Hybrid Model for Medical Image Segmentation: Integration between Divide and Conquer Algorithm and Convolutional Neural Network

Salman Hanif - 13523056
Informatics Engineering Study Program
School of Electrical Engineering and Informatics
Bandung Institute of Technology, Ganesha street 10 Bandung
E-mail: samanhanif702@gmail.com , 13523056@std.stei.itb.ac.id

***Abstract*— Medical image segmentation is very important for making accurate diagnoses and planning treatments. But high-resolution images are hard to work with because regular Deep Learning models don't have enough memory or processing power. This study suggests a hybrid methodology for the segmentation of skin lesions from dermatoscopic images, employing a combination of the Divide and Conquer (D&C) algorithm and a Convolutional Neural Network (CNN). This model aims to alleviate resource limitations by partitioning extensive images into smaller sub-images. A U-Net architecture then processes each sub-image. The results of the segmentation for each sub-image are then put back together to make the full mask. The D&C approach helps with memory management and speeds up processing, and the CNN's ability to extract local features makes sure that lesion patterns can be found even when the global context is broken. The goal of this study is to show that this hybrid architecture can work and be useful in solving the problems of high-resolution medical image segmentation. This will make it possible to build more scalable AI-based diagnostic systems.**

Keywords : Medical Image Segmentation, Divide and Conquer, Convolutional Neural Network, U-Net, Deep Learning

I. INTRODUCTION

Medical image segmentation is a basic and very important step in many clinical settings. These encompass disease diagnosis, pre-operative planning, monitoring therapeutic progression, and anatomical research [1]. The ability to accurately identify and separate anatomical structures or pathologies (like lesions, tumors, or specific organs) from medical images such as Magnetic Resonance Imaging (MRI) or dermatoscopy images can significantly boost the efficiency and accuracy of medical decision-making. The accuracy of segmentation affects not only the quality of the first diagnosis but also the quality of care for each patient. This lets doctors do more precise procedures and lower the risks. In dermatology, separating skin lesions from dermatoscopy images is very important for finding melanoma and other skin conditions early on. The edges of these lesions are often hard to see, they can be different shapes and colors, and they often look like healthy skin structures around them [2]. Because of this, this job is very hard, even for dermatologists. This shows how important it is to have automated systems to help with this.

As imaging technology improves, the resolution of medical images keeps getting better and better, which means there is a lot of data and a lot of detail. Deep Learning, particularly Convolutional Neural Networks (CNNs), has shown outstanding efficacy in numerous image segmentation tasks, including those in the medical field. However, traditional CNN architectures frequently face considerable difficulties when processing high-resolution images [3]. When trying to process very high-resolution images in their entirety, the Graphics Processing Unit (GPU) memory limits and heavy computational loads are big problems that often cause out-of-memory errors or processing times that are too long. This makes researchers lower the resolution of images or use less effective patching methods, which could cause them to lose important details that are necessary for accurate diagnosis and make the segmentation less accurate overall.

To address these limitations and fully exploit the capabilities of high-resolution medical image data, this research introduces a hybrid methodology that strategically combines the Divide and Conquer (D&C) algorithm with a Convolutional Neural Network (CNN) for medical image segmentation, initially concentrating on skin lesion segmentation. The D&C strategy is used to split up big input images into smaller, easier-to-handle patches or sub-images. A CNN model, specifically the U-Net architecture, processes each sub-image separately. This model has been very successful in biomedical pixel-to-pixel segmentation tasks. After that, the segmentation results from each sub-image are put back together to make a full segmentation mask of the original image. This method makes use of CNN's strengths in deep local feature extraction and pixel classification, while also making good use of the memory and processing power needed for high-resolution images. This study seeks to establish the preliminary viability and efficacy of this hybrid architecture in delivering a more scalable, efficient, and pragmatic solution for high-resolution medical image segmentation challenges, potentially facilitating the advancement of more resilient and dependable AI-driven diagnostic systems in clinical settings.

II. THEORETICAL FOUNDATION

This research is built on a robust theoretical framework derived from computer science, particularly in the areas of computer vision and machine learning, with a focused application in medical image analysis.

A. Medical Image Segmentation

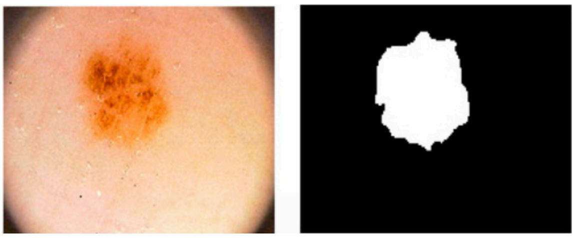

Fig. 1 Dermastopy Skin Lesion Segmentation
Source : www.mdpi.com/2075-4418/9/3/72

Image segmentation is the process of breaking a digital image into several homogeneous segments or sets of pixels based on things like colour, intensity, or texture [1]. The primary aim of segmentation is to convert the image representation into a more significant and analysable format, particularly in a clinical setting. In the medical field, segmentation's goal is to accurately separate and locate certain anatomical structures (like organs, bones, and blood vessels) or pathological anomalies (like tumours, lesions, and inflammatory areas) from the background or other structures [4]. The precision of segmentation holds significant consequences, as it directly influences diagnosis, the formulation of therapeutic strategies (including radiotherapy and surgical interventions), and the assessment of a patient's therapeutic response. In dermatology, segmenting skin lesions from dermatoscopic images is an essential procedure for the prompt identification and characterisation of melanoma and other dermatological disorders. In this context, lesion boundaries are frequently indistinct, exhibiting morphological and pigmentation heterogeneity, and may mimic healthy skin structures [2]. This complexity emphasises the necessity of creating automated systems that can enhance diagnostic accuracy.

B. ***Divide and Conquer (D&C) Algorithm***

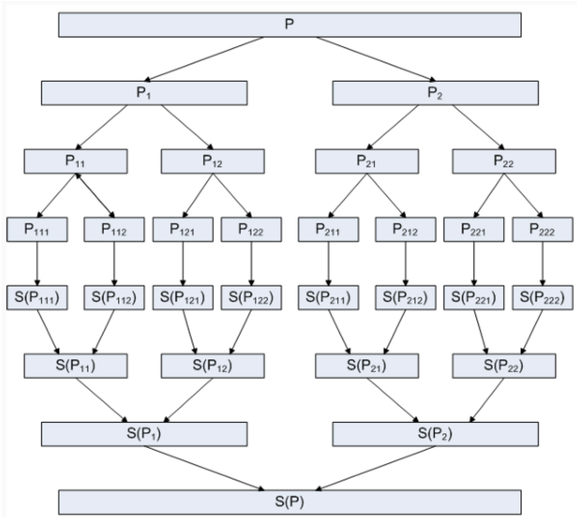

Fig. 2 Divide and Conquer Algorithm Ilustration
Source :
https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2024-2025/

The Divide and Conquer (D&C) algorithm is a reliable and effective way to solve a wide range of computational problems. This method methodically divides a big problem into a number of smaller, similar sub-problems, solves each of these sub-problems, and then combines the partial solutions to get the final answer to the original problem [5]. There are three main steps in this process:

1. Divide

   The first problem is broken down into two or more smaller problems that are much easier to solve. This division should ideally lead to sub-problems that are of the same kind as the main problem.

2. Conquer

   Then, each subproblem is solved again. If a subproblem is small enough to be solved directly (the base case), it is solved without any more recursion.

3. Combine

   The solutions from each sub-problem are put together to make a complete solution to the main problem.

When it comes to processing high-resolution medical images, using D&C is very smart. Modern medical images often have very large spatial dimensions, like 4000x4000 pixels or more, even in 3D formats. This can be more than what Graphics Processing Units (GPUs) can handle, which are often used to train deep learning models. D&C lets you break these huge images down into smaller, easier-to-handle pieces or patches. You can work on each patch on its own, and then put the results of each patch back together to make a complete segmented image. This method effectively reduces the memory and processing limitations that are often faced when dealing with very large image datasets [6].

C. *Convolutional Neural Network (CNN)*

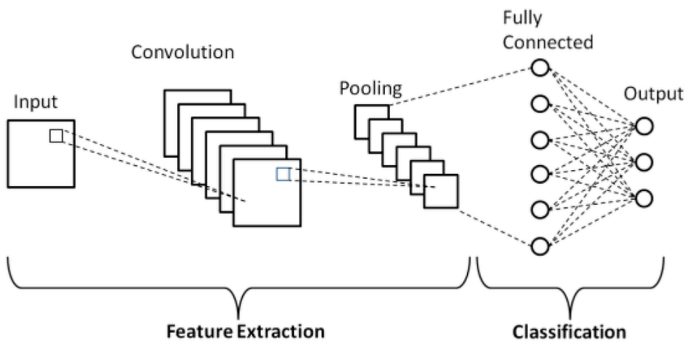

Fig. 3 Convutional Neural Network Diagram
Source : www.researchgate.com

A Convolutional Neural Network (CNN) is a type of deep neural network that is best for processing data with a grid-like structure, like images, videos, or time-series data [7]. CNNs have changed the field of computer vision because they can automatically extract and learn hierarchical feature representations from raw data, which means that manual feature engineering is no longer necessary.

There are a few important layers in the basic CNN architecture.

1. Convolutional Layers

   This layer is the most important part of CNN operations. It uses a series of small filters (or kernels) that scan the input image. Each filter is made to find certain local spatial patterns, like edges, textures, or colour gradients.

2. Pooling Layers

   After the convolutional layers, pooling layers usually lower the resolution of the feature maps by making their spatial dimensions smaller. This lowers the number of parameters, makes the model work more efficiently, and lowers the chance of overfitting. One of the most common pooling operations is Max Pooling, which finds the highest value in a small window.

3. Fully Connected Layers

   These layers are at the end of the network in classification CNN architectures. They get the high-level features that have been pulled out and use them to make the final prediction, like figuring out what objects are in a picture.

When it comes to image segmentation tasks, CNN architectures are changed so that the output has the same spatial dimensions as the input image and each pixel is put into a certain category. In this situation, architectures like U-Net are very useful and have been very successful.

D. U-Net Architecture

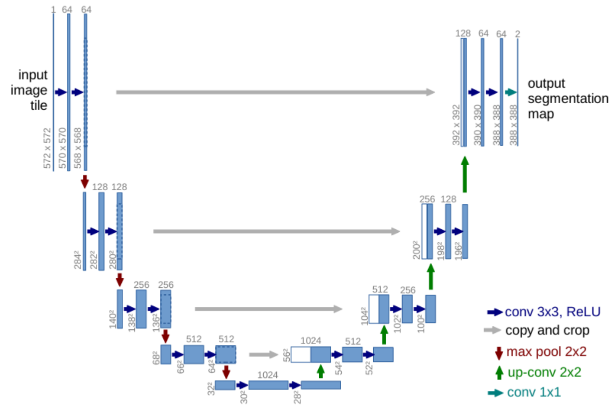

Fig. 4 U-Net Architecture
Source : https://arxiv.org/pdf/1505.04597

U-Net is one of the most popular and best-performing CNN architectures for splitting up images, especially in biomedical settings. Ronneberger, Fischer, and Brox first came up with this architecture in 2015 [8]. It is called "U-Net" because its schematic representation looks like the letter "U." U-Net's main strength is that it can make very accurate and detailed segmentations, even when it doesn't have a lot of training data to work with.

U-Net is made up of two paths that aren't the same:

1. Contracting Path

   This part works like a feature extractor, just like regular classification CNN architectures. The encoder path gradually reduces the spatial dimensions of the image by using repeated convolutional blocks and max pooling operations. At the same time, it adds more features to deeper channels. Its main goal is to get a lot of contextual information and high-level semantic features.

2. Expanding Path

   This part is in charge of making a high-resolution segmentation mask from the features that the encoder found. This part has up-sampling operations, like transposed convolution, that make the spatial dimensions bigger, and then convolutional layers.

3. Skip Connections

   Skip connections are an important part of U-Net that makes it work so well. These connections move feature maps from encoder layers to decoder layers at the same resolution level. This mechanism lets the decoder path get back small spatial details and location information that may have been lost during the pooling process in the encoder path. The outcome is an exceptionally precise segmentation mask featuring distinct and accurate object boundaries, which is a crucial element in medical image segmentatio.

This research aims to create a more scalable, efficient, and reliable medical image segmentation system to deal with the problems of high-resolution images by combining the Divide and Conquer Algorithm for handling large amounts of data and computational efficiency with U-Net's superior pixel-to-pixel image segmentation capabilities that take into account the details and context of the image. This new combination is likely to improve AI-based medical diagnosis

III. IMPLEMENTATION

A. Architecture for Hybrid Systems

The segmentation system makes use of a hybrid architecture that capitalises on each component's advantages. This method is intended to effectively handle high-resolution medical images, which can be difficult for traditional deep learning models to process because of memory and processing constraints.

In general, there are three primary stages to the workflow:

- Making smaller patches out of the input image.
- Applying a U-Net model to segment every patch.
- Putting the divided patches back together to form a full mask.

Here, the divide and conquer algorithm is used, which makes it possible to process smaller patches of data more quickly than processing an entire image in the same manner.

B. Dataset

ISIC 2016: Skin Lesion Analysis Towards Melanoma Detection is the dataset used in this study. This dataset is a component of an international challenge to create algorithms for the analysis of dermatoscopy images. In particular, the dataset includes:

- Original dermatoscopy pictures with different sizes (usually 767 x 1022 pixels with three RGB channels).
- Binary ground truth segmentation masks that correspond to each original image and indicate the location of skin lesions.
- Each image's metadata, which only serves as background information about the dataset and isn't used directly in this study.

A total of 1279 training images and their ground truth masks were used for training.

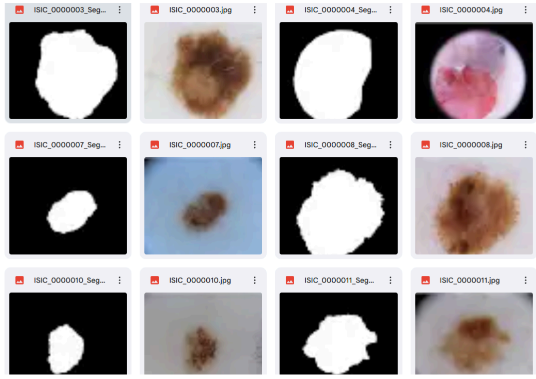

Fig. 5 Images and Segmentation Mask Dataset
Source : Dataset ISIC 2016

C. Data Pre-Processing

The purpose of data pre-processing is to apply the D&C "Divide" strategy and get the images ready to fit the input format needed by the U-Net model. The following actions were taken:

1. Normalization

   After reading each image and mask result, it is resized to the desired 256 x 256 pixel size and normalised to binary values. A channel dimension is added, and the binary mask's values are guaranteed to be either 0 or 1.

2. Patching (divide stage)

   The processed images are split into 128x128 pixel patches, creating four sub-units from a single image that the U-Net can process.

3. Dataset splitting

   The prepared image samples are transformed into sets of image patches (x_patches) and mask patches (y_patches), which are subsequently separated into sets for training (80%) and validation (20%).

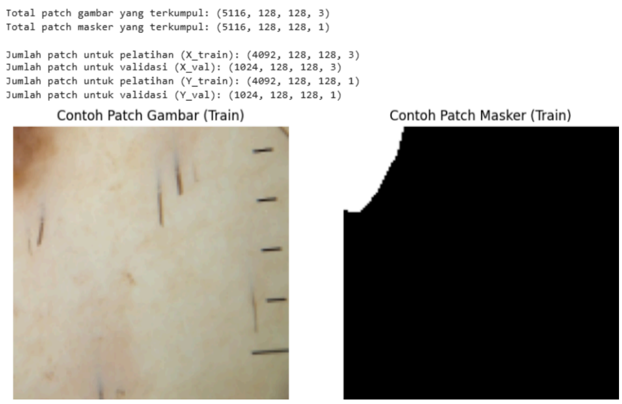

Fig. 6 Pre-processing image

Source : https://github.com/salmaanhaniif/MedicalImageSegmentation_DnC_on_CNN

D. *Design of the U-Net Model*

The U-Net architecture is the CNN model utilised for each patch's segmentation task. Because of its demonstrated ability to handle biomedical segmentation tasks—particularly the recovery of precise spatial details through its skip connections mechanism—U-Net was selected.

The specifications of the implemented U-Net architecture are as follows:

- Image patches with dimensions of 128, 128, and 3 are accepted by the input layer.
- Four convolutional and max pooling blocks with progressively more filters (64, 128, 256, and 512) make up the encoder path.
- Bottleneck: The central region that contains 1024 filters and two convolutional layers.
- Decoder Path: Consists of four convolutional and upsampling blocks with decreasing filters (512, 256, 128, 64).
- Skip Connections: At each corresponding level, features from the encoder and the decoder are concatenated.
- Output Layer: A binary segmentation mask of dimensions (128, 128, 1) is generated by a 1x1 convolutional layer with a sigmoid activation function.
- With a learning rate of 1e-4, the Adam optimiser is used to compile the model. Dice_coef and accuracy are the performance metrics tracked, and binary_crossentropy (or dice_loss, if it was modified) is used as the loss function.

E. Training Process

The patched training dataset (x_train and y_train) was used to train the U-Net model. There were 20 epochs in the training process, with a batch size of 32.

- The model was trained for 20 epochs, and the batch size was set to 32, which controls how many patches are processed before the weights of the model are updated..
- The model was trained for 20 epochs..

At the conclusion of each epoch, model performance was tracked on the validation set (x_val and y_val). The two metrics below show that in the early epochs, the dice coefficient increased and the loss graph decreased significantly, but after a certain number, the dice coefficient gradually decreased. Adding more epochs could still increase the accuracy of the model.

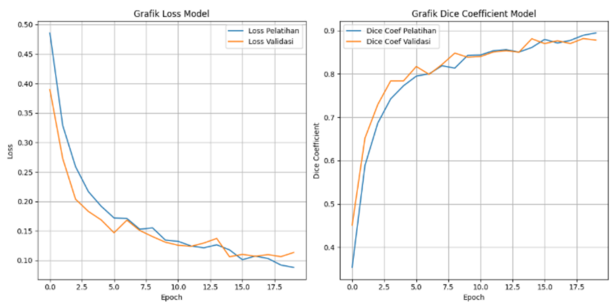

Fig. 7 Model Performance after training based on epoch

Source : https://github.com/salmaanhaniif/MedicalImageSegmentation_DnC_on_CNN

F. Recontruction

The target images to be segmented were separated once the model had been successfully trained. The "Conquer" and "Combine" stages of the D&C stages were then demonstrated by performing inference and reconstruction to generate the full segmentation results of the original input images.

The following were the steps involved in the model demonstration:

1. [Pre-Processing] An unused medical image is taken. This intact image is resized to the target size (in this case 256x256 pixels)

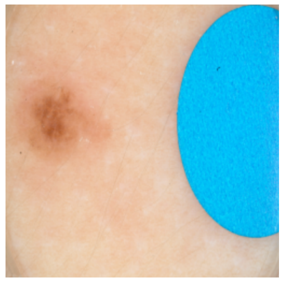

Fig. 8 Normalized image (256x256 pixel)

Sumber : https://github.com/salmaanhaniif/MedicalImageSegmentation_DnC_on_CNN

2. [Divide] The image is then split into several patches (128x128 pixels) using the same function as the pre-processing stage.

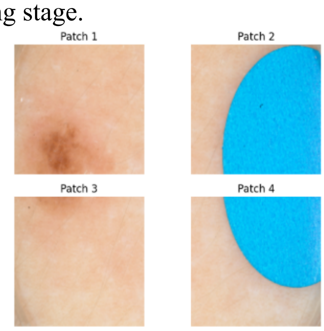

Fig. 9 Divide stage to the image

Source : https://github.com/salmaanhaniif/MedicalImageSegmentation_DnC_on_CNN

3. [Conquer] Each generated patch is then fed into a trained U-Net model to obtain its segmentation mask prediction. The probability output of the model is converted into a binary mask (value 0 or 1) with a threshold of 0.5.

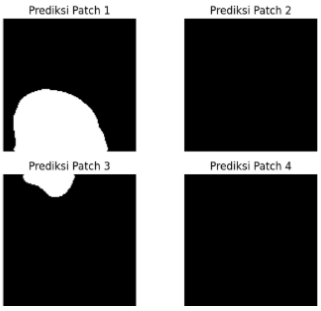

Fig. 10 Conquer Stage to the Patches
Source :
https://github.com/salmaanhaniif/MedicalImageSegmentation_DnC_on_CNN

4. [Combine] All predicted segmentation masks from each patch are then merged back together to form a complete segmentation mask of the original image. In this prototype, the merging is done by placing each patch mask at its original position on the normalized initial size mask canvas.

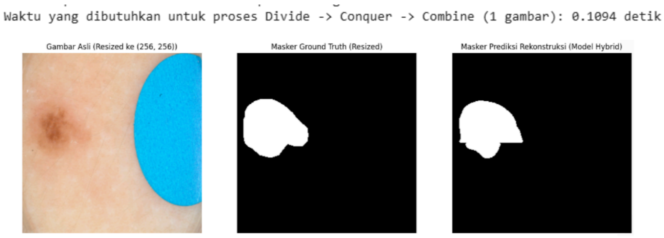

Fig. 11 Experiment Result
Source :
https://github.com/salmaanhaniif/MedicalImageSegmentation_DnC_on_CNN

IV. RESULT AND DISCUSSION

1. Result Analysis

The validation set (x_val), which is made up of image patches that were not used for training, was used to assess the trained model. Accuracy and Dice Coefficient were the main performance indicators employed.

```
Loss pada set validasi: 0.1952
Dice Coefficient pada set validasi: 0.8500
Akurasi pada set validasi: 0.9333
```

Fig. 12 Evaluation results for model number
Source :
https://github.com/salmaanhaniif/MedicalImageSegmentation_DnC_on_CNN

A Dice Coefficient value of 0.8500, as shown in the above results, suggests a respectably high level of overlap between the ground truth masks and the model's predicted masks. The model's accuracy of 0.9333 indicates that the majority of pixels were correctly classified. Although the loss is still fairly high, the model has successfully learnt segmentation patterns, as evidenced by the loss value of 0.1952, which is still comparatively above 0.15.

The model's performance can be intuitively understood by visualising the segmentation results. An example comparison between the ground truth mask, the original image, and the model-reconstructed predicted mask is shown in Figure 11.

The segmentation results visually demonstrate that the model has a reasonably high success rate in identifying and demarcating skin lesion areas. Overall, the model has been successful in predicting the locations of the lesions, despite the possibility of some flaws at the segmentation boundaries or on complex lesions. This shows that coherent segmentation masks can be produced through patch-based training and subsequent reconstruction..

2. Discussion

Medical image segmentation has been successfully accomplished by combining the Divide and Conquer Algorithm with a Convolutional Neural Network (U-Net). The U-Net model can process each segment of an image efficiently by dividing it into smaller patches. The computational efficiency of this method is demonstrated by the successful reconstruction of the segmentation mask from predicted patches, as shown by a processing time of 0.1094 seconds for an entire image. We anticipate that this will greatly improve the analysis of medical images in large files.
This prototype has certain drawbacks even though its results are encouraging. The quality of the final segmentation, particularly at lesion boundaries, may be impacted by the small size of the training data and the application of a straightforward merging technique. Future research opportunities are presented by the lack of a more thorough comparison of scalability with alternative approaches or on different hardware.

V. Conclusion

This study effectively illustrates the preliminary viability and efficacy of a hybrid Divide and Conquer (D&C) strategy using a Convolutional Neural Network (U-Net) for the segmentation of skin lesions from high-resolution dermatoscopy images. The system overcome computational memory constraints and produced coherent segmentation masks by first decomposing images using D&C and then reassembling the U-Net's predictions.
The model's capacity to detect lesions is validated by the quantitative findings and qualitative illustrations. D&C's added value in facilitating useful high-resolution medical image analysis is further highlighted by the quick processing time required to segment an entire image. Despite the dataset's relatively low resolution, the processing time demonstrates the model's viability and efficiency potential. This study offers a framework for effective high-resolution medical image segmentation, which could lead to more effective segmentation model construction for upcoming

high-resolution datasets, particularly with additional division and conquer strategy optimisation.

VI. Suggestion

Several avenues for further research can be investigated in light of the research's limitations and findings. Training the model on the full ISIC 2016 training dataset or on larger and more diverse datasets, as well as taking into account more sophisticated data augmentation techniques, can all enhance the model's performance.

One crucial step in reducing artefacts and improving the calibre of segmentation boundaries is the implementation and assessment of various Divide and Conquer tactics. Another promising avenue for more extensive clinical applications is the development of systems that combine segmentation with diagnostic classification tasks (e.g., benign/malignant).

Acknowledgment

First of all, I would like to express my gratitude to God Almighty who has given me guidance during the process of writing this paper. I would also like to thank all the lecturers of the IF2211 course, especially Dr. Nur Ulfa who has taught class K01 for one semester, not to forget also Dr. Rinaldi Munir who has provided and published lecture materials for his students on his website. I hope this paper can also provide benefits for the readers. Thank you.

References

[1] A. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, Dec. 2017.

[2] M. E. Celebi *et al.*, "Automated detection of melanoma using digital image processing: A review," *Skin Research and Technology*, Feb. 2012.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

[4] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision,* 1992.

[5] R. Munir, "Algoritma Divide and Conquer (Bagian 1)," [Online]. Available: https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2024-2025/07-Algoritma-Divide-and-Conquer-(2025)-Bagian1.pdf

[6] J. Chen *et al.*, "Medical image segmentation based on deep learning: A review," *IEEE Access*, 2020.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, May 2015.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

STATEMENT

I hereby declare that the paper I wrote is my own writing, not an adaptation or translation of someone else's paper, and is not plagiarized.

Bandung, 22 June 2025

Salman Hanif - 13523056