Detecting the Width of Pap Smear Cytoplasm Image Based on GLCM Feature

by Nita Merlina

Smart Innovation, Systems and Technologies 182

Yu-Dong Zhang
Tomonoby Senjyu
Chakchai SO–IN
Amit Joshi *Editors*

Smart Trends in Computing and Communications: Proceedings of SmartCom 2020

SIST 182

Springer

Smart Innovation, Systems and Technologies

Volume 182

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, Google Scholar and Springerlink ****

More information about this series at http://www.springer.com/series/8767

Yu-Dong Zhang · Tomonoby Senjyu ·
Chakchai SO–IN · Amit Joshi
Editors

Smart Trends in Computing and Communications: Proceedings of SmartCom 2020

Springer

Editors
Yu-Dong Zhang
School of Informatics
University of Leicester
Leicester, England, UK

Chakchai SO–IN
Khon Kaen University
Khon Kaen, Thailand

Tomonoby Senjyu
University of the Ryukyus
Nishihara, Japan

Amit Joshi
Global Knowledge Research Foundation
Gujarat, India

ISSN 2190-3018 ISSN 2190-3026 (electronic)
Smart Innovation, Systems and Technologies
ISBN 978-981-15-5223-6 ISBN 978-981-15-5224-3 (eBook)
https://doi.org/10.1007/978-981-15-5224-3

Preface

This SIST volume contains the papers presented at the SmartCom 2020: Fourth International Conference on Smart Trends for Computing and Communications. The conference was held during January 24–25, 2020, at Hotel Novotel, Siam Square, Bangkok, and organized by Global Knowledge Research Foundation with support from Knowledge Chamber of Commerce and Industry and InterYIT IFIP—International Federation for Information Processing.

The books presents the state-of-the-art as well as emerging topics pertaining to information, computer communications and effective strategies for its implementation for engineering and managerial applications.

The conference attracts a large number of high-quality submissions and stimulates the cutting-edge research discussions among many academic pioneering researchers, scientists, industrial engineers, students from all around the world and provides a forum to researchers on proposing new technologies, sharing their experiences and discussing future solutions for design infrastructure for ICT, providing common platform for academic pioneering researchers, scientists, engineers and students to share their views and achievements, enriching technocrats and academicians by presenting their innovative and constructive ideas and focusing on innovative issues at international level by bringing together the experts from different countries.

Research submissions in various advanced technology areas were received, and after a rigorous peer review process with the help of program committee members and 56 external reviewers for 262 papers from 20 different countries including Australia, China, Malaysia, Indonesia, Thailand, Bangladesh, Japan, South Korea and Sri Lanka, etc., out of which 50 were accepted with an acceptance ratio of 0.19.

This event's success was possible only with the help and support of our team and organizations. With immense pleasure and honor, we would like to express our sincere thanks to the authors for their remarkable contributions, all the Technical Program Committee members for their time and expertise in reviewing the papers within a very tight schedule and the publisher Springer for their professional help. This is the fourth conference of the series SmartCom in which proceedings is published as a CCIS volume by Springer. We are overwhelmed by our two

distinguished scholars and appreciate them for accepting our invitation to deliver keynote speeches to the conference and six technical session chairs for analyzing the research work presented by the researchers. Most importantly, we are also grateful to our local support team for their hard work for the conference. This series has already been made a continuous series which will be hosted at different locations every year.

This event's success was possible only with the help and support of our team and organizations. With immense pleasure and honor, we would like to express our sincere thanks to the authors for their remarkable contributions, all the Technical Program Committee members for their time and expertise in reviewing the papers within a very tight schedule and the publisher Springer for their professional help. This is the fourth conference of the series SmartCom in which proceedings is published as a SIST volume by Springer. We are overwhelmed by our three distinguished scholars and appreciate them for accepting our invitation to deliver keynote speeches including Prof. Mike Hinchey, President IFIP; Prof. Milan Tuba, Singidunum University, Serbia; and Dr. ChakChai So-In, Khon Kaen University, Thailand, in the conference and further six technical session chairs for analyzing the research work presented by the researchers. Most importantly, we are also grateful to our local support team for their hard work for the conference. This series has already been made a continuous series which will be hosted at different locations every year.

Bangkok, Thailand

Yu-Dong Zhang
Tomonoby Senjyu
Chakchai SO–IN
Amit Joshi

Contents

About the Editors

Prof. Yu-Dong Zhang received his B.S. and M.S. degrees from Nanjing University of Aeronautics and Astronautics in 2004 and 2007, prior to completing his Ph.D. in Signal and Information Processing at Southeast University in 2010. From 2010 to 2012, he worked at Columbia University as a postdoc. From 2012 to 2013, he worked as a research scientist at Columbia University and New York State Psychiatric Institute. From 2013 to 2017, he worked as a Full Professor and doctoral advisor at the School of Computer Science and Technology, Nanjing Normal University. Since 2018, he has been a Full Professor at the Department of Informatics, University of Leicester, UK.

Prof. Tomonoby Senjyu received his B.S. and M.S. degrees in Electrical Engineering from the University of the Ryukyus, Nishihara, Japan, in 1986 and 1988, respectively, and his Ph.D. in Electrical Engineering from Nagoya University, Japan, in 1994. He is currently a Full Professor at the Department of Electrical and Electronics Engineering, University of the Ryukyus. His research interests include renewable energy, power system optimization and operation, power electronics, and advanced control of electrical machines.

Prof. Chakchai SO–IN has been with the Department of Computer Science at Khon Kaen University since 2010. Dr. SO–IN received his B.E. and M.E. degrees from Kasetsart University (KU), Bangkok, Thailand, in 1999 and 2001, respectively. He also completed M.S. and Ph.D. degrees in Computer Engineering at the Department of Computer Science and Engineering, Washington University, in St. Louis (WUSTL), MO, USA, in 2006 and 2010. He has authored more than 100 publications in prominent journals and proceedings, together with 10 books. His research interests include mobile computing/sensor networks, Internet of Things, computer/wireless/distributed networks, cyber security, intelligent systems and the future Internet. He is a senior member of the IEEE and ACM.

Dr. Amit Joshi is currently Director of the Global Knowledge Research Foundation, and an entrepreneur and researcher. Holding a B.Tech. in Information Technology and M.Tech. in Computer Science and Engineering, he has conducted extensive research in the areas of cloud computing and cryptography in medical imaging, with a focus on analyzing current government strategies and global forums concerning security. Dr. Joshi has more than 10 years of experience in academia and industry, having worked at prestigious organizations. He is an active member of the ACM, IEEE, CSI, AMIE, IACSIT Singapore, IDES, ACEEE, NPA, and many other professional societies. Currently, he is the International Chair of InterYIT at the International Federation of Information Processing (IFIP, Austria). He has presented and published more than 50 papers in national and international journals/conferences of the IEEE and ACM, and has edited more than 20 books for Springer, ACM, and other reputed publishers. He has also organized more than 40 national and international conference and workshops in association with the ACM, Springer, and the IEEE, e.g. in India, the UK, Thailand, and Europe.

Chapter 22
Detecting the Width of Pap Smear Cytoplasm Image Based on GLCM Feature

Nita Merlina, Edi Noersasongko, Pulung Nurtantio, M. A. Soeleman, Dwiza Riana, and Sri Hadianti

Abstract Color image segmentation on cytoplasm Pap smear single cell image identified as normal condition is an interesting subject to study. It is caused by the image limitation and morphological transformation complexity of the cell structural part. Feature analysis on cytoplasm area is an important thing in the process of biomedical image analysis because of the noise and complex background and the bad cytoplasm contrast as well. Thus, an analysis on the feature area on cytoplasm automatically is an urge thing to do to identify Pap smear normal cell image based on feature analysis on cytoplasm area in single cell image identified as normal condition. The purpose of this research is to analyze how far the process color image segmentation on cytoplasm by using normal single cell image is able to produce features of texture and form analysis. To analyze the form of cytoplasm, this research used RGB color to HSV color conversion method which produces metric and eccentricity value. It is then continued to the process of threshold image and counting the wide area by changing threshold into binary image. On the other side, to analyze the texture, this research applied an analysis using gray-level co-occurrence matrix (GLCM) using *K*-means method to produce contrast, correlation, energy, and homogeneity parameters. The result of the research is the segmentation outcome to Pap smear normal single cell image sample to get metric, eccentricity, contrast, correlation, and energy features.

22.1 Introduction

Women are generally susceptible to cervical cancer, which ranks fourth as a very deadly disease [1]. Papanicolaou test also called Pap smear or Pap test is a health test method that can help prevent cervical cancer. The main purpose of a Pap smear

N. Merlina (✉) · E. Noersasongko · P. Nurtantio · M. A. Soeleman
Universitas Dian Nuswantoro, Semarang, Indonesia
e-mail: nita@mhs.dinus.ac.id

N. Merlina · D. Riana · S. Hadianti
STMIK Nusa Mandiri, Jakarta, Indonesia

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications: Proceedings of SmartCom 2020*, Smart Innovation, Systems and Technologies 182,
https://doi.org/10.1007/978-981-15-5224-3_22

is to detect cell abnormalities that may occur or before the cancer develops. Correct interpretation of microscopic examination of cells and tissues is very important for the final diagnosis judgment of the disease [2]. However, the Pap smear testing process has weaknesses, especially in many developing countries where the number of pathologists who can examine slides is inadequate. In Indonesia, with a population of productive age women spread across 32 provinces with limited facilities and limited human resources, it is estimated that 80% of the coverage of examinations will be completed in five years with 7,992,486 Pap smear tests per year [3].

The classification of cervical cells in Pap smear images is an interesting thing because of image limitations and the complexity of morphological alterations in the structural parts of cells [4]. Pap smear image processing has been performed for single cell types and non-overlapping. Image segmentation process is expected to identify cytoplasm and nucleus images [5–7]. Segmentation is needed to define the area of interest (ROI) in a normal single cell image and is the basis of an automatic cervical cancer screening system. Efficient image segmentation enables extraction of significant information and streamlines image data for further analysis. Weak segmentation results in weak results during image analysis [8]. Analysis of features in the cytoplasm area in normal Pap smear cell images is interesting. That is due to image limitations and the complexity of morphological changes in the structural parts of cells. Analysis of features in the cytoplasm area is important in the process of analyzing biomedical images, and it is caused by background noise and complex and weak cytoplasm contrast. So, it is necessary to do an analysis of the features of the area in the cytoplasm by automation to identify the normal Pap smear cell image based on the analysis of the features of the cytoplasm area in a single cell image that is detected normally. The purpose of this research is to see the extent of the process of analyzing feature areas in the cytoplasm by using a normal single cell image so that features of texture and shape analysis can be produced. The results of this study are expected to facilitate the identification of cytoplasm images from Pap smear images.

This paper is divided into several sections. Section 22.2, related work, discusses about the methods used in the research. Section 22.3 describes the results and discussion, which will be enclosed with conclusions and further research plans.

22.2 Related Work

There are many methods used to identify the area of cytoplasm image, and this study attempts to do different things, which is to propose a segmentation method for cytoplasm area in a single Pap smear image that utilizes the GLCM feature. This GLCM feature has not been widely used in previous research. In the initial detection of a Pap smear image, a single cell image is easy to observe [9]. In the previous research, cytoplasm detection process has been successfully carried out by

color classification using *K*-means and joint shape matching. However, this research is limited only to determine the presence of cytoplasm and not to determine the extent of cytoplasm in detail [10, 11]. The gray-level co-occurrence matrix (GLCM) previous studies have been used, but they use a dataset from a nearby hospital and only discuss cervical cancer related without the Pap smear cell being used [12].

22.3 Method

Data processing techniques use the MATLAB application program on color and binary image matrices as well as analyzing textures. In order to determine the area in cytoplasm image, the initial step of this research is to convert RGB color to HSV color (hue, saturation, value). The HSV color model was formulated by searching the RGB color cube along the gray axis of the color wheel, where this model is more widely used than RGB in depicting color sensations. After the HSV color conversion is generated, the next step is to determine the threshold image and calculate the area of the area by changing the threshold image into a binary image that aims to calculate objects in a digital image in a simple way from the original image. In this process, we need a boundary value called the threshold value. Image intensity values that are more than or equal to the threshold value will be changed to white (1), while image intensity values less than the threshold value will be changed to black (0) [13]. Furthermore, to analyze the texture of the image of the area value that has been segmented by the *K*-means method and the researcher uses the Gray-Level Co-Occurrence Matrix (GLCM) so that the parameters obtained are contrast, correlation, energy, and homogeneity. The results of this research are the results of segmentation of a single Pap smear normal cell image sample so that it can be done well and is able to get the features of metric, eccentricity, contrast, correlation, and energy [14], where the texture has characteristics that can be extracted to distinguish an image with another image. Figure 22.1 shows the image of a single cell used in this study.

In this study, a research flow diagram is shown in Fig. 22.2 on the digital image of papers.

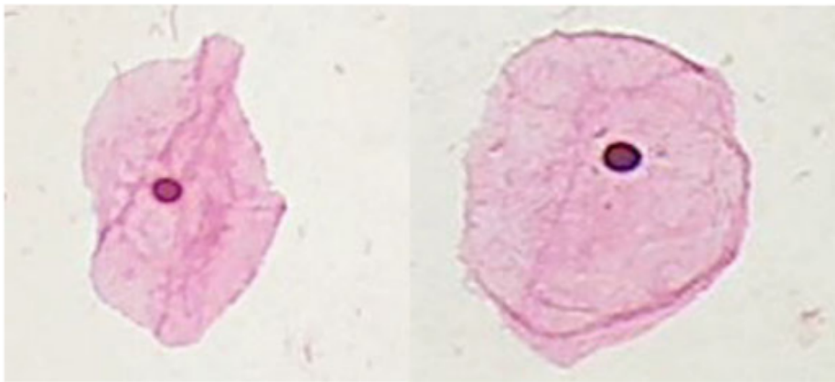

Fig. 22.1 Example of a single cell image

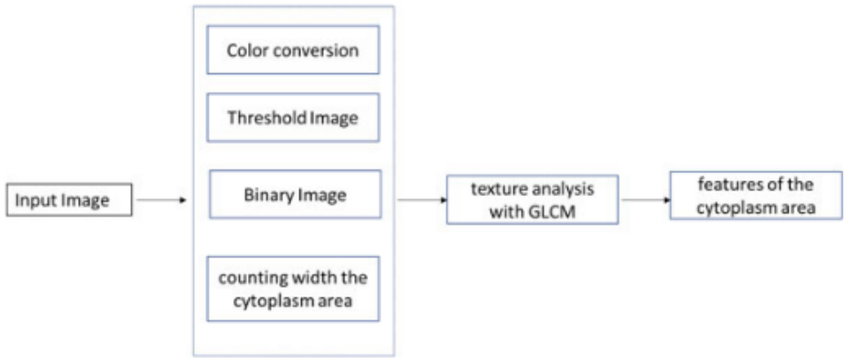

Fig. 22.2 Research stages

22.4 Results and Discussion

22.4.1 Result

The stages of the research consisting of five stages of the process are evaluated on traditional Pap smear images. All of these images were obtained through a Logitech camera (Logitech HD C525 Webcam) adjusted to an optical microscope (Olympus CH20). 40× amplification is used, and the results are saved in JPEG format. The following are the results of the process of segmentation methods in each.

22.4.1.1 Convert RGB Color to HSV Color

The result of the conversion of RGB color into HSV color aims to detect the cytoplasm so as to produce an image as shown in Fig. 22.3, where the cytoplasm has a color that is more dominant than the nucleus.

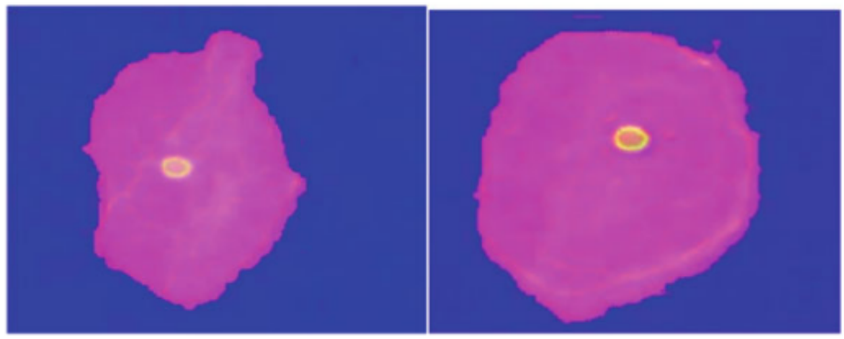

Fig. 22.3 Image results from color conversion

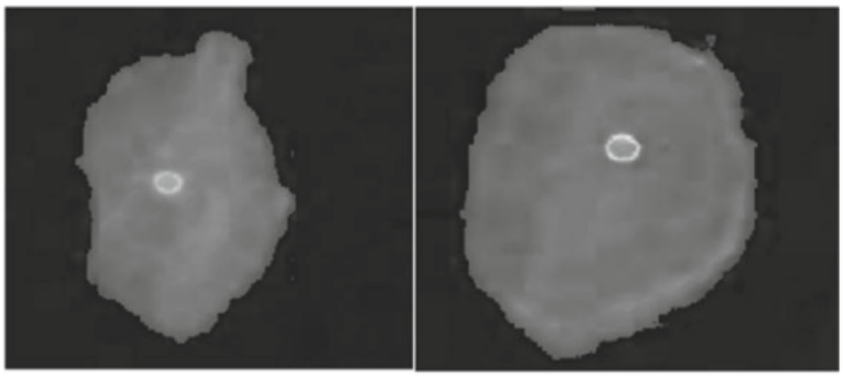

Fig. 22.4 Image from threshold

22.4.1.2 Threshold Image

When analyzing an area in the cytoplasm, the initial process is to determine the edge of the object from a single normal cell. The results of edge detection in normal single cell images are taken from images that have been converted to HSV color, and Fig. 22.4 shows images of edge detection.

22.4.1.3 Binary Imagery

The calculation process of the attributes that exist in a single normal Pap smear cell can be simply done by the process of conversion to binary images; after the conversion process is carried out, the area calculation process can be carried out, along with pictures that present the results of binary images (Figs. 22.5, 22.6, and 22.7).

Fig. 22.5 Binary image

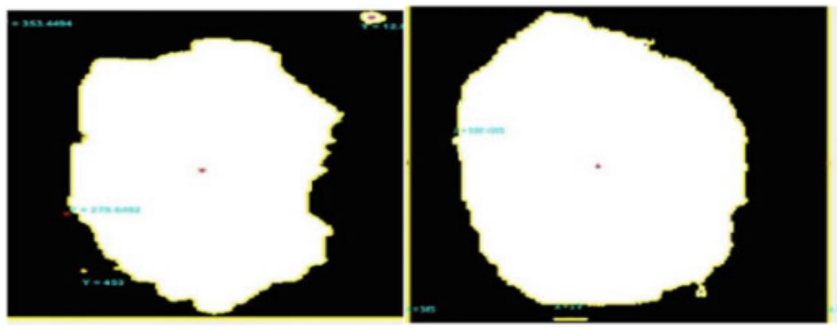

Fig. 22.6 Width binary image

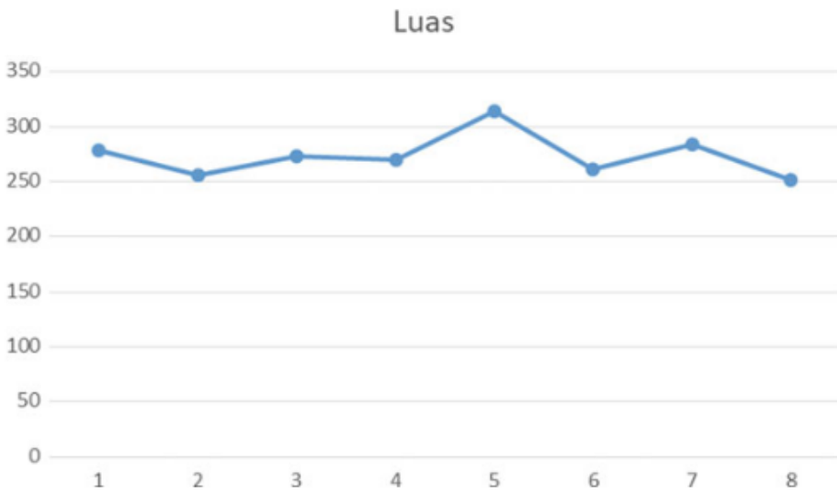

Fig. 22.7 Chart width of binary image

22.4.1.4 Gray-Level Co-occurrence Matrix (GLCM)

The process of this GLCM is to analyze the texture of the cytoplasm area, so that this research results in segmentation of a single Pap smear normal cell image sample by producing metric, eccentricity, contrast, correlation, and energy features. The following are the formulas for calculating the metric feature, eccentricity, contrast, correlation, and energy:

$$\text{Contrast} = \sum_{i_1} \sum_{i_2} (i_1 - i_2)^2 p(i_1, i_2) \tag{22.1}$$

$$\text{Homogeneity} = \sum_{i_1} \sum_{i_2} \frac{p(i_1, i_2)}{1 + |1_1 - i_2|} \tag{22.2}$$

$$\text{Energy} = \sum_{i_1} \sum_{i_2} p^2(i_1, i_2) \tag{22.3}$$

Table 22.1 Texture of the cytoplasm area

Metric	Eccentricity	Contrast	Correlation	Energy	Homogeneity
0.01	1	0.01	0.99	0.57	1
0.01	0.46	0.01	0.98	0.49	0.99
0.01	1	0.01	0.99	0.49	1
0.02	1	0.01	0.99	0.51	1
0.95	0.7	0.01	0.99	0.53	1
0.78	0.7	0.01	0.88	0.95	1
0.01	0.73	0.01	0.98	0.52	0.99
0.67	0.53	0.01	0.99	0.5	1

$$\text{Entropy} = -\sum_{i_1}\sum_{i_2} p(i_1, i_2)\log p(i_1, i_2) \tag{22.4}$$

$$\text{Eccentricity} = e = \sqrt{1 - \frac{b^2}{a^2}} \tag{22.5}$$

$$\text{Metric} = M = \frac{4\pi \times A}{C^2} \tag{22.6}$$

Table 22.1 gives the results of the GLCM calculations that have been carried out.

22.4.2 Discussion

In this study, we segmented Pap smear images by taking ten image from a single normal goal, but there are two images that are not segmented perfectly; this is because there is noise in the image, so the segmentation is done using the method taken less than perfect.

22.5 Conclusion

The results of this study are to see the extent of the cytoplasm color image segmentation process using normal single cell images so that features of texture and shape analysis can be produced. To analyze the shape of the cytoplasm, this study uses the method of converting RGB color conversion into HSV color conversion that produces metric and eccentricity values and then proceeds with the process to determine the threshold image and calculate the area of the area by changing the threshold image which becomes a binary image. For texture analysis, the analysis is done using the gray-level co-occurrence matrix (GLCM) method which uses the *K*-means method

to obtain contrast, correlation, energy, and homogeneity parameters. The results of this research are the results of segmentation of normal Pap smear single cell image samples so that they can be done well and are able to get the features of metric, eccentricity, contrast, correlation, and energy. The limitation of this study is that the number of samples used needs to be increased so that it will produce more feature data to get more extensive information. This research is a preliminary study for cytoplasm recognition. Further research will focus on segmentation of the cytoplasm in single Pap smear cells by adding new features besides GLCM. Another thing to do is research for group Pap smears.

References

1. Irawan, F.: Statistik penderita kanker di Indonesia (2018)
2. Riana, D., Widyantoro, D.H., Mengko, T.L.: Extraction and classification texture of inflammatory cells and nuclei in normal pap smear images. In: 2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (2015)
3. Kusuma, F.: Tes Pap Dan Cara Deteksi Dini Kanker Serviks Lainnya, Departemen Obstetri dan Ginekologi Fakultas Kedokteran Universitas Indonesia RSUPN Dr. Cipto Mangunkusumo, Prevention and Early Detection of Cervical Cancer (PEACE), Jakarta (2012)
4. Plissiti , S.E., Dimitrakopoulos, P., Sfikas, G., Nikou, C., Krikoni, O., Charchanti, A.: Dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. In: 25th IEEE International Conference on Image Processing (2018)
5. Lu, Z., Carneiro, G., Bradley, A.P.: Automated nucleus and cytoplasm segmentation of overlapping cervical cells. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, vol 8149. Lecture Notes in Computer Science, pp. 452–460 (2013)
6. Zhao, L., Li, K. Wang, M., Yin, J., Zhu, E., Wu, C., Wang, S., Zhu, C.: Automatic cytoplasm and nuclei segmentation for color cervical smear image using an efficient gap-search MRF 46. Comput. Biol. Med. **71**, 46 (2016). ISSN 00104825
7. Riana, D., Hidayanto, A.N., Widyantoro, D.H., Mengko, T.L.R., Kalsoem, O.: Segmentation of overlapping cytoplasm and overlapped areas in Pap smear images. In: Information, Intelligence, Systems & Applications (IISA), pp. 1–5 (2017)
8. Sun, G., Li, S., Cao, Y., Lang, F.: Cervical cancer diagnosis based on random forest. Int. J. Performability Eng. (2017)
9. Riana, D., Ramdhani, Y., Prasetio, R.T., Hidayanto, A.N.: Improving hierarchical decision approach for single image classification of pap smear. Int. J. Electr. Comput. Eng. **8**, 5415–5424 (2018)
10. Riana, D., Tohir, H., Hidayanto, A.N.: Segmentation of Overlapping Areas on Pap Smear Images with Color Features Using K-Means and Otsu Methods (2018). https://ieeexplore.ieee.org/xpl/conhome/8767368/proceeding
11. Song, Y., Qin, J., Lei, B., He, S., Choi, K.-S.: Joint shape matching for overlapping cytoplasm segmentation in cervical smear images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) Venice, Italy, pp. 191–194 (2019)
12. Nehra, S., Raheja, J.L., Butte, K., Zope, A.: Detection of cervical cancer using GLCM and support vector machines, pp. 49–53. https://ieeexplore.ieee.org/xpl/conhome/8768798/proceeding

13. Pamungkas, A.: Pemrograman Matlab (2019). [Online]. Available: https://pemrogramanmatlab.com/pengolahan-citra-digital/segmentasi-citra
14. Bhan, S., Vyas, A., Mishra, G.: Supervised segmentation of overlapping cervical pap smear images. In: 2016 International Conference on Signal Processing and Communications (2016)

Author Index

Y.-D. Zhang et al. (eds.), *Smart Trends in Computing and Communications: Proceedings of SmartCom 2020*, Smart Innovation, Systems and Technologies 182,
https://doi.org/10.1007/978-981-15-5224-3

Detecting the Width of Pap Smear Cytoplasm Image Based on GLCM Feature

ORIGINALITY REPORT

22%	20%	15%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

9%

★ "ICT Systems and Sustainability", Springer Science and Business Media LLC, 2023

Publication

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		