Cervical Precancerous Classification System based on Texture Features and Support Vector Machine

1st Yessi Jusman
Department of Electrical Engineering, Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
*yjusman@umy.ac.id

2nd Brilian Permata Sari
Department of Electrical Engineering, Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

3rd Slamet Riyadi
Department of Information Technology Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

2021 1st International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS) | 978-1-6654-0546-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICE3IS54102.2021.9649687

***Abstract*— Cervical cancer is one of the female reproductive health diseases being a significant issue globally because of the large number of new cases and deaths, particularly among women in developing countries. Cervical cancer can be avoided if detected early. The Pap smear screening procedure is used in industrialized nations to detect cervical cancer early. However, limited human resources, a significant time commitment, high prices, and insufficient infrastructure make it less successful in developing countries. With three types of cervical cell images: Normal, Low-grade Squamous Intraepithelial Lesion (LSIL), and High-grade Squamous Intraepithelial Lesion (HSIL), this study offers a classification system for cervical cell images using an image processing technique called Gray Level Co-occurrence Matrix (GLCM) and a Support Vector Machine (SVM) classification method (HSIL). With HSIL class as positive data and LSIL and Normal as negative data, the classification system used three SVM models: Cubic, Quadratic, and Fine Gaussian. SVM classification accuracy was 97.5 percent for 3.54s using the GLCM feature extraction approach.**

Keywords*— Cervical Cell, GLCM, Cubic SVM, Quadratic SVM, Fine Gaussian SVM

I. INTRODUCTION

Cervical cancer is a malignant disease originating from the cervix—the lower third of the uterus, cylindrical in shape, protruding and communicating with the vagina through the external uterine os. However, this disease can be prevented by early detection and adequate treatment [1]. In developed countries, early detection of the Pap Smear test effectively reduces the incidence and mortality rate of invasive cervical cancer. Unfortunately, the implementation of this early detection test faces many obstacles, particularly in the screening process. The difficulties are related to the Pap smear accuracy, material collection techniques, Pap smear examination being less practical since it can only be performed by trained personnel, and interpretation of results taking longer time due to a lengthy and complex procedure resulting in reasonably high examination costs. Furthermore, screening is hampered by a lack of human resources, procedural and geographic barriers, and a lack of women who should be screened [2].

Therefore, an effective screening method that does not require manual microscopic pathological assessment is necessary. There is currently an artificial intelligence system for image processing of cervical cells with a computer that is applied for early diagnosis of cervical precancerous lesions in the prevention of cervical cell cancer. The technology can deliver a percentage of true positive and negative numbers more accurate and reliable. The exploration of the machine learning sytem for cervical cancer is published at [3]. Performance analysis of the system for cervical cancer detection can be explored in review [4], [5], [6]. The system based on the cytology images [7], [8], spectra cells [9], colposcopy [10] and microscope electron [11], have been presented. A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images is presented [12]. A review of computational methods for cervical cells segmentation and abnormality classification is also presented in [13]. For cervical cells classification system, texture of the cells is different between normal and abnormal cells [14]. Gray Level Co-occurrence Matrix (GLCM) algorithm is used to extract the texture features on the cells. Several studies employed the GLCM method on the cervical cell images achieved an accuracy of 95% [15], while the GLCM method with SVM classification reached an accuracy value above 80% [16], [17], [18], [19], [20].

Support Vector machine (SVM) has evolved as an effective classification model. SVM is the most well-known machine learning approach. For classification and regression, SVM has the best mathematical model. This strong mathematical framework opens up possibilities for research in the large field of classification and regression. A study examines the various SVM computational models and surveys their applicability for image classification [21].

Based on the limitation of the manual screening and the previous research in the cervical precancerous classification using machine learning, this research developed a system to improve the performance of the system by applying image processing technique and the SVM based machine learning technique. A combination of the GLCM for feature extraction method and classifying it using the SVM method are proposed.

II. METHOD

A flow diagram represents the steps taken to achieve the research objectives presented in Figure 1.

A. Data Colection

This study received 291 cervical cell imaging data from the Universiti Sains Malaysia Hospital, including 61 HSIL photos, 161 LSIL images, and 69 Normal images. The image retrieval process and data have been validated, and a code of ethics has been established at the university's code of ethics institution.

B. Preprocessing

Cervical cell images must be processed to increase their quality in order to yield more accurate results. Image improvement to sharpen, cropping, resizing the image to a size of 150x150 pixels with a MATLAB image batch processor, and augmentation with three times alterations in flip vertical, flip horizontal, and flip horizontal-vertical were all performed throughout the preparation phase.

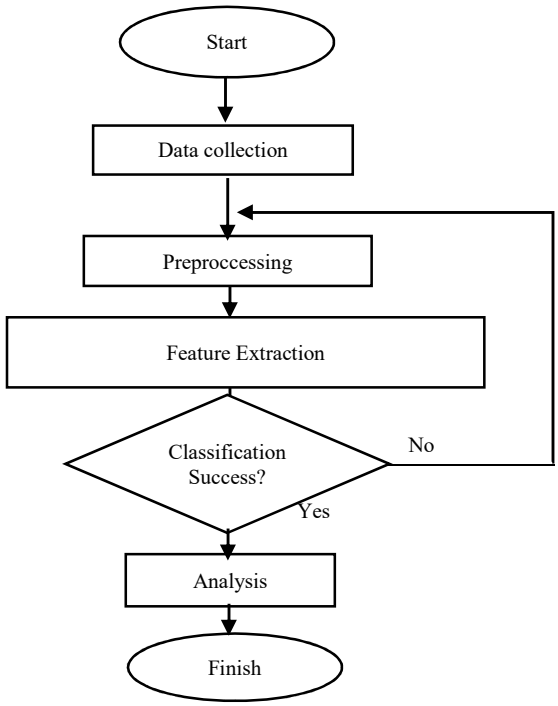
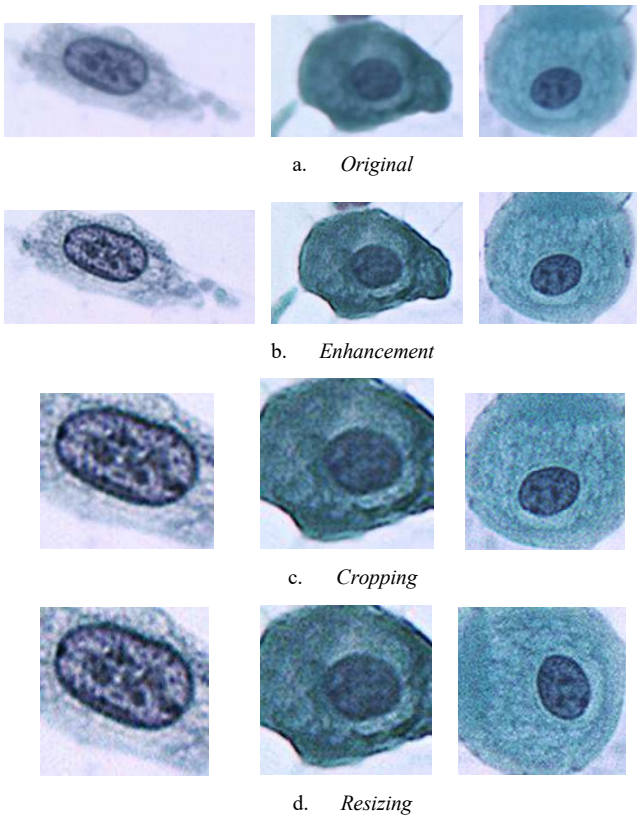


Fig. 1. Research Method Flow



a. *Original*

b. *Enhancement*

c. *Cropping*

d. *Resizing*

Fig. 2. Preprocessing Cervical Cells: a. Original; b. Enhancement; c. Cropping; d. Resizing

Before augmentation, the initial images were 291 consisting of 61 HSIL, 161 LSIL, and 69 Normal images. After augmentation, the total images became 1,164, comprising 244 HSIL, 644 LSIL, and 276 Normal images. The testing data employed 116 images encompassing 24 HSIL images, 64 LSIL images, and 28 Normal images.

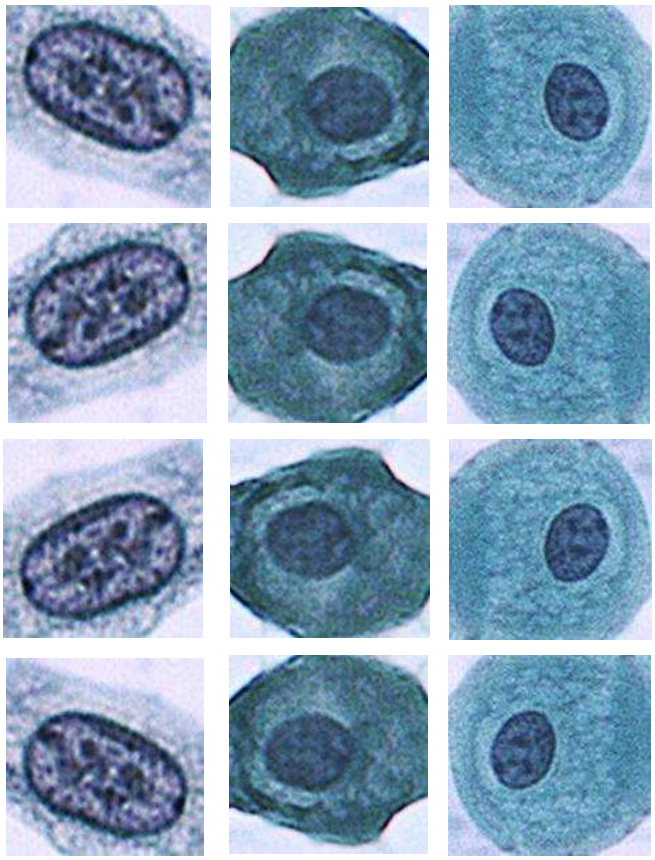


Fig. 3. Cervical Cell Image Augmentation

C. *Feature Extraction Stage*

The feature extraction used in this research was the second-order texture analysis method of Gray Level Co-occurrence Matrix (GLCM). This method determines the features of contrast, correlation, energy, and homogeneity at each angle of 0◦, 45◦, 90◦, and 135◦ by determining the pixel distance (D) of 100, as well as quantization values (Q) of 16.

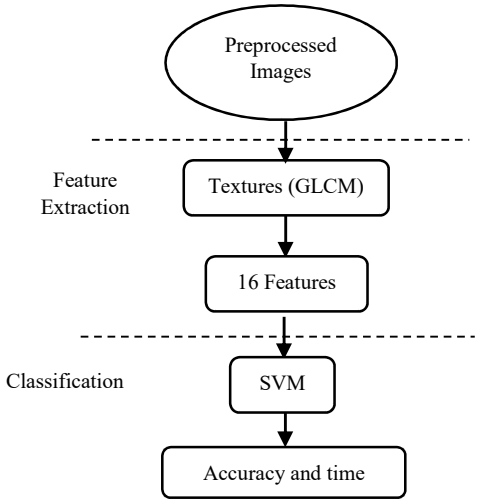


Fig. 4. GLCM and SVM Methods

D. Classification Stage

The SVM was used to classify the GLCM feature extraction results. SVM was developed to find the best hyperplane separating two classes with the greatest margin. SVM was enhanced by the addition of a kernel function for mapping data to a higher dimension (feature space). In this study, the kernel functions Cubic SVM, Quadratic SVM, and Fine Gaussian SVM were compared.

Polynomial SVM (*Cubic* & *Quadratic*)

$$K(x,y) = (x,y+c)^d \quad (1)$$

Gaussian SVM (*Fine Gaussian*)

$$K(x,y) = exp((-||x- [y||] ^2)/(2. \sigma ^2)) \quad (2)$$

III. Results and Discussion

The GLCM feature extraction results extracted the features of the training images and transformed them into 16 feature values from each type of cervical cell image. They were then analyzed using MATLAB's Classification Learner toolbox and the SVM classification algorithm. The 10-fold cross-validation technique was used in this cervical cell testing images, yielding accuracy and running time data for each training dataset.

A. Feature Extraction Results

The following demonstrates the average and standard deviation of each type of cervical cell image (HSIL, LSIL, Normal) using the GLCM method at the values of D=100 and Q=16.

TABLE I RESULTS OF GLCM FEATURE EXTRACTION

Features	Angel	HSIL	LSIL	Normal
Contrast	**0**	18,23 ± 9,8	13,99 ± 10	8,76 ± 7,7
	45	17,42 ±16,3	14,59 ±12	10,78 ±11,6
	90	23,42 ±13,7	16,29 ±9,7	10,53 ±9,7
	135	17,43 ±16,3	14,58 ±12	10,79 ±11,6
Correlation	**0**	0,04 ±0,3	0,02 ±0,2	0,00 ±0,2
	45	-0,28 ±0,3	-0,32 ±0,2	-0,31 ±0,3
	90	-0,26 ±0,2	-0,17 ±0,2	-0,12 ±0,2
	135	-0,28 ±0,3	-0,32 ±0,2	-0,31 ±0,3
Energy	**0**	0,08 ±0,1	0,05 ±0,1	0,09 ±0,1
	45	0,14 ±0,2	0,09 ±0,1	0,13 ±0,1
	90	0,08 ±0,1	0,05 ±0,1	0,09 ±0,1
	135	0,14 ±0,2	0,08 ±0,1	0,13 ±0,1
Homogeneity	**0**	0,46 ±0,1	0,45 ±0,1	0,52 ±0,1
	45	0,47 ±0,2	0,42 ±0,2	0,48 ±0,2
	90	0,42 ±0,1	0,41 ±0,1	0,50 ±0,2
	135	0,47 ±0,2	0,41 ±0,2	0,48 ±0,2

Table 1 shows the outcomes of feature extraction using statistical science, including the average value and standard deviation for the GLCM features. The purpose of this calculation was to determine the difference in feature values between each type of cervical cell image.

B. Classification Results

The classification process required system learning by extracting cervical cell image features to be classified correctly as a learning model. One of the results of the GLCM dataset determination was in the form of accuracy and running time data, as shown in the Table 3.

TABLE III CERVICAL CELL TRAINING ACCURACY RESULTS WITH SVM CLASSIFICATION

Dataset	Cubic		Quadratic		Fine Gaussian	
	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)
Run 1	96.5	7.97	76.2	7.50	96.2	5.99
Run 2	94.4	4.52	74.1	3.51	95.6	2.05
Run 3	94.7	3.99	76.7	3.51	96.6	1.48
Run 4	94.9	3.99	76.3	3.51	94.2	2.04
Run 5	95.3	4.49	76	3.51	96.6	1.84
Run 6	96.9	4.00	77.4	3.01	98.1	1.61
Run 7	95.1	4.49	76.5	3.66	97.5	3.54
Run 8	95.7	3.99	76.1	3.51	96.9	1.48
Run 9	95.8	4.50	76	4.01	96.9	3.49
Run 10	95.8	5.49	76.9	3.50	96.7	1.47
Average	95.51	4.75	76.22	3.92	96.53	2.50
STD	0.788	1.22	0.87	1.28	1.06	1.45

Table III depicts that the average accuracy value of each model is above 75%, as in the Fine Gaussian model with the highest accuracy value of 96.53% for 2.501s. Furthermore, the Cubic model reached 95.51% for 4.746s. The lowest average accuracy value from this test was still in a reasonably good percentage, namely the Quadratic model of 76.22% for 3.924s.

One of the classification results from the GLCM dataset was in ROC Graph data. It was utilized to analyze the classification learning results of data training. The ROC graph results from the results of 10x run on the Cubic SVM type obtained AUC values ranging from 0.90 – 1.00, hence categorized in an excellent classification as presented in Figure 5. Based on Figure 6, the AUC value ranged from 0.89 to 0.91. It indicates that the Quadratic SVM class classification can be diagnosed as a good classification for an AUC value of 0.89 and excellent classification at an AUC value above 0.90. The classification results on the Fine Gaussian SVM obtained AUC values of 0.90-1.0, which can be diagnosed as an excellent classification as presented in Figure 7.

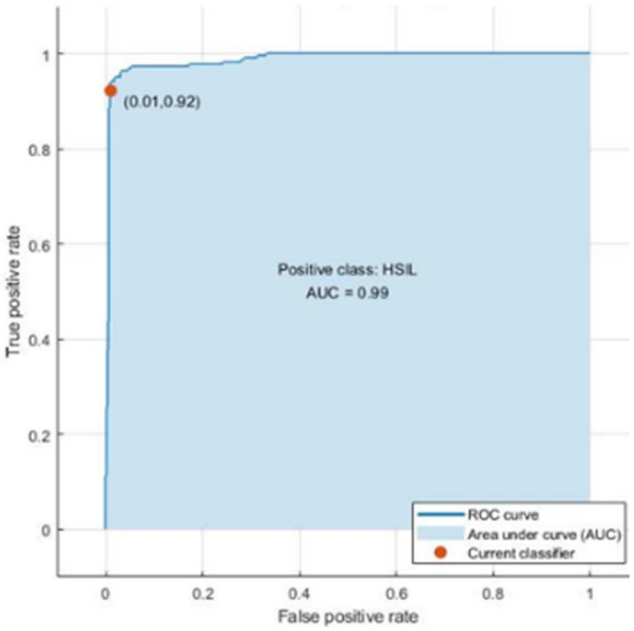


Fig. 5. ROC Graph on Cubic SVM

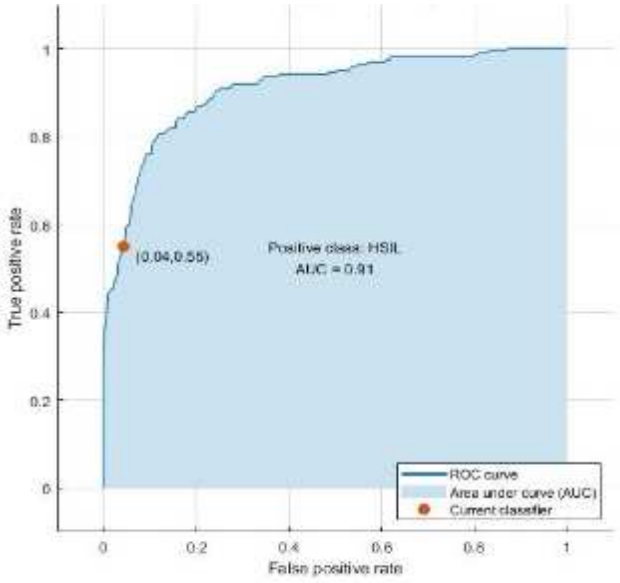


Fig. 6. ROC Graph on Quadratic SVM

C. Performance Analysis

The best performance results from testing with the GLCM feature extraction method had an accuracy value of 96.9% for 4.00s by the Cubic SVM model, 77.4% for 3.01s by Quadratic SVM, and 97.5% for 3.54s by Fine Gaussian SVM, the model with the highest accuracy.

TABLE IV PERFORMANCE ANALYSIS OF THE SVM MODEL

GLCM	Accuracy (%)	Running time (s)
Cubic	96.9	4.00
Quadratic	77.4	3.01
Fine Gaussian	97.5	3.54

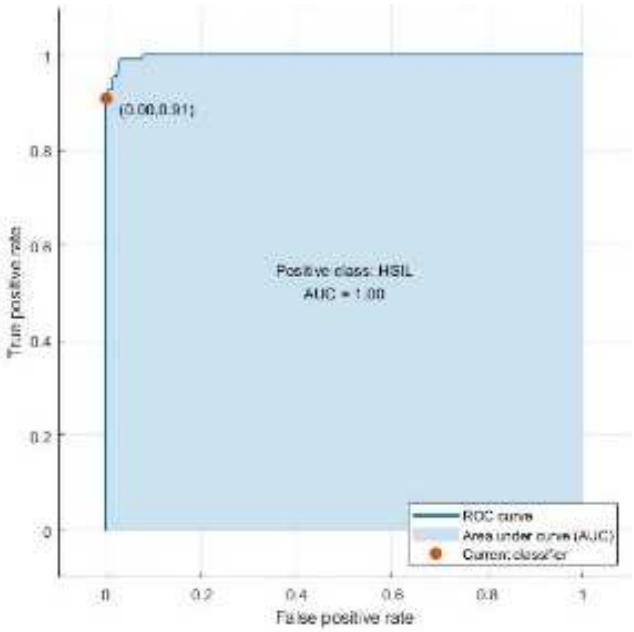


Fig. 7. ROC Graph on Fine Gaussian SVM

D. Confusion Matrix Testing Data

TABLE V CONFUSION MATRIX OF TESTING DATA CLASSIFICATION

Model	Confusion Matrix
Cubic SVM	ACTUAL: HSIL [12, 12, 0]; LSIL [6, 46, 12]; NOR [0, 17, 11]; PREDICTED: HSIL, LSIL, NOR
Quadratic SVM	ACTUAL: HSIL [8, 12, 4]; LSIL [4, 50, 10]; NOR [0, 16, 12]; PREDICTED: HSIL, LSIL, NOR
Fine Gaussian SVM	ACTUAL: HSIL [4, 20, 0]; LSIL [6, 46, 12]; NOR [0, 17, 11]; PREDICTED: HSIL, LSIL, NOR

Table V illustrates that 12 HSIL, 46 LSIL, and 11 normal images in the Cubic SVM model are appropriately classified according to their classes. However, 12 HSIL images were misclassified as LSIL, six LSIL images were misclassified as HSIL, 12 LSIL images were misclassified as Normal, and 17 Normal images were misclassified as LSIL.

The Quadratic SVM model could classify the testing images into the correct category following their classes, encompassing 8 HSIL images, 50 LSIL images, and 12 Normal images. Unfortunately, the system misclassified 12 HSIL images to LSIL, four HSIL images to Normal, four

LSIL images to HSIL, ten LSIL images to Normal, and 16 Normal images to LSIL.

Moreover, the Fine Gaussian SVM model could properly categorize four HSIL images but improperly classify 20 HSIL images to LSIL. Regarding LSIL images, 46 were appropriately classified, but six were misclassified as HSIL, and 12 became Normal. In the Normal images, 11 were appropriately categorized, but 17 were misclassified as LSIL.

IV. CONCLUSIONS

A system for classification of cervical precancerous is developed in this study. It applied a preprocessing algorithm consisting of enhancement, cropping, resizing, and augmentation techniques. The feature extraction process by the GLCM method was utilized as input in the Classification Learner app. The best SVM classification performance with the feature extraction GLCM method was 97.5% for 3.54s. It is recommended for classification system research based on image processing techniques and Support Vector Machine (SVM) for cervical cell images to add treatment at the preprocessing stage on the input images to have more optimal results. Therefore, the images can provide more optimal results at testing in term of accuracy value and time processing. Next research can improve the accuracy by proposing other algorithms.

ACKNOWLEDGMENT

This research is supported by Universitas Muhammadiyah Yogyakarta and a research project grant from the Ministry of Research and Technology of the Republic of Indonesia.

REFERENCES

[1] Organization, W.H., Global strategy to accelerate the elimination of cervical cancer as a public health problem. 2020.

[2] Wang, P., et al., Automatic cell nuclei segmentation and classification of cervical Pap smear images. Biomedical Signal Processing and Control, 2019. 48: p. 93-103.

[3] Singh, S.K. and A. Goyal, Performance Analysis of Machine Learning Algorithms for Cervical Cancer Detection. International Journal of Healthcare Information Systems and Informatics (IJHISI), 2020. 15(2): p. 1-21.

[4] Jusman, Y., S.C. Ng, and N.A. Abu Osman, Intelligent screening systems for cervical cancer. The Scientific World Journal, 2014. 2014.

[5] Lu, J., et al., Machine learning for assisting cervical cancer diagnosis: An ensemble approach. Future Generation Computer Systems, 2020. 106: p. 199-205.

[6] Chitra, B. and S.S. Kumar, Recent advancement in cervical cancer diagnosis for automated screening: a detailed review. Journal of Ambient Intelligence and Humanized Computing, 2021.

[7] Sulaiman, S.N., et al., Improvement of features extraction process and classification of Cervical cancer for the NeuralPap system. Procedia Computer Science, 2015. 60: p. 750-759.

[8] Sari, B.P. and Y. Jusman. Classification System for Cervical Cell Images based on Hu Moment Invariants Methods and Support Vector Machine. in 2021 International Conference on Intelligent Technologies (CONIT). 2021.

[9] Jusman, Y., et al., Automated cervical precancerous cells screening system based on Fourier transform infrared spectroscopy features. Journal of biomedical optics, 2016. 21(7): p. 075005.

[10] Saini, S.K., et al., ColpoNet for automated cervical cancer screening using colposcopy images. Machine Vision and Applications, 2020. 31(3): p. 15.

[11] Jusman, Y., et al., A protocol for Enhanced imaging and Quantification of Cervical Cell Under Scanning electron Microscope. International Journal of Artificial Intelligence Research, 2019. 3(2).

[12] William, W., et al., A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. Computer methods and programs in biomedicine, 2018. 164: p. 15-22.

[13] Conceição, T., et al., A review of computational methods for cervical cells segmentation and abnormality classification. International journal of molecular sciences, 2019. 20(20): p. 5114.

[14] Jusman, Y., et al., Computer-aided screening system for cervical precancerous cells based on field emission scanning electron microscopy and energy dispersive x-ray images and spectra. Optical Engineering, 2016. 55(10): p. 103110.

[15] Jusman, Y., et al., A system for detection of cervical precancerous in field emission scanning electron microscope images using texture features. Journal of Innovative Optical Health Sciences, 2017. 10(02): p. 1650045.

[16] Nehra, S., et al. Detection of cervical cancer using GLCM and support vector machines. in 2018 6th Edition of International Conference on Wireless Networks & Embedded Systems (WECON). 2018. IEEE.

[17] Usha, R. and K. Perumal, SVM classification of brain images from MRI scans using morphological transformation and GLCM texture features. International journal of computational systems engineering, 2019. 5(1): p. 18-23.

[18] Jusman, Y., et al. Feature Extraction Performance to Differentiate Spinal Curvature Types using Gray Level Co-occurrence Matrix Algorithm. in 2020 3rd International Conference on Information and Communications Technology (ICOIACT). 2020.

[19] Huang, P., et al., Classification of cervical biopsy images based on LASSO and EL-SVM. IEEE Access, 2020. 8: p. 24219-24228.

[20] Lin, D., et al., Biomedical image classification based on a cascade of an SVM with a reject option and subspace analysis. Computers in biology and medicine, 2018. 96: p. 128-140.

[21] Chandra, M.A. and S.S. Bedi, Survey on SVM and their application in image classification. International Journal of Information Technology, 2018.