

## Segmentation of the Lungs on X-Ray Thorax Image with CNN Architecture U-Net

Teddi Pranata, Anita Desiani<sup>2\*</sup>, Bambang Suprihatin, Herlina Hanum, Filda Efriliyanti

*Department of Mathematics of Mathematics and Natural Science, Universitas Sriwijaya*

*\*anita\_desiani@unsri.ac.id*

### ABSTRACT

Lungs are one of the most important parts of the human body. They are very susceptible to various disorders and diseases. For this reason, it is necessary to detect or diagnose the lungs. In this study, we present a method for lung segmentation using the CNN method U-Net architecture. The initial stage was pre-processed did a 1-1 correspondence to equalize the amount of training data and testing data and resized the image so all images have the same size. The process continued with the CLAHE (Contrast Limited Adaptive Histogram Equalization), and after that, the segmentation process was carried out according to the method. This study used a dataset from the Kaggle website. The results used the CNN method of the U-Net architecture in data get an average accuracy of 91.68%, sensitivity 92.80%, and specificity 89.15%, precision 95.07, and F1-Score 93.92%. Based on the performance evaluation results, it was concluded that the method proposed in the study is great and valid in the lungs segmentation on X-Ray Thorax images.

**Keywords:** CNN, U-Net, Lung, Segmentation

### 1. INTRODUCTION

One of the most important organs in the human body is the lungs, where this organ has two main parts, namely the right and left lungs separated by the mediastinum. Lungs have a main function, namely, for gas exchange between atmospheric air and blood [1]. Apart from the importance of this organ in the human body, this organ is also prone to various health problems. This is because the interaction of the lungs with the outside world is only limited by the respiratory tract. When there is a disturbance in the lungs, the respiratory process will be hampered and the supply of oxygen to the blood will also be disrupted, among the disorders and diseases in the lungs that often occur are tuberculosis, pneumonia, lung cancer, which are currently increasing, COVID -19.

For this reason, it is necessary to treat and diagnose early so that the condition does not get worse and lead to death. In diagnosing problems that occur in the lungs, generally, the tool or technique used is to perform x-rays with x-rays on the patient's chest (thorax), the image data generated from this technique was called a thorax x-ray image. This technique was chosen because it is economical and easy to use. However, on the x-ray image of the thorax, there are other parts such as the heart, respiratory tract, blood vessels, lymph nodes, ribs, collarbone, and the upper part of the spine [2]. To overcome this, a process is needed to get the important parts to be

analyzed from an image and remove areas that are not used automatically [3]. The process is image segmentation.

Image segmentation is a process of dividing the image into homogeneous areas based on certain similarity criteria between the level of a pixel and its neighboring levels [3]. Several studies have carried out lung segmentation, including: The study by Chung et al [4] used the Atlas-Based Sparse Shape Composition method by obtained the accuracy and F1 Score values of 66% and 74%, respectively. Another study was conducted by Zhai et al [5] which used the Graph Cuts method with an F1 Score of 76%. Another study was also carried out by Saad & Hamid [6] with Edge Detection and Morphology of dilation and erosion by producing a Jaccard-Similarity value of 80.9%. Research by Mardhiyah & Harjoko [7] which used the Geometric Active Contour method produced an accuracy value of 84.19%, sensitivity of 62.88%, and specificity of 94.05%. In the study of Chung et al [4] and Zhai et al [5], the accuracy value and F1 Score obtained were still relatively small. In the study of Saad & Hamid [6] the Jaccard-Similarity value obtained was quite good, but still below 90%. In the research of Mardhiyah & Harjoko [7] the accuracy and specificity values obtained were quite good, but the sensitivity values obtained were still relatively small. These studied still used a conventional method. The conventional method has not been able to accurately distinguish an object from other objects, because the conventional method only studied the feature representation superficially [8]. To overcome this, it needs other methods that can study the feature representation in-depth to get accurate segmentation results.

One of the segmentation methods that currently being developed is Deep Learning. Deep Learning is a subfield of machine learning that attempts to abstract high levels of data by utilizing a hierarchical architecture [9]. Convolution Neural Network (CNN), often referred to as ConvNet, has a deep feed-forward architecture and has a remarkable ability to generalize in a better way compared to a fully connected network [10]. Several studies using the CNN method include: Study by Rouhi et al [11] in breast tumor segmentation get 96.47% for accuracy, 96.47% for sensitivity, and 95.94% for specificity. Another study by Wang & Deng [8] in brain tumor segmentation used Cascaded Anisotropic Convolutional Neural Network with an F1-Score value of 90.50%. The study by Thaha et al [12] in segmenting brain tumors used the Enhanced Convolutional Neural Network resulted in accuracy, sensitivity, and specificity values of 92%, 90%, and 87%, respectively. The research Rouhi et al [11] on breast tumor segmentation and Thaha et al [12] on brain tumor segmentation, the accuracy, sensitivity, and specificity values obtained were quite good, which was above 90%. However, the two studies did not include precision values and F1 scores. In the research of Wang & Deng [8] the value of the F1 Score generated is above 90%, but does not calculate the value of other performance evaluations. The performance evaluation values obtained in these studies were quite good because the CNN method has a strong ability to process large input data [13]. Based on the advantages and disadvantages of the three studies, this study used the U-Net Architecture Convolutional Neural Network method to evaluate the performance of the model in the form of accuracy, sensitivity, specificity, precision, and F1-score.

## 2. MATERIAL AND METHODS

### 2.1 COLLECTION AND DESCRIPTION OF DATA

In this study, the dataset used was obtained from the Kaggle website via the link <https://www.kaggle.com/nikhilpandey360/chest-xray-masks-and-labels>. In this dataset, there were 1600 images divided into 3 groups. The first group consisted of 800 images named CXR, the second group consisted of 704 images named masks, and the third group consisted of 96 images named test. In this study, only data in the CXR folder was used as training data and data in the masks folder as ground truth data. The training and testing image samples used in this study were shown in Table 1.

TABLE 1.  
Samples of Thorax Image and ground truth

Image Index	Data Training	Data Testing
Image 1		
Image 2		
Image 3		

### 2.2 PRE-PROCESSING DATA

At this stage, the image data used will correspond 1-1 so that the number of training data and testing data was the same. Then, the image changed using the image resize to  $256 \times 256$ .

Then the image was converted into CLAHE (Contrast Limited Adaptive Histogram Equalization) which was used to increase the contrast of an image by changing its intensity value, operates on a small area, called a tile used bilinear interpolation to eliminate regional boundaries so that small neighboring areas are smoothed [14]. The results of the image conversion to CLAHE were shown in Figure 1.

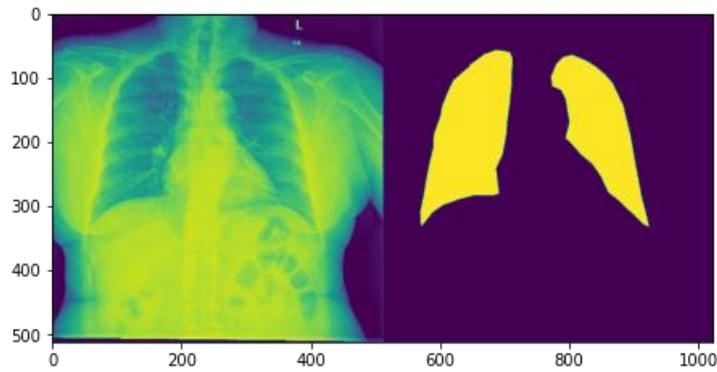


FIGURE 1. CLAHE results on the image

### 2.3 CNN

Convolutional Neural Network (CNN) is a development of the Multilayer Perceptron (MLP) which designed to process two-dimensional data. CNN is included in the type of Deep Neural Network because it has a high network depth and is widely applied to image data. In the case of image classification, MLP was not suitable for use because it did not store spatial information from image data and considers each pixel to be an independent feature, resulting in poor results [15].

### 2.4 U-NET ARCHITECTURE

U-Net architecture is build on a Fully Convolutional Network (FCN) and modified in such a way that it results in better segmentation in medical imaging. The two main advantages of U-Net over FCN are that U-Net more symmetrical and there a skip connection between the downsampling (contraction) and upsampling (expansion) lines that apply the union operator rather than addition. The purpose of this skip connection was to provide local information to global information during upsampling [16]. An illustration of the U-Net architectural process used in this research was shown in Figure 2.

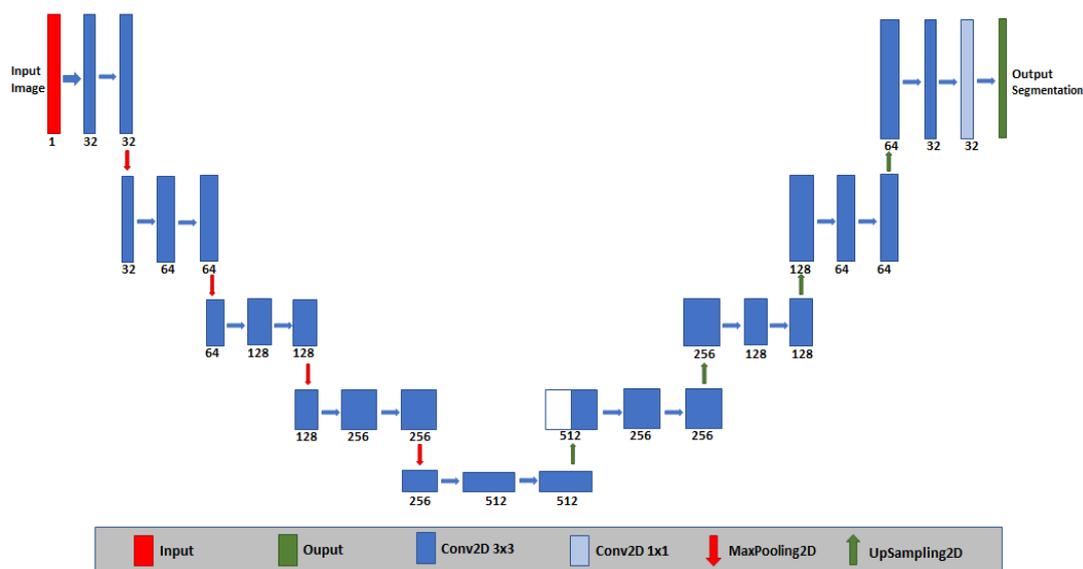


FIGURE 2. U-Net Graphics

Based on Figure 2, the image segmentation in this study used the U-Net architecture which consisted of a contraction and expansion process. The phase began with a 2D Convolution double process with a  $3 \times 3$  filter and a ReLU activation function, which brings the total feature map to 32, followed by Max Pooling with a  $2 \times 2$  kernel. The contraction stage in this study consisted of 4 blocks with the number of features maps increasing 2 times for each block. Then proceed with the fifth block which was the link between the contraction and expansion processes with the same stages as the first block without the pooling process. Next, the expansion stage began with the up-sampling process with 2D convolution transpose with a  $2 \times 2$  filter, then continued with the same process as the first block without the pooling process. This expansion phase consisted of 4 blocks with the number of feature maps reduced by 2 for each block and ending with ten blocks, with a 2D convolution process with  $1 \times 1$  filters and sigmoid activation.

## 2.5 CONVOLUTION LAYER

This operation applies the output function as the Feature Map of the input image. These inputs and outputs could be seen as two arguments of real value [17]. Formally the convolution operation could be written with equation (1) [18]:

$$a_{i,j} = \left( \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} (c_{u+i,v+j} \times k_{u+i,v+j}) \right) + b_q \quad (1)$$

## 2.6 POOLING LAYER

The pooling layer often directly follows the convolution layer on CNN. The image would be divided into several parts according to the size of the layer that had been determined. There were two methods in the Pooling Layer, namely Max Pooling (with the maximum function) and Average Pooling (with the average function). In this research, the method used was Max Pooling.

## 2.7 ReLU ACTIVATION

ReLU (Rectified Linear Unit) activation was an activation layer in the CNN model that applies the function  $f(x) = \max(0, x)$  which means this function did thresholding with a zero value to the pixel value in the image input. This activation made all pixel values that were less than zero in an image would be made 0 [19].

## 2.8 SIGMOID ACTIVATION

The sigmoid activation function had values in the range of 0 to 1 [19]. Sigmoid activation function was defined in equation (2) [19] and could be represented in Figure 3:

$$y_{\text{Sigmoid}} = \frac{1}{(1 + e^{-x})} \quad (2)$$

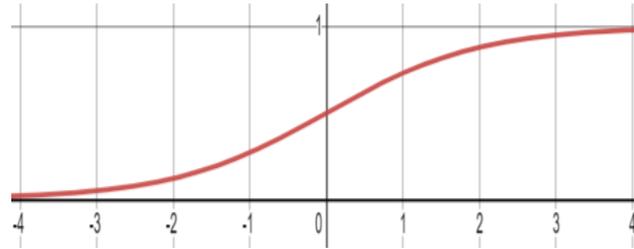


FIGURE 3. Sigmoid Activation Function Graph

## 2.9 UPSAMPLING LAYER

Upsampling layer was a learning-based method that enhances low-resolution images reconstructed from information from an external high-resolution image set [20].

## 2.10 LOSS FUNCTION

Loss function is function to evaluate the predictable results on training. This study used the binary loss function in equation (3) [21]:

$$L_{BCE} = \sum_x - (T_x \log P_x + (1 - T_x) \log (1 - P_x)) \quad (3)$$

## 2.11 TESTING

The testing process was the final process of the whole research system. The testing process was carried out to test the classification accuracy by assessing the index generated by the CNN model that had been trained [22].

## 2.12 EVALUATION OF MODEL PERFORMANCE

The results of the method used in thorax image segmentation could give a conclusion about how well the U-Net model performs in lung segmentation. The performance measured and used in this study are as follows:

- Accuracy value defined as the accuracy of data that is classified correctly after testing the segmentation results [23]. Formally the accuracy can be written with equation (4) [24].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

- Sensitivity (Recall) defined as the ratio of selected relevant items to the number of relevant items available. Formally the accuracy can be written with equation (5) [24].

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \quad (5)$$

- Specificity is a measure to see the ability of the architecture to identify TN labels for negative actual cases [25]. Formally the accuracy can be written with equation (6) [24].

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

- Precision can be interpreted as a match between requests for information and answers to requests [23]. Formally the accuracy can be written with equation (7) [24].

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

- F1-Score is the harmonic average between Sensitivity and Precision. Formally the accuracy can be written with equation (8) [24].

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \quad (8)$$

### 3. RESULTS AND DISCUSSION

In this research, the dataset used is the dataset with the training and validation ratio is 90:10. The data was trained using the U-Net architecture with the optimum function of Adam Optimizer with a learning rate of 0.00001, batch size 16 and epoch 30. The results of the training are shown in Figures 4 and 5.

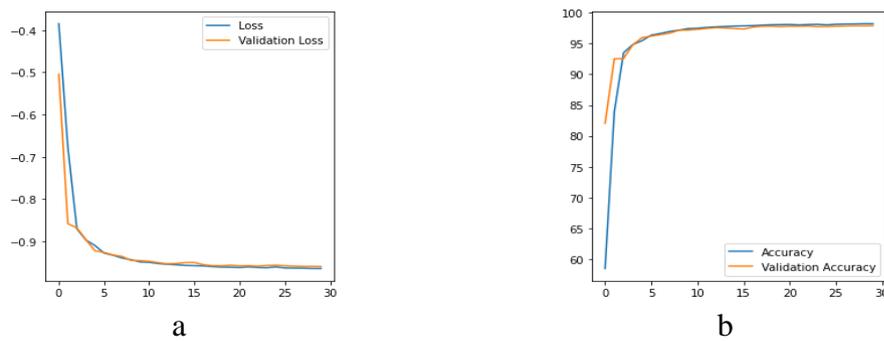
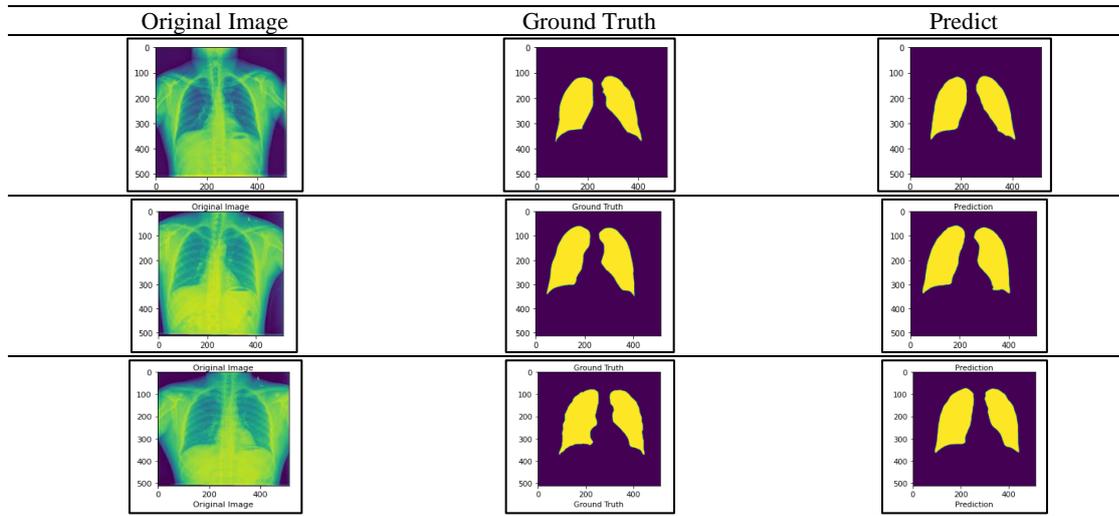


FIGURE 4. Graph Accuracy and Loss with U-Net

Based on figure 4a, the value of loss and validation loss decreased significantly at epochs 2 and 3 and at epoch 15 the values of loss and validation loss had started to be constant. With the graph of loss and validation loss that continues to decreased, it shows that the value of the training error in the U-Net model used quite good. Based on Figure 4b, the value of accuracy and validation of accuracy increased significantly at epochs 2 and 3 and epoch 10, the values of accuracy and validation accuracy had started to be constant. With graphs of accuracy and validation accuracy that continues to increased, it shows that the accuracy of the training on the U-Net model used quite good. After the training process was carried out, the results of lung segmentation would be compared with ground truth, as shown in Table 2.

TABLE 2.  
 Comparison of Segmentation Prediction Results with Ground Truth



Based on Table 2, it could be seen that the segmentation results obtained using the U-Net architecture are very close to the ground truth. From the prediction results, the values obtained are True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP), each of which was shown in Table 3.

TABLE 3.  
 Confusion Matrix

		Predict	
		Yes	No
Actual	Yes	1083384	84064
	No	56152	461588

Based on Table 3, it could be determined that the value of TP (True Positive) was 1083384, FN (False, Negative) was 84064, FP (False Positive) is 56152, and TN (True Negative) is 461588. And from the confusion matrix in Table 3, the values of Accuracy, Sensitivity (Recall), Specificity, Precision, and F1 Score can be determined. To calculate the values of accuracy, sensitivity, specificity, precision and F1 Score respectively, you can use equations (4), (5), (6), (7), and (8).

So that based on equations (4), (5), (6), (7), and (8) the value of the model's performance evaluation was obtained. The accuracy of the model used to segment the lungs is very good, which is indicated by an accuracy value of 91.68%. The model's ability to predict lung labels as information was very good, as indicated by a sensitivity value of 92.80%. The ability of the model to predict the background as the desired information was good as indicated by the specificity value of 89.15%. The accuracy of the desired information with very good prediction results is indicated by a precision value of 95.07%, and an F1 score of 93.92% describing the average harmonic between recall and precision. To compare the performance evaluation of the proposed method, it would be compared with other similar studies, it can be seen in Table 4.

TABLE 4.  
Comparisons with other journals

Journal	Method	Accuracy	Sensitivity	Specificity	Precision	F1 Score
Mardhiyah & Harjoko [7]	Geometric Active Contour	89.19%	62.88%	94.05%		
Wei et al [26]	Fully Automatic	<b>98.60%</b>				
Tang [27]	U-Net				80.40%	84.20%
Vidal et al [28]	Transfer Learning	88.13%	82.67%	<b>98.32%</b>	90.64%	83.66%
CNN U-Net Architecture (Proposed Method)	U-Net	91.68%	<b>92.80%</b>	89.15%	<b>95.07%</b>	<b>93.92%</b>

Based on Table 4, Wei et al [26] research with the Fully Automatic method had the highest accuracy value among others, namely 98.60%, but this study only described the accuracy results without the sensitivity, specificity, precision, and F1 scores. In Vidal et al [28] research with the Transfer Learning method it had the highest specificity value of 98.32%. The research by Mardhiyah & Harjoko [7] with the Geometric Active Contour method get the specificity of 94.05%, but other performance values were evaluated in these two studies still lower than the proposed method. In research Tang [27] the Precision and F1 Score obtained was still lower than the proposed method and in this study did not describe the values of accuracy, sensitivity and specificity. In the proposed method the sensitivity, precision and F1 Score values generated were higher than other studies, but the accuracy and specificity result produced were quite good, which are above 88%

#### 4. CONCLUSION

The proposed method consisted of 2 stages of image processing starting from image pre-processing and image segmentation. The results of accuracy, sensitivity, specificity, precision and F1-score on this study were above 85%. It can be concluded that U-Net as proposed method was great for lungs segmentation on chest X-Ray images. The proposed method had a great ability to separate lungs as foreground and other features on image as background. The proposed method was also balance when it used to segment between foreground and background. Form the sensitivity and specificity resulted, it can be see that the proposed method was better to segment foreground or lungs than to segment background. Based all the performance resulted, the U-Net method is great method in lung segmentation on chest X-Ray images.

#### ACKNOWLEDGEMENTS

The author would like to thank the Kaggle project team for making this image database publicly available on the Internet. To Computation laboratory of FMIPA Universitas Sriwijaya which has provided facilities for this research. The author also thank to Talenta Inovasi Kemendikbud Dikti program 2021 which has provided funding assistance for the study.

## REFERENCES

- [1] A. E. Oja *et al.*, “Trigger-happy resident memory CD4 + T cells inhabit the human lungs,” *Mucosal Immunol.*, vol. 11, no. 3, pp. 654–667, 2018, doi: 10.1038/mi.2017.94.
- [2] S. Lika Aprilia, “Rontgen Thorax,” 2021. <https://helohehat.com/pernapasan-rontgen-dada/>.
- [3] A. Desiani, B. Suprihatin, S. Yahdin, A. I. Putri, and F. R. Husein, “Bi - path Architecture of CNN Segmentation and Classification Method for Cervical Cancer Disorders Based on Pap - smear Images,” *IAENG Int. J. Comput. Sci.*, vol. 48, no. 3, p. 37, 2021.
- [4] H. Chung, H. Ko, S. J. Jeon, K. H. Yoon, and J. Lee, “Automatic Lung Segmentation with Juxta-Pleural Nodule Identification Using Active Contour Model and Bayesian Approach,” *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, no. October 2017, pp. 1–13, 2018, doi: 10.1109/JTEHM.2018.2837901.
- [5] Z. Zhai, M. Staring, and B. C. Stoel, “Lung vessel segmentation in CT images using graph cuts,” *STATS*, no. March, pp. 1–7, 2016, doi: 10.1117/12.2216827.
- [6] M. N. Saad and H. A. Hamid, “Image Segmentation for Lung Region in Chest X-ray Images using Edge Detection and Morphology,” *IEEE Int. Conf. Controy Syst.*, no. November, pp. 28–30, 2014.
- [7] A. Mardhiyah and A. Harjoko, “Metode Segmentasi Paru-paru dan Jantung Pada Citra X-Ray Thorax,” *IJEIS (Indonesian J. Electron. Instrum. Syst.*, vol. 1, no. 2, pp. 35–44, 2011, doi: 10.22146/ijeis.1961.
- [8] M. Wang and W. Deng, “Deep visual domain adaptation: A survey,” *Neurocomputing*, vol. 312, no. April, pp. 135–153, 2018, doi: 10.1016/j.neucom.2018.05.083.
- [9] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016, doi: 10.1016/j.neucom.2015.09.116.
- [10] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, “Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach,” *Procedia Comput. Sci.*, vol. 132, pp. 679–688, 2018, doi: 10.1016/j.procs.2018.05.069.
- [11] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, “Benign and malignant breast tumors classification based on region growing and CNN segmentation,” *Expert Syst. Appl.*, vol. 42, no. 3, pp. 990–1002, 2015, doi: 10.1016/j.eswa.2014.09.020.
- [12] M. M. Thaha, K. P. M. Kumar, B. S. Murugan, S. Dhanasekeran, P. Vijayakarhick, and A. S. Selvi, “Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images,” *J. Med. Syst.*, vol. 43, no. 9, p. 294, 2019, doi: 10.1007/s10916-019-1416-0.
- [13] A. Desiani, M. Erwin, B. Suprihatin, S. Yahdin, A. I. Putri, and F. R. Husein, “Bi-path Architecture of CNN Segmentation and Classification Method for Cervical Cancer Disorders Based on Pap-smear Images,” *IAENG Int. J. Comput. Sci.*, vol. 48, no. 3, pp. 1–9, 2021.
- [14] R. D. P. Olvera, E. M. Zerón, J. C. P. Ortega, J. M. R. Arreguín, and E. G. Hurtado, “A Feature Extraction Using SIFT with a Preprocessing by Adding CLAHE Algorithm to Enhance Image Histograms,” *Proc. - 2014 IEEE Int. Conf. Mechatronics, Electron. Automot. Eng. ICMEAE 2014*, pp. 20–25, 2015,

- doi: 10.1109/ICMEAE.2014.41.
- [15] W. S. Eka Putra, “Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101,” *J. Tek. ITS*, vol. 5, no. 1, 2016, doi: 10.12962/j23373539.v5i1.15696.
- [16] I. Bagus, L. Mahadya, M. Sudarma, I. N. S. Kumara, and A. Optimizer, “Resonance Imaging Dengan Menggunakan Metode U-NET,” *Majalah Ilmiah Teknologi Elektro*, vol. 19, no. 2, pp. 151–156, 2020.
- [17] S. Albawi, T. A. M. Mohammed, and S. Alzawi, “Layers of a Convolutional Neural Network,” *Ieee*, p. 16, 2017.
- [18] W. Chen, B. Yang, J. Li, and J. Wang, “An approach to detecting diabetic retinopathy based on integrated shallow convolutional neural networks,” *IEEE Access*, vol. 8, pp. 178552–178562, 2020, doi: 10.1109/ACCESS.2020.3027794.
- [19] S. Sharma, S. Sharma, and A. Athaiya, “Activation Functions in Neural Networks,” *Int. J. Eng. Appl. Sci. Technol.*, vol. 04, no. 12, pp. 310–316, 2020, doi: 10.33564/ijeast.2020.v04i12.054.
- [20] J. Lin, D. Liu, H. Yang, H. Li, and F. Wu, “Convolutional Neural Network-Based Block Up-Sampling for HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3701–3715, 2019, doi: 10.1109/TCSVT.2018.2884203.
- [21] U. Ruby, P. Theerthagiri, Jacob Jeena, and Vamsidhar, “Binary cross entropy with deep learning technique for Image classification,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5393–5397, 2020, doi: 10.30534/ijatcse/2020/175942020.
- [22] K. H. Mahmud *et al.*, “Classification of plant leaf diseases based on improved convolutional neural network,” *Neurocomputing*, vol. 3, no. 1, pp. 467–488, 2019, doi: 10.1016/j.eswa.2018.11.008.
- [23] S. Yahdin, A. Desiani, N. Gofar, K. Agustin, and D. Rodiah, “Application of the Relief-f Algorithm for Feature Selection in the Prediction of the Relevance Education Background with the Graduate Employment of the Universitas Sriwijaya,” *Comput. Eng. Appl.*, vol. 10, no. 2, pp. 71–80, 2021, [Online]. Available: <https://comengapp.unsri.ac.id/index.php/comengapp/article/view/369>.
- [24] B. Oltu, B. K. Karaca, H. Erdem, and A. Özgür, “A systematic review of transfer learning based approaches for diabetic retinopathy detection,” 2021, [Online]. Available: <http://arxiv.org/abs/2105.13793>.
- [25] D. M. W. Powers, “Evaluation Evaluation a Monte Carlo study,” pp. 1–5, 2015, [Online]. Available: <http://arxiv.org/abs/1504.00854>.
- [26] Y. Wei, G. Shen, and J. J. Li, “A fully automatic method for lung parenchyma segmentation and repairing,” *J. Digit. Imaging*, vol. 26, no. 3, pp. 483–495, 2013, doi: 10.1007/s10278-012-9528-9.
- [27] Y. Tang, “XLSor : A Robust and Accurate Lung Segmentor on Chest X-Rays Using Criss-Cross Attention and Customized Radiorealistic Abnormalities Generation,” pp. 457–467, 2019.
- [28] L. Vidal, J. De Moura, J. Novo, and M. Ortega, “Multi-stage transfer learning for lung segmentation using portable X-ray devices for patients with COVID-19,” *Expert Syst. Appl.*, vol. 173, p. 8, 2021, doi: 10.1016/j.eswa.2021.114677.