

Research Article

Vaginal Secretions Epithelial Cells and Bacteria Recognition Based on Computer Vision

Shaozhi Guo ¹, Haoyuan Guan ², Jianzhong Li ^{1,3}, Yonghua Liao ⁴,
Weiwen Zhang ⁵ and Shengqiang Chen ⁶

¹School of Automation, Guangdong University of Technology, Guangzhou 511436, China

²Guangdong Institute of Scientific & Technical Information, Guangzhou, China

³Research Institute of Integrated Circuit Innovation, Guangdong University of Technology, Guangzhou 511436, China

⁴School of Basic Medical Sciences, Guangzhou Medical University, Guangzhou 511436, China

⁵Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

⁶Department of Applied Psychology, Guangzhou Medical University, Guangzhou 511436, China

Correspondence should be addressed to Weiwen Zhang; zhangweiwen@gzhmu.edu.cn and Shengqiang Chen; shengqiangchen@gzhmu.edu.cn

Received 15 September 2020; Revised 23 October 2020; Accepted 10 November 2020; Published 2 December 2020

Academic Editor: Zhan Li

Copyright © 2020 Shaozhi Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The examination of vaginal secretions epithelial is of great importance to female reproductive health. The results of manual detection will be affected by subjective judgment and fatigue. In this paper, an identification method based on computer vision is proposed to identify vaginal secretions epithelial cells and some bacteria. The experimental results demonstrate that the method can significantly distinguish the vaginal secretions epithelial cells from the bacteria and virus.

1. Introduction

Vaginitis is one of the most common diseases in women. It brings burning sensation and pain to the affected areas, and, even worse, it will cause symptoms of general discomfort and reduce the immune function of people, resulting in other complications [1]. The examination of the vaginal secretions is necessary to help to discover vaginal lesions early and to implement timely and effective treatment to patients. The vaginal secretions are mainly examined using Gram staining. After the steps of making smear, drying, fixing, and staining, with counting the number of cells and bacteria in the smear, the state of vaginal microecological environment can be preliminarily judged [2, 3]. Moreover, the health situation of reproductive system can be predicted.

The results of Gram staining microscopy of vaginal secretions provide important information for the diagnosis of vaginal health [4]. However, counting results relies

heavily on the professionalism, as well as the state of the operators. The results provided by different operators would be quite different. And even for the same operator, the results may not be the same at different time. Therefore, an automatic system should be required to detect efficiently cells and bacteria in vaginal secretions Gram stained smear images.

The automatic recognition of cells or bacteria has been attracting the attention of scholars from all over the world. Yi-De Ma et al. used the logical and morphological feature of blood cell images to count the number of cells [5]. Jacey-Lynn Minoi et al. developed an image processing technique on smart mobile App, which can detect and isolate colonies on real environment by shape, color, and other morphological features [6]. However, with the above methods, it may be difficult to achieve good recognition results for the objects with fuzzy edge and low contrast with the background, especially those with irregular shape and area. Aslı Genctav et al. proposed an

unsupervised method based on a multiscale hierarchical segmentation algorithm and binary classifier for the separation of nuclei in cervical cell image [7]. Kuan Li et al. proposed a method named Radiating Gradient Vector Flow (RGVF) Snake that can more accurately extract the contours of nucleus and cytoplasm in cervical smear images [8]. However, both works are carried out on images with relative simple content and could not provide a good effect in smear images of vaginal secretions with complex content including impurities. Machine learning algorithms have been also introduced to detect cell in recent years [9–11]. However, these algorithms need an experimental training step, which is difficult to achieve due to the lack of label data of cell images. Moreover, indeterminacy of manual annotation of class label can greatly affect the accuracy of machine learning algorithms [12].

The content of vaginal secretions smear images includes cytoplasts, nuclei, cells (completely composed by cytoplasm and nucleus), bacteria, and impurities. In this paper, we propose a dual process and nucleus-cytoplasm cross verification method to recognize vaginal secretions epithelial cells in vaginal secretions smear images. The number of the cells can be counted, which works as the feedback. In addition, the method can generate pixel-level label data of vaginal secretions epithelial cells and bacteria rapidly and efficiently, which can be used in model training process of machine learning or deep learning, reducing the manual burden of labeling medical images.

The rest of this paper is organized as follows. Section 2 analyses the smear images. The realization of recognition method is shown in Section 3. Then, Section 4 gives the experimental results. Section 5 presents the discussion and conclusion.

2. Image Analysis

The vaginal secretions smear images used in this paper were provided by the Third Affiliated Hospital of Guangzhou Medical University. The vaginal secretions were observed with gram stain in white back-light illumination. Figure 1 shows several kinds of objects in the smear images, which can be divided into targets and jamming information. According to the recognition requirements, the targets include nuclei, cells, and bacteria, while the jamming information includes cytoplasts, impurities, and background.

The features of the objects in vaginal secretions smear images are explained in the following:

- (1) Nucleus. Nucleus may be inside a cell as a part of it, or outside. Its color is dark purple after Gram stain, and its shape is ideally round or oval but sometime strange shape. Generally, the area is between 40 and 250 pixels.
- (2) Bacterium. A bacterium is much smaller than nucleus with slender shape, in dark purple, randomly distributed in the image and may appear in the regions of background or a cell.
- (3) Cell. A cell consists of nucleus and cytoplasm with an area greater than 500 pixels, of which the color is red

distinct from nucleus. Due to the uneven distribution of intracellular substances, there is Gaussian noise in the region of cytoplasm [13].

- (4) Cytoplast. The only difference between a cell and a cytoplast is that there is no nucleus in the cytoplast.
- (5) Impurity. In addition to all kinds of physiological objects, there are impurities in the cell smear, such as bubbles, black spots, and staining agent residues, which are close to the cell or nucleus in morphology.
- (6) Background. The image light source is white back-light. Thus, the background is a uniform and high brightness area in white and is brighter than other areas.

3. Recognition Method

The recognition targets are cells and bacteria. However, in order to recognize effective cells, it is necessary to recognize the cytoplast as well as the nucleus. Some bacteria and nuclei may overlap with the cytoplasmic, and the image information of bacteria and nuclei will be covered up when the cytoplasmic image information is enhanced for extracting the cytoplasmic regions. Therefore, in order to obtain the information of bacteria and cells more effectively and simultaneously, a dual process structure is adopted. The regions of cells and bacteria are separately extracted in one process and the regions of cytoplasm are extracted in the other one [14, 15]. After that, the regions of nuclei and cytoplasm are verified mutually to distinguish cells from cytoplasts and impurities. Figure 2 outlines the overall flow.

The original image is a color image using RGB color model to store data. There are three decimal values from 0 to 1 representing the brightness of red, green, and blue in each pixel. The recognition results are reflected by generated binary images marking the regions of interest (ROI) with true value and other regions with false value.

3.1. Process for Cells and Bacteria. The steps to detect nuclei and bacteria can be found as follows.

Step 1. Separating the red component from a color image as gray-scale image for weakening cytoplasm information: the cytoplasm is stained red, which hardly blocks the light in the red. As a result, the cytoplasm information in the red component of original image is less than that in the blue and green components. In conclusion, using red component for graying can effectively weaken the information of cytoplasm, highlight the information of nuclei and bacteria, and reduce the amount of data to 1/3.

Step 2. Segment background by threshold segmentation: the gray values of background and cytoplasm is higher than that of bacteria and nuclei. Therefore, threshold segmentation is used to segment ROI:

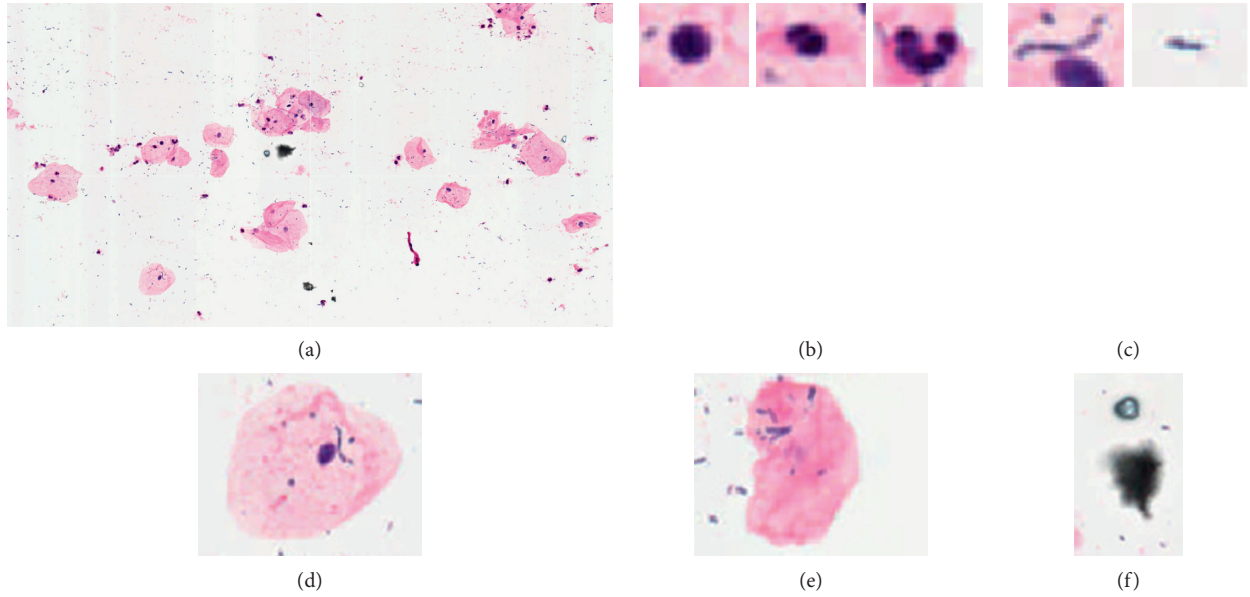


FIGURE 1: A part of a smear image and the content of it: (a) original image, (b) nuclei, (c) bacteria, (d) cell, (e) cytoplasm, and (f) impurities.

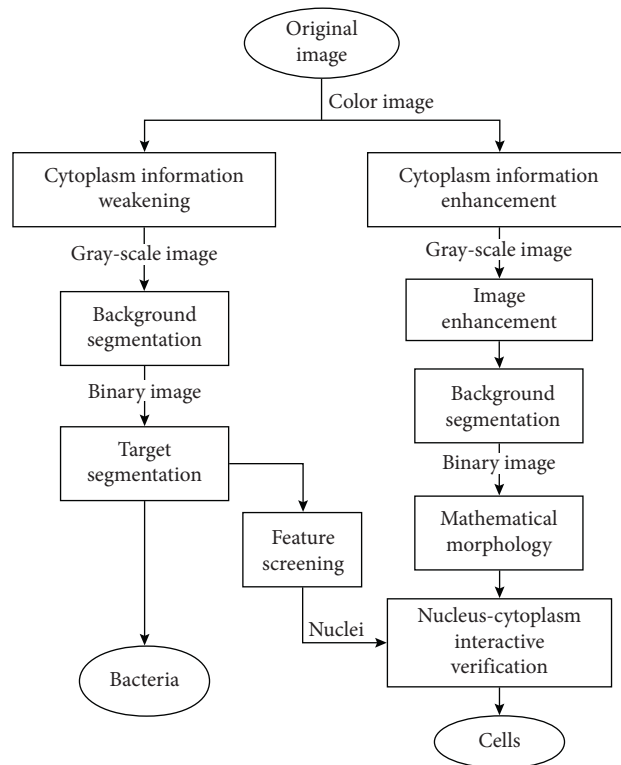


FIGURE 2: Overall flow.

$$B(x, y) = \begin{cases} 1, & \text{if } f(x, y) \leq T, \\ 0, & \text{if } f(x, y) > T, \end{cases} \quad (1)$$

where f is the gray-scale image from Step 1 and T is a fixed threshold value.

Step 3. Isolate bacteria for targets segmentation through morphological operation and complement operation: first, it is primary to label the connected components in the binary image from Step 2 forming a label matrix. And then, the opening arithmetic of mathematical morphology is used to eliminate small connected components:

$$L_N = (L \ominus b) \oplus b, \quad (2)$$

where b is the structuring element, L is the label matrix, \ominus is the erosion operation, and \oplus is the dilation operation. The ROI binary image of bacteria can be generated by complement operation:

$$S = C_{\{|l| \in L\}} \{l_n | l_n \in L_N\},$$

$$B_B(x, y) = \begin{cases} 1, & \text{if } L(x, y) \in S, \\ 0, & \text{else.} \end{cases} \quad (3)$$

Step 4. Filter targets through area feature: as the area of a cell is between 50 and 250 pixels, it is helpful to calculate the connected component areas and remove the regions with more than 250 pixels or fewer than 50 pixels:

$$L'_N(x, y) = \begin{cases} L_N(x, y), & \text{if } 50 < n_{L_N(x, y)} < 250, \\ 0, & \text{else,} \end{cases} \quad (4)$$

$$B_N(x, y) = \text{sgn}(L'_N(x, y)),$$

where n_i is the number of i label and sgn is the symbolic function.

After the above steps, the binary image of bacteria B_B and that of nuclei B_N are obtained.

3.2. Process for Cytoplasm. The steps to obtain the regions of cytoplasm are described as follows.

Step 1. Gray the original image using green component for cytoplasm information enhancement.

Step 2. Enhance image with Gaussian filtering: there is Gaussian noise in the region of cytoplasm due to the uneven staining. It is necessary to apply Gaussian filtering, and the result is shown in Figure 3. The image enhancement step is not applied in process for cells and bacteria. In contrast to cytoplasm containing a large number of moistures, which lead to the uneven staining, bacteria can be evenly stained. In fact, the Gaussian filter will undermine the contour information of bacteria in image instead as can be seen in Figure 3.

Step 3. Segment Background by weighted Otsu [16] threshold segmentation: since the cytoplasm is not as dark as the nucleus and its gray value is related to the staining effect, the weighted Otsu algorithm is used to obtain the adaptive segmentation threshold, which should be in the valley part of the gray histogram:

$$T_w = w \times \arg\max [p_B(\mu_B - \mu)^2 + p_O(\mu_O - \mu)^2],$$

$$B(x, y) = \begin{cases} 1, & \text{if } f(x, y) \leq T_w, \\ 0, & \text{else,} \end{cases} \quad (5)$$

where f is the gray-scale image from Step 2, w is the weight, p_B and μ_B are the probability distribution and the gray mean of background regions, and p_O and μ_O are that of object regions. Through numerous experiments, when the weight is 1.1, the cytoplasm regions can be segmented from the background very well.

Step 4. Remove irrelevant regions by mathematical morphology: opening arithmetic is used to remove the bacteria regions in the binary image from Step 3. The edges of ROIs can be deburred as well through the operation:

$$B_p = (B \ominus b) \oplus b, \quad (6)$$

where B is the binary image and b is the structuring element.

After the above steps, the binary image of cytoplasm B_p is generated.

3.3. Nucleus-Cytoplasm Cross Verification. In fact, the binary image of cytoplasm includes the regions of cells, cytoplasts, free nuclei without enough cytoplasm, and impurities, while the binary image of nuclei includes the regions of nuclei and impurities with the area between 50 and 250 pixels. To distinguish cells from cytoplasts and meanwhile eliminate the interference of impurities, the nucleus-cytoplasm cross verification is designed as follows.

Step 1. Label the connected components in the binary image of cytoplasm so that the corresponding region can be operated by the label value.

Step 2. Set up a set of regions containing nuclei: the elements in the set are the label values of regions with nuclei:

$$S_N = \{l_N | l_N \in L_p \cdot B_N\}. \quad (7)$$

Step 3. Set up a set of regions with enough cytoplasm area: after removing the regions of nuclei from the regions of cytoplasm, the quantity of each label value is calculated as the area of corresponding region. Then, those label values with the quantity greater than 500 will become the element of the set:

$$S_p = \{l_p | l_p \in L_p \cdot (1 - B_N), n(l_p) > 500\}, \quad (8)$$

where $n(i)$ is the quantity of the label value i .

Step 4. According to the sets, reserve the regions with nucleus and with enough area of cytoplasm:

$$S_C = S_N \cap S_p,$$

$$B_C(x, y) = \begin{cases} 1, & \text{if } L_p(x, y) \in S_C, \\ 0, & \text{else.} \end{cases} \quad (9)$$

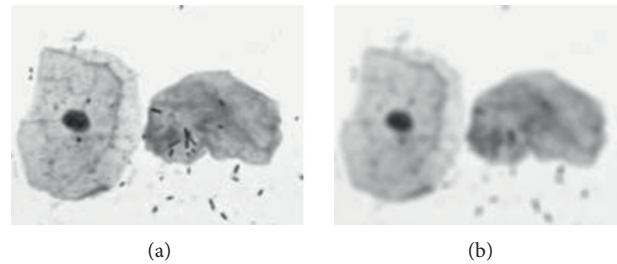


FIGURE 3: The effect of Gaussian filtering: (a) before filtering and (b) after filtering.

After all of the above steps, the binary images of cells and bacteria are separately obtained. The number of cells and bacteria can be obtained by counting the number of connected components.

4. Experimental Result

An image with a size of 1920 pixels * 1024 pixels is selected for the experiment. The results of each step in the process for cells and bacteria are shown in Figures 4–6. Those in the process for cytoplasm are shown in Figures 7 and 8, and the results of cross verification are shown in Figure 9.

As shown in Figure 4, the regions of cytoplasm are similar to the background region in the gray-scale image of (b), which is formed of the red component of the original image of (a). Therefore, it is possible to segment the regions without background and cytoplasm of (c) using a fixed threshold.

As shown in Figure 5, by opening arithmetic of mathematical morphology, the regions of bacteria are removed leaving the regions without bacteria of (a), while the regions of bacteria of (b) are formed of those eliminated regions.

In view of the fact that the area of nucleus is between 50 and 250 pixels, area filtering is used to remove the regions of impurities, and the result is shown in Figure 6.

Figure 7 shows the results of the enhancement of cytoplasm information, in which the gray value of the regions of cytoplasm is lower than that of background regions after enhancement of cytoplasm by graying image using the green component and noise reduction by Gaussian filtering.

For the sake of segmentation of background, the weighted Otsu threshold is adopted, which, as shown in Figure 8, is at the valley of the gray histogram of (a). As a result, the segmentation regions of (b) remain the regions of cytoplasm very well. After a morphological opening operation to remove regions of bacteria, the regions of cytoplasm of (c) are obtained.

As shown in Figure 9, the cross verification firstly removes the regions of nuclei from regions of cytoplasm and reserves the regions with an area greater than 500 pixels of (a). Then, the regions of nuclei are used again to reserve the regions with nuclei, the regions of cells of (b).

Marking the original image with the mark of region contours of cells in red, that of bacteria in blue, and that of nuclei in green, the effect is shown in Figure 10, where the bacteria and cells are significantly distinguished from cytoplasts, free nuclei, and impurities.

The results of each object are shown in Figures 11–16. The results of free nucleus in Figure 11 suggest that the nuclei are not mistaken for bacteria, and the free nuclei will not be misidentified as cells during cross verification due to the insufficient of cytoplasm. The results of bacterium in Figure 12 show that bacteria can be identified correctly. The results of cell in Figure 13 indicate that, as the regions of cells include both the regions of nuclei and cytoplasm, it can be identified correctly, and the regions of bacteria can be obtained without hindrance. The results of cytoplast in Figure 14 show that as a result of the lack of nucleus, the regions of cytoplasts are removed in cross verification. The results also demonstrate that cells and cytoplasts can be distinguished well with the cross verification. The results of impurities in Figures 15 and 16 suggest that, in the case that the areas of impurities are similar to the cells, the impurities would be removed as nuclei, due to the fact that regions of the impurities are reserved in the regions of both nuclei and cytoplasm resulting in insufficient area of cytoplasm in cross verification. In other cases, the impurities would be removed like cytoplasts for the lack of regions of nuclei owing to the area filtering step.

5. Discussion and Conclusion

For assisting in the diagnosis of vaginal health, a dual process and nucleus-cytoplasm cross verification method is proposed in this paper to identify bacteria and cells in vaginal secretions Gram stained smear images. The proposed method can effectively distinguish bacteria and cells from impurities, free nuclei, and cytoplasts and extract regions of bacteria and cells, respectively, which can be used to count the generate pixel-level label data of vaginal secretions epithelial cells and bacteria.

The proposed method was designed according to the features of nuclei, bacteria, cells, cytoplasts, and impurities after Gram stain. Therefore, the method is not applicable in the cases of images made with other techniques or containing other objects. The method is considered to be able to recognize other cells in different smear images with some adjustments, provided the images belong to the applicable cases. Because of the constant color feature of Gram stain and the morphology difference between cells, by modifying the upper and lower limit of nucleus area of Step 4 in process for cells and bacteria and adjusting the threshold for cytoplasm area of Step 3 in nucleus-cytoplasm cross verification, the proposed method can be suitable for a new smear image.

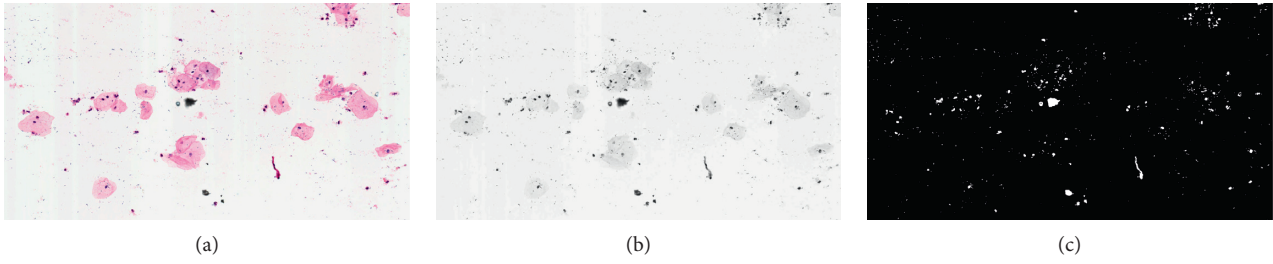


FIGURE 4: Results of separating regions without background and cytoplasm: (a) original image, (b) gray-scale image with weakened cytoplasm, and (c) regions without background and cytoplasm.



FIGURE 5: Results of separating regions of bacteria: (a) regions without bacteria and (b) regions of bacteria.

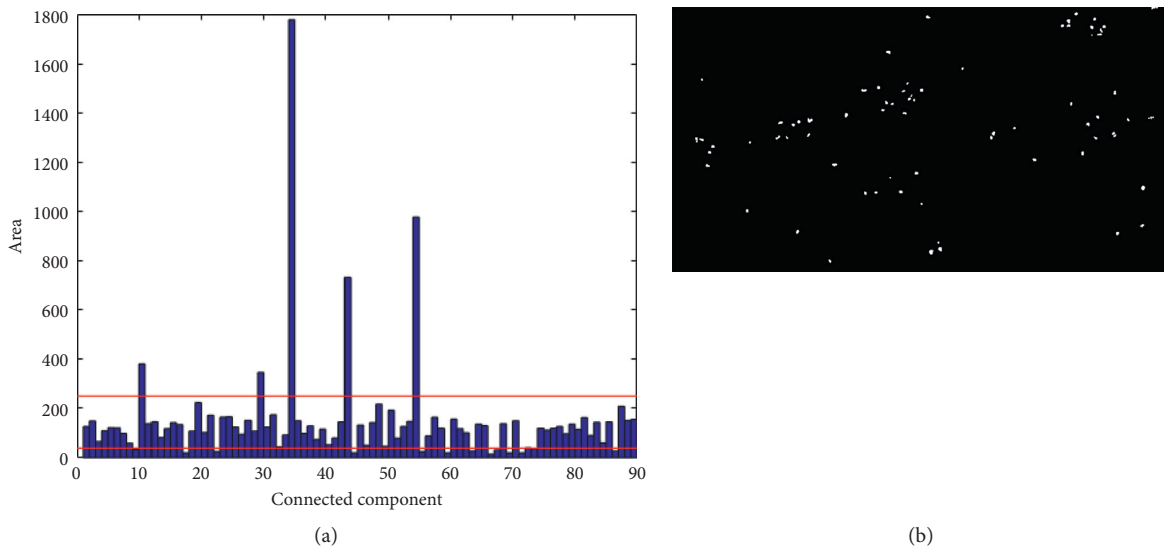


FIGURE 6: Result of area filtering: (a) area histogram of the regions without bacteria and (b) regions of nuclei.

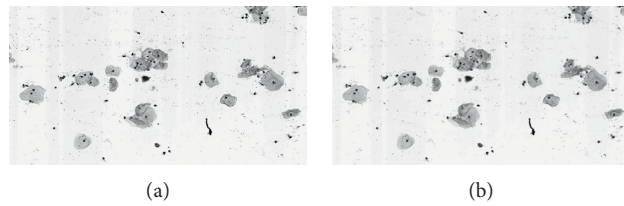


FIGURE 7: The results of enhancement of cytoplasm information: (a) gray-scale image with enhanced cytoplasm and (b) image after Gaussian filtering.

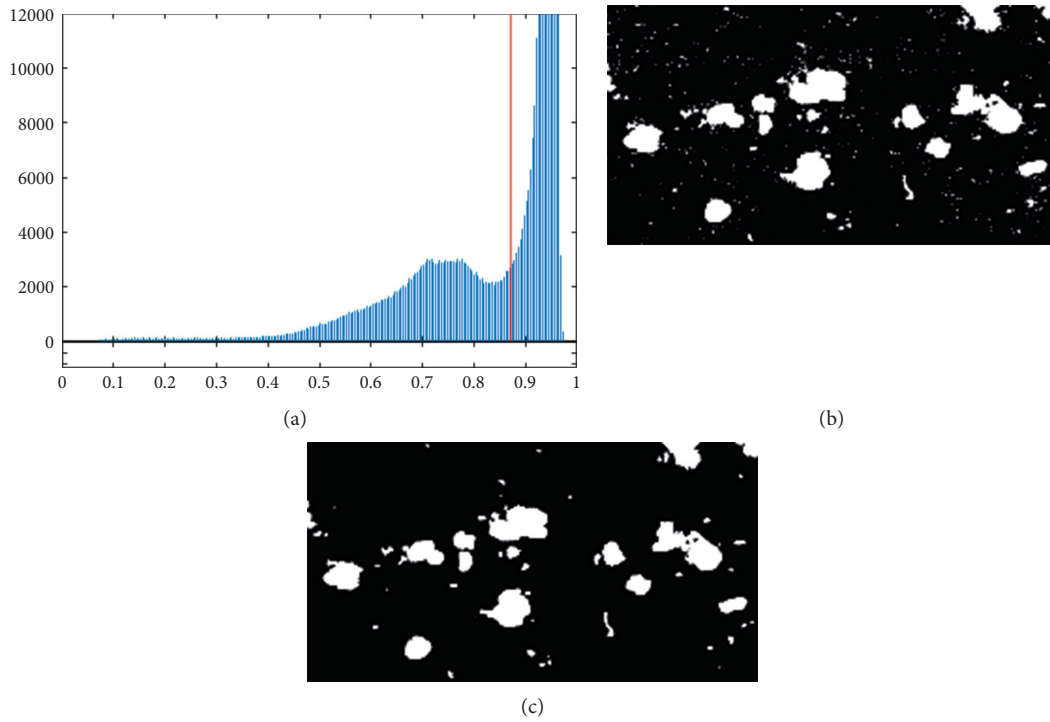


FIGURE 8: Result of segmentation: (a) gray histogram and threshold, (b) segmentation regions, and (c) regions of cytoplasm.



FIGURE 9: Result of cross verification: (a) the regions with area greater than 500 pixels and (b) the region of cells.

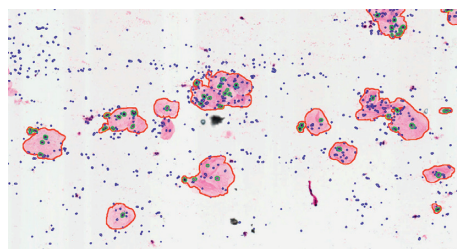


FIGURE 10: Final effect image.

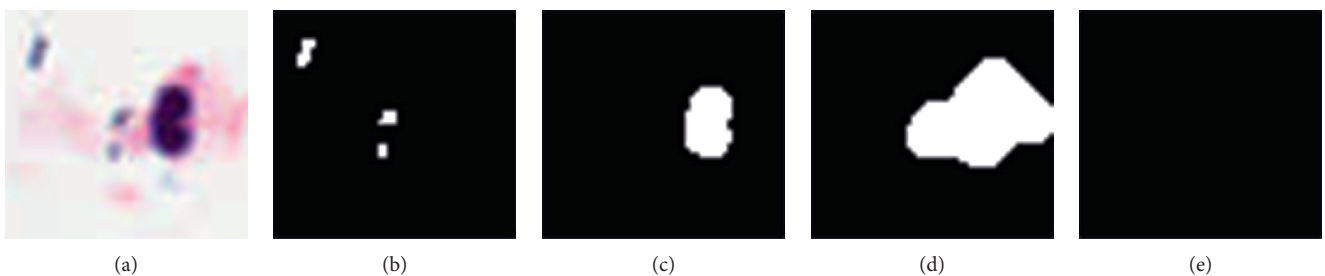


FIGURE 11: The results of free nucleus: (a) original image, (b) regions of bacteria, (c) region of nucleus, (d) region of cytoplasm, and (e) region of cell.

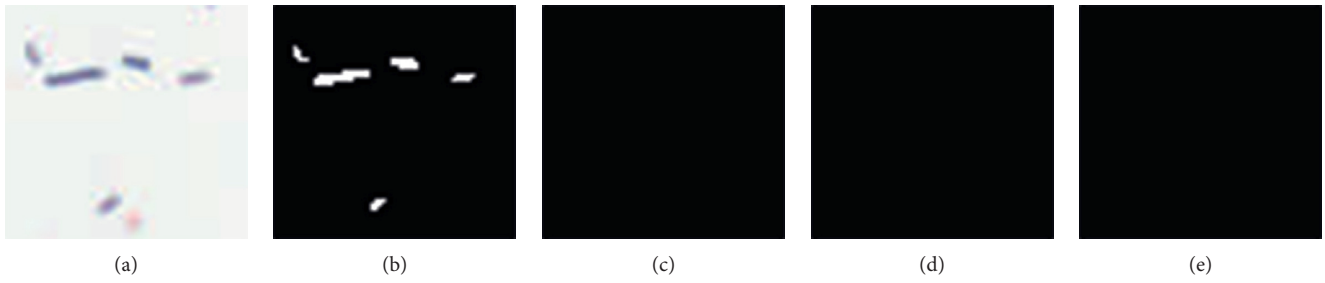


FIGURE 12: The results of bacterium: (a) original image, (b) regions of bacteria, (c) region of nucleus, (d) region of cytoplasm, and (e) region of cell.

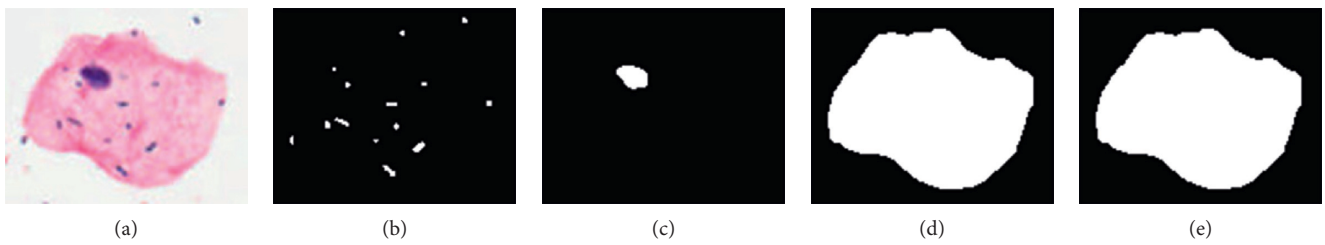


FIGURE 13: The results of cell: (a) original image, (b) regions of bacteria, (c) region of nucleus, (d) region of cytoplasm, and (e) region of cell.

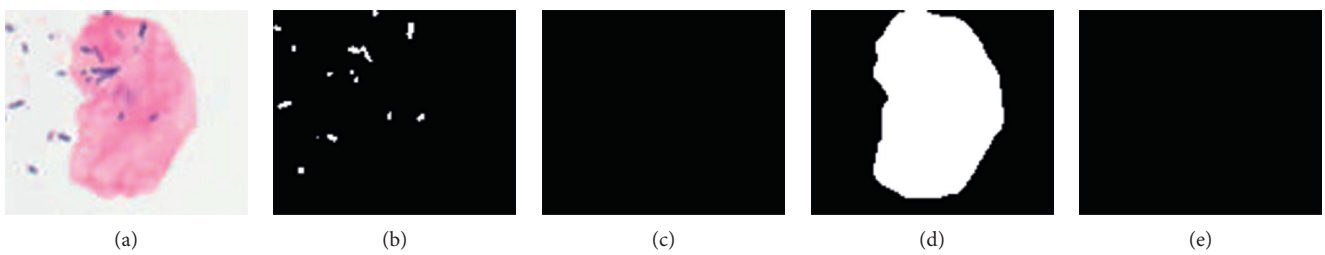


FIGURE 14: The results of cytoplasm: (a) original image, (b) regions of bacteria, (c) region of nucleus, (d) region of cytoplasm, and (e) region of cell.

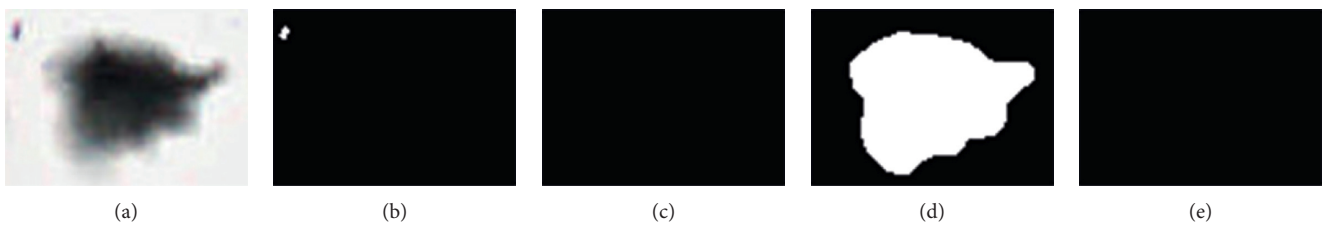


FIGURE 15: The results of large impurity: (a) original image, (b) region of bacterium, (c) region of nucleus, (d) region of cytoplasm, and (e) region of cell.

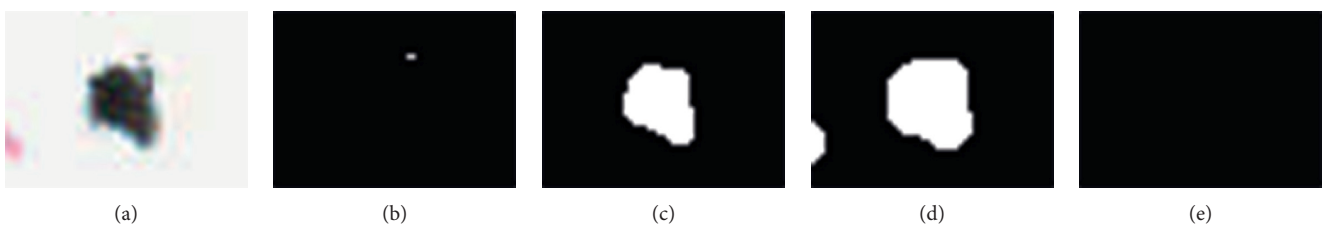


FIGURE 16: The results of nucleus like impurity: (a) original image and (b) region of bacterium (c) region of nucleus, (d) region of cytoplasm, and (e) region of cell.

Data Availability

The raw data required to reproduce these findings cannot be shared at this time as the data are concerned with personal conceal.

Disclosure

Shaozhi Guo and Haoyuan Guan are co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Shaozhi Guo and Haoyuan Guan contributed equally to this work.

Acknowledgments

This work was supported by the Key-Area Research & Development Program of Guangdong Province under grant no. 2019B010145001 and the Science & Technology Program of Guangdong under grant no. 2019A141401005.

References

- [1] S. M. Kovachev, "Obstetric and gynecological diseases and complications resulting from vaginal dysbacteriosis," *Microbial Ecology*, vol. 68, no. 2, pp. 173–184, 2014.
- [2] J. Ravel, P. Gajer, Z. Abdo et al., "Vaginal microbiome of reproductive-age women," *Proceedings of the National Academy of Sciences*, vol. 108, no. Supplement_1, pp. 4680–4687, 2011.
- [3] C. A. Spiegel, R. Amsel, and K. K. Holmes, "Diagnosis of bacterial vaginosis by direct gram stain of vaginal fluid," *Journal of Clinical Microbiology*, vol. 18, no. 1, pp. 170–177, 1983.
- [4] R. P. Nugent, M. A. Krohn and S L. Hillier, "Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation," *Journal of Clinical Microbiology*, vol. 29, no. 2, pp. 297–301, 1991.
- [5] Y. Ma, R.-L. Dai, L. Lian, and Z.-F. Zhang, "An counting and segmentation method of blood cell image with logical and morphological feature of cell," *Chinese Journal of Electronics*, vol. 11, no. 1, 2002.
- [6] J. L. Minoi, T. T. Chiang, T. Lim, Z. Yusoff, and A. Zulharnain, "Mobile vision-based automatic counting of bacteria colonies," in *2016 International Conference on Information and Communication Technology (ICICTM)*, IEEE, Kuala Lumpur, Malaysia, April 2017.
- [7] A. Gençtav, S. Aksoy, and S. Önder, "Unsupervised segmentation and classification of cervical cell images," *Pattern Recognition*, vol. 45, no. 12, pp. 4151–4168, 2012.
- [8] K. Li, Z. Lu, W. Liu, and J. Yin, "Cytoplasm and nucleus segmentation in cervical smear images using radiating gvf snake," *Pattern Recognition*, vol. 45, no. 4, pp. 1255–1264, 2012.
- [9] J. Zhang, S. Lu, X. Wang et al., "Automatic identification of fungi in microscopic leucorrhoea images," *Journal of the Optical Society of America A*, vol. 34, no. 9, p. 1484, 2017.
- [10] L. Zhou, L. Wang, Z. Gao, and J. Zhang, "HEp-2 cell image classification with convolutional neural networks," in *2014 1st Workshop on Pattern Recognition Techniques for Indirect Immunofluorescence Images*, IEEE, Stockholm, Sweden, December 2014.
- [11] C.-J. Tseng, C.-J. Lu, C.-C. Chang, and G.-D. Chen, "Application of machine learning to predict the recurrence-proneness for cervical cancer," *Neural Computing and Applications*, vol. 24, no. 6, pp. 1311–1316, 2014.
- [12] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *Computer Vision & Pattern Recognition CVPR 2011*, IEEE, Providence, RI, USA, August 2011.
- [13] W. Liping and L. Renren, "Red blood cell recognition based on bp neural networks," *Computer & Digital Engineering*, vol. 46, no. 3, 584 pages, 2018.
- [14] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 532–550, 1987.
- [15] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–168, 2004.
- [16] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.