Jurnal Nasional Teknik Elektro dan Teknologi Informasi
Volume 14 Number 1 February 2025

Translation of article 10.22146/ jnteti.v14i1.15262

Comparison of U-NET and ELU-NET for Pancreatic Cancer Medical Image Semantic Segmentation

Algi Fari Ramdhani[1], Yudi Widhiyasana[1], Setiadi Rachmat[1]

[1] Informatics Engineering Study Program, Department of Computer and Informatics Engineering, Politeknik Negeri Bandung, Kabupaten Bandung Barat, Jawa Barat 40559, Indonesia

[Submitted: 23 September 2024, Revised: 13 December 2024, Accepted: 1 February 2025]
(Corresponding Author: Algi Fari Ramdhani (email: algi.fari.tif420@polban.ac.id)

ABSTRACT — Medical image analysis for semantic segmentation using deep learning technology has been extensively developed. One of the notable architectures is U-NET, which has demonstrated high accuracy in segmentation tasks. Further advancements have led to the development of ELU-NET, which aims to enhance model efficiency. ELU-NET achieves relatively good accuracy; however, further comparative analysis of both models is necessary. The comparison between these models is based on accuracy, storage usage, and processing time in performing semantic segmentation of pancreatic cancer images. The pancreatic cancer images utilized in this study are sourced from the PAIP 2023 Challenge, consisting of hematoxylin and eosin (H&E)-stained images. Experiments were conducted by varying the number of filters and model depth for both architectures. The evaluation was performed using a dataset of 57 pancreatic cancer images. The experimental results indicated that U-NET achieved the highest accuracy at 92.8%, slightly outperforming ELU-NET, which attained 89.7%. However, ELU-NET is significantly more efficient in terms of storage usage (8.1 MB for ELU-NET compared to 93.31 MB for U-NET) and processing time (4.0 s for ELU-NET and 5.3 s for U-NET). Although ELU-NET exhibited slightly lower accuracy than U-NET, it surpassed U-NET considerably in terms of storage efficiency (by 85.21 MB) and processing speed (by 1.3 s). These findings suggest that ELU-NET is not superior to U-NET in accuracy. However, given the storage size ratio of 1:11.51 and the processing time ratio of 1:1.325 between ELU-NET and U-NET, the 3.1% accuracy difference represents a reasonable trade-off.

KEYWORDS — U-NET, ELU-NET, Lightweight Model, Semantic Segmentation.

I. INTRODUCTION

According to the World Health Organization (WHO), cancer remains one of the leading causes of mortality worldwide [1]. Various types of cancer pose serious threats to global health, necessitating targeted interventions and treatment strategies. Among the most lethal types is pancreatic cancer, which is associated with a particularly low survival rate—approximately 10% [2]. This indicates that only a small proportion of patients survive beyond five years after the initial diagnosis. One of the primary reasons for this low survival rate is the lack of adequate treatment, especially in terms of early detection.

Early detection of pancreatic cancer plays a crucial role in improving patient survival rates [3]. Pancreatic cancer often remains asymptomatic in its early stages, leading to diagnosis only at advanced stages when treatment becomes significantly more challenging and less effective. Early detection facilitates the identification of cancer at an earlier phase, allowing for more timely and effective treatment interventions. Raising awareness of early detection and advancing diagnostic technologies are essential steps in increasing survival rates among pancreatic cancer patients.

Cancer detection is a pivotal step in the overall cancer treatment process. Accurate and timely detection enables doctors to establish precise diagnoses, which serve as the foundation for effective treatment strategies. One of the most critical aspects of cancer detection is early diagnosis, as it significantly enhances the probability of successful treatment. However, despite its importance, early detection faces numerous challenges in practical implementation. A major obstacle is the predominance of manual analysis performed by pathologists [4], [5]. Manual detection introduces challenges related to subjectivity and the inherent limitations of human expertise. Subjectivity arises as medical image analysis relies heavily on the knowledge and experience of individual pathologists. Two pathologists examining the same image may arrive at different diagnoses, influenced by their varying levels of expertise and interpretation of the image.

Beyond subjectivity, another significant limitation is the physical and mental constraints of pathologists. Like other highly precise and concentration-intensive professions, pathologists cannot work continuously without experiencing fatigue [6]. Long hours spent analyzing complex medical images may lead to decreased concentration and accuracy, ultimately impacting the quality of diagnostic assessments. This fatigue factor becomes even more critical in scenarios where pathologists are required to analyze a large volume of cases, increasing the risk of diagnostic errors.

To overcome the challenges of manual cancer detection, technological approaches have been explored [7]. The advancement of technology has facilitated the development of automated early cancer detection systems using deep learning [8]. A specific branch of deep learning tailored for image analysis is convolutional neural networks (CNNs), which have been extensively researched in medical image analysis [9]. CNN-based image analysis has demonstrated promising accuracy levels.

CNNs require medical images for analysis, with one of their key applications being semantic segmentation. To achieve semantic segmentation, CNNs are designed based on specific architectures. One of the widely used CNN architectures for semantic segmentation is U-NET [10]. U-NET was specifically designed for semantic segmentation in medical images and was originally developed for segmenting differential interference

contrast (DIC)-stained HeLa cell images. U-NET has demonstrated superior accuracy in semantic segmentation compared to other models within similar research contexts.

While U-NET achieves high accuracy, it demands substantial computational resources [11]. Specifically, U-NET comprises approximately 31 million parameters, reflecting its complexity and significant computational requirements. This large number of parameters enables U-NET to capture intricate details in segmentation tasks, which serves as a major advantage. However, this also results in higher memory consumption and longer execution times, reducing its efficiency for real-world applications.

In pursuit of a more lightweight yet effective model, ELU-NET was developed. ELU-NET builds upon U-NET while reducing the number of parameters significantly without compromising segmentation performance [12]. One of the innovations introduced in ELU-NET is the incorporation of deep skip connections, allowing for more efficient image information processing. In a prior study, U-NET was integrated with the VGG16 architecture in the encoder section, leading to a model with only 600,000 parameters—a drastic reduction compared to the 31 million parameters in U-NET.

VGG16 is a deep learning architecture originally designed for image classification tasks and has proven highly effective in handling various types of image data [13]. By incorporating VGG16 into ELU-NET, the model achieves a balance between high accuracy and parameter efficiency, enabling broader applicability in resource-constrained environments.

Despite the reduced number of parameters, ELU-NET does not exhibit diminished accuracy compared to U-NET. In certain studies, ELU-NET integrated with VGG16 has achieved a Dice coefficient accuracy of 0.96, surpassing U-NET's accuracy of 0.92. The Dice coefficient is a commonly used evaluation metric in image segmentation research, alongside the Jaccard index, as it effectively measures the alignment between model predictions and ground truth labels [14].

The impressive performance of ELU-NET was derived from a dataset containing brain organ images, enabling the model to achieve superior segmentation accuracy over U-NET. Nevertheless, to comprehensively validate ELU-NET's advantages, further comparisons using diverse medical image datasets are required. This ensures that ELU-NET's superiority in accuracy and efficiency is consistently demonstrated across different conditions and data types. Therefore, in this study, ELU-NET and U-NET are being compared using pancreatic cancer images from the PAIP 2023 Challenge. This challenge provides hematoxylin and eosin (H&E)-stained pancreatic cancer images, and since U-NET has already been proven effective for segmenting H&E images [15], this comparison will offer crucial insights into the relative performance of U-NET and ELU-NET in this specific context.

In this study, the evaluation of both models is not limited to segmentation accuracy but also considers the number of parameters, model file size, and inference time. This comprehensive evaluation provides deeper insights into the strengths and weaknesses of each architecture, as well as their potential clinical applications where a balance between accuracy and efficiency is required.

II. METHODOLOGY

This section discusses aspects related to the research, including research data, image processing, model training, and model testing

A. RESEARCH DATA

This study used data provided by the organizers of the PAIP 2023 Challenge. To obtain the data, registration is required to participate in the challenge and complete a form to access the data. The challenge registration is conducted on the website https://2023paip.grand-challenge.org/. After registering for the challenge, an account must be registered on the page http://wisepaip.org/challenge2023 using the same email as on the challenge page. The data are then downloaded according to the research needs. The PAIP 2023 Challenge is the first challenge series from PAIP that uses a dataset consisting of images of pancreatic cancer or tumors.

The data specifications provided by the challenge organizers are shown in Table I. There are 53 training data samples with labels, consisting of 50 pancreatic and 3 colorectal images. These 53 data samples have the same dimensions, namely 1,024 × 1,024. The images include magnifications of 0.250 mpp (micron per pixel) and 0.500 mpp. Since the majority of the images are pancreatic images with a magnification of 0.250 mpp, this study used pancreatic images with a magnification of 0.250 mpp.

The validation and test data from the PAIP 2023 Challenge were not used in this study because these datasets did not contain annotation labels required for the model training stage. Without labels, the data cannot be effectively used for training or validating model performance, as labels are essential for measuring the model's prediction accuracy against the ground truth. Therefore, this study only utilized the training data provided by the PAIP 2023 Challenge, consisting of pancreatic images with a magnification of 0.250 mpp. This training data includes labels that enable the model to learn specific patterns relevant to pancreatic image segmentation tasks. A complete specification of the data used in this study, including the number and dimensions of the images, is presented in Table II.

An example of an image used in this study is shown in Figure 1 and Figure 2. Figure 1 is an H&E-stained image from a patient diagnosed with pancreatic cancer. Darker-colored spots representing cells can be observed; however, not all dark-colored cells are cancerous. This can be seen in Figure 2, which is the annotation label of Figure 1. The white areas in Figure 2 represent cancerous cells, while the black areas represent the background or non-cancerous cells.

In accordance with Table II, the training and validation data consist of 36 images. The training and validation data were randomly shuffled during training to avoid the occurrence of overfitting. At the same time, ten images were used for testing. Image preprocessing was subsequently carried out on these 46 images.

B. IMAGE PREPROCESSING

In this study, image preprocessing was conducted primarily to enhance image quality. This process affects the accuracy of the trained model. High-quality images help the model recognize important patterns, thereby improving its performance in medical image segmentation. One commonly used image preprocessing technique is image normalization. Normalization can be performed using various methods, one of which is histogram equalization [16]. This technique aims to enhance image contrast by distributing pixel intensity more evenly, making image details more prominent. The histogram equalization process in this study was performed using the OpenCV library. OpenCV is a commonly used image-processing library in the field of computer vision.

Jurnal Nasional Teknik Elektro dan Teknologi Informasi
Volume 14 Number 1 February 2025

TABLE I
PAIP 2023 DATA SPECIFICATION

Data Classification	Image Size (Pixel)	Number of Pancreatic Images	Number of Colon Images
Training	1,024 × 1,024	50	3
Validation		10	-
Testing		20	20
Total		80	3

TABLE II
RESEARCH DATA SPECIFICATION

Data Classification	Image Size (Pixel)	Number of Pancreatic Images
Training	1,024 × 1,024	36
Validation		
Testing		10
Total		46

The training data used in this study is relatively small, consisting of only 36 images. These images were used for both training and validation. The division of images into training and validation data was performed randomly, with a training-to-validation ratio of 80:20. Thus, out of 36 images, approximately 29 were used for training, while the remaining 7 were used for validation. Due to the limited amount of data, image augmentation was performed to increase the number of available images.

Image augmentation in this study was performed by applying several simple transformations, including vertical flipping, horizontal flipping, and a combination of both [17]. By applying these transformations, each original image generated three new images after augmentation, increasing the total number of images from 36 to 144. Each augmented image retained the original dimensions of 1,024 × 1,024 pixels. An example of an image augmentation result is shown in Figure 3.

Subsequently, an image with dimensions of 1,024 × 1,024 was cropped into 16 smaller sections, each with dimensions of 256 × 256 pixels. These cropped images were then used as training or validation data. Before being used for model training, the images were shuffled and converted into a NumPy array. This conversion ensured that the images were in a format compatible with the model. Consequently, the images were ready to be used as training data for the U-NET and ELU-NET models.

C. MODEL TRAINING

In this study, model training included several important stages, from model creation using the Keras library to model training and saving the trained model. Model preparation involved designing various model variations, each tested to determine the best configuration. The variations included model depth and the number of initial filters in the convolution process, ensuring that each model type received the same variations.

Model depth is a crucial attribute because, in the ELU-NET study, a depth of 5 was found not to be the optimal depth [12]. This prompted further research to confirm whether depth 5 was indeed the best or if another depth would be more effective. Meanwhile, the number of initial filters was considered an additional attribute because another study suggested that U-NET models with fewer initial filters could achieve higher accuracy than those with larger filter sizes [11].

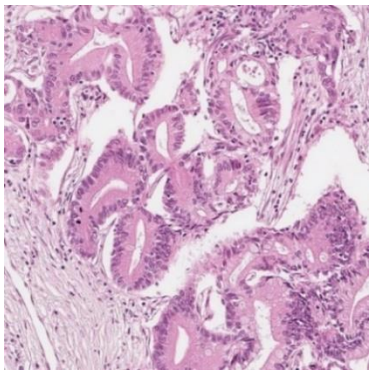

Figure 1. H&E-tinted Image on Pancreas

Figure 2. Label for Figure 1.

In the U-NET architecture, the proposed model depth is 5 [10], with an initial convolutional filter size of 64. Conversely, in the ELU-NET architecture, the model depth is set to 5 with an initial filter size of 8 [12]. Based on these studies, to determine the best model using data from the PAIP 2023 Challenge, several model versions with varying depths and initial filter numbers were created. A detailed description of the tested model versions is presented in Table III.

In the U-NET architecture, the proposed model depth is 5 [10], with an initial convolutional filter size of 64. Conversely, in the ELU-NET architecture, the model depth is set to 5 with an initial filter size of 8 [12]. Based on these studies, to determine the best model using data from the PAIP 2023 Challenge, several model versions with varying depths and initial filter numbers were created. A detailed description of the tested model versions is presented in Table III.

The model depth in this study refers to the number of down-sampling stages in the model's contractive path. In the U-NET architecture, this depth determines the number of convolutional layers at each level before up-sampling in the expansive path. Examples of different depths in the U-NET architecture are shown in Figure 4, with architectures presented in ascending depth order (4, 5, and 6). In Figure 4, the value of n represents the number of initial filters at each convolutional stage. Consequently, the highest number of filters in the deepest convolutional layer is $8n$ for a depth of 4.

The model training process was conducted using Kaggle Notebooks, which provides a T4 graphics processing unit (GPU) to accelerate model training. Before training commenced, the model needed to be compiled. At this stage, several model configurations were incorporated, including the addition of the Adam optimizer with a learning rate of 1×10^{-4} [18], the binary cross-entropy loss function [12], and accuracy metrics to evaluate model performance. Once the model was

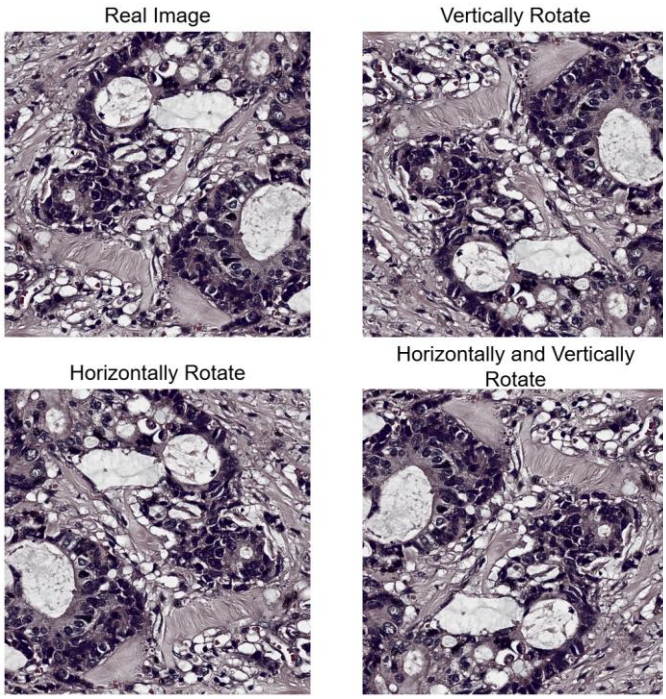

Figure 3. Image augmentation with flip.

compiled, the next step was to train the model using the predefined configurations.

During the training phase, several key parameters were utilized, including a batch size of 8, 40 epochs, a validation split of 0.2, and an early stopping callback. Early stopping was implemented to prevent overfitting, meaning that the training process would be halted if the validation loss started to increase. An increase in validation loss during training indicates that the model is beginning to overfit the training data [19].

The entire model training process was conducted using the aforementioned hyperparameters. The primary differences among the tested models lay in the number of initial filters and the model depth. The variations in initial filter numbers and model depths explored in this study are presented in Table III. By testing these different combinations, this study aims to identify the most optimal model configuration for medical image segmentation in the PAIP 2023 Challenge.

D. MODEL TESTING

Model testing was conducted after all models were developed in accordance with the specifications listed in Table III. This testing phase is crucial for evaluating the model's performance in semantic segmentation. Once the model was constructed, the next step was to test it using the prepared test dataset. The model was utilized to predict input images, and these predictions were then compared with the ground truth. Ideally, the model should generate predictions that accurately correspond to the ground truth.

The comparison between the predicted results and the ground truth was performed using the Jaccard index, a commonly used metric for measuring accuracy in semantic segmentation [20]. The Jaccard index quantifies the similarity between two images, specifically between the regions predicted by the model and the actual regions in the ground truth. The formula for computing the Jaccard index is presented in (1). The Jaccard index ranges from 0 to 1, where a value of 1 indicates that the model perfectly predicts the ground truth,

TABLE III
MODEL VARIATION OF U-NET AND ELU-NET

Architecture Type	Number of Filters	Number of Depths
U-NET	32	4
		5
		6
	64	4
		5
		6
	128	4
		5
		6
ELU-NET	4	4
		5
		6
	8	4
		5
		6
	16	4
		5
		6

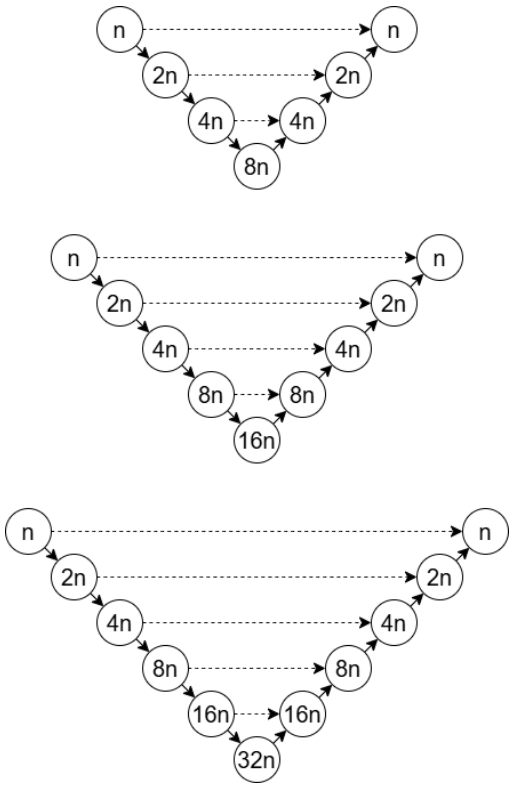

Figure 4. Example of the number of depths of the U-NET architecture.

whereas a value of 0 signifies that the model fails to predict the ground truth entirely. Technically, the Jaccard index is computed by calculating the ratio of the intersection area to the union area of the predicted image and the ground truth image.

$$J(A,B) = \frac{(A \cap B)}{(A \cup B)} = \frac{TP}{TP+FN+FP}. \quad (1)$$

The accuracy calculation of semantic segmentation using the Jaccard index can be performed with the aid of a confusion matrix. In the context of semantic segmentation, the model's predicted output is a binary image, where a pixel value of 0 represents the background, and a pixel value of 1 represents the

region of interest (ROI). The ground truth image is also a binary image, with values of 0 and 1 corresponding to the same classifications as the predicted image. Accordingly, the comparison between the predicted output and the ground truth is conducted by computing the confusion matrix. This matrix facilitates the classification of each pixel in the image into four categories: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The arrangement of the confusion matrix values for semantic segmentation is shown in Table IV.

The testing process was conducted by randomly selecting 57 images with dimensions of 256 × 256 pixels from the test dataset. Each image was then processed by the model to generate predictions, which were subsequently compared with the ground truth. After predicting all 57 images, the Jaccard index was computed for each image, and the mean of these values was taken as the final accuracy score of the tested model. This average value provides an overall representation of the model's capability in performing segmentation on the test dataset.

Apart from accuracy, several other attributes are assessed in model evaluation, including the number of parameters, model file size, and execution time required to make predictions on 57 images. The number of parameters and file size are part of the model properties that provide information about its complexity and physical size. Meanwhile, execution time measures the model's efficiency in making predictions, which is a crucial factor when deploying the model in real-world applications.

Execution time testing was conducted using the central processing unit (CPU) provided by Kaggle. The use of the CPU during the evaluation phase ensures that the model is not only accurate but also efficient in terms of processing time. The execution time evaluation mechanism starts from loading the pre-trained model to predicting 57 medical images, with time measurement beginning at model loading and ending after the final image prediction.

By integrating accuracy assessment, the number of parameters, file size, and execution time, a comprehensive evaluation of each model's performance can be obtained. This holistic evaluation is then used to determine the optimal model for the semantic segmentation of a pancreatic cancer medical image.

Accuracy calculation in this study is limited to the Jaccard index. Other accuracy metrics, such as the Dice coefficient, were not utilized due to resource constraints in the evaluation process.

III. RESEARCH RESULTS

The models developed with various configurations, including different initial filter numbers and depths, were evaluated based on multiple performance metrics. These metrics include prediction accuracy, which was compared with the ground truth. Besides accuracy measured using the Jaccard index, other evaluation metrics include the number of parameters used by the model, model size, and execution time required to predict 57 images.

The results presented in Table V indicate that some data points are missing due to computational resource limitations. These constraints are particularly evident in the U-NET architecture with a depth of 6, where no results were obtained for all filter variations. Additionally, for U-NET with an initial filter count of 128, the model could not be constructed for all intended depth levels. In contrast, for the ELU-NET architecture, all combinations of filter numbers and depths were successfully built and tested, demonstrating that this model is more flexible and applicable under limited computational resources compared to U-NET.

TABLE IV
CLASSIFICATION OF CONFUSION MATRIX VALUES IN SEMANTIC SEGMENTATION

		Ground Truth Value	
		Positive (1)	***Negative (0)***
Prediction Value	***Positive (1)***	True Positive	False Positive
	Negative (0)	False Negative	True Negative

Models highlighted in blue represent the models recommended by the respective researchers. Based on the evaluation, the ELU-NET model outperforms U-NET in terms of accuracy, parameter count, execution time, and model size. Despite a parameter reduction of up to 97%, an execution time of 11.6 s, and a file size of 372.61 MB, ELU-NET maintains comparable accuracy to U-NET.

Overall, the U-NET model achieves higher accuracy than ELU-NET. The best-performing U-NET model achieved an accuracy of 0.928 (92.8%) with 32 filters and a depth of 5, whereas the best ELU-NET model attained an accuracy of 0.897 (89.7%) with 8 filters and a depth of 5. This accuracy difference was achieved by U-NET at the cost of greater resource utilization, including a higher parameter count, which led to increased execution time and model size. ELU-NET, with a 3.1% lower accuracy, reduced the number of parameters by 91.6%, resulting in a decreased execution time of 1.3 s and a model size of 85.21 MB. While these reductions in parameters, execution time, and model size are significant, they come with the trade-off of a slight reduction in accuracy. This trade-off is reasonable, given that the primary goal of ELU-NET development was to create a more efficient version of U-NET.

Focusing solely on accuracy, the U-NET model outperforms ELU-NET. This advantage is reflected in U-NET's ability to produce more precise predictions for medical image segmentation tasks. However, this accuracy advantage comes at a significant computational cost. U-NET requires a longer execution time and has a larger model size than ELU-NET due to its complex architecture, which incorporates a greater number of parameters to capture image details more comprehensively.

On the other hand, ELU-NET provides a more efficient approach in terms of execution time and model size by leveraging an enhanced skip connection mechanism from U-NET, namely the deep skip connection. Although ELU-NET yields slightly lower accuracy than U-NET, the difference is not substantial. The accuracy gap between the best models of each architecture, as reported in Table V, is approximately 3.1%.

To validate these findings, a visual inspection of the model predictions was conducted. This qualitative evaluation involved comparing the model-generated predictions against the ground truth. Although visual inspection is not the primary evaluation method, it is essential for ensuring that ELU-NET produces results closely aligned with U-NET.

TABLE V
MODEL TESTING RESULT

Architecture	Number of Filters	Number of Depth	Number of Parameters	Accuracy (Jaccard Index)	Execution Time (s)	Model Size (MB)
U-NET	32	4	1,926,149	0.903	4.3	23.26
		5	7,760,645	0.928	5.3	93.31
		6	-	-	-	-
	64	4	7,698,437	0.889	12.3	92.53
		5	31,032,837	0.829	15.6	372.61
		6	-	-	-	-
	128	4	-	-	-	-
		5	-	-	-	-
		6	-	-	-	-
ELU-NET	4	4	52,053	0.867	2.9	0.916
		5	163,109	0.851	3.9	2.25
		6	580,337	0.824	5.2	7.47
	8	4	205,769	0.854	3.4	2.76
		5	644,697	0.897	4.0	8.1
		6	2,313,537	0.830	6.3	28.26
	16	4	818,193	0.812	5.2	10.1
		5	2,570,417	0.867	5.5	31.2
		6	9,238,529	0.857	9.1	111.34

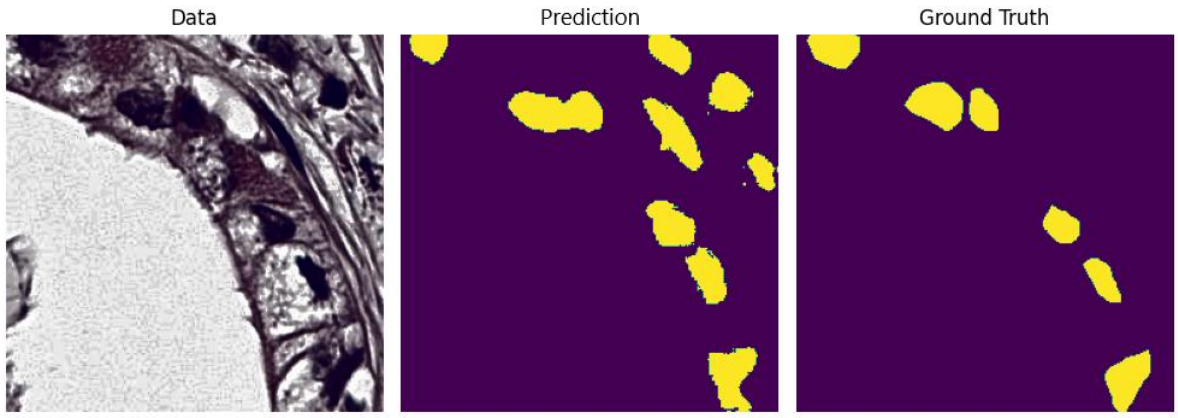

Figure 5. Prediction results with ELU-NET.

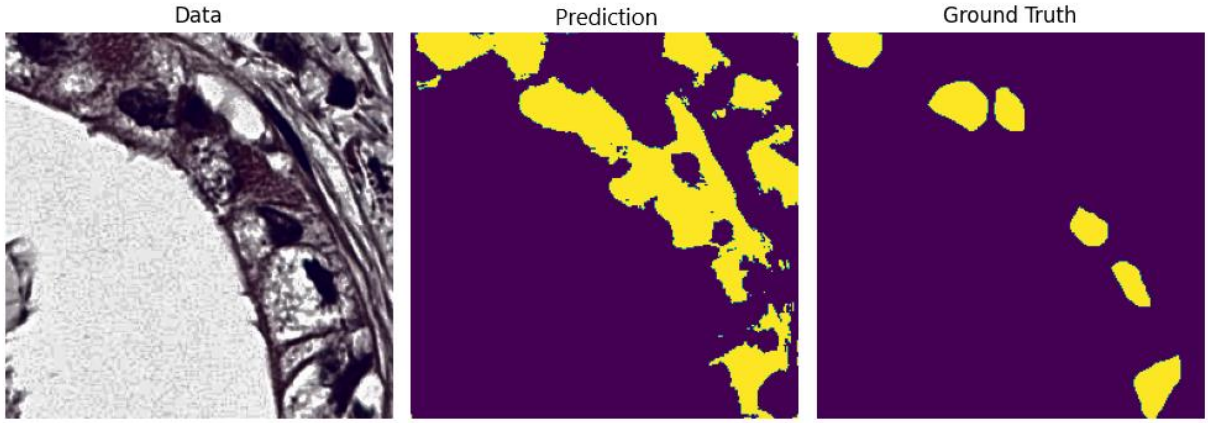

Figure 6. Prediction results with U-NET.

In Figure 5, the segmentation results generated by ELU-NET closely resemble the actual ground truth. Similarly, Figure 6 shows that U-NET's predictions are also near the ground truth. However, there are differences in how the two models handle specific regions. For example, ELU-NET performs well in certain image areas but misclassifies some small regions as cancerous cells. Conversely, U-NET also makes good predictions but, in some cases, exhibits over-segmentation by predicting areas that should not be classified as cancer cells.

Based on this qualitative assessment, the differences in predictions between the two models are not visually significant. Although U-NET achieves higher numerical accuracy, ELU-

NET still produces reasonably comparable predictions while offering computational efficiency, which should not be overlooked.

The differences in test results between the two models are primarily attributed to the development of "skip connections". The ELU-NET architecture represents an extension of U-NET that specifically focuses on enhancing the "skip connections" of the U-NET model. Previous studies, such as U-NET++, also explored modifications to U-NET, particularly in terms of skip connections. U-NET++ introduced various model configurations with different depths, demonstrating that the influence of "skip connections" allows shallower models to achieve accuracy comparable to deeper models. Previous study introduced modifications by adding and removing certain skip connections, similar to the approach taken in the U-NET++ architecture [12].

IV. CONCLUSION

Both architectures, U-NET and ELU-NET, achieved high accuracy in semantic segmentation. The best-performing models from each architecture attained accuracy exceeding 89%. Among the best models from each architecture, U-NET demonstrated higher accuracy compared to ELU-NET. However, this superior accuracy came at the cost of greater resource utilization. In contrast, ELU-NET was able to produce a model with lower resource consumption, albeit with slightly reduced accuracy compared to U-NET. This accuracy difference is considered reasonable, given that the primary goal of ELU-NET is to develop a more resource-efficient variant of U-NET while maintaining satisfactory accuracy. This is evident from the best ELU-NET model, which achieved an accuracy of 89.7%.

The dataset used in this study was relatively small for training a deep learning model, necessitating further research, particularly in collecting additional research data comprising H&E-stained pancreatic cancer images. Additionally, alternative data augmentation methods could be explored to enhance the dataset. However, from a different perspective, limited data availability could serve as an avenue for innovation in developing deep learning architectures capable of learning effectively from minimal data.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

AUTHORS' CONTRIBUTIONS

Conceptualization, Algi Fari Ramdhani, Yudi Widhiyasana, and Setiadi Rachmat; methodology, Algi Fari Ramdhani, Yudi Widhiyasana, and Setiadi Rachmat; software, Algi Fari Ramdhani; validation, Yudi Widhiyasana, and Setiadi Rachmat; formal analysis, Yudi Widhiyasana; investigation, Algi Fari Ramdhani and Yudi Widhiyasana; resources, Yudi Widhiyasana and Setiadi Rachmat; data curation, Yudi Widhiyasana; writing-original drafting, Algi Fari Ramdhani; writing-reviewing and editing, Yudi Widhiyasana and Setiadi Rachmat; visualization, Yudi Widhiyasana; supervision, Yudi Widhiyasana and Setiadi Rachmat; project administration, Yudi Widhiyasana and Setiadi Rachmat; funding acquisition, Algi Fari Ramdhani.

ACKNOWLEDGMENT

The authors express their gratitude to Politeknik Negeri Bandung for providing financial support for this research. Appreciation is also extended to the organizers of the PAIP 2023 Challenge for providing the research dataset and for the following acknowledgment: "De-identified pathology images and annotations used in this research were prepared and provided by the Seoul National University Hospital through a grant from the Korea Health Technology R&D Project via the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI18C0316)."

REFERENCES

[1] World Health Organization (WHO), "Cancer." Access date: 16-Jun-2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer

[2] J. Cai *et al.*, "Advances in the epidemiology of pancreatic cancer: Trends, risk factors, screening, and prognosis", *Cancer Lett.*, vol. 520, pp. 1–11, Nov. 2021, doi: 10.1016/j.canlet.2021.06.027.

[3] J.S. de Moor *et al.*, "Cancer survivors in the United States: Prevalence across the survivorship trajectory and implications for care," *Cancer Epidemiol. Biomarkers Prev.*, vol. 22, no. 4, pp. 561–570, Apr. 2013, doi: 10.1158/1055-9965.EPI-12-1356.

[4] S. Bharati, P. Podder, and M.R.H. Mondal, "Artificial neural network based breast cancer screening: A comprehensive review," *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, vol. 12, pp. 125–137, May 2020.

[5] D. Veiga-Canuto *et al.*, "Comparative multicentric evaluation of inter-observer variability in manual and automatic segmentation of neuroblastic tumors in magnetic resonance images," *Cancers*, vol. 14, no. 15, pp. 1–15, Aug. 2022, doi: 10.3390/cancers14153648.

[6] M.H. Aziz and A.A. Abdulla, "Computer-aided diagnosis for the early breast cancer detection," *UHD J. Sci. Technol.*, vol. 7, no. 1, pp. 7–14, Jan. 2023, doi: 10.21928/uhdjst.v7n1y2023.pp7-14.

[7] A. Rakhlin *et al.*, "Breast tumor cellularity assessment using deep neural networks," in *2019 IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, 2019, pp. 371–380, doi: 10.1109/ICCVW.2019.00048.

[8] M. Torres-Velazquez, W.-J. Chen, X. Li, and A.B. McMillan, "Application and construction of deep learning networks in medical imaging," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 2, pp. 137–159, Mar. 2021, doi: 10.1109/TRPMS.2020.3030611.

[9] E. Sudarshan *et al.*, "Deep learning for the detection and classification of brain tumors using CNN," *AIP Conf. Proc.*, vol. 2971, no. 1, Jun. 2024, Art no. 020031, doi: 10.1063/5.0196072.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. Wells., A. Frangi, Eds., Cham, Swiss: Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[11] S. Arvind, J.V. Tembhurne, T. Diwan, and P. Sahare, "Improvised light weight deep CNN based U-Net for the semantic segmentation of lungs from chest x-rays," *Results Eng.*, vol. 17, pp. 1–9, Mar. 2023, doi: 10.1016/j.rineng.2023.100929.

[12] Y. Deng, Y. Hou, J. Yan, and D. Zeng, "ELU-Net: An efficient and lightweight U-Net for medical image segmentation," *IEEE Access*, vol. 10, pp. 35932–35941, Mar. 2022, doi: 10.1109/ACCESS.2022.3163711.

[13] K. Manasa and G.V. Murthy, "Skin Cancer Detection Using VGG-16," *Eur. J. Mol. Clin. Med.*, vol. 8, no. 1, pp. 1419–1426, Jan. 2021.

[14] S. Jha *et al.*, "Neutrosophic image segmentation with Dice Coefficients," *Measurement*, vol. 134, pp. 762–772, Feb. 2019, doi: 10.1016/j.measurement.2018.11.006.

[15] R. Awan *et al.*, "Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images," *Sci. Rep.*, vol. 7, pp. 1–12, Dec. 2017, doi: 10.1038/s41598-017-16516-w.

[16] S. Roy, K. Bhalla, S. Patel, and R. Patel, "Statistical analysis of histogram equalization techniques for medical image enhancement: A brief study," *SSRN Electronic Journal*, Mar. 2022, doi: 10.2139/ssrn.4049614.

[17] X. Zhang *et al.*, "FlipCAM: A feature-level flipping augmentation method for weakly supervised building extraction from high-resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–17, Jan. 2024, doi: 10.1109/TGRS.2024.3360276.

[18] I. Saputra and D.A. Kristiyanti, *Machine Learning untuk Pemula*. Bandung, Indonesia: Informatika, 2022.

[19] H. Li *et al.*, "Keeping deep learning models in check: A history-based approach to mitigate overfitting," *IEEE Access*, vol. 12, pp. 70676–70689, May 2024, doi: 10.1109/ACCESS.2024.3402543.

[20] D. Ogwok and E.M. Ehlers, "Jaccard index in ensemble image segmentation: An approach," in *Proc. 2022 5th Int. Conf. Comput. Intelli. Intell. Syst. (CIIS '22).*, 2022, pp. 9–14, doi: 10.1145/3581792.3581794.