Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15318
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Cervical Cancer Classification Using Multi-Directional GLCM Shape-Texture Features in LBC

Surmayanti[1)], Irohito Nozomi[2)], Febri Aldi[3)*]
[1,2,3)]Universitas Putra Indonesia YPTK Padang, Indonesia
[1)]surmayanti@upiyptk.ac.id, [2)]irohito_nozomi@ upiyptk.ac.id , [3)]febri_aldi@ upiyptk.ac.id

Submitted :Sep 7, 2025 | **Accepted** : Oct 4, 2025 | **Published** : Oct 10, 2025

Abstract: Cervical cancer is still one of the leading causes of death in women worldwide, especially in developing countries with a high prevalence of Human Papillomavirus (HPV) infection. Early detection through Pap smears and Liquid-Based Cytology (LBC) has proven to be effective in finding precancerous lesions. However, manual interpretation of cytology images is often subjective and prone to misdiagnosis, so a more objective and reliable computer-based system is needed. This study proposes a method of classification of cervical cancer on LBC-based Pap smear images by combining shape and texture features. The preprocessing stage is carried out by converting the RGB image to the L*a*b color space, then segmenting the cell nucleus using K-Means Clustering. The extracted shape features are Metric and Eccentricity, while texture features are calculated using the Gray Level Co-occurrence Matrix (GLCM) with a multi-directional and multi-distance approach. In addition to the standard features, Dissimilarity, Entropy, and Maximum Probability are added. The classification was carried out using a Support Vector Machine (SVM) with k-fold cross-validation. Testing was conducted on 550 LBC images divided into four Bethesda classes: NILM, LSIL, HSIL, and SCC. The proposed method achieved the highest recall value of 85% from the HSIL class, the highest precision of 86% from the LSIL class, and an F1-Score of 75% from the NILM class. Meanwhile, the SCC class had the lowest evaluation value due to data imbalance and similarity in morphological patterns. The results demonstrate that a combination of extended shape and texture features can improve the representation of cervical cell characteristics. Although its accuracy is lower than deep learning approaches such as ResNet50 (95%), this method is computationally lighter and has potential for application in regional hospitals with limited infrastructure. Further research is needed to explore larger datasets and compare its performance with modern deep learning models.

Keywords: Cervical cancer; LBC; GLCM; feature extraction; SVM

INTRODUCTION

Cervical cancer is one of the major health problems that threatens the survival of women worldwide (Arbyn et al., 2020). According to Global Cancer Observatory (GLOBOCAN) data released by the World Health Organization (WHO), more than 500,000 new cases of cervical cancer are recorded every year, with more than half of them leading to death (D. Singh et al., 2023). In Indonesia itself, cervical cancer ranks second after breast cancer in terms of prevalence, and is one of the highest causes of cancer death in women (Wahidin et al., 2022).

In fact, cervical cancer is classified as a type of cancer that can be prevented and detected early, especially through routine screening programs such as Pap smears (Mishra et al., 2021). In its development, the Liquid-Based Cytology (LBC) method has replaced conventional Pap smears because it is able to produce a cleaner, more even, and less contaminated cervical cell image (Strander et al., 2007). This makes LBC more ideal for digital processing and analysis using an automated approach based on computer technology (Ikeda et al., 2021).

However, the process of interpreting LBC results still relies heavily on medical personnel or cytology pathologists, who manually observe the shape and structure of cells under a microscope. This process is not only time-consuming, but also prone to subjective errors due to fatigue or limitations of experience, especially in distinguishing between precancerous lesions and early-stage cervical cancer (Díaz del Arco & Saiz Robles, 2024).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15318
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Therefore, there is an urgent need for an automated system that is able to identify and classify cervical cells accurately, objectively, and efficiently.

One promising approach in the classification of medical images is by extracting shape and texture features, which represent morphological characteristics and intensity distribution patterns in cell images (Plissiti et al., 2011). In the context of texture, the Gray Level Co-occurrence Matrix (GLCM) method has been widely used to capture spatial relationships between pixels and describe image textures statistically (Huang et al., 2014). GLCM produces a number of important features such as contrast, correlation, energy, and homogeneity that have proven useful in identifying abnormal patterns in cell structure (Chaddad & Tanougast, 2017).

However, the implementation of GLCM in previous studies has often been limited to only one direction and a certain distance, so it has not been able to fully represent the texture complexity of the LBC image (Garg & Dhiman, 2021). The texture of the nucleus of cancer cells can have a varied orientation and scale, depending on the type of lesion or stage of the disease. For this reason, a more comprehensive approach is needed, namely by developing multi-directional and multi-distance based GLCM to capture more complete and rotation-invariant texture information.

In addition, the expansion of GLCM statistical features such as the addition of entropy, dissimilarity, and angular second moment (ASM) is also believed to be able to strengthen the system's ability to distinguish diagnostic classes (Merlina et al., 2021) based on Bethesda's classification system, namely: NILM (Negative for Intraepithelial Lesion or Malignancy), LSIL (Low-Grade Squamous Intraepithelial Lesion), HSIL (High-Grade Squamous Intraepithelial Lesion), SCC (Squamous Cell Carcinoma).

Despite the rapid development of deep learning methods for medical image classification, such approaches often require very large datasets and high computational resources, which may not be available in many healthcare or research settings, particularly in developing countries. In contrast, handcrafted feature-based methods such as shape and texture analysis offer a more lightweight and interpretable alternative, making them highly relevant when data and computational capacity are limited. By combining the extraction of shape features and the development of multi-directional and multi-distance GLCM methods, this study aims to develop a classification system that is able to automatically identify cervical cell abnormalities in LBC images and assist medical personnel in early detection of cervical cancer, which can ultimately improve the effectiveness of treatment and reduce mortality rates.

LITERATURE REVIEW

The literature review discusses previous research related to the detection of cervical cancer using Pap smears and LBC. Various methods have been applied, ranging from segmentation, extraction of shape and texture features, to machine learning-based classification. This study confirms the importance of multi-feature integration to improve diagnostic accuracy.

Table 1. Previous Research

Author	Methods Used	Datasets Used	Research Results
(Merlina et al., 2021)	Color image segmentation, RGB to HSV conversion, thresholding, shape feature extraction (metric, eccentricity) and texture (GLCM), classification with K-means	Normal single-cell Pap smear image	Successfully extracts metric features, eccentricity, contrast, correlation, and energy from the cell's cytoplasm
(Raga Permana & Setiawan, 2024)	GLCM and L*a*b feature extraction, classification with Random Forest, Decision Tree, Extra Trees	Colposcopy images of CIN1, CIN2, CIN3 categories	98% accuracy, 97% sensitivity, 98% specificity with an average run time of 26.20 s using GLCM feature
(T. G. Singh & Karthik, 2023)	Segmentation, image enhancement, feature extraction, classification with SVM, Random Forest, Enhanced CNN (ECNN)	Pap smear image	ECNN achieves 92% accuracy, higher than RF (87%) and SVM (90%)
(Rastogi et al., 2023)	Binary classification using EfficientNet-B7 with empirical resolution and global pooling	3 independent Pap smear datasets (including Herlev)	94% accuracy on Herlev and up to 99% on other datasets, with heatmap visualization for interpretability

*name of corresponding author

Author	Methods Used	Datasets Used	Research Results
(Alsalatie et al., 2023)	Automatic feature extraction with AlexNet, DarkNet19, NasNet; PCA for dimension reduction; classification with SVM and RF using ALO and PSO	Pap smear image (7 classes: whole cell, cytoplasm, nucleus)	SVM+PSO achieves 99.5% accuracy, better than RF (98.9%); Focus on cell tissue (not just the nucleus) improves accuracy
(Kaur et al., 2025)	Transfer learning used 16 pre-trained CNN models (VGG16, VGG19, ResNet50, ResNet101, DenseNet, MobileNet, Inception, etc.) for direct classification of pap smear images without segmentation	Herlev and Sipakmed	ResNet50 achieves 95% accuracy for 2-class and 7-class (Herlev) classifications; VGG16 achieved 99.95% accuracy for class 2 and 5 classification (Sipakmed); DenseNet121 achieved 97.65% accuracy for the 3-class classification. Effective transfer learning for automatic classification of cervical cancer
(Attallah, 2023)	CAD model based on feature integration of many domains without segmentation, using 3 lightweight DL models for spatial feature extraction, plus statistical and texture feature extraction from spatial and time-frequency domains. PCA is used for the incorporation of DL and artificial features.	The specific name of the dataset is not mentioned, but it uses a pap smear imagery dataset	Accuracy reaches 100% with 35 main components using qualitative SVM. The combination of DL and artificial features from various domains significantly improves accuracy, more efficient than previous CAD methods

Based on the results of a review of seven previous studies, it can be concluded that most of the approaches used in the classification of cervical cancer tend to rely on deep learning-based methods, such as transfer learning with pre-trained CNN models, or a combination of deep and handcrafted features in complex CAD systems. Although these methods show high accuracy, most of these studies use conventional Pap smear datasets (such as Herlev or Sipakmed), do not explore textural features in depth, and rarely explicitly incorporate cell shape features. In addition, many of the approaches used require high computing resources and are not suitable for data and infrastructure constraints.

The research to be conducted offers a different and more efficient approach by developing a method of extraction of texture features using a multi-directional and multi-distance based GLCM, which has rarely been discussed in depth in previous studies. In addition, the study also integrates the morphological shape features of the cell nucleus, such as eccentricity, elongation, and nucleus-cytoplasm ratio, to enrich the classification information. Another uniqueness is the use of Liquid-Based Cytology-based Pap smear images which have more modern characteristics and are representative of the latest medical practices, but are still rarely used in automatic classification research. By focusing the classification on four categories according to the Bethesda system (NILM, LSIL, HSIL, and SCC), this study makes a real contribution that is more applicable and relevant to clinical needs. Thus, the novelty of the research lies in the combination of optimized classical methods, a thorough exploration of GLCM features, and its application to LBC data with a lightweight, measurable, and easy-to-implement approach.

METHOD

The Research Methods section explains the systematic steps taken, starting from image pre-processing, segmentation, extraction of shape and texture features, classification, and evaluation. Each step is structured to produce an accurate representation of cervical cells, thus supporting the process of identifying cervical cancer based on LBC images. Therefore, this research framework serves as a workflow that integrates LBC-based Pap smear image processing to identify potential cervical cancer based on a combination of shape and texture features. The design of the research framework can be seen in Figure 1.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15318
e-ISSN : 2541-2019
p-ISSN : 2541-044X

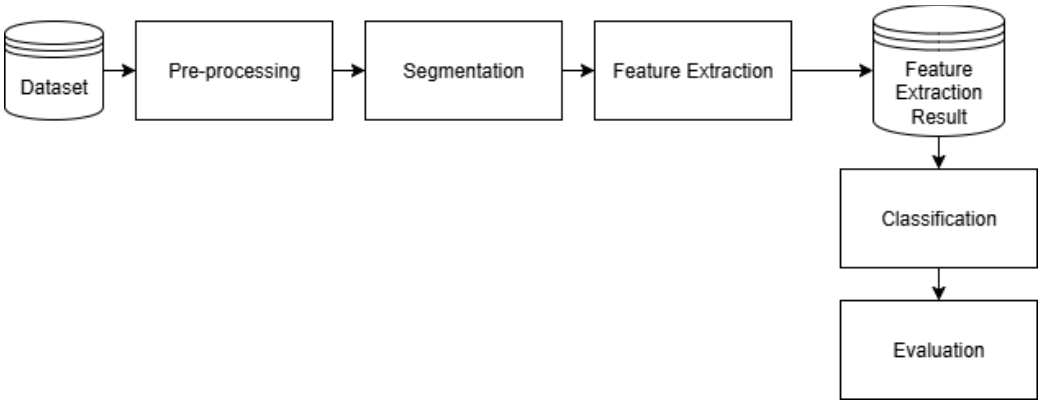


Fig. 1 Research Stages

Dataset

The dataset used in this study is a collection of LBC-based Pap Smear images in JPG format, which are classified based on the Bethesda system into four classes, namely NILM, LSIL, HSIL, and SCC. Data were obtained from the Kaggle site with a total of 550 images, consisting of 163 NILM images, 113 LSIL images, 200 HSIL images, and 74 SCC images. The entire dataset was used thoroughly in the training and testing process by applying cross-validation techniques, so that the data distribution is more balanced and the model evaluation results are more reliable.

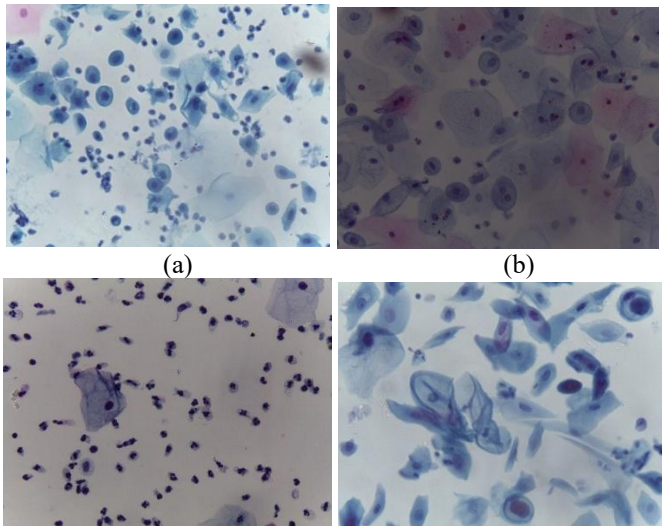


Fig. 2 LBC Pap Smear Image (a) HSIL, (b) LSIL, (c) NL, (d) SCC

Pre-processing

In the pre-processing stage, the initial image in the RGB color space is first converted to the L*a*b color space. This transformation aims to separate the brightness component (L) from the color information (a and b) so that the color differences between objects in the image can be analyzed more effectively. The L*a*b color space is considered superior in the segmentation process compared to RGB because it is more adaptive to lighting variations and is able to represent colors more consistently.

Segmentation

The next stage is segmentation using the K-Means Clustering algorithm with the number of clusters set to two. The selection of the number of clusters is adjusted with the aim of separating the area of the cell nucleus from the other parts of the image, namely the background and cytoplasm. By taking advantage of the difference in color distribution on channels a and b, K-Means can group pixels based on their similarity in color features.

After the segmentation process, a cluster label image is generated that still needs cleaning. Therefore, binarization is carried out to change the segmented image into a binary image (black-and-white), followed by hole filling to close the empty area inside the cell nucleus object that should be solid. Furthermore, the removal of small objects is performed using the bwareaopen morphology operation to eliminate noise or irrelevant small particles. This set

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

of processes ensures that the results of cell nucleus segmentation are intact, clear, and ready to use at the feature extraction stage.

Feature Extraction
Feature extraction is done on two main types of features, namely shapes and textures:

Shape Features
Shape features are extracted from the results of cell nucleus segmentation, including Metric (circularity) to measure how rounded the cell nucleus is and Eccentricity describing how oval or elliptical the shape of the cell nucleus is.

Texture Features GLCM
Grayscale imagery of the cell core is used to construct GLCM matrices in the directions of 0°, 45°, 90°, and 135°, as well as pixel spacing 1, 2, and 4. From each GLCM, texture features are calculated, including Contrast, Correlation, Energy, Homogeneity, Dissimilarity, Entropy, and Maximum Probability. The value of each feature is calculated from each GLCM and averaged to obtain a more stable and orientation-independent representation of global textures.

Feature Database Formation
All the extracted shape and texture features are compiled into a numerical database with a row size as many as the number of images, and a column as many as the number of features. This database is further used as an input for the classification process.

Classification
The feature extraction results then form a feature dataset that is used as input for the classification stage. This study uses the SVM method based on Error-Correcting Output Codes (ECOC) to handle multiclass classification. The training and testing processes were conducted using the k-fold cross-validation technique with a k value of 5 to increase the reliability of the evaluation results.

Evaluation
The final stage is the evaluation of system performance, which is carried out by calculating accuracy, precision, recall, and F1-score metrics for each class, as well as displaying a confusion matrix and ROC (Receiver Operating Characteristic) curve along with AUC (Area Under Curve) values. With this flow, the research framework is expected to be able to demonstrate the extent to which the combination of shape and texture feature extraction can improve classification accuracy in cervical cancer identification based on LBC Pap Smear images.

RESULT

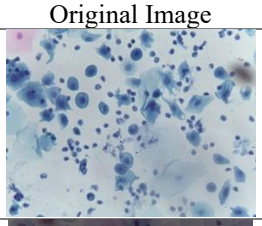
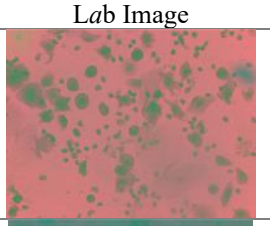
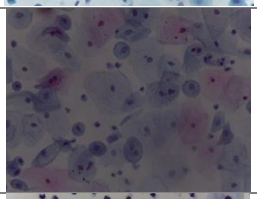
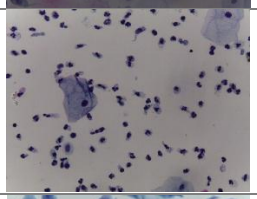
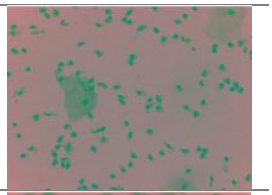
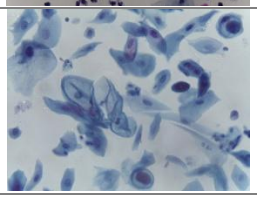
The results section presents research findings from each stage carried out, starting from image preprocessing, segmentation, extraction of shape and texture features, to test image identification. The results obtained are presented in the form of tables and analyses, to demonstrate the effectiveness of the method in classifying cervical cells.

Pre-processing
The research process begins with the pre-image processing stage, where the pap smear image is converted from RGB format to the L*a*b color space to separate the brightness and color components. This conversion makes the segmentation process easier as the color differences between cells become more pronounced. The results of the pre-processing can be seen in Table 2.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15318
e-ISSN : 2541-2019
p-ISSN : 2541-044X

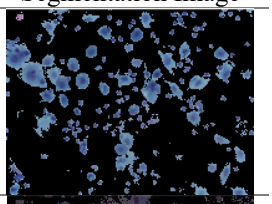
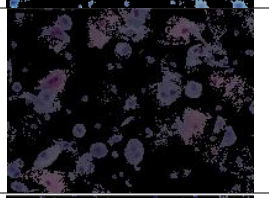
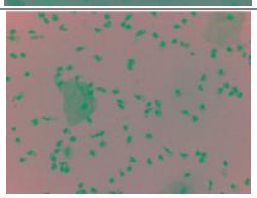
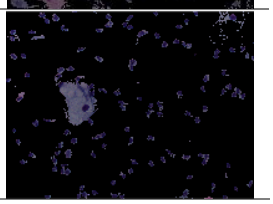
Table 2. L*a*b Images

Image Categories	Original Image	L*a*b Image
HSIL		
LSIL		
NL		
SCC		

Segmentation (K-Means Clustering)

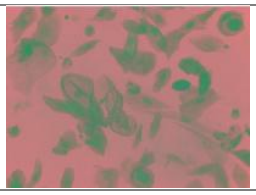
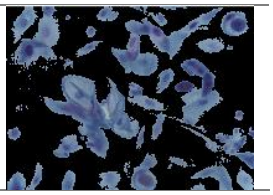
The next stage is segmentation using the K-Means Clustering method with the number of clusters of two, to separate the cell objects from the background. The results of segmentation showed that this method was able to extract cell objects well, reduce noise, and maintain the shape of the cell nucleus that was the focus of the analysis. The results of segmentation are presented in Table 3.

Table 3. Segmentation Image Results

Image Categories	L*a*b Image	Segmentation Image
HSIL		
LSIL		
NL		

*name of corresponding author

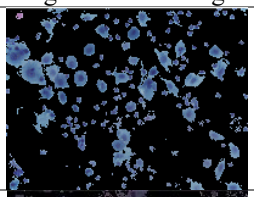
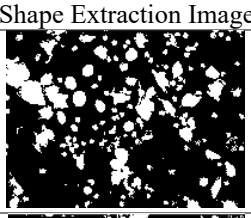
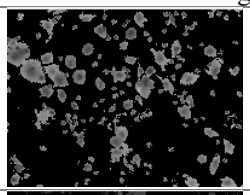
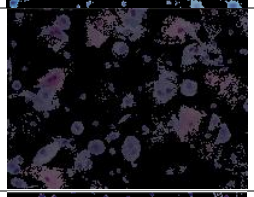
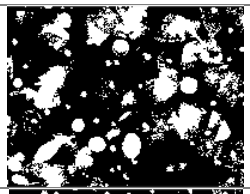
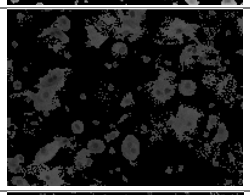
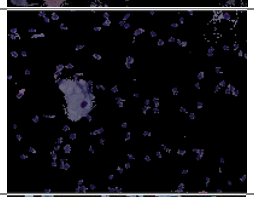
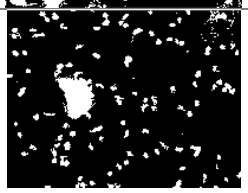
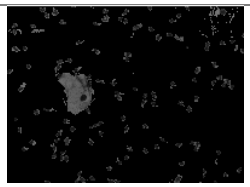
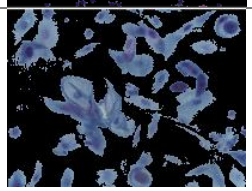
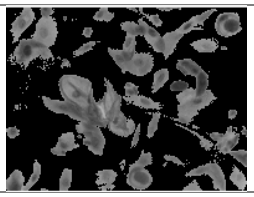
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15318
e-ISSN : 2541-2019
p-ISSN : 2541-044X

SCC		

Texture and Shape Feature Extraction

After segmentation, shape features such as Metric and Eccentricity are extracted, which represent compactness and elliptical shapes in the nucleus of the cell. Texture feature extraction using the GLCM method with the calculation of Contrast, Correlation, Energy, and Homogeneity features. This research was developed by adding Dissimilarity, Entropy, and Maximum Probability features to make texture representation richer.

Table 4. Shape and Texture Extraction Image Results

Image Categories	Segmentation Image	Shape Extraction Image	Texture Extraction Image
HSIL			
LSIL			
NL			
SCC			

Results of Database Extraction of Shape and Texture Features

The results of the calculation of all features are presented in the form of a numerical database that contains a combination of shape and texture features for each image. The values presented in Table 5 show characteristic variations between cytology categories (HSIL, LSIL, NILM, SCC), where abnormal cell categories generally have significantly different Metric and Eccentricity values than normal cells. In texture features, Contrast, Dissimilarity, and Entropy values tend to be higher in abnormal cells due to the more varied distribution of pixel intensity, while Energy and Homogeneity tend to be higher in normal cells that have more uniform patterns.

Table 5. Results of Database Extraction of Shape and Texture Features

Metric	Eccentricity	Contrast	Correlation	Energy	Homogeneity	Dissimilarity	Entropy	Maximum Probability
0.07264260	0.826761582	0.599080285	0.855687404	0.604193844	0.958327082	0.16427992	1.478638329	0.770872886
0.36789140	0.949480123	0.49348982	0.855765908	0.687543203	0.964271914	0.13778113	1.221587541	0.826016246
0.25959508	0.960706659	0.643450189	0.863785355	0.547636016	0.95207934	0.1833996	1.656950177	0.730327599
0.41374162	0.637277781	0.590760654	0.86872805	0.51371765	0.953229742	0.174003396	1.765869953	0.705338537
0.53982291	0.905284458	0.678033733	0.857127135	0.515023789	0.947147768	0.198057934	1.763474022	0.705889273
0.46845406	0.933692025	0.756005499	0.856706927	0.492809373	0.943735528	0.214173236	1.854258767	0.688694228

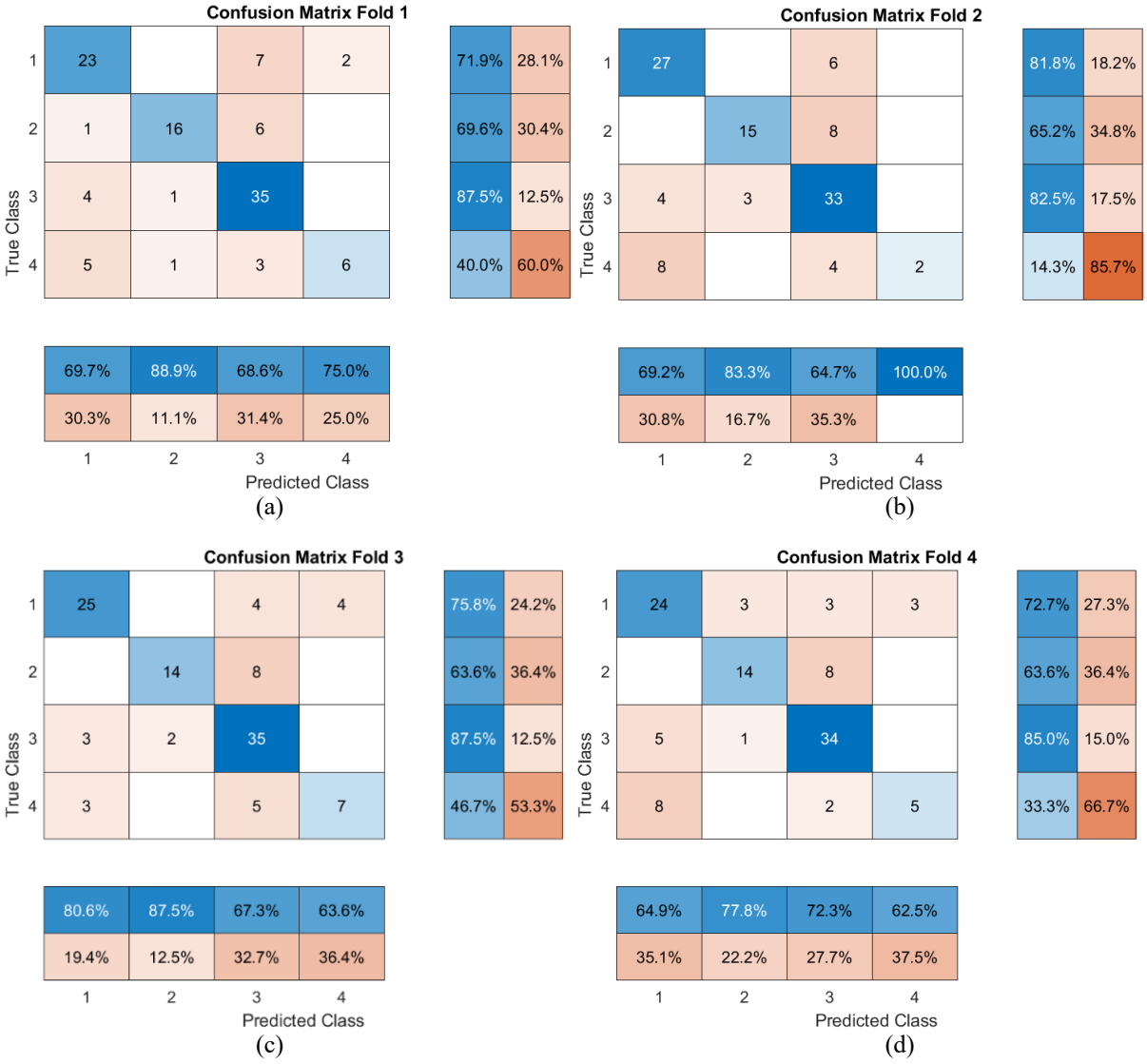
*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15318
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Metric	Eccentricity	Contrast	Correlation	Energy	Homogeneity	Dissimilarity	Entropy	Maximum Probability
0.78147397	0.787346454	0.531341047	0.862434554	0.603068166	0.956773724	0.158870177	1.502089626	0.770516273
0.62516106	0.648164923	0.535119491	0.872234986	0.662697797	0.965922579	0.138794359	1.288429254	0.809934258
0.62565980	0.87791886	0.535064704	0.862742721	0.659161073	0.96575066	0.140195415	1.298287707	0.807455156
0.62957667	0.815035761	0.542668326	0.873756517	0.660447009	0.966242669	0.139031021	1.283173296	0.808193792
0.46254364	0.927951156	0.510643401	0.886605961	0.571770684	0.962648262	0.143208556	1.580254466	0.747847536
...	...	...	...	...	...	...	...	...
0.76861862	0.75951038	0.538702219	0.859838668	0.60930661	0.957865944	0.159040436	1.462563067	0.774640771

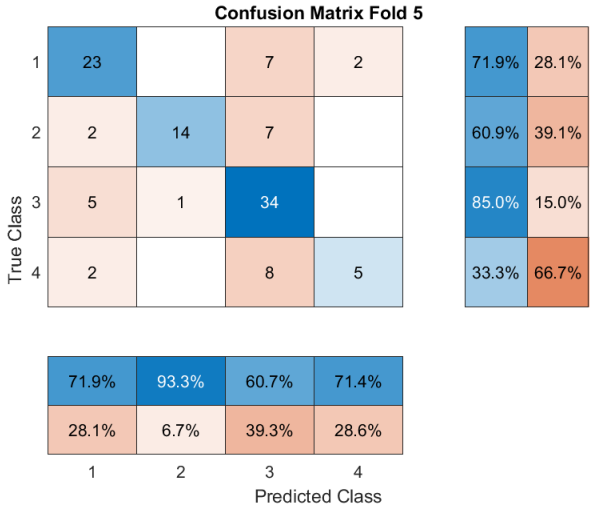
Classification

To determine the classification performance for each cervical cell category, the test results using the Support Vector Machine (SVM) method are presented in the form of a confusion matrix. This visualization shows the distribution of correct and incorrect predictions for each class: HSIL, LSIL, NILM, and SCC, thus providing a clearer picture of the model's ability to recognize the characteristic patterns of each specific class.



*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15318
e-ISSN : 2541-2019
p-ISSN : 2541-044X



(e)

Fig. 3 Confusion Matrix (a) Fold 1, (b) Fold 2, (c) Fold 3, (d) Fold 4, (e) Fold 5

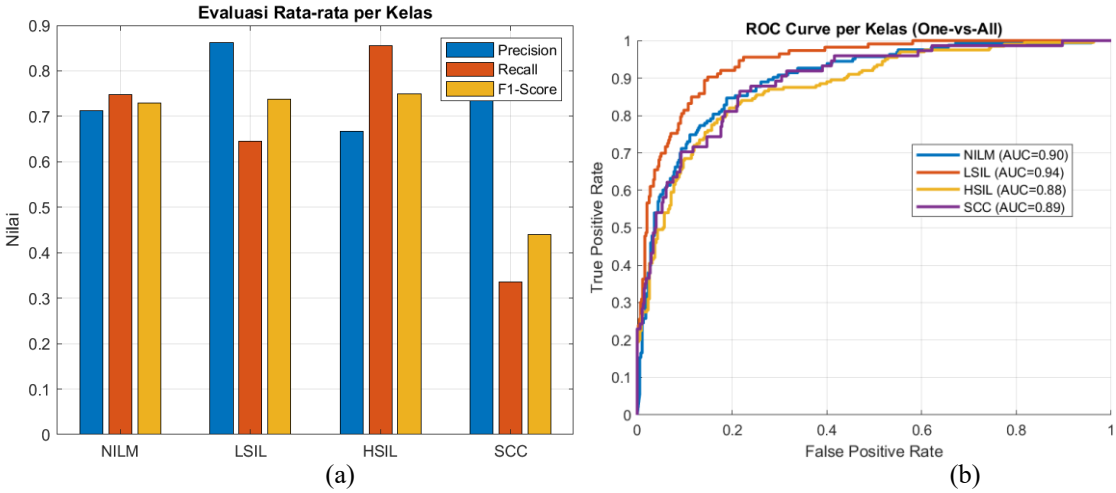
Based on the confusion matrix shown in the figure, it can be seen that the SVM model is able to recognize the HSIL class with a fairly high level of accuracy, while the performance for the NILM and LSIL classes is in the moderate category, with several prediction errors, especially for the HSIL class. As for the SCC class, the classification success rate is still low because most of the SCC images are incorrectly identified as HSIL. This indicates that although the shape and texture feature extraction method is able to improve accuracy in general, further development is needed, especially in handling classes with an unbalanced amount of data to achieve more optimal identification results.

Evaluation

To evaluate the performance of the classification model, precision, recall, and F1-score metrics were calculated for each class (HSIL, LSIL, NILM, and SCC). Additionally, the ROC (Receiver Operating Characteristic) curve and AUC (Area Under Curve) values are displayed to assess the model's ability to differentiate each class using a one-vs-all approach. The evaluation results are presented in Table 6 and visualized graphically in Figure 4.

Table 6. Results

Class	Precision	Recall	F1-Score
HSIL	0.71	0.75	0.73
LSIL	0.86	0.65	0.74
NILM	0.67	0.86	0.75
SCC	0.75	0.34	0.44



(a) (b)

Fig. 4 (a) Interclass Average Evaluation, (b) ROC Curve between Classes (One vs. All)

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Based on the evaluation results, the model performed quite well in the HSIL, LSIL, and NILM classes, with F1-scores ranging from 0.73 to 0.75. Meanwhile, the SCC class showed a low recall value (0.34), indicating that many SCC images were misclassified into other classes, despite a relatively high precision value (0.75). This indicates that the model is better at avoiding SCC prediction errors, but less sensitive in recognizing all true SCC images.

The ROC curve results showed high AUC values for all classes (0.88 to 0.94), indicating the model has good discrimination ability in distinguishing between classes. Overall, the integration of shape and texture features with the SVM method was able to provide fairly stable classification performance, although additional strategies are still needed to improve accuracy, especially for the SCC class, which has less data than other classes.

DISCUSSIONS

The results showed that converting images from RGB to L*a*b color space improved the separation of color information, facilitating the identification of cell nuclei. The advantage of the L*a*b color space lies in the separation of luminance and chromaticity, which allows for clearer detection of color variations between cells. The segmentation stage using the K-Means Clustering method successfully extracted cell nuclei, minimized background noise, and maintained the morphological shape of the nuclei. This segmentation accuracy is crucial because it forms the basis for the validity of the shape and texture features extracted in the subsequent stages.

In shape feature extraction, the Metric and Eccentricity parameters showed consistent variations between normal and abnormal cells. Abnormal cell nuclei tended to be more elongated (higher eccentricity) and less rounded (lower metric), consistent with morphological changes reported in the cervical cytology literature. Meanwhile, GLCM-based texture features (contrast, correlation, energy, homogeneity, dissimilarity, entropy, and maximum probability) enriched the representation of cell nucleus characteristics. For example, high entropy values in abnormal cells reflect greater texture complexity, while normal cells exhibit higher energy and homogeneity, consistent with Attallah's (2023) study, which emphasized the importance of texture features in cervical cytology classification.

Classification results using SVM with cross-validation showed varying performance across classes. HSIL, LSIL, and NILM demonstrated fairly good F1-scores (0.73–0.75), while the SCC class had a low recall value (0.34), indicating that many SCC images were misidentified as other classes. However, the AUC values for all classes remained high (0.88–0.94), indicating the model's discriminatory ability was quite stable despite inter-class variation.

When compared with deep learning-based studies, the accuracy of this study is still lower. For example, Kaur's (2025) study reported up to 95% accuracy using ResNet50, while Rastogi (2023) achieved similar results using transfer learning. This difference can be explained by two main factors: the smaller dataset size (550 LBC images) and the use of handcrafted features, which tend to be less flexible than automated deep learning representations. However, the advantage of this feature-based approach is its lower computational requirements, making it easier to implement in regional hospital settings with limited infrastructure (Attallah, 2023).

Practically, the results of this study demonstrate that, although its performance does not yet match that of CNN methods, the combination of shape and texture features still provides an informative representation of cell morphology. This makes it an efficient solution for initial screening applications in healthcare facilities with limited resources. With the development of larger datasets and the integration of hybrid methods (handcrafted + deep learning), the potential for improving classification accuracy remains significant.

CONCLUSION

This study successfully developed a cervical cancer identification method in LBC-based Pap smear images through a combination of shape and texture feature extraction. Pre-processing, including conversion to the Lab* color space and segmentation using K-Means Clustering, proved effective in separating cell nuclei from the background, reducing noise, and preserving nuclear shape. The extraction of shape features (Metric, Eccentricity) and texture features based on GLCM using a multi-directional and multi-distance approach, coupled with additional features (Dissimilarity, Entropy, and Maximum Probability), provided a richer and more comprehensive representation of cell nuclear morphology. Classification results demonstrated that this feature combination was able to effectively differentiate cervical cytology diagnostic categories (NILM, LSIL, HSIL, and SCC) and demonstrated potential for supporting efficient and adaptive Computer-Aided Diagnosis (CAD) systems.

Although the achieved accuracy is still below that of deep learning-based methods such as CNN or ResNet reported in previous studies, this feature-based approach offers advantages in computational efficiency and ease of implementation in healthcare facilities with limited infrastructure. For further research, testing on larger and more diverse datasets, exploration of other classification algorithms (e.g., Random Forest, Gradient Boosting, or ensemble learning), and direct comparison with deep learning approaches are recommended. Integrating handcrafted features with deep learning representations also has the potential to improve accuracy while maintaining efficiency, thus optimizing the system's support for early cervical cancer diagnosis.

*name of corresponding author

ACKNOWLEDGMENT

The author would like to thank the Putra Indonesia University YPTK Padang for providing full support in funding this research through the SIMLIT UPI Batch 5 grant of 2025. This support is very meaningful so that this research can be carried out well. The author also expressed his appreciation to all parties who have helped, both directly and indirectly, in the process of preparing this research.

REFERENCES

Alsalatie, M., Alquran, H., Mustafa, W. A., Zyout, A., Alqudah, A. M., Kaifi, R., & Qudsieh, S. (2023). A New Weighted Deep Learning Feature Using Particle Swarm and Ant Lion Optimization for Cervical Cancer Diagnosis on Pap Smear Images. *Diagnostics*, *13*(17), 2762. https://doi.org/10.3390/diagnostics13172762

Arbyn, M., Weiderpass, E., Bruni, L., Sanjosé, S. de, Saraiya, M., Ferlay, J., & Bray, F. (2020). Estimates of incidence and mortality of cervical cancer in 2018: A worldwide analysis. *The Lancet Global Health*, *8*(2), e191–e203. https://doi.org/10.1016/S2214-109X(19)30482-6

Attallah, O. (2023). Cervical Cancer Diagnosis Based on Multi-Domain Features Using Deep Learning Enhanced by Handcrafted Descriptors. *Applied Sciences*, *13*(3), 1916. https://doi.org/10.3390/app13031916

Chaddad, A., & Tanougast, C. (2017). Texture Analysis of Abnormal Cell Images for Predicting the Continuum of Colorectal Cancer. *Analytical Cellular Pathology*, *2017*(1), 8428102. https://doi.org/10.1155/2017/8428102

Díaz del Arco, C., & Saiz Robles, A. (2024). Advancements in Cytological Techniques in Cancer. In *Handbook of Cancer and Immunology* (pp. 1–46). Springer, Cham. https://doi.org/10.1007/978-3-030-80962-1_385-1

Garg, M., & Dhiman, G. (2021). A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants. *Neural Computing and Applications*, *33*(4), 1311–1328. https://doi.org/10.1007/s00521-020-05017-z

Huang, X., Liu, X., & Zhang, L. (2014). A Multichannel Gray Level Co-Occurrence Matrix for Multi/Hyperspectral Image Texture Representation. *Remote Sensing*, *6*(9), 8424–8445. https://doi.org/10.3390/rs6098424

Ikeda, K., Oboshi, W., Hashimoto, Y., Komene, T., Yamaguchi, Y., Sato, S., Maruyama, S., Furukawa, N., Sakabe, N., & Nagata, K. (2021). *Characterizing the Effect of Processing Technique and Solution Type on Cytomorphology Using Liquid-Based Cytology*. https://dx.doi.org/10.1159/000519335

Kaur, H., Sharma, R., & Kaur, J. (2025). Comparison of deep transfer learning models for classification of cervical cancer from pap smear images. *Scientific Reports*, *15*(1), 3945. https://doi.org/10.1038/s41598-024-74531-0

Merlina, N., Noersasongko, E., Nurtantio, P., Soeleman, M. A., Riana, D., & Hadianti, S. (2021). Detecting the Width of Pap Smear Cytoplasm Image Based on GLCM Feature. In Y.-D. Zhang, T. Senjyu, C. SO–IN, & A. Joshi (Eds.), *Smart Trends in Computing and Communications: Proceedings of SmartCom 2020* (pp. 231–239). Springer. https://doi.org/10.1007/978-981-15-5224-3_22

Mishra, G. A., Pimple, S. A., & Shastri, S. S. (2021). An overview of prevention and early detection of cervical cancers. *Indian Journal of Medical and Paediatric Oncology*, *32*, 125–132.

Plissiti, M. E., Nikou, C., & Charchanti, A. (2011). Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images. *Pattern Recognition Letters*, *32*(6), 838–853. https://doi.org/10.1016/j.patrec.2011.01.008

Raga Permana, Z. Z., & Setiawan, A. W. (2024). Classification of Cervical Intraepithelial Neoplasia Based on Combination of GLCM and L*a*b* on Colposcopy Image Using Machine Learning. *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 035–040. https://doi.org/10.1109/ICAIIC60209.2024.10463256

Rastogi, P., Khanna, K., & Singh, V. (2023, August 8). *Classification of single-cell cervical pap smear images using EfficientNet—Rastogi—2023—Expert Systems—Wiley Online Library*. https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.13418

Singh, D., Vignat, J., Lorenzoni, V., Eslahi, M., Ginsburg, O., Lauby-Secretan, B., Arbyn, M., Basu, P., Bray, F., & Vaccarella, S. (2023). Global estimates of incidence and mortality of cervical cancer in 2020: A baseline analysis of the WHO Global Cervical Cancer Elimination Initiative. *The Lancet Global Health*, *11*(2), e197–e206. https://doi.org/10.1016/S2214-109X(22)00501-0

*name of corresponding author

Singh, T. G., & Karthik, B. (2023). Accurate Cervical Tumor Cell Segmentation and Classification from Overlapping Clumps in Pap Smear Images. In S. N. Singh, S. Mahanta, & Y. J. Singh (Eds.), *Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology* (pp. 659–673). Springer Nature. https://doi.org/10.1007/978-981-99-1699-3_46

Strander, B., Andersson-Ellström, A., Milsom, I., Rådberg, T., & Ryd, W. (2007). Liquid-based cytology versus conventional Papanicolaou smear in an organized screening program. *Cancer Cytopathology*, *111*(5), 285–291. https://doi.org/10.1002/cncr.22953

Wahidin, M., Febrianti, R., Susanty, F., & Hasanah, S. R. (2022, March 1). *Twelve Years Implementation of Cervical and Breast Cancer Screening Program in Indonesia—PMC*. https://pmc.ncbi.nlm.nih.gov/articles/PMC9360967/

*name of corresponding author