

Segmentation of Squamous Columnar Junction on VIA Images using U-Net Architecture

Akhlar Wista Arum¹, Siti Nurmaini^{2*}, Dian Palupi Rini¹, Patiyus Agustiansyah³, Muhammad Naufal Rachmatullah²

¹*Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya, Indonesia*

²*Intelligent System Research Group, Universitas Sriwijaya, Indonesia*

³*Department of Obstetric and Gynecology, Division Oncology of Gynecology, Faculty of Medicine, Universitas Sriwijaya, Indonesia*

*wistaarum16@gmail.com

ABSTRACT

Cervical cancer is the second most common cancer that affects women, especially in developing countries including Indonesia. Cervical cancer is a type of cancer found in the cervix, precisely in the Squamous Columnar Junction (SCJ). Early screening for cervical cancer can reduce the risk of cervical cancer. One of the popular screening tool methods for the detection of cervical pre-cancer is in the Visual Inspection with Acetic Acid (VIA) method. This is due to the level of effectiveness, convenience, and low cost. VIA examination is done by applying 3-5% acetic acid to the cervical area. After applying acetic acid, a lesion called Acetowhite (AW) will be seen. AW is a precancerous lesion surrounding the SCJ. Early screening of the SCJ region will facilitate the detection of AW lesions in future studies. This method detect and segment unstructured and small patterns even though the amount of data is limited with good accuracy results. This paper proposes a method for the detection and segmentation of the SCJ region on VIA images using U-Net. This study is the first research conducted using the CNN method to perform segmentation tasks in the SCJ region. The proposed method is applied to nine different models. By using number of filter variations such as default, downfilter and upfilter and post processing methods Fix Threshold and Otsu Threshold are tested to find the best performance results. The best performance results was achieved by U-Net upfilter architecture with 90.86%, 56.5%, 75.69%, 34.09%, 41.24%, and 56.91% for Pixel Accuracy, Mean IoU, Mean Accuracy, Dice coefficient, Precision, and Sensitivity respectively. It is hoped that in the future, an easy, inexpensive, and effective automatic cervical pre-cancer detection device can be developed and accessed widely.

Keywords: Cervical Pre-cancer, Screening VIA, SCJ, U-Net.

1. INTRODUCTION

Visual Inspection with Acetic Acid (VIA) is one of the early cervical cancer screening methods recommended by WHO for developing countries. Currently, VIA is a popular screening tool for early detection of cervical cancer in developing

**Akhiar Wista Arum, Siti Nurmaini, Dian Palupi Rini,
Patiyus Agustiansyah, Muhammad Naufal Rachmatullah**
**Segmentation of Squamous Columnar Junction on VIA Images
using U-Net Architecture**

countries because of its effectiveness, convenience, and low cost [1], [2]. However, the VIA examination is subjective. Due to it depends on the experience and expertise of the operator to interpret the examination results [3].

VIA examination is done by applying 3-5% acetic acid to the cervical area [4][5]. Epithelial tissue that has been smeared with acetic acid will then turn white, called Acetowhite (AW). AW lesions are cervical pre-cancerous lesions detected during the examination. The basic characteristic of AW lesions is that they are always in the SCJ (Columnar junction) region [6], so identification of the CNS is very important to determine whether this tissue is an AW lesion or not.

Direct identification of SCJ and AW lesions is very difficult due to the similarities between AW lesions and other types of cervical disease [7]. This impediment can lead to misdiagnosis and therapy. Therefore, it is necessary to build a computer assistant to overcome this problem. With the automation of cervical pre-cancer early detection, it is hoped to help the diagnostic become more accurate and objective.

Many studies have been carried out on the automation of detection and classification of cervical pre-cancer. Several studies have focused on detection and classification based on risk factors present in the cervix [8]. Another study used image data with several methods of early screening for cervical cancer, including the Pap smear method [9]–[11], HPV test [12], and using images from colposcopy [13]–[15]. Where the method cannot be fully applied in developing countries due to limited resources and expensive costs [2], [16].

Only a few studies were focused on detecting and analyzing cervical pre-cancerous lesions using the VIA test image. Research [5], Jun Liu et al, segmented the Acetowhite region (Pre-cancerous lesions) using two clustering algorithms, namely Chan-Vese Level Set Algorithm (CV-LSA) and Modified Chan-Vese Level Set Algorithm (MCV-LSA). The performance results displayed are sensitivity, specificity, and Jaccard index (JI) with the values of 89.13%, 89.31%, and 75.81%, respectively. However, there are still many False-Negatives detected.

Further research, Kudva et al conducted an android-based study to screen for cervical cancer [17]. Kudva et al used machine learning methods to generate models on android devices. The results are quite good, with an average performance of more than 97%. However, to be implemented on android devices with k-fold validation k90 will require more time. Moreover, the detected Hausdorff distance is still large, so that it is less efficient to be applied to android devices.

Other studies continue to be carried out [3], [18]. However, due to various limitations, there are no robust performance results that can be applied as a cervical cancer screening tool, particularly in developing countries. One of such limitation is the small amount of data. So, it is necessary to build a detection automation model that can be applied even though the amount of owned data is small. The possible method is to use the Deep Learning (DL) method.

The DL method based on Convolutional Neural Network (CNN) has been known to make a diagnosis in the medical field quickly and more accurately when compared to humans. In addition, the small amount of data is not a problem in implementing the CNN method. One of the CNN methods that is widely used in medical diagnosis is U-Net [19]–[21]. U-Net is one of the most widely used segmentation methods. It can detect and segment unstructured and small patterns even though the amount of data is limited with good accuracy results [22], [23].

The proposed research focuses on automatic segmentation of the SCJ region using the U-Net architecture because it is assumed that the location of cervical pre-cancerous lesions is in contact with SCJ. By detecting and segmenting SCJ in early research, it will be easier to build a pre-cancer lesion detection model in future studies.

2. MATERIAL AND METHODS

There are 4 parts to the SSK segmentation process: a. Data Preparation, b. Architecture U-Net, c. Post Processing and d. Performance and Evaluation. Process detail on each of the stages described in the following subsection:

2.1 DATA PREPARATION

The data used in this study were obtained from an Indonesian hospital in Palembang with 444 images. The cervical image taken is an IVA image, which is the condition of the cervical region after drops of acetic acid so that the color changes can be seen. Furthermore, the image with the visible SSK was labeled by the expert as ground truth. Furthermore, the labeled raw data is carried out by pre-processing the data. At this stage, a specular reflection (SR) process will be carried out to eliminate the white bias caught during image capture due to the wet cervical region [4].

Images that have gone through the data pre-processing process are then ground-truth. Ground-Truth is used as an annotation of data in the training model process to calculate the performance results of the segmentation process carried out. Figure 1 shows the pre-processed image (a) and an example of the resulting raw data and Ground-Truth image (b).

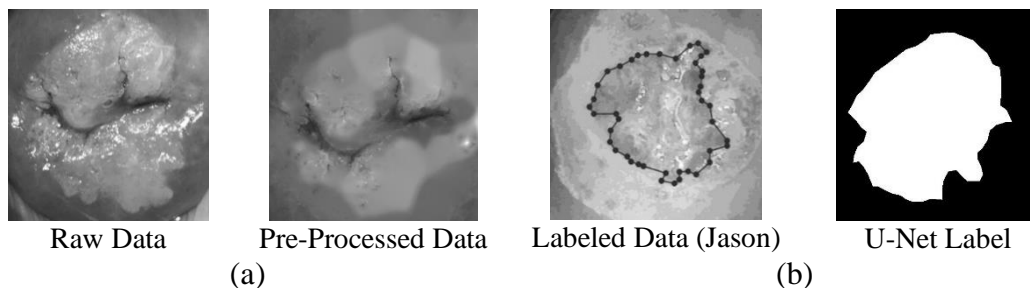


FIGURE 1. (a) Pre-Processed Data, (b) Ground-Truth Result

2.2 U-NET ARCHITECTURE

The U-Net-based CNN method is one of the semantic segmentation algorithms that can specifically find the detection object accurately. U-Net architecture is proposed as a segmentation method in this study because it has been proven to be able to complete segmentation tasks in the medical field well, including cardiac fetal

**Akhiar Wista Arum, Siti Nurmaini, Dian Palupi Rini,
Patiyus Agustiansyah, Muhammad Naufal Rachmatullah**
Segmentation of Squamous Columnar Junction on VIA Images
using U-Net Architecture

segmentation [24], liver variety segmentation [19], and Brain Tumor segmentation [20]. U-Net is an end-to-end fully convolution network type architecture that contains a convolution layer without a fully connected (dense) layer.

U-Net architecture generally has two layers, namely the convolution layer and the pooling layer. The convolution layer processes the matrix value (kernel or filter) and then changes it based on the filter's value. The process in the convolution layer is described in (1):

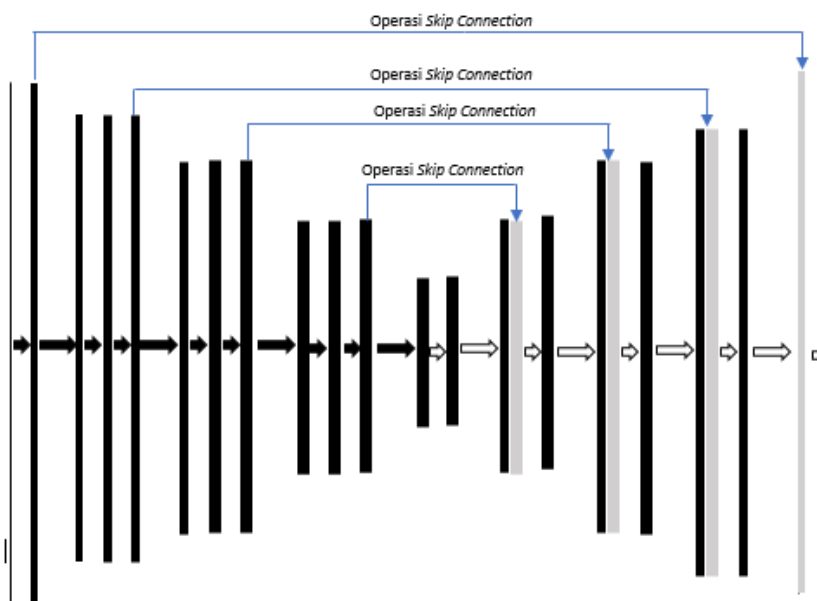
$$G[m, n] = (f * h)[m, n] = \sum_j \sum_k h[j, k] f[m - j, n - k] \quad (1)$$

where $[m, n]$ is the input images, and f is a filter. For calculating the process of convolution operator by using (2):

$$n_{out} = \left\lfloor \frac{n_{in} + 2p - k}{s} \right\rfloor + 1 \quad (2)$$

where n_{out} is the output features, n_{in} The input features, k a convolution kernel size, p a convolution padding size, and s a convolution stride size. The pooling layer used by the U-Net model is max pooling. The function of this pooling layer is to reduce the size of the map feature so that there are fewer parameters. The max-pooling process selects the maximum pixel value from the features map and is aggregated into a single features map. One of the important processes in the pooling layer is reducing the image resolution to low-resolution images.

The U-Net architecture is divided into two main parts: the encoder/contraction path and the decoder/symmetric expansion path. Each part has different functions, such as; the encoder or contraction function is to capture the context contained in the images. In contrast, the decoder or symmetric expanding path functions to localize objects using convolution transformation. The symmetric expanding path stage includes the up-sampling operation of the result of the contraction path. Figure 2 is the standard U-Net architecture used in this study.



- ➡ : 3 x 3 2D convolution + ReLU (pre-training)
- ➡➡ : 2 x 2 Max Polling
- ↻ : 3 x 3 2D convolution + ReLU
- ↻↻ : 3 x 3 Transpose 2D convolution (2 shit) + ReLU

FIGURE 2. Standard Architecture U-Net

Figure 2 illustrates the standard UNet architecture used in this study. In addition this study also examines the effect of the number of filters on the segmentation results. The number of filters in each convolution layer will be increased and also reduced. The number of filters in each convolution layer is shown in Table 1.

TABLE 1.
Architecture U-Net.

Layer Name	Number of Filter		
	Default	Down Filter	Up Filter
Conv Layer 1	64	8	128
Conv Layer 2	128	16	256
Conv Layer 3	256	32	512
Conv Layer 4	512	64	1024
Conv Layer 5	1024	128	2048
Conv Layer 6	512	64	1024
Conv Layer 7	256	32	512
Conv Layer 8	128	16	256
Conv Layer 9	64	8	128
Conv Layer 10	1	1	1

After the U-Net model is successfully built, the loss function value is calculated to determine how good the model is at making predictions. If the model's prediction is closer to the true value, the loss value will show the minimum value. Otherwise, the loss value will be maximum if the prediction is really far from the original value. In this study, the loss function used is binary cross-entropy.

Binary Cross Entropy, also known as Log loss, is a loss function commonly used to predict two-class (binary) problems. Binary Cross Entropy is the negative mean of the log of the corrected probability prediction [8]. Mathematically, binary cross entropy can be calculated by equation (3):

$$\text{Log loss} = \frac{1}{N} \sum_{i=1}^N - (y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)) \quad (3)$$

where P_i is the probability of class 1, and $((1 - p_i))$ is the probability of class 0.

2.3 POST-PROCESSING

Post-processing is a step to improve the accuracy of the segmentation process. The post-processing stage makes each image pixel is to 0 for the background and 1 for the foreground. The two best post-processing methods [24] were compared to obtain optimal results in SSK segmentation. The post-processing method used is global thresholding (fixed thresholding), and Otsu thresholding. Fix Thresholding and Otsu can be used to make segmentation results more precise.

Fix thresholding has a threshold value of 127, meaning that if the pixel value is below 127, it will be converted to 0 (background/black) and the pixel value above 127 will be converted to 1 (foreground/white). Otsu thresholding involves iterating through all threshold values. It aims to get the minimum variance from foreground and background pixel distribution.

2.4 PERFORMANCE AND EVALUATION

The results of the model testing will be validated using several performance metrics, including pixel accuracy (PA), Mean intersection over union (MIoU) and dice score or F-1 score [40]. PA is calculated based on the ratio between the number of correctly classified pixels and the number of pixels. The PA is shown in (4):

$$PA = \frac{\sum_{x=1}^{N_{cls}} N_{xx}}{\sum_{x=1}^{N_{cls}} \sum_{y=1}^{N_{cls}} N_{xy}} \quad (4)$$

where N_{cls} is number of class and N_{xy} is number of pixels in x class that were predicted as y class. The Mean Intersection over Union (MIoU) is also known as the jaccard index. This metric is used to calculate the intersection percentage between the labeled mask and the predicted output. The IoU is calculated from the grades of each class and then the overall grades are averaged. IoU metrics are very effective and very straightforward. MIoU is described in (5):

$$MIoU = \frac{1}{N_{cls}} \sum_{x=1}^{N_{cls}} \frac{N_{xx}}{\sum_{y=1}^{N_{cls}} N_{xy} + \sum_{y=1}^{N_{cls}} N_{yx} - N_{xx}} \quad (5)$$

The Dice coefficient also known as F1-Score is a measured used in the evaluation of segmentation and semantic segmentation. It is given by (6):

$$DSC = \frac{2TP}{2TP+FP+FN} \quad (6)$$

Where TP is the number of true positive, FP is the number of false positive, and FN is the number of false negative. An dice coefficient reaches its best value at 1 and the worst score at 0. Overall, the stages of the method used in this study can be seen in Figure 3.

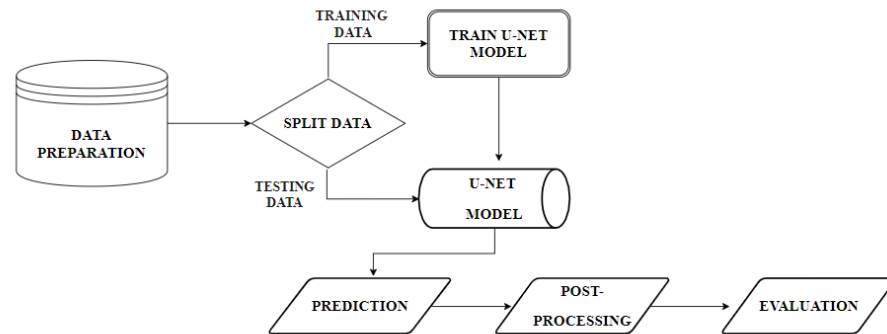


FIGURE 3. Flowchart Diagram

3. RESULT AND DISCUSSION

In this study, 3 segmentation models were built using the U-Net architecture, namely the default U-Net model, the U-Net Down-filter model and the U-Net Up-filter model. Up-filter UNET architecture was constructed assuming that the more filters used, the more/diverse features produced but the training time will be longer. On the other hand, if a few filters are used, the training time required will be short but the resulting features are limited.

The results of the three models were tested using 3 different cases, namely without using post processing and using the fix method and Otsu thresholding as the post-processing stage. The fix threshold method uses a value of 127, while the Otsu method uses the threshold value by drinking variations from intra-class. In general, the hyperparameters on the UNet architecture use a loss function binary cross-entropy and the Adam optimizer optimization method. The learning rate value is determined at 10^{-5} , the epoch is 100 iterations and the batch size value is 64.

From the experimental results, it is known that the performance of the U-Net model on the default architecture, up and down filters is obtained around 11-25% for Pixel Accuracy (PA), 5-13% for Mean IoU (M-IoU), 43-53% for Mean Accuracy (MA) and 9-11% for Dice Score (DCS). The performance of the U-Net model without post-processing stages gives unsatisfactory results. This is because the prediction results of segmentation were blurry (See Figure 4), so it is necessary to add a post-processing method. The post-processing method used is Otsu and Fix Threshold and is carried out after the prediction results are obtained.

**Akhiar Wista Arum, Siti Nurmaini, Dian Palupi Rini,
Patiyus Agustiansyah, Muhammad Naufal Rachmatullah**
Segmentation of Squamous Columnar Junction on VIA Images
using U-Net Architecture



FIGURE 4. (a) Groundtruth Label dan (b) Result of Model U-Net no post-processing

The U-Net model which was built with added post-processing method gave better results. The results obtained have increased compared to the U-Net model without post-processing. The segmentation prediction results look clearer with cleaner black-and-white image patterns and pixels. The comparison of the predicted results of the U-Net no post-processing segmentation, Otsu threshold and fix threshold can be seen in Table 2. The performance results in the U-Net Default, Up and Down Filter model experiments are obtained around 89-91% for PA, 49-56 % for MIoU, 72-84% for MA and 9-34% for DCS. More detailed performance results can be seen in Table 3.

TABLE 2.
Comparison of segmentation prediction results


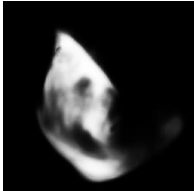


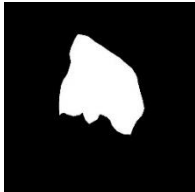
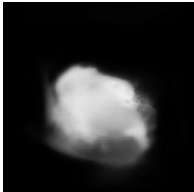
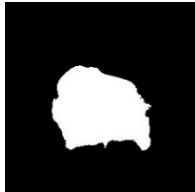

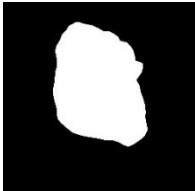
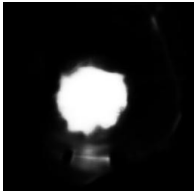


Groundtruth Label	U-Net No Post-Processing	U-Net Otsu Threshold	U-Net Fix Threshold
			
			
			

TABLE 3.
Result U-Net Model.

Case	Label	Pixel Accurac y	Mean IoU	Mean Accuracy	Precision	Recall	Dice Coefficient
Unet_Default_Model_Epo chs-100_Batch-64	No Post	13.68	7.16	44.25	5.42	79.33	9.48
Unet_Default_Model_Epo chs-100_Batch-64	Fix Threshold	91.54	56.32	73.59	43.36	51.76	32.82
Unet_Default_Model_Epo chs-100_Batch-64	Otsu	89.73	55.76	77.28	36.98	61.56	33.39
Unet_Downfilter_100_Bat ch_64	No Post	24.55	13.43	53.46	6.83	86.54	11.78
Unet_Downfilter_100_Bat ch_64	Fix Threshold	90.09	55.86	76.81	34.52	60.5	33.56
Unet_Downfilter_100_Bat ch_64	Otsu	79	49.19	84.94	22.11	91.17	30.53
Unet_Upfilter_100_Batch _64	No Post	11.15	5.79	43.19	5.27	79.82	9.22
Unet_Upfilter_100_Batch _64	Fix Threshold	91.28	55.58	72.91	44.78	50.61	30.96
Unet_Upfilter_100_Batch _64	Otsu	90.86	56.5	75.69	41.24	56.91	34.09

Based on Table 3. It is known that the best results obtained in this experiment are on the U-Net Up-Filter Model using the Otsu Threshold post-processing method. The best performance results obtained were 90.86% for PA, 56.5% for MIoU, 75.69% for MA and 34.09 for DCS. The results obtained for the segmentation case using U-Net are not satisfactory, especially for the MIoU and DCS values. The performance value for this segmentation task can still be improved. However, as the first experiment in the topic of SCJ segmentation in Indonesia. The results of this experiment give hope that research on the development of computer assistance models for automatic detection and classification of cervical pre-cancer can be carried out. Improvements in performance results and the application of other deep learning methods will be carried out to produce robust models that can be applied in the medical field.

4. CONCLUSION

The best performance results are shown from the Pixel Accuracy, Mean IoU, Mean Accuracy, Dice coefficient, Precision and Sensitivity values, namely 90.86%, 56.5%, 75.69%, 34.09%, 41.24%, and 56.91%. Although the resulting performance value is still not optimal, this experiment gives hope that research in the field of automatic cervical pre-cancer detection can continue to be carried out to get better results

REFERENCES

- [1] N. V. Q. Huy *et al.*, "The value of visual inspection with acetic acid and Pap

**Akhiar Wista Arum, Siti Nurmaini, Dian Palupi Rini,
Patiyus Agustiansyah, Muhammad Naufal Rachmatullah
Segmentation of Squamous Columnar Junction on VIA Images
using U-Net Architecture**

- smear in cervical cancer screening program in low resource settings – A population-based study,” *Gynecol. Oncol. Reports*, vol. 24, no. February, pp. 18–20, 2018.
- [2] A. Srivastava, P. Sinha, P. Vatsal, F. Khatoon, and N. Lal, “Visual Inspection with Acetic Acid Versus Papanicolaou Test in Cervical Cancer Screening,” *Indian J. Gynecol. Oncol.*, 2020.
- [3] J. Liu, Y. Peng, and Y. Zhang, “A Fuzzy Reasoning Model for Cervical Intraepithelial Neoplasia Classification Using Temporal Grayscale Change and Textures of Cervical Images During Acetic Acid Tests,” *IEEE Access*, 2019.
- [4] V. Kudva, K. Prasad, and S. Guruvare, “Detection of Specular Reflection and Segmentation of Cervix Region in Uterine Cervix Images for Cervical Cancer Screening,” *Irbm*, vol. 38, no. 5, pp. 281–291, 2017.
- [5] J. Liu, L. Li, and L. Wang, “Acetowhite region segmentation in uterine cervix images using a registered ratio image,” 2017.
- [6] K. Gutiérrez-fragoso, H. G. Acosta-mesa, N. Cruz-ramírez, and R. Hernández-jiménez, “Optimization of Classification Strategies of Acetowhite Temporal Patterns towards Improving Diagnostic Performance of Colposcopy,” *Hindawi*, vol. 2017, 2017.
- [7] V. Kudva, K. Prasad, and S. Guruvare, “Automation of detection of cervical cancer using convolutional neural networks,” *Crit. Rev. Biomed. Eng.*, vol. 46, no. 2, pp. 135–145, 2018.
- [8] J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, “Machine learning for assisting cervical cancer diagnosis : An ensemble approach,” *Futur. Gener. Comput. Syst.*, vol. 106, pp. 199–205, 2020.
- [9] M. Sharma, S. Kumar Singh, P. Agrawal, and V. Madaan, “Classification of Clinical Dataset of Cervical Cancer using KNN,” *Indian J. Sci. Technol.*, vol. 9, no. 28, 2016.
- [10] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, “A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images,” *Comput. Methods Programs Biomed.*, vol. 164, pp. 15–22, 2018.
- [11] Y. Song *et al.*, “Accurate cervical cell segmentation from overlapping clumps in pap smear images,” *IEEE Trans. Med. Imaging*, vol. 36, no. 1, pp. 288–300, 2017.
- [12] K. M. A. Adweb, N. Cavus, and B. Sekeroglu, “Cervical Cancer Diagnosis Using Very Deep Networks over Different Activation Functions,” *IEEE Access*, vol. 9, pp. 46612–46625, 2021.
- [13] M. N. Asiedu *et al.*, “Development of algorithms for automated detection of cervical pre-cancers with a low-cost, point-of-care, Pocket Colposcope HHS Public Access,” *IEEE Trans Biomed Eng*, vol. 66, no. 8, pp. 2306–2318, 2019.
- [14] B. Bai, Y. Du, P. Liu, P. Sun, P. Li, and Y. Lv, “Detection of cervical lesion region from colposcopic images based on feature reselection,” *Biomed. Signal Process. Control*, vol. 57, p. 101785, 2020.
- [15] H. L. Holgersti-medicalcom, “Automatic detection of multi-level acetowhite regions in RGB color images of the uterine cervix,” vol. 5747, pp. 1004–1017, 2005.
- [16] T. Xu, E. Kim, and X. Huang, “ADJUSTABLE ADABOOST CLASSIFIER AND PYRAMID FEATURES FOR IMAGE-BASED CERVICAL CANCER

- DIAGNOSIS Computer Science and Engineering Department , Lehigh University , Bethlehem , PA , USA ;,” pp. 281–285, 2015.
- [17] K. V., P. K., and G. S., “Andriod Device-Based Cervical Cancer Screening for Resource-Poor Settings,” *J. Digit. Imaging*, vol. 31, no. 5, pp. 646–654, 2018.
- [18] V. Kudva, K. Prasad, and S. Guruvare, “Hybrid Transfer Learning for Classification of Uterine Cervix Images for Cervical Cancer Screening,” *J. Digit. Imaging*, vol. 33, no. 3, pp. 619–631, 2020.
- [19] H. H. Son, P. C. Phuong, T. Van Walsum, and L. Manh Ha, “Liver Segmentation on a Variety of Computed Tomography (CT) Images Based on Convolutional Neural Networks Combined with Connected Components,” *VNU J. Sci. Comput. Sci. Commun. Eng.*, vol. 36, no. 1, pp. 25–37, 2020.
- [20] M. Aghalari, A. Aghagolzadeh, and M. Ezoji, “Brain tumor image segmentation via asymmetric/symmetric Unet based on two-pathway-residual blocks,” *Biomed. Signal Process. Control*, 2021.
- [21] J. Wu and C. Hicks, “Breast cancer type classification using machine learning,” *J. Pers. Med.*, vol. 11, no. 2, pp. 1–12, 2021.
- [22] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” pp. 1–8.
- [23] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” vol. 8828, no. c, pp. 1–12, 2016.
- [24] M. N. Rachmatullah, S. Nurmaini, A. I. Sapitri, A. Darmawahyuni, B. Tutuko, and F. Firdaus, “Convolutional neural network for semantic segmentation of fetal echocardiography based on four-chamber view,” *Bull. Electr. Eng. Informatics*, vol. 10, no. 4, pp. 1987–1996, 2021.