Computer Engineering and Applications Vol. 14, No. 2, June 2025

Cervical Pre-cancer Classification Using MLP Based on Hybrid Features from GLCM, LBP, and MobileNetV2

Nugroho Suhandono[1], Siti Nurmaini[2]

Department of Computer Engineering,Faculty of Computer Science, Universitas Sriwijaya
**nugroho.suhandono@gmail.com*

ABSTRACT

The early and accurate diagnosis of cervical intraepithelial neoplasia lesions (CIN), particularly in a resource-limited environment, is paramount in helping to control the rising epidemic of cervical cancer. This research offers a hybrid classification model that merge texture features like Gray Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP), alongside semantic features from MobileNetV2. These features, after being extracted, are merged and supplied to a Multilayer Perceptron (MLP) for multiclass classification into Normal, CIN1, CIN2, or CIN3. The model was trained and evaluated using a 5-fold stratified cross-validation technique on an IARC dataset that contains 200 cases of colposcopy images. The experimental results illustrate that the model developed with a stratified k-fold cross-validation performed consistently well with high performance, average accuracy reported as 86.75% ± 2.62% and Cohen's kappa 0.7963 ± 0.0524 showed substantial to almost perfect in agreement across folds. The best performance was recorded for Fold 4 achieving 90.31% accuracy, while maintaining robust F1-scores across all classes. This hybrid approach offers a promising direction for developing efficient and accurate computer-aided diagnosis (CAD) systems for cervical lesion classification.

Keywords: cervical intraepithelial neoplasia lesions (CIN), GLCM, LBP, MobileNetV2, MLP, colposcopy images.

1. INTRODUCTION

Cervical cancer remains a major public health challenge, particularly in low-resource settings where access to regular screening and expert medical evaluation is limited [4]. Cervical precancer detection is critically important for diagnosing whether a patient has cervical intraepithelial neoplasia (CIN). Identifying CIN as early as possible can prevent cervical cancer through more timely and effective treatment. Because Cervical cancer is one of the main causes of cancer deaths among women throughout the world [1]. Therefore, preventive measures as early as possible by detecting CIN are very important to prevent it from becoming cervical cancer

CIN is classified into three stages: CIN 1, which represents mild changes and often resolves on its own; CIN 2, which indicates moderate changes and carries a risk of progressing to cancer if left untreated; and CIN 3, which signifies severe changes and poses a very high risk of developing into cervical cancer. Early detection of CIN is vital because it allows for treatment before the condition

advances to cancer. To that end, many methods both conventional and AI-based have been developed to detect CIN. Conventional approaches rely on medical professionals visually analyzing colposcopic images, a process that is highly subjective and can lead to inconsistent diagnoses. so. AI could aid clinicians in making decisions, reducing their workload as well as the likelihood of misdiagnoses [7]. Moreover, many regions lack adequate equipment for CIN detection altogether. Therefore, there has been a lot of research and development to overcome the limitations of tools to detect CIN in remote or isolated areas, replacing medical personnel who diagnose CIN through colposcopy images subjectively. The use of artificial intelligence can replace the position of medical personnel in diagnosing whether it is CIN or not objectively. Various algorithms are used to process colposcopy image data to detect CIN, including CNN, KNN, and SVM [6,8,9,10]. However, the limitations of images as a data source are still an obstacle in training this artificial intelligence to obtain better accuracy. In this paper we propose using GLCM, LBP, and mobilenetv2 as feature extraction then we classifies with multilayer perceptron for detect CIN 1, CIN 2, CIN 3, and Normal cases.

2. LITERATURE STUDY

2.1 Multiclass Classification Problem

Basically, multilayer feedforward neural networks is method for classification of multiclass problem. As multilayer feedforward neural networks, MLP can be applied in multiclass for our case are normal or CIN which has four label classes. The label representation are normal, CIN1, CIN2, and CIN3.

2.2 Previous Study

Various classification methods developed and used to solve this problem. The classification of colposcopic images encounters several difficulties that complicate the creation of automatic detection models. A key problem is the significant variability in the quality and consistency of images caused by variations in lighting, resolution, and the occurrence of artifacts like blood, mucus, or light reflections (glare) that can disrupt visual analysis. Moreover, the visual distinctions among lesion grades like CIN1, CIN2, and CIN3 are frequently quite subtle, which makes them challenging to identify, even for specialists, not to mention for automated systems. Class imbalance presents a significant issue, as data from the normal class is significantly more plentiful than that of abnormal classes like CIN2 or CIN3, leading to models being skewed towards the majority class. The labels employed are frequently subjective, depending on physicians' interpretation, and not all data is backed by histopathological confirmation.

Texture-related features like Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP) are commonly utilized to analyze the spatial and textural properties of cervical tissue, which frequently reflect lesion traits such as acetowhite epithelium, mosaic patterns, or punctation. Besides these manually crafted features, deep features obtained from pretrained convolutional neural networks like MobileNetV2 have demonstrated effectiveness in grasping high-level semantic details from the images. These characteristics, utilized separately or together, are subsequently fed into a classification model like a Multi-Layer Perceptron (MLP) to differentiate normal from abnormal cervical tissues or to categorize the severity of

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

lesions (e.g., CIN1, CIN2, CIN3). This combined method leverages the advantages of both classic texture analysis and contemporary deep learning, MLP, facilitating more precise and widely applicable predictions, particularly when dealing with restricted datasets.

3. METHODOLOGY

In this research we using feature extraction methodology such as GLCM, LBP, and mobilenetv2. Data with number not significant lead us to using image augmentation. The extraction feature output will use to classify using multilayer perceptron or MLP.

3.1 DATA

The dataset used was obtained from IARC which consists of 200 cases. The dataset consists of several classes, namely normal, CIN 1, CIN 2, and CIN 3, each of which can be seen as follows. The following is four picture of a normal colposcopy that does not have any symptoms of pre-cervical cancer.

Figure 1. Normal

And the next four colposcopy image which is classified as cervical intraepithelial neoplasia 1

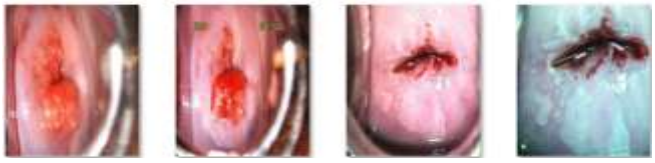

Figure 2. CIN 1

And upcoming four colposcopy image which is classified as cervical intraepithelial neoplasia 2

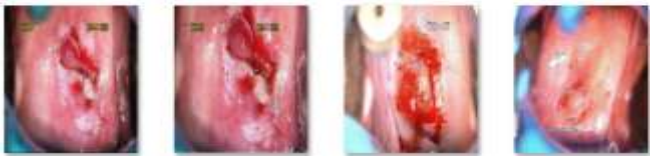

Figure 3. CIN2

The last one about four picture of a colcoscopy which is classified as cervical intraepithelial neoplasia 3

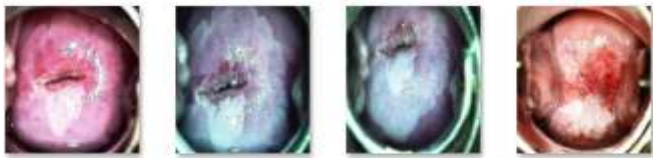

Figure 4. CIN 3

These images are part of the images that will be used as a dataset for data input according to their respective classes.

3.2 PREPROCESSING

3.2.1 IMAGE AUGMENTATION

Each image is made into 3 variations, namely the original, flipped horizontally, and rotated 90 degrees. The purpose of adding data (augmentation) is to make the model more robust against rotation and symmetry and reduce overfitting. Then the image is changed to a size of 224x224 pixels and changed to grayscale for LBP and GLCM purposes.

3.3 GLCM

GLCM is a texture-based feature extraction that is often used in the classification process [5].The Gray Level Co-occurrence Matrix (GLCM) is a statistical method used to examine the texture of an image by considering the spatial relationship between pairs of pixels. For a given image I, GLCM is defined as a matrix C(i,j) where each element represents the frequency with which two pixels with gray levels i and j occur adjacent to each other at a specified spatial offset (Δx,Δy). The matrix can be normalized so that its elements represent joint probabilities:

$$P(i,j) = \frac{C(i,j)}{\sum_{i,j} C(i,j)} \tag{1}$$

GLCM computes how often a pixel with intensity i occurs in a specific spatial relationship (e.g., distance d and angle θ) with a pixel of intensity j. Commonly used directions include 0°, 45°, 90°, and 135°, with a distance typically set to 1. By iterating through the image, a matrix of co-occurrence counts is built, typically of size p×p, where p is the number of gray levels (often reduced via quantization). The normalized GLCM matrix P(i,j) serves as a basis for calculating various second-order texture descriptors. From the GLCM, a set of texture features known as **Haralick features** can be extracted to quantify texture characteristics such as contras, dissimilarity, homogeneity, energy, correlation, ASM (angular second moment).

3.4 Local Binary Pattern (LBP)

By thresholding each pixel's neighborhood and treating the result as a binary number, the Local Binary Pattern (LBP) texture descriptor encapsulates an image's

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

local spatial structure. The LBP operator creates a binary code depending on whether each neighbor's intensity is higher or lower than the core pixel in a grayscale image by comparing each pixel with its surrounding neighbors.

3.5 MobileNetV2

MobileNetV2 is a convolutional neural network architecture proposed by Sandler et al. [2] designed for efficient on-device vision applications. It introduces two novel components: inverted residual blocks and linear bottlenecks, which significantly reduce the number of parameters and computational cost while maintaining competitive accuracy on large-scale image classification benchmarks such as ImageNet. Unlike traditional convolutional networks that expand feature dimensions progressively, MobileNetV2 utilizes an expansion depthwise projection structure in its bottleneck layers. Specifically, each block expands the input feature maps to a higher dimension using a 1×1 convolution, processes spatial information through a depthwise separable convolution, and then projects the output back to a lower-dimensional space via another 1×1 convolution. A skip connection is added when the input and output dimensions match, forming an inverted residual.

Additionally, MobileNetV2 avoids using non-linear activation (e.g., ReLU) at the projection layer, preserving valuable information in low-dimensional embeddings. The final feature map is processed via a global average pooling layer, producing a compact and discriminative 1280-dimensional feature vector. In this study, MobileNetV2 is employed as a feature extractor, where the pretrained weights on ImageNet are utilized. The top classification layer is removed (include_top=False), and the model outputs a 1280-length vector via global average pooling (pooling='avg')

3.6 Multilayer Perceptron

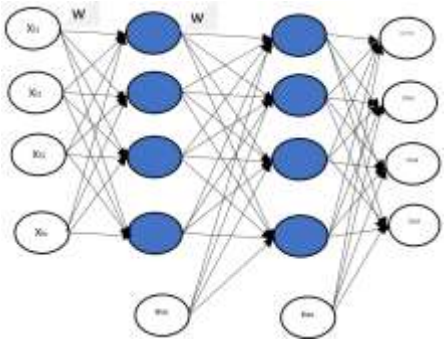

Figure 5. MLP with two hidden layer

Structure of MLP Explanation

- **Inputs (Xi1, Xi2, ..., Xin)**: These represent extracted features from images (LBP, GLCM, and MobileNetV2 features).

- **Hidden Layers (Blue nodes)**: Fully connected dense layers that learn to combine and transform features.
- **Output Layer (Normal, CIN1, CIN2, CIN3)**: A softmax layer that gives the class probabilities.
- **Weights (W)** and **Bias** nodes: Essential elements in neural network computation (learned during training).

3.7 Cross Validation

Cross-validation is a strong method for assessing a machine learning model's performance by dividing the dataset into several subsets, ensuring that the model can generalize effectively to new data. Rather than depending on one train-test division, which may lead to bias or variance based on the data's segmentation, cross-validation methodically cycles through various training and validation sets. Specifically, Stratified K-Fold cross-validation, applied in our project, divides the dataset into K equally sized folds while maintaining the class distribution in every fold — an essential aspect for imbalanced datasets such as those found in medical imaging. In every iteration, one fold serves as the validation set, while the other K–1 folds are utilized for training. This procedure is carried out K times to ensure that each data point is utilized for training and validation precisely one time. Once all folds are assessed, the outcomes are averaged to provide a more dependable estimation of the model's overall effectiveness. This method aids in reducing overfitting, lessens bias in performance measurements, and optimizes the utilization of limited data, which is especially important in fields like cervical lesion classification where labeled data may be rare. So we used K is 5 for this experiment.

3.8 Classifier Analysis

The problem when we built classifier model to predict classification issue is how we know the measurement of that model is capable to implementation. In this [3] explained that kappa statistic is one of solution to evaluate the model classifier. The scale of kappa is indicator value to represent that model is good enough or not.
Interpretation of cohen kappa K, can view in this below:

Kappa	Agreement
< 0	less than chance agreement
0.01–0.20	Slight agreement
0.21– 0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement (perfect)

4. RESULT AND DISCUSSION

This result of computation MLP with feature extraction by GLCM, LBP, and mobilenetv2. You can see below.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Trend of experimental training accuracy and validation accuracy then train loss and validation loss, in which, from top to bottom and from left to right, are training accuracy and validation accuracy of fold 1–5, same too with train loss and validation loss in right side.

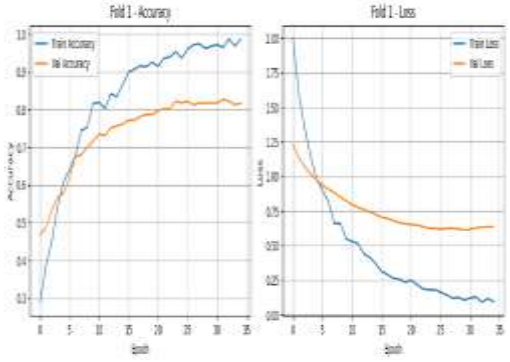

figure 6. acc and loss fold 1

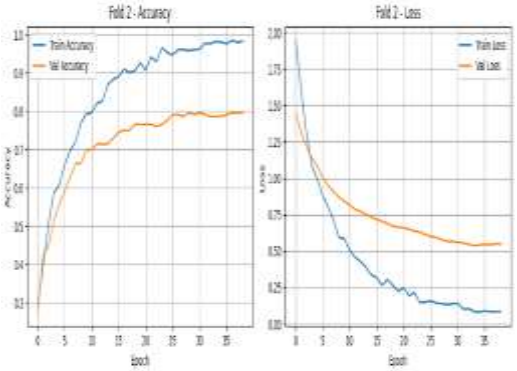

figure 7. acc and loss fold 2

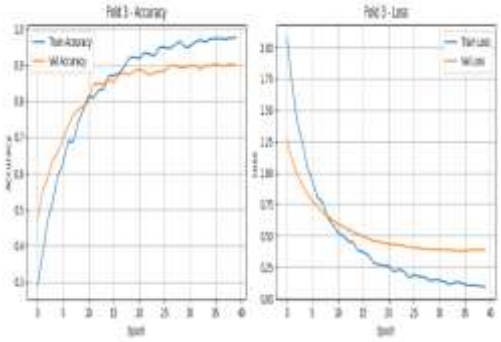

Figure 8. acc and loss fold 1

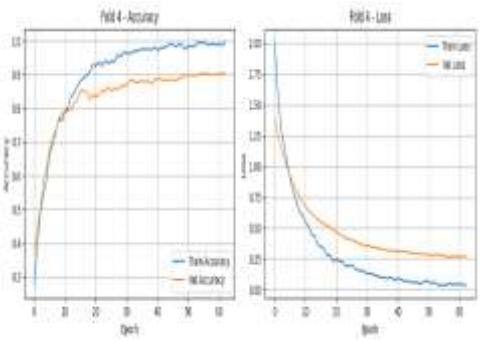

Figure 9. acc and loss fold 4

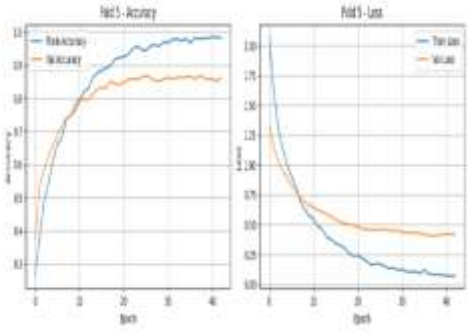

figure 10. acc and loss fold 5

Table 1. Fold 1

	precision	Recall	F1-score
CIN 1	0.82	0.82	0.82
CIN 2	0.79	0.52	0.63
CIN3	0.73	0.88	0.80
Normal	0.93	0.90	0.91

Table 2. Fold 2

	precision	Recall	F1-score
CIN 1	0.76	0.68	0.71
CIN 2	0.82	0.64	0.72
CIN3	0.81	0.96	0.88
Normal	0.89	0.89	0.89

Table 3. Fold 3

	Precision	Recall	F1-score
CIN 1	0.88	0.81	0.85
CIN 2	0.93	0.64	0.76
CIN3	0.86	0.92	0.89
Normal	0.89	0.96	0.92

Table 4. Fold 4

	precision	Recall	F1-score
CIN 1	0.88	0.95	0.91
CIN 2	0.85	0.77	0.81
CIN3	0.90	0.96	0.93
Normal	0.93	0.89	0.90

Table 5. Fold 5

	precision	Recall	F1-score
CIN 1	0.93	0.76	0.84
CIN 2	0.88	0.71	0.79
CIN3	0.92	0.90	0.91
Normal	0.84	0.96	0.89

Table 6. accuracy and cohen kappa each fold

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
accuracy	0.8376	0.8367	0.8827	0.9031	0.8776
Cohen kappa	0.7461	0.7241	0.8488	0.8514	0.8111
Degree of agreement	substantial	substantial	Perfect	Perfect	perfect

The evaluation includes four classes: CIN1, CIN2, CIN3, and NORMAL, assessed using precision, recall, f1-score, and overall accuracy metrics.

In Fold 1, the model achieved an accuracy of 83.76%, with the highest performance in the NORMAL (f1-score 0.91) and CIN1 (0.82) classes, while the weakest performance was observed in the CIN2 class with an f1-score of 0.63, primarily due to a low recall of 0.52. This indicates that many CIN2 cases were missed, leading to a high false negative rate. Nevertheless, the precision for CIN2 remained relatively good at 0.79, suggesting that while predictions were correct when made, the model struggled to capture all actual CIN2 cases. The model achieved cohen kappa 0.7461 it is means degree of agreement substantial.

In Fold 2, the accuracy slightly improved to 83.67%, with a significant boost in the CIN3 class, which achieved a high f1-score of 0.88 due to an impressive recall of

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

0.96. CIN1 and CIN2 still showed relatively lower f1-scores (0.71 and 0.72), indicating that the model continued to struggle in distinguishing lower-grade lesions. The NORMAL class maintained a consistent performance with an f1-score of 0.89, reinforcing the trend that this class is the easiest for the model to recognize. The model achieved cohen kappa 0.7241 it is means degree of agreement substantial.

In Fold 3, the model experienced a notable performance increase, reaching an accuracy of 88.27%. F1-scores for all classes improved, with the highest values found in the NORMAL (0.92) and CIN3 (0.89) classes, demonstrating the model's strong ability to classify normal and high-grade lesion cases. The precision for CIN2 reached 0.93, but its recall remained low (0.64), again highlighting the imbalance between the model's detection capability and prediction confidence for CIN2. The model achieved cohen kappa 0.8488 it is means degree of agreement almost perfect.

Next, Fold 4 achieved the best results with the highest accuracy of 90.31%. In this fold, the model performed consistently well across all classes. F1-scores for CIN1, CIN3, and NORMAL all exceeded 0.90, with CIN3 reaching a recall of 0.96. This indicates that the model was able to capture classification patterns very effectively, even for the previously challenging CIN2 class, which attained a respectable f1-score of 0.81. This fold illustrates the model's maximum potential in multi-class cervical lesion classification. The model achieved cohen kappa 0.8514 it is means degree of agreement almost perfect or perfect.

In Fold 5, the model recorded an accuracy of 87.76%, with the highest precision seen in CIN1 (0.93) and CIN3 (0.92). The best recall was observed in the NORMAL class (0.96), while CIN2 once again demonstrated the weakest performance in terms of recall (0.71) and f1-score (0.79). This reaffirms the common challenge in detecting CIN2, which, even in clinical practice, is a transitional category that is often difficult to distinguish visually and texturally. The model achieved cohen kappa 0.8111 it is means degree of agreement almost perfect or perfect.

Overall, the model achieved an average accuracy of 86.75% ± 2.62% and average cohen kappa 0.7963 ± 0.0524 (substantial), indicating good stability and consistency across folds. The average macro f1-score approached 0.85, reflecting a balanced performance across all classes despite minor fluctuations, particularly in CIN2. A general pattern across the five folds suggests that the model is highly reliable in recognizing advanced lesions (CIN3) and normal conditions but needs improvement in detecting intermediate lesions like CIN2, which are clinically known to be transitional stages between mild and severe abnormalities.

In the end, this model demonstrates strong potential for use as a clinical decision support system based on colposcopic images, with further refinements needed, particularly in learning better representations for minor or ambiguous classes.

Strategies such as data balancing, class-specific augmentation, or ensemble methods could be considered to further enhance the model's generalization and robustness.

5. CONCLUSION

The five-fold cross-validation results demonstrate that the model achieves consistently high performance in classifying cervical lesion images, with an average accuracy of 86.75% ± 2.62% and an average cohen kappa 0.7963 ± 0.0524 (substantial almost perfect agreement). Across all folds, the model shows strong capability in identifying NORMAL and CIN3 classes, indicated by high precision, recall, and f1-scores. However, the model consistently struggles with the CIN2 class, which presents the lowest recall and f1-score values in most folds. This suggests that CIN2 remains the most challenging class to classify accurately, likely due to its intermediate and overlapping features with other lesion grades. In summary, the best performance is observed in Fold 4, with an accuracy of 90.31%, showing that the model can achieve excellent classification under optimal conditions.

ACKNOWLEDGEMENTS

We would like to thank to IARC for their support by providing the data image case servical precancerous with total data are 200 cases and Isys research group to focus research in precancerous cervical.

REFERENCE

[1]. Jorge Aguilar-Santiago, José Trinidad Guillen-Bonilla, Mario Alberto García-Ramírez, Maricela Jiménez-Rodríguez. Identification of Lacerations Caused by Cervical Cancer through a comparative study among texture-extraction techniques. Applied Sciences 13 (14), 8292, 2023.
[2]. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 4510–4520.
[3]. Anthony J. Viera, MD; Joanne M. Garrett, PhD. "Understanding Interobserver Agreement: The Kappa Statistic". Fam Med 2005
[4]. Siti Nurmaini, Firdaus Firdaus, Patiyus Agustiyansyah, Annisa Darmawahyuni Anggun Islami, Akhiar Wista Arum, Muhammad Naufal Rachmatullah, Bambang Tutuko, Ade Iriani Sapitri, Rizal Sanif, Irawan Sastradinata, Legiran Legiran, "Robust assessment of cervical precancerous lesions from pre- and post-acetic acid cervicography by combining deep learning and medical guidelines," informatic in medicine, vol 52, 2024
[5]. Dian Candra Rini Novitasari, Ahmad Hanif Asyhar, Muhammad Thohir, Ahmad Zaenal Arifin, Hanimatim Mu'jizah, Ahmad Zoebad Foeady, "Cervical Cancer Identification Based Texture Analysis Using GLCM-KELM on Colposcopy Data," IEEE, 2020

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[6]. Shalini Nehra, Jagdish Lal Raheja, Kaustubh Butte, Ameya Zope, "Detection of Cervical Cancer using GLCM and Support Vector Machines," IEEE, 2018.
[7]. Leila Allahqoli, Antonio Simone Laganà , Afrooz Mazidimoradi, Hamid Salehiniya, Veronika Günther, Vito Chiantera, Shirin Karimi Goghari, Mohammad Matin Ghiasvand, Azam Rahmani, Zohre Momenimovahed and Ibrahim Alkatout, "Diagnosis of Cervical Cancer and Pre-Cancerous Lesions by Artificial Intelligence: A Systematic Review," diagnostics, 2022.
[8]. Mercy Nyamewaa Asiedu, Anish Simhal, Usamah Chaudhary, Jenna L. Mueller, Christopher T. Lam, John W. Schmitt, Gino Venegas, Guillermo Sapiro, Nimmi Ramanujam, "Development of algorithms for automated detection of cervical pre-cancers with a low-cost, point-of-care, Pocket Coposcope," IEEE, 2017.
[9]. Gengyou Peng, Hua Dong, Tong Liang, Ling Li, Jun Liu, "Title Diagnosis of cervical precancerous lesions based on multimodal feature changes," Computers in Biology and Medicine, 2021.
[10]. Peng Chen, Fobao Liu, Jun Zhang, Bing Wang, "MFEM-CIN: A Lightweight Architecture Combining CNN and Transformer for the Classification of Pre-Cancerous Lesions of the Cervix," IEEE, Engineering in Medicine and Biology, 2024.