Fuzzy Systems and Data Mining X
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA241412

# Classification Using U-Net CN on Multi-Resolution CT Scan Image

Sugiyarto Surono [a,1], Muhammad Rivaldi [a] and Nursyiva Irsalinda [a]
[a] *Department of Mathematics, FAST UAD Yogyakarta Indonesia*
*ORCiD ID: Sugiyarto Surono https://orcid.org/0000-0001-6210-7258*
Nursyiva Irsalinda https://orcid.org/0000-0001-7622-5060

**Abstract.** Image processing has become a central topic in the era of big data, particularly within computer vision, due to the growing volume and diverse resolutions of images. Low-resolution images introduce uncertainty, underscoring the need for high-performance classification methods. Convolutional Neural Networks (CNN), especially the U-Net architecture, are widely applied for pixel-level segmentation due to their encoder-decoder structure. This study applied U-Net on a CT scan image dataset to segment lung images, followed by a CNN classifier to classify lung cancer stages (I, II, IIIa, IIIb). The U-Net model outperformed standard CNNs, achieving 99% in accuracy, precision, sensitivity, and F1 score, compared to the conventional CNN's 97%, 95%, 97%, and 96%, respectively.

**Keywords.** U-Net; Image Segmentation; CNN; Image Classification; Multi-resolution Image Processing

## 1. Introduction

Technological developments have increased the amount of data in the form of numerical data, sound data, video data and data in the form of images. The image data is generated by various tools such as cameras with different specifications. Not all image capture tools have good specifications resulting in poor-quality images [1]. These problems also tend to increase the amount of image data, which is more diverse and cause the classification method to produce low accuracy values [2]. This provides a great opportunity for researchers to explore research related to image data quality improvement techniques which implies a quality classification model that is needed to achieve high performance. As stated in [3], the multi-resolution image classification model tends to enhance the effectiveness of the image classification process when image data, obtained from various optical sensors with differing resolutions, increases [3]. This makes the model more flexible since image classification applies to images with different resolutions.

Computer Vision frequently plays a central role in image classification, with Deep Learning as a commonly utilized approach in its development [4]. According to [5], Deep Learning methods have achieved success across numerous fields. Deep Learning, or Deep Neural Networks (DNNs), refers to intricate, multi-layered Artificial Neural Networks (ANNs) and encompasses commonly used models like Convolutional Neural Networks (CNNs), which are among the most popular techniques in the field of Deep

[1] Corresponding Author, Sugiyarto Surono, FAST UAD Yogyakarta Indonesia; Email: sugiyarto@math.uad.ac.id

Neural Networks [6]. Convolutional Neural Networks (CNNs) are widely applied in image processing and computer vision tasks because they can automatically extract hierarchical features from image data, minimize computational complexity by sharing parameters, and deliver exceptional performance in tasks like image classification, object detection, and segmentation [7]. CNN has shown significant advancements, particularly in tasks related to image processing and computer vision [8], making it well-suited for the rapid developments in Artificial Intelligence technology, which has recently gained popularity, especially in image processing applications. Image processing offers numerous advantages, including its use in classification, clustering, segmentation, and other. The technique for segmenting images based on information about each image pixel used to predict the category of each predicted image pixel called Semantic segmentation.

Semantic segmentation provides pixel-level categorical information, which has proven valuable in numerous real-world applications, including self-driving cars, pedestrian detection, disability identification, therapy planning, and computer-assisted diagnosis [9]. According to [10], it assigns a label to each pixel in an image. This pixel-level semantic information data enables intelligent systems to comprehend spatial positions and make critical decisions. Deep CNNs have been extensively applied in the medical field to address challenges in biomedical segmentation [11]. Recently, CNNs, particularly in the context of segmentation models [12], have seen significant advancements, with U-Net CN being one notable architecture [13]. U-Net, introduced in 2015, utilizes an encoder-decoder architecture that employs skip connections to merge high-level semantic features from the low-level decoders in the encoder [12]

Biomedical image processing is a rapidly developing field that encompasses the acquisition of biomedical signals, the generation of images processing and display for medical diagnosis are derived from these signals. Clinical imaging systems combine hardware and software, with the growing number of medical imaging sensors attracting significant attention [14]. Notably, image classification techniques are commonly used in medical imaging to analyze and interpret medical images [15]. As highlighted in [16], lung cancer is the most lethal type of cancer, with non-small cell lung cancer (NSCLC) often diagnosed at advanced stages. In 2018, lung cancer accounted for 18.4% of all cancer-related deaths worldwide [17]. Hence, there is a critical need for methods to detect NSCLC early, enabling more effective treatment.

An essential process in training machine learning models is tunneling. This process allows the model to achieve good results through parameter selection or hyperparameters [18], [19]. In previous research, the CNN model applied ADAM optimization to training the model. The result of selecting the right hyperparameter increases the speed of training the model [20].

This study explores the use of the CN U-Net classification method for multi-resolution CT scan images of lung cancer. Building on the work in [21], which suggested using multi-resolution image data for image processing, this approach focuses on multi-resolution image classification. The first part of the study discusses lung image segmentation using the U-Net architecture. Next, the application of the CNN classification method for determining the stages of non-small cell lung cancer (NSCLC) is presented.

## 2. Dataset

The data is sourced from NSCLC-Radiomics, accessible at https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics. The data used is licensed under CC BY 3.0 and publicly accessible. A sample image from the NSCLC-Radiomics dataset containing CT scan image data, lung mask images, and metadata from patients with Non-Small Cell Lung Cancer (NSCLC) is shown in Figure 1. Furthermore, the NSCLC-Radiomics data were obtained from 421 non-small cell lung cancer (NSCLC) patients with 4 grades of disease. The image data were of the Dicom data type, and approximately 20 lung CT scan images and 1 lung mask image were used for each patient. Next, the data were then tested using the CN U-Net classification method. The data contains 1773 images with 338 class I training data, 480 class II, 436 class IIIa, and 469 class IIIb. In comparison, the 1:9 test data has 197 images with 91 class I data, 29 class II, Class IIIa 77, and IIIb 120. The data contained low-severity Stage I Cancer, moderate-severity Stage II Cancer, medium-severity Stage IIIa Cancer, and high-severity Stage IIIb Cancer.

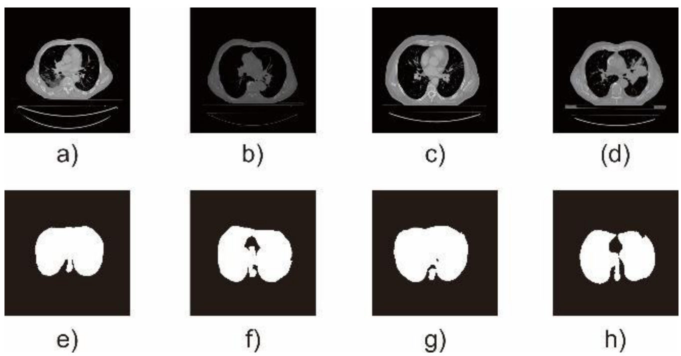


**Figure 1.** CT scan and mask preview. a) - d) is the CT scan images of stage I, II, IIIa, and IIIb respectively of NSCLC. e) – h) is the lung masks for image a), b), c), and d) respectively.

## 3. Method

This section explains how the dataset passes through the data preparation stage before being processed by the U-Net Network. The lungs in the image dataset were segmented using the U-Net network and later cropped in the lung area. The segmentation results were then classified with CNN.

### *3.1. Data Preparation*

Define Before the segmentation stage, the data were prepared to ensure that the input could also be used in the network. The first step was to resize the resolution of the entire CT scan and mask the images to ensure they were uniform. The second step was to normalize the data using the min-max scaler.

- Resize Data

  Multi-resolution data has been resized to a uniform resolution of 256x256. This allows the data to enter the U-Net network for segmentation.

The min-max scaler is a normalization method used to transform data values into a range of probabilities, typically from 0 to 1, based on the minimum and maximum values of the data [22]. This technique is represented by Equation 1.

$$x_{scaled} = \frac{x_i - x_{min}}{x_{max} - x_{min}}, \tag{1}$$

### 3.2. U-Net Segmentation

As deep neural networks evolve, the semantic segmentation model's performance improves. This has led to several developments to overcome limitations in this area [23]. According to [24], U-Net is a semantic segmentation network architecture built on fully convolutional networks, consisting of two main layers: the encoder and the decoder. In this study, U-Net architecture was applied to data that underwent the preparation stage, with its structure illustrated in Figure 2 and detailed in Tables 1 and 2.

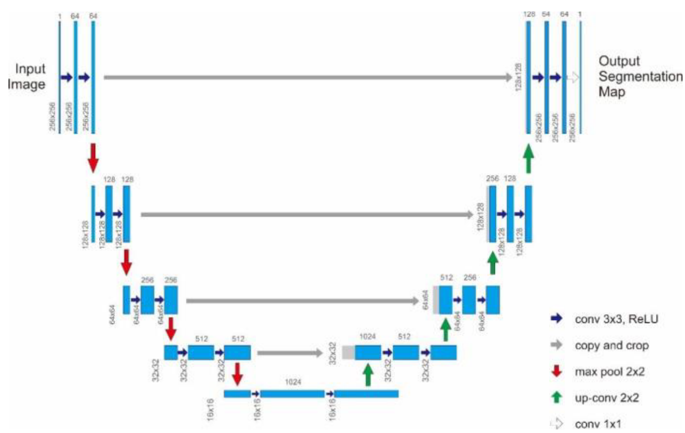


**Figure 2.** U-Net architecture for image segmentation..

### 3.3. CNN Classification

Once the image data was segmented and cropped, it was subjected to a classification process using CNN, a technique that has seen significant growth in the field of computer vision. As noted in [25], CNN consists of several layers of neural network connections that require minimal systematic processing. Due to its crucial role, CNN has been widely applied in various image processing tasks, including image recognition, segmentation, and object detection [26].

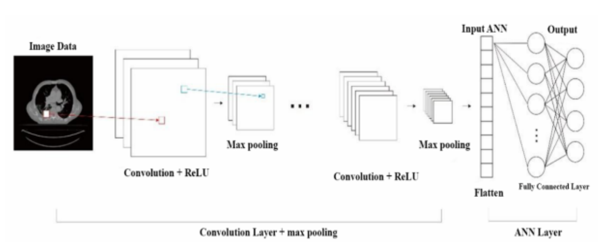


**Figure 3.** CNN architecture for image Classification.

## 4. Discussion And Experimental Result

### *4.1. Experimental Setup*

The proposed method was implemented using Python in a Jupyter Notebook environment on a computer with a 2-core Intel(R) Xeon(R) CPU running at 2.20GHz and 32GB of RAM.

### *4.2. Training Process*

Image segmentation and classification task, respectively. In the 20 epochs' U-Net training stage, the CT scan multi-resolution image data were used as the input data, while the mask image was utilized as the label for the prediction. Furthermore, the input image was cropped based on the appropriate area on the mask image. Meanwhile, in the CNN testing stage with 4 epochs, the image data was trimmed, and the predicted mask generated on the U-Net network was utilized as input and output in the form of 4 classes, namely stages I, II, IIIa, and IIIb. This model is trained with Adam optimization, setting hyperparameter bath size = 16 and learning rate = 0.001.

### *4.3. Result*

The results demonstrate the performance of the U-Net CN model in classifying image data and its comparison with the CNN model used for classifying lung cancer stages. In addition to this comparison, the U-Net segmentation model was assessed by evaluating how effectively the multi-resolution data was resized to 256×256. The evaluation was based on the loss value, along with metrics such as accuracy, precision, recall, and F1 score. The computational time was also measured to assess the performance of the proposed classification model. A confusion matrix was employed to determine the accuracy, precision, recall, and F1 score values, providing insights into the model's prediction performance. Table 1 and Figure 4 illustrate the loss value changes for each epoch during the application of the segmentation model, showing both validation and training loss over 20 epochs.

TABLE 1 SEGMENTATION MODEL PERFORMANCES

| Epoch | Dice Loss | Dice Coefficient | Time (seconds) |
|---|---|---|---|
| 1 | 0.4580 | 0.5420 | 42 |
| 2 | 0.2764 | 0.7236 | 25 |
| 3 | 0.1133 | 0.8867 | 25 |
| 4 | 0.0862 | 0.9138 | 27 |
| 5 | 0.0725 | 0.9275 | 26 |
| 6 | 0.0583 | 0.9417 | 26 |
| 7 | 0.0458 | 0.9542 | 26 |
| 8 | 0.0376 | 0.9624 | 27 |
| 9 | 0.0308 | 0.9692 | 27 |
| 10 | 0.0231 | 0.9769 | 26 |
| 11 | 0.0199 | 0.9801 | 27 |
| 12 | 0.0179 | 0.9821 | 27 |
| 13 | 0.0161 | 0.9839 | 26 |
| 14 | 0.0149 | 0.9851 | 27 |
| 15 | 0.0136 | 0.9864 | 27 |
| 16 | 0.0126 | 0.9874 | 26 |
| 17 | 0.0119 | 0.9881 | 26 |

| | | | |
|---|---|---|---|
| 18 | 0.0113 | 0.9887 | 27 |
| 19 | 0.0106 | 0.9894 | 26 |
| 20 | 0.0101 | 0.9899 | 26 |
| | | | 542 |

The training results show slight dice loss, and the dice coefficient is close to 1.0000. This indicates that the model has a small error, and the processed pixel values have perfect consistency so that the model can distinguish the predictions of each class well.

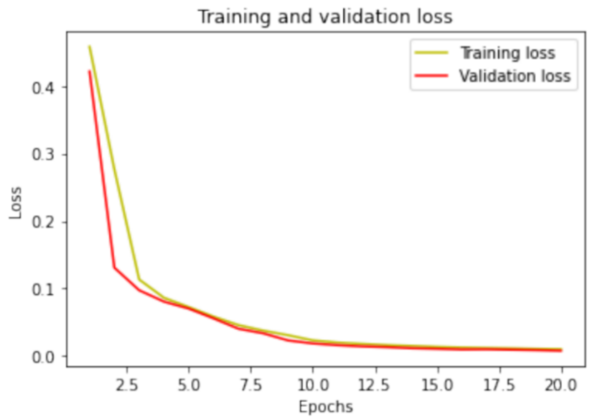


**Figure 4.** Training loss of the U-Net segmentation model.

Figure 4 shows the prediction results of the mask image generated by the model trained on the ratio and resolution type of the dataset. It also shows the results of cropping the CT scan image in the area corresponding to the pixels of the mask image. It was observed from the change in the loss value of each epoch in the training data that as the training ratio increased, the difference in loss from one generation to another became more stable. After obtaining the cropped image in the U-Net CN model using the predicted mask image, the CNN was used for classification. In contrast, the image data was used without performing the segmentation process. The confusion matrix of the U-Net CN and CNN models showing the numbers of correct predictions is also shown in Figure 6.

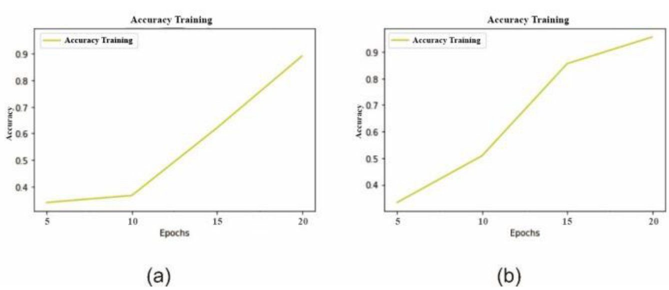


**Figure 5.** Accuracy for classification results using training data. a) CNN training accuracy. b) U-Net CN training accuracy.

Figure 5 (b) shows that there is an increase in the accuracy at epoch 5. Compared to the CNN model, the U-Net CN model has a significant increase in accuracy that occurs from epoch 5 until epoch 20.

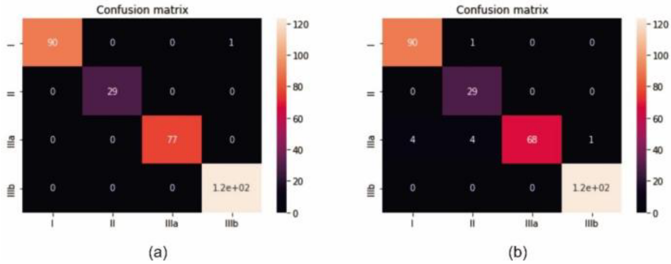

**Figure 6.** Confusion matrix for classification results using testing data. a) U-Net CN confusion matrix. b) CNN confusion matrix.

Figure 6 shows the confusion matrices of two CNN models and U-Net CN model. The results show that U-Net CN has better testing accuracy. This is due to the more negligible data error bias. After obtaining the value of the confusion matrix, the model will be calculated using several metrics. Based on the results of the matrix calculation, the classification can be determined. Table 2 and Table 3 show the classification results of data testing.

TBALE 2. U-NET CN PERFORMANCE USING TESTING DATA

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| I | 99% | 100% | 99% | 100% |
| II | 99% | 100% | 100% | 100% |
| IIIa | 99% | 100% | 100% | 100% |
| IIIb | 99% | 99% | 100% | 99% |

TBALE 3 CNN PERFORMANCE USING TESTING DATA

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| I | 97% | 96% | 99% | 97% |
| II | 97% | 85% | 100% | 92% |
| IIIa | 97% | 100% | 88% | 94% |
| IIIb | 97% | 99% | 100% | 100% |

Based on Tables 2 and Table 3 it can be seen that the metrics produced by the U-Net CN model are much better by producing a larger average than the CNN model. This aligns with the results from the confusion matrix of the U-Net CN model, which exhibits minimal prediction bias. The improvement in accuracy can be attributed to the use of an appropriate segmentation method. U-Net CN uses encoder-decoder based segmentation so that the image to be classified becomes clearer through this process. The pixel points will be uniformed with their respective classes through an encoder-decoder process. Due to the two stages, the encoder, the time required to train the model is somewhat time consuming. From the results of this training, the model that has been created can be made to predict future events and can reduce these cases. This aligns with the findings from research on adaptive moment estimation, which aims to minimize squared error in the backpropagation algorithm [26].

## 5. Conclusions

The analysis led to the following conclusions: During the data testing phase, the U-Net CN classification method demonstrated outstanding performance, achieving an average of 99% in accuracy, precision, sensitivity, and F1 scores, outperforming the CNN method, which had an average of 96.25%.Various problems, including the choice of the CNN model, the activation function, and hyperparameters in the tunneling step, influence the acquisition of model metrics. Considering the limitations of this study, we suggest that further research use segmentation methods, parameters, and activation functions that can increase accuracy and speed up training time.

## Acknowledgment

The authors are grateful to the Ministry of Education and Culture for funding the PUDPT research for the fiscal year 2022 with contract number 029/PB.PDUPT/BRIn.LPPM/VI/2022.

## Funding Statement

This research was funded by the Ministry of Education and Culture for the fiscal year 2022 with the contract number 157/E5/PG.02.00.PT/2022,1989.10/LL5-INT/PG.02.00/2022, 029/PB.PDUPT/BRIn.LPPM/VI/2022.

## Conflicts of Interests

This research has no conflict of interest regarding the publication of this manuscript.

## References

[1] Singh S, Mittal N, Singh H. Multifocus image fusion based on the multiresolution pyramid and bilateral filter. IETE Journal of Research. 2020 Jan; 68(4): 2476–2487, doi: 10.1080/03772063.2019.1711205

[2] Peryanto, A., Yudhana, A. & Umar, R. Klasifikasi Citra Menggunakan Convolutional Neural Network dan K Fold Cross Validation. J. Appl. Informatics Comput. 4, 45–51 (2020).

[3] Duarte, D., Nex, F., Kerle, N. & Vosselman, G. Multi-resolution feature fusion for image classification of building damages with convolutional neural networks. Remote Sens. 10, 10 (2018).

[4] Voulodimos, A., Doulamis, N., Doulamis, A. & Protopapadakis, E. Deep learning for computer vision: A brief review. Comput. Intell. Neurosci. (2018).

[5] Soydaner, D. A comparison of optimization algorithms for deep learning. Int. J. Pattern Recognit. Artif. Intell. 34, (2020).

[6] Albawi, S., Bayat, O., Al-Azawi, S. & Ucan, O. N. Social touch gesture recognition using convolutional neural network. Comput. Intell. Neurosci. (2018).

[7] Sharma, N., Jain, V., & Mishra, A. (2018). An Analysis of Convolutional Neural Networks for Image Classification. *Procedia Computer Science*, *132*, 377–384. https://doi.org/10.1016/j.procs.2018.05.198

[8] Khan, A., Sohail, A., Zahoora, U. & Qureshi, A. S. A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 53, (2020).

[9] Hao, S., Zhou, Y., & Guo, Y. (2020). A Brief Survey on Semantic Segmentation with Deep Learning. *Neurocomputing*, *406*, 302–321. https://doi.org/10.1016/j.neucom.2019.11.118

[10] Hao, S., Zhou, Y. & Guo, Y. A brief survey on semantic segmentation with deep learning. Neurocomputing 406, (2020).

[11] Salehi, S. S. M., Erdogmus, D. & Gholipour, A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. Int. Work. Mach. Learn. Med. imaging (2017).

[12] Huang, H. et al. Unet 3+: A full-scale connected unet for medical image segmentation. ICASSP 2020-2020 IEEE Int. Conf. Acoust. Speech Signal Process. (2020).

[13] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. Int. Conf. Med. image Comput. Comput. Interv. (2015).

[14] Nair, R. R. & Singh, T. Multi ‐ sensor medical image fusion using pyramid ‐ based DWT: a multi ‐ resolution approach. IET Image Process. (2019).

[15] Yadav, S. S. & Jadhav, S. M. Deep convolutional neural network based medical image classification for disease diagnosis. J. Big Data (2019).

[16] Berghmans, T., Durieux, V., Hendriks, L. E. & Dingemans, A. M. Immunotherapy: from advanced NSCLC to early stages, an evolving concept. Front. Med. (2020).

[17] Arbour, K. C. & Riely, G. J. Systemic therapy for locally advanced and metastatic non–small cell lung cancer. Jama (2019).

[18] Elbaz, K., Yan, T., Zhou, A. & Shen, S.-L. Deep learning analysis for energy consumption of shield tunneling machine drive system, Tunnelling and Underground Space Technology. Elsevier (2022)

[19] Shen, S., Elbaz, K. & Shaban, W. M. Real-time prediction of shield moving trajectory during tunnelling. Acta Geotech (2022) doi:https://doi.org/10.1007/s11440-022-01461-4.

[20] Yan, T., Shen, S.-L., Zhou, A. & Chen, X. Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm. J. Rock Mech. Geotech. Eng. (2022).

[21] Soomro, T. A. et al. Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. Expert Syst. Appl. (2019).

[22] Phaladisailoed, T. & Numnonda, T. Machine learning models comparison for bitcoin price prediction. 2018 10th Int. Conf. Inf. Technol. Electr. Eng. (2018).

[23] Zheng, S. et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. Proc. IEEE/CVF Conf. Comput. Vis. pattern Recognit. (2021).

[24] Xiao, Z., Liu, B., Geng, L., Zhang, F. & Liu, Y. Segmentation of lung nodules using improved 3D-UNet neural network. Symmetry (Basel). (2020).

[25] Ting, F. F., Tan, Y. J. & Sim, K. S. Convolutional neural network improvement for breast cancer classification. Expert Syst. Appl. (2019).

[26] Yao, P. et al. Fully hardware-implemented memristor convolutional neural network. Nature (2020).