Hellwich, O., Hufnagl, P., 2017. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. Comput. Med. Imaging Graphics, 2–13.
Song, Y., Zhang, L., Chen, S., Ni, D., Lei, B., Wang, T., 2015. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning. IEEE Trans. Biomed. Eng. 62, 2421–2433, http://dx.doi.org/10.1109/TBME.2015.2430895.
Staniewicz, L., Midgley, P.A., 2015. Machine learning as a tool for classifying electron tomographic reconstructions. Adv. Struct. Chem. Imaging 1, 9, http://dx.doi.org/10.1186/s40679-015-0010-x.
Tsai, M.H., Chan, Y.K., Lin, Z.Z., Yang-Mao, S.F., Huang, P.C., 2008. Nucleus and cytoplast contour detector of cervical smear image. Pattern Recognit. Lett. 29, 1441–1453.
Ushizima, D., Bianchi, A.G.C., Carneiro, C., 2014. Segmentation of subcellular compartiments combining superpixel representation with voronoi diagrams. Overlapping Cervical Cytology Image Segmentation Challenge – IEEE ISBI, 1–2.
Vyas, N., Sammons, R.L., Addison, O., Dehghani, H., Walmsley, A.D., 2016. A quantitative method to measure biofilm removal efficiency from complex biomaterial surfaces using sem and image analysis. Sci. Rep., 6, http://dx.doi.org/10.1038/srep32694 https://www.nature.com/articles/srep32694.
van der Walt, S., Schönberger, J., Nunez-Iglesias, J., Boulogne, F., Warner, J., Yager, N., Gouillart, E., Yu, T., 2014. scikit-image: image processing in Python. PeerJ 2, e453.
Wang, B., Brown, D., Gao, Y., Salle, J.L., 2015. March: multiscale-arch-height description for mobile retrieval of leaf images. Inf. Sci. 302, 132–148 http://www.sciencedirect.com/science/article/pii/S0020025514007282.
WHO, 2018. Comprehensive cervical cancer prevention and control: a healthier future for girls and women (accessed 11.15.18) http://www.who.int/reproductivehealth/topics/cancers/en/.
Wong, K.K., Wang, L., Wang, D., 2017. Recent developments in machine learning for medical imaging applications. Comput. Med. Imaging Graphics 57, 1–3.
Wu, H.S., Barba, J., Gil, J., 1998. A parametric fitting algorithm for segmentation of cell images. IEEE Trans. Biomed. Eng. 45, 400–407, http://dx.doi.org/10.1109/10.661165.
Yen, J.C., Chang, F.J., Chang, S., 1995. A new criterion for automatic multilevel thresholding. IEEE Trans. Image Process. 4, 370–378, http://dx.doi.org/10.1109/83.366472.
Yu, L., Guo, Y., Wang, Y., Yu, J., Chen, P., 2016. Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks. IEEE Trans. Biomed. Eng., 1886–1895, http://dx.doi.org/10.1109/TBME.2016.2628401.
Zhang, L., Kong, H., Chin, C.T., Liu, S., Chen, Z., Wang, T., Chen, S., 2014. Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts. Comput. Med. Imaging Graphics 38, 369–380 http://www.sciencedirect.com/science/article/pii/S0895611114000305.