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

Exploration U-Net Architecture for Cervical Precancerous Lesions Segmentation

Akhiar Wista Arum, Muhammad Naufal Rachmatullah*, Bambang Tutuko, Firdaus, Annisa Darmawahyuni, Ade Iriani Sapitri, Anggun Islami, Dea Agustria Ananda

[1]*Intelligent System Research Group, Universitas Sriwijaya, Palembang, Indonesia*
** naufalrachmatullah@gmail.com*

ABSTRACT

The automatic analysis of images for the early detection of cervical cancer relies on the segmentation of cervical precancerous lesions. This paper investigates the incorporation of various CNN-based backbones into a U-Net model for improved segmentation accuracy. A set of twelve backbones was tested, including VGG16, VGG19, ResNet50, ResNext50, EfficientNetB7, InceptionResNetv2, DenseNet201, InceptionV3, MobileNet V2, SE-ResNet50, SE-ResNext50, and SE-Net154. Evaluation metrics were computed using Intersection over Union, pixel accuracy, and Dice coefficient. The findings demonstrate that U-Net with EfficientNetB7 backbone outperforms all other models with an IoU of 73.13%, pixel accuracy of 89.92%, and a Dice coefficient of 77.64%. These results were visually confirmed; segmentation outputs were examined, showing accurate delineation of lesion borders. The dominating performance of EfficientNetB7 was observed to be due to high feature extraction efficiency coupled with powerful spatial information representation. The study is, however, limited by a lack of clinical validation and expert evaluation from trained medical personnel. The results demonstrate the effectiveness of combining the U-Net architecture with advanced CNN backbones towards designing automated systems to analyze medical images.

Keywords: Segmentation, U-Net, Convolutional Neural Network, Cervical Precancerous Lesions

1. INTRODUCTION

Once advances in AI began to emerge, their application within computer vision technology progressed, granting the ability to automatically capture, assess, and understand information from images and videos[1]. This will also significantly impact medical imaging, which is currently an essential part of modern healthcare [2]. One of artificial intelligence's most common medical imaging uses is to help diagnose diseases like cervical cancer [3]. Artificial intelligence methods, especially deep learning, have been widely applied to detect the severity of cervical cancer, or to classify and segment early cervical cancer methods, such as the visual inspection of acetic acid (VIA), the Pap smear, or a colposcopy [4] - [7]. Deep learning methods can be implemented to perform image segmentation [8]. Image segmentation aims to separate important parts in medical images so they can be easily analyzed. In the context of cervical precancerous lesions, image segmentation can identify critical areas such as lesions and other structures or areas related to the possibility of cervical precancerous lesions [9]-[10].

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

Deep learning models have been widely used to segment lesions [11], [12]. Several studies have been conducted related to lesion segmentation, including utilizing a variety of CNN methods to achieve optimal results in segmenting lesions [13] []. Yu et al. implemented a combination of several deep learning methods, including R-CNN, ASPP, and EfficientNet, to obtain a lesion segmentation result [14]. Toshihiro et al. applied a U-Net model to segment lesions in images before and after acetic acid application [7]All the studies focused only on segmenting the lesions and using colposcopy images. While segmenting other parts for precancerous lesions is also very important, using colposcopy data is not applicable in Indonesia.

Thus, this study suggests using a deep learning technique, U-Net with CNN, to segment columnar, lesion, and cervical regions. In addition, the segmentation process is based on primary cervicography data from Indonesia. Unfortunately, the dataset for cervical images, or cervicography, is scarce in Indonesia. The first step in building a segmentation model is obtaining public data from the IARC, part of the WHO [15].

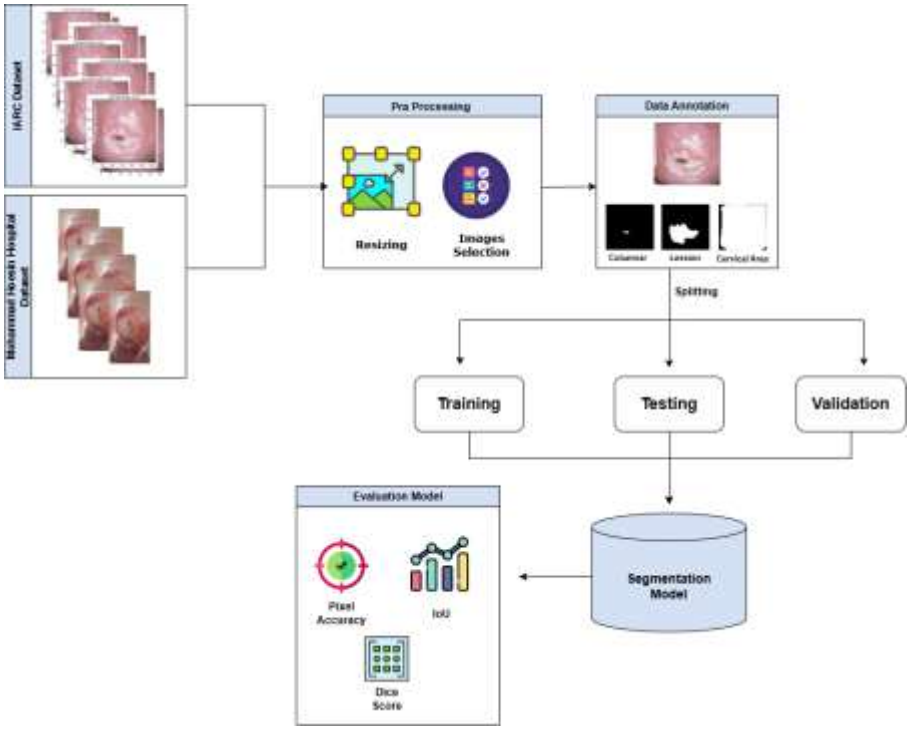


Figure 1. Research Workflow.

2. MATERIAL AND METHOD

The objective of establishing a framework based on the segmentation model is to enhance research organization within the designed structure. The initial step involves acquiring the data. For this study, the data were obtained from IARC's public datasets along with the private datasets of Mohammad Hoesin Palembang Hospital, which included the VIA-based early detection cervical cancer image datasets. The VIA image dataset comprises two classes: abnormal and normal precancerous cervical image clusters.

Following the data collection, the next step is data pre-processing. Image data processing entails cleaning as well as resizing them to the required dimensions and formats. The next step is the creation of ground truths which comprises image

annotations that will serve as reference data throughout the model training phase. In the third stage, the dataset is generated. Upon completing the annotation phase, the output will yield a ground truth dataset that consists of images organized in a directory intended for model training.

. The ground truth data will be divided into train, validation, and test datasets. In the fourth stage, we will build a model using the U-Net architecture. The model training process and backbone tuning in the U-Net architecture section are performed simultaneously. The goal of this process is to enhance the model evaluation results. The last stage of this research is model evaluation. After completing the U-Net model training and tuning the backbone, each model will be tested with the testing dataset to assess the model's performance. The evaluation results of each model will be compared to determine the best model.

2.1 DATA ACQUISITION

Data acquisition is the process of collecting and converting data, which involves various steps to ensure that the data used in analysis or model training is clean, consistent, and ready for use. The data used in this study came from a combination of data from the International Research for Cancer/World Health Organization [15] and Mohammad Hoesin Hospital, Palembang. The raw data obtained was previously divided into abnormal and normal data. The image data used for training is the post-IVA image obtained after applying acetic acid. The model training process on segmentation uses normal and abnormal data, with 168 normal and 217 abnormal data.

2.2 PRE-PROCESSING DATA

Data pre-processing is the initial process performed to improve the quality of input data before it is used in model training and testing. The present study utilizes two pre-processing techniques: Image selection and resizing. In this research, the raw data selection process selects raw images that cannot be used or duplicate data. The objective is to ensure optimal data quality before entering the training model process [16]. Resizing is changing the image's width and height dimensions to a uniform size.

a.Different Lighting Image b. Image with Interference c. Different Viewing

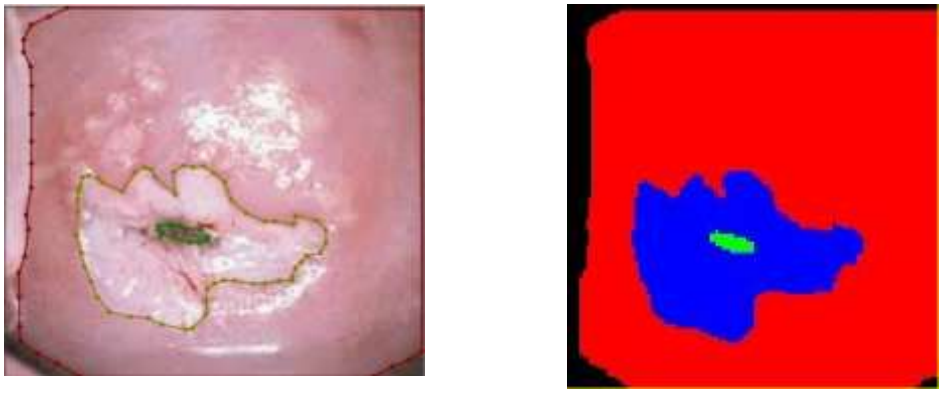
Figure 2. Image Selecting Process.

According to the dataset that has been collected, not all data is considered to be of good quality and meets established requirements. This is because the datasets obtained still vary, such as different viewing angles, blurred images, different lighting, and disturbances that can interfere with the results of data training. The image selection technique can maximize the data quality before entering the model training process. Figure 2 shows some examples of images that are unclear and cannot be used.

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

2.3 ANNOTATION DATA

Object annotation is a process used to label specific objects in the image to identify important areas, such as the cervical region, the columnar, and lesions. The labelling process is facilitated by Roboflow annotation software [17], which allows users to annotate objects manually or semi-automatically to produce accurate ground truth data. Figure 3. (a) shows the three objects labelled in this data: cervical area, columnar and lesion.



a. Annotating b. Ground Truth

Figure 3. Image Annotating Process.

In multi-class semantics, each pixel in a segmentation object has a unique array value that refers to a specific object class. When converted to image format, human vision cannot directly recognize these array values as they are only numerical representations. Therefore, to provide a more precise visualization, each object class in the segmentation image is given a different color according to its category, as shown in Figure 3. (b) After the annotation process is complete, the resulting dataset is then divided into three subsets: 80% training data, 10% validation data and 10% test data.

2.4 SEGMENTATION MODEL

The model segmentation process in this study uses the U-Net architecture. The U-Net architecture consists of two main parts: the encoder and decoder. The encoder is responsible for extracting important features from the image through convolution and downsampling (pooling) operations. At the same time, the decoder reconstructs spatial information by upsampling and convolution to produce a predicted segmentation map that has the same size as the original image [18],[19].

This research utilizes the U-Net architecture and several CNN-based backbones for better feature extraction [20]. The choice of CNN backbone is based on its proven ability to effectively extract deep features, especially when using a pretrained model. The CNN backbones used in this study include; Vgg16, Vgg19, ResNet50, ResNext50, EfficientNetb7, Inception ResNetv2, DenseNet201, Inceptionv3, Mobilenetv2, SE-ResNet50, SE-ResNext50, and SENet154 .

TABLE 1.
The Model Parameters Configurations.

Parameter	Value
Epoch	100; 500
Optimizer	Adam
Learning Rate	0.0001
Batch Size	8; 16

Hyperparameter tuning techniques are also carried out using training and validation data in the model training process. The objective is to find the parameters of the model that produce the best evaluation results. The model parameters that produce evaluation results are attached in Table 1.

2.5 MODEL EVALUATION

Performance measurement is a process of evaluation and analysis to determine the extent to which a system, process, or entity meets its goals and standards. This process involves various metrics and methods to comprehensively evaluate the model's performance. In this segmentation, the performance of the model process can be assessed using several metrics, including the IoU, pixel accuracy, and the dice coefficient.

2.5.1 Intersection Over Union

In image processing segmentation and object detection, one of the evaluation metrics is the Intersection over Union (IoU). The metric takes the overlapping area of both the predicted and actual object or region, assesses their conjunction, and then splits it using the aggregate area of both regions as the divisor[21]. This metric is calculated based on the given equation, as follows:

$$\text{IoU} = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{1}$$

Where:

- Area of Overlap: The area of overlap between prediction and ground truth.
- Area of Union: The combined area of prediction and ground truth.

2.5.2 Pixel Accuracy

Pixel Accuracy is one of the performance metrics used to evaluate model performance in image segmentation. It works by measuring the percentage of pixels correctly classified by the model and comparing them to all pixels in the image [22]. It is calculated based on the number of correctly predicted pixels divided by the total number of pixels. This metric has limitations in handling class imbalance and small details in the image.

2.5.3 Dice Coefficient

Dice Coefficient is one of the metrics calculated as twice the area of overlap of the predicted image and ground truth map divided by the total number of pixels.

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

The smaller the intersection over union value, the smaller the dice coefficient value [23]. This metric is calculated based on the given equation, as follows:

$$\text{Dice Coefficient} = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

Where:

- $|\boldsymbol{A} \cap \boldsymbol{B}|$: Number of pixels of part of the intersection
- $|\boldsymbol{A}|$: Number of pixels in the first set (ground truth)
- $|\boldsymbol{B}|$: Number of pixels in the second set (segment result)

3. RESULT AND DISCUSSION

This study uses twelve U-Net model architectures with various CNN backbones to segment cervical precancerous lesion images. The backbones used include various architectures, namely: VGG16, VGG19, ResNet50, ResNeXt50, EfficientNetB7, InceptionResNetV2, DenseNet201, InceptionV3, MobileNetV2, SE-ResNet50, SE-ResNeXt50, and SENet154. Each backbone is integrated into the U-Net encoder structure to evaluate its effect on segmentation performance. The evaluation is based on three main metrics: Intersection over Union (IoU), Pixel Accuracy, and Dice Coefficient.

Performance metrics across all models show no considerable differences. This indicates how all models are comparably proficient in feature extraction and representation from images of precancerous lesions. As indicated in Table 2, the U-Net model incorporating the EfficientNetB7 CNN backbone achieved superior performance with the best results. This model obtained the highest IoU value of 73.13%. IoU, a standard metric in segmentation exercises, assesses the overlap of marked regions and actual regions of interest, including areas of interest and the ground truth. Achieving a value of 73.13% suggests that the model performed quite confidently in identifying and sophisticatedly drawing the target object within the image.

Moreover, the U-Net model with EfficientNetB7 CNN backbone also holds the pixel accuracy record by achieving 89.92%. Pixel accuracy gauges the number of accurately identified pixels in the image of a cervical precancerous lesion against the total number of pixels. High pixel accuracy suggests a stronger capability of the model in predicting the image's pixels. The last metric, Dice Coefficient, also applies, but it still stands out the most regarding this model's performance, which is better than others.

The U-Net model with EfficientNetB7 CNN backbone achieves a dice coefficient value of 77.64%. This high value of the Dice coefficient shows its capacity to measure the degree of resemblance between the two areas, termed the predicted area and the actual value area, with equal importance to both.

TABLE 2.
Comparison of The Best Models on Segmentation.

Backbone	Pixel Accuracy	IoU (%)				Dice Coefficient (%)			
		CA	Lession	Cervix	Avarage	CA	Lession	Cervix	Avarage
Vgg-19	86.24	57.11	61.99	85.83	68.31	58.98	66.93	91.15	72.35
Vgg-16	86.93	57.68	59.99	86.28	67.98	59.78	64.85	91.56	72.06
ResNet50	87.07	65.81	65.54	86.02	72.46	67.53	71.73	91.42	76.89
ResNext50	88.25	68	63.97	87	73	70.35	68.95	92.08	77.13
EfficientNetb7	**89.92**	**63.11**	**67.59**	**88.69**	**73.13**	**65.7**	**73.93**	**93.3**	**77.64**
InceptionResNetv2	84.44	55.59	52.5	82.94	63.68	58.43	54.48	89.69	67.53
DenseNet201	87.58	64.18	61.21	86.78	70.72	65.29	65.37	92.08	74.25
Inceptionv3	87.43	61.4	62.92	86.68	70.33	63.01	67.7	92.06	74.26
MobileNetv2	87.55	62.43	65.45	86.35	71.41	65.19	71.44	91.81	76.15
SE-ResNet50	87.62	65.78	49.82	86.84	67.48	67.48	49.95	92.16	69.85
SE-ResNext50	89.1167	67	64.38	88.24	72.9	68.28	69.3	92.88	76.82
SENet154	89.19	65.18	62.25	87.9	71.78	67.36	67.77	92.78	75.97

Other models like U-Net with ResNet50 and SE-ResNext50 backbones have competitive performance, but they still do not outperform U-Net models with efficientNetB7 CNN backbones. Also, models with IncepionResNetv2, VGG16, and VGG19 backbones performed poorly on all evaluation metrics relative to the other models. The poor performance of these three models is due to the VGG and Inception ResNetv2's deeper architecture, which is inefficient for complex feature representation. The U-Net model with EfficientNetB7 as a backbone demonstrated superior performance in segmentation of cervical precancerous lesions due to its compound scaling architecture which strategically increases the depth, width, and resolution of the image efficiently. This enables the model to capture necessary unique feature information without adding much computational burden.

Figure 4 illustrates the results of image segmentation for cervical precancerous lesions and their visualizations on all tested models. Each model's visualization results testify to its ability in classifying the lesion region and normal cervical tissues. The U-Net model with EfficientNetB7 backbone demonstrates segmentation results where the contours are aligned with actual cervical precancerous lesion boundaries. These visualisation results are comparable to the performance values of IoU, pixel accuracy, and Dice coefficient of the U-Net model with EfficientNetB7 backbone, which outperforms the performance values of other models.

Raw Image **True Ground Truth**

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

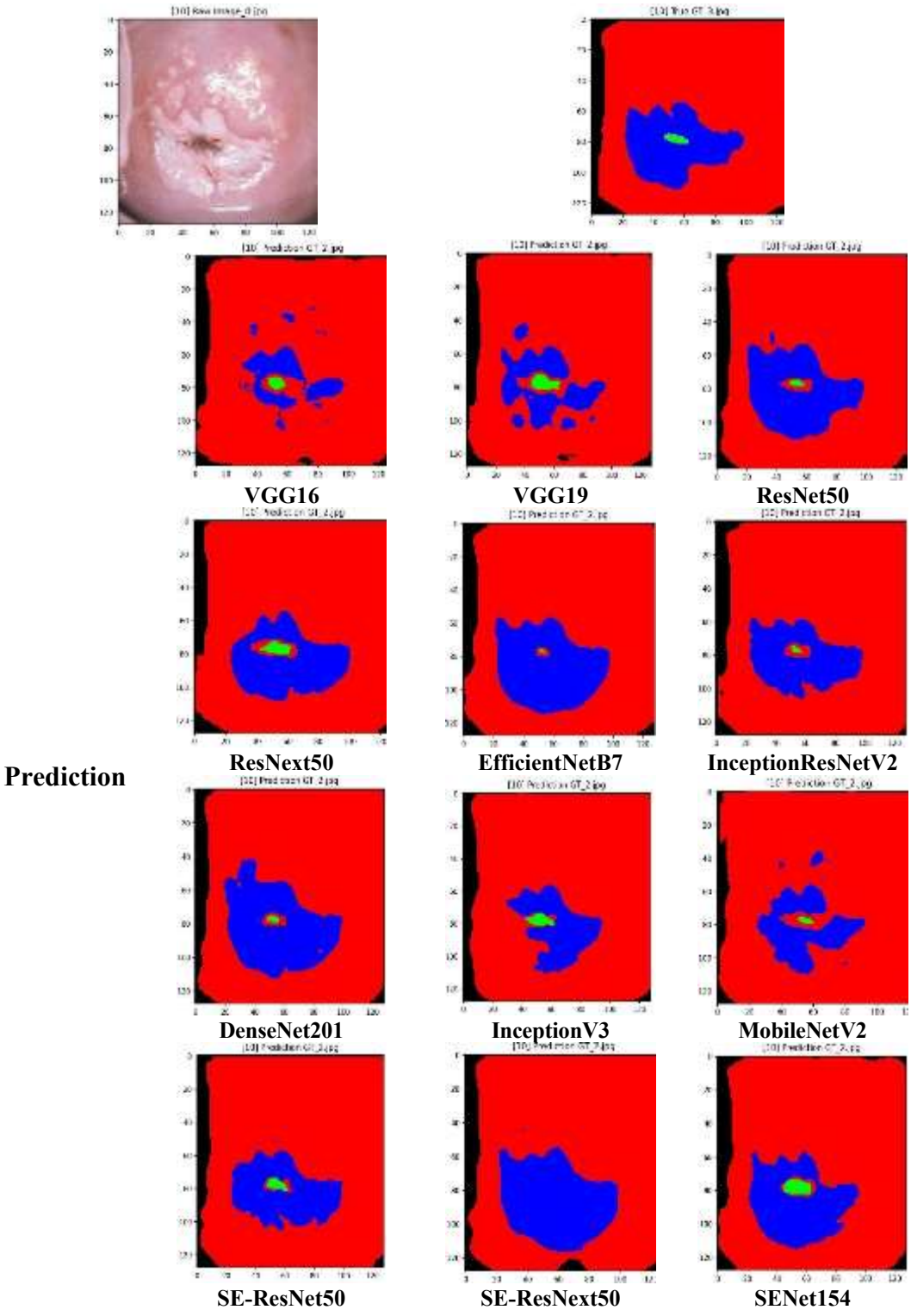


FIGURE 4. The Visualization of Cervical Precancerous Lesion Image Segmentation.

4. CONCLUSION

Based on the results of testing twelve U-Net models using different CNN backbones, it can be concluded that backbone selection significantly influences the performance of the segmentation model. The U-Net model with EfficientNetB7 backbone performs best, with the highest IoU value of 73.13%, pixel accuracy of

89.92%, and dice coefficient of 77.64%. These values indicate that the model is accurate in detecting the correct pixels (as indicated by pixel accuracy) and excellent in capturing the overall shape of the target object in the image, including complex boundaries and contours. These results confirm that the EfficientNetB7 architecture can extract features efficiently and capture spatial information more accurately than other backbones.

To our knowledge, this study is one of the earliest studies that comprehensively evaluates the impact of twelve different CNN backbones integrated into the U-Net architecture for cervical pre-cancerous lesion segmentation using clinical image data. The addition of EfficientNetB7 as a backbone in this context represents a novel contribution, showing significant improvement over traditional backbones such as VGG16 and InceptionResNetv2.

Notwithstanding the encouraging findings, this study suffers from several limitations. The model was trained and validated on a relatively small and domain-specific dataset, which may limit its ability to generalize to different types of cervical images or wider clinical environments. In addition, the assessment was based on standard quantitative metrics and did not include any qualitative assessment from specialists, which would provide a deeper understanding of its clinical utility.

ACKNOWLEDGEMENTS

We thank Prof. Siti Nurmaini and the Intelligent System Research Group (ISysRG), Faculty of Computer Science, Universitas Sriwijaya, Indonesia, for supporting the infrastructure.

REFERENCES

[1] D. Kitaguchi, N. Takeshita, H. Hasegawa, and M. Ito, ‘Artificial intelligence-based computer vision in surgery: Recent advances and future perspectives’, Ann Gastroenterol Surg, vol. 6, no. 1, pp. 29–36, 2022, doi: 10.1002/ags3.12513.

[2] I. Rizwan I Haque and J. Neubert, ‘Deep learning approaches to biomedical image segmentation’, Jan. 01, 2020, Elsevier Ltd. doi: 10.1016/j.imu.2020.100297.

[3] L. Allahqoli et al., ‘Diagnosis of Cervical Cancer and Pre-Cancerous Lesions by Artificial Intelligence: A Systematic Review’, Diagnostics, vol. 12, no. 11, pp. 1–32, 2022, doi: 10.3390/diagnostics12112771.

[4] E. Hussain, L. B. Mahanta, K. A. Borbora, H. Borah, and S. S. Choudhury, ‘Exploring explainable artificial intelligence techniques for evaluating cervical intraepithelial neoplasia (CIN) diagnosis using colposcopy images’, Expert Syst Appl, vol. 249, Sep. 2024, doi: 10.1016/j.eswa.2024.123579.

[5] S. L. Tan, G. Selvachandran, W. Ding, R. Paramesran, and K. Kotecha, ‘Cervical Cancer Classification From Pap Smear Images Using Deep Convolutional Neural Network Models’, Interdiscip Sci, vol. 16, no. 1, pp. 16–38, Mar. 2024, doi: 10.1007/s12539-023-00589-5.

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

[6] Y. R. Park, Y. J. Kim, W. Ju, K. Nam, S. Kim, and K. G. Kim, 'Comparison of machine and deep learning for the classification of cervical cancer based on cervicography images', Sci Rep, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-95748-3.

[7] T. Shinohara, K. Murakami, and N. Matsumura, 'Diagnosis Assistance in Colposcopy by Segmenting Acetowhite Epithelium Using U-Net with Images before and after Acetic Acid Solution Application', Diagnostics, vol. 13, no. 9, May 2023, doi: 10.3390/diagnostics13091596.

[8] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, 'Image Segmentation Using Deep Learning: A Survey', IEEE Trans Pattern Anal Mach Intell, vol. 44, no. 7, pp. 3523–3542, 2022, doi: 10.1109/TPAMI.2021.3059968.

[9] I. Rizwan I Haque and J. Neubert, 'Deep learning approaches to biomedical image segmentation', Inform Med Unlocked, vol. 18, p. 100297, 2020, doi: 10.1016/j.imu.2020.100297.

[10] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, 'Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges', J Digit Imaging, vol. 32, no. 4, pp. 582–596, 2019, doi: 10.1007/s10278-019-00227-x.

[11] H. Ahmadzadeh Sarhangi, D. Beigifard, E. Farmani, and H. Bolhasani, 'Deep learning techniques for cervical cancer diagnosis based on pathology and colposcopy images', Jan. 01, 2024, Elsevier Ltd. doi: 10.1016/j.imu.2024.101503.

[12] H. Jiang et al., 'A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation', May 01, 2023, Elsevier Ltd. doi: 10.1016/j.compbiomed.2023.106726.

[13] J. Kim, C. M. Park, S. Y. Kim, and A. Cho, 'Convolutional neural network-based classification of cervical intraepithelial neoplasias using colposcopic image segmentation for acetowhite epithelium', Sci Rep, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-21692-5.

[14] H. Yu et al., 'Segmentation of the cervical lesion region in colposcopic images based on deep learning', Front Oncol, vol. 12, Aug. 2022, doi: 10.3389/fonc.2022.952847.

[15] Visual Inspection with Acetic Acid (VIA) Image Bank, 'International Agency for Research on Cancer (IARC)'. Accessed: Dec. 18, 2024. [Online]. Available: https://screening.iarc.fr/cervicalimagebank.php

[16] S. Vahidi, 'A New Method for Resizing the Images', May 2022, doi: 10.13140/RG.2.2.15756.59521/1.

[17] Roboflow, 'Data Annotation Platform '. Accessed: Sep. 06, 2024. [Online]. Available: https://www.roboflow.com/

[18] M. Krithika alias AnbuDevi and K. Suganthi, 'Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET', Diagnostics, vol. 12, no. 12, p. 3064, 2022.

[19] L. Cai, J. Gao, and D. Zhao, 'A review of the application of deep learning in medical image classification and segmentation', Ann Transl Med, vol. 8, no. 11, 2020.

[20] X. Liu, Z. Deng, and Y. Yang, 'Recent progress in semantic image segmentation', Artif Intell Rev, vol. 52, pp. 1089–1106, 2019.

[21] X. Zheng, Q. Lei, R. Yao, Y. Gong, and Q. Yin, 'Image segmentation based on adaptive K-means algorithm', EURASIP J Image Video Process, vol. 2018, no. 1, pp. 1–10, 2018.

[22] M. A. Moll, H. S. Baird, and C. An, 'Truthing for pixel-accurate segmentation', in DAS 2008 - Proceedings of the 8th IAPR International Workshop on Document Analysis Systems, 2008, pp. 379–385. doi: 10.1109/DAS.2008.47.

[23] R. R. Shamir, Y. Duchin, J. Kim, G. Sapiro, and N. Harel, 'Continuous dice coefficient: a method for evaluating probabilistic segmentations', arXiv preprint arXiv:1906.11031, 2019.

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