

# A Study on the Performance Comparison of the Fundus Image Generation Model

Yong Suk Kim <sup>1\*</sup>

<sup>1\*</sup> Konyang University, Gwanjeodong-ro, Seo-gu, Daejeon, South Korea.

<sup>1\*</sup> yongsuk@konyang.ac.kr

## Article Info

Page Number: 527 – 537

Publication Issue:

Vol. 71 No. 3s (2022)

## Abstract

Recently, in the medical field, research using medical data based on artificial intelligence image processing technology has been actively conducted. However, research in the medical field using deep learning has many difficulties in obtaining data because personal information and medical information exist. This causes a lot of time and economic losses in conducting research. Therefore, in this study, a medical image generation study was conducted to realize the characteristics of lesions based on abnormal medical images among medical data to activate deep learning research in the medical field. In this study, three lesions, including normal images, were screened using the 'Ocular Disease Intelligent Recognition' open dataset, and a total of 356 fundus images were used. The deep learning model used for generating fundus images in this study is Res U-Net, which adds Residual Block to the U-Net structure that generates existing fundus images, and produces an image similar to the actual fundus image. In addition, for the performance comparison between the existing U-Net model and the Res U-Net model in this study, the fundus images generated by each model were quantitatively compared and evaluated through three image similarity indicators and ophthalmologist verification. Comparative evaluation results showed that Res U-Net showed higher results in all image similarity indicators than conventional models, especially Fréchet inception distance (FID) showed 8 times better performance. As a result of the ophthalmologist's verification of the generated fundus image, the average area under the curve (AUC) of the four lesions was U-Net 0.7705, Res U-Net 0.7415, indicating that the image generated in this study was more similar to the original image. As a future study, the addition and removal of lesions and patient information generation models will be studied based on fundus images.

**Keywords:** Fundus image, res U-Net, ophthalmologist, deep learning, image generation.

## Article History

Article Received: 22 April 2022

Revised: 10 May 2022

Accepted: 15 June 2022

Publication: 19 July 2022

## 1. Introduction

The development of image processing technology using deep learning has been an opportunity for many studies in various fields (Dhamo et al., 2020). Many studies are also being conducted in the medical field, and in particular, cases of deep learning studies using medical images have been published one after another (Shen et al., 2017). However, there are many difficulties in obtaining medical image data required for deep learning model research in the medical field (Kim et al., 2020). To proceed with the study using medical imaging, many legal procedures, including the Institutional Review Board (IRB), must be passed. It has a complex procedure, and it takes a lot of time to pass all the procedures and proceed with the study (Dyrbye et al., 2007). In addition, de-identification through anonymity and data verification of medical personnel, including doctors, must be performed. This causes a lot of time and economic losses when conducting research. In addition, there is a data imbalance of abnormal data (data with lesions) compared to normal data, and it is absolutely difficult to secure data. The quality and quantity of the data determine the performance of the deep learning model (Sun et al., 2017). Research in the medical field using artificial intelligence should solve the difficulties of data acquisition, and efficient research should be conducted by accumulating high-quality medical data. Therefore, in this paper, a model for generating fundus images among medical images is studied by developing the existing medical image generation research, and this is compared with the existing generation model through three image similarity indicators and ophthalmologists.

## **2. Related Works**

### **2.1. U-Net**

U-Net is a deep learning network developed for fast and accurate image segmentation in the biomedical field. Based on image feature extraction and segmentation performance, it was first used in cell structure segmentation of electron microscopes (Ronneberger et al., 2015). The U-Net is named U-Net because the shape of the network is composed of a 'U' shape with a left-right symmetrical structure. The U-Net consists of a contraction process and an expansion process. In the contraction process, the feature extraction of the input image and information between the proximity pixels are grasped through Down Sampling. Each layer of the contraction process consists of a combination of Convolution layers, and as the learning progresses, the size of the Feature Map is halved and the number of channels is doubled. The end of the contraction process leads to Skip Connection connected to the extension path. The skip connection serves to prevent loss of image information when connecting from a contraction process to an extension process. In the expansion process, it is a section that combines feature extraction information of an image obtained during the contraction process through Up Sampling. The extension path also consists of a combination of convolutional layers in each layer, and as opposed to the contraction process, the size of the feature map doubles and the number of channels halved each time the learning progresses. Because U-Net exhibits excellent performance for image feature extraction, it is also used in studies on image generation, including image segmentation studies. In 2018, a study was conducted to extract features of images through U-Net to generate images (Esser et al., 2018).

### **2.2. Residual Block**

In a deep learning network, there are problems in that the weight value of model learning

changes abnormally as the steps of the network increase and become deeper. Gradient Vanishing, in which Gradient gradually decreases, is Gradient Exploding, in which the Gradient gradually increases and the weighted value gives off. Because of these problems, deep learning networks are degraded. One of the ways to solve this problem is to use Residual Block. The learning process of the existing deep learning network sequentially proceeds with the accumulated convolution operation. On the other hand, Residual Block is a method of adding previously received values when transferring parameters by configuring the existing convolutional layer on a block-by-block basis (He et al., 2016). By adding blocks with simple calculations, it is possible to improve performance by solving problems that occur as the deep learning network becomes deeper.

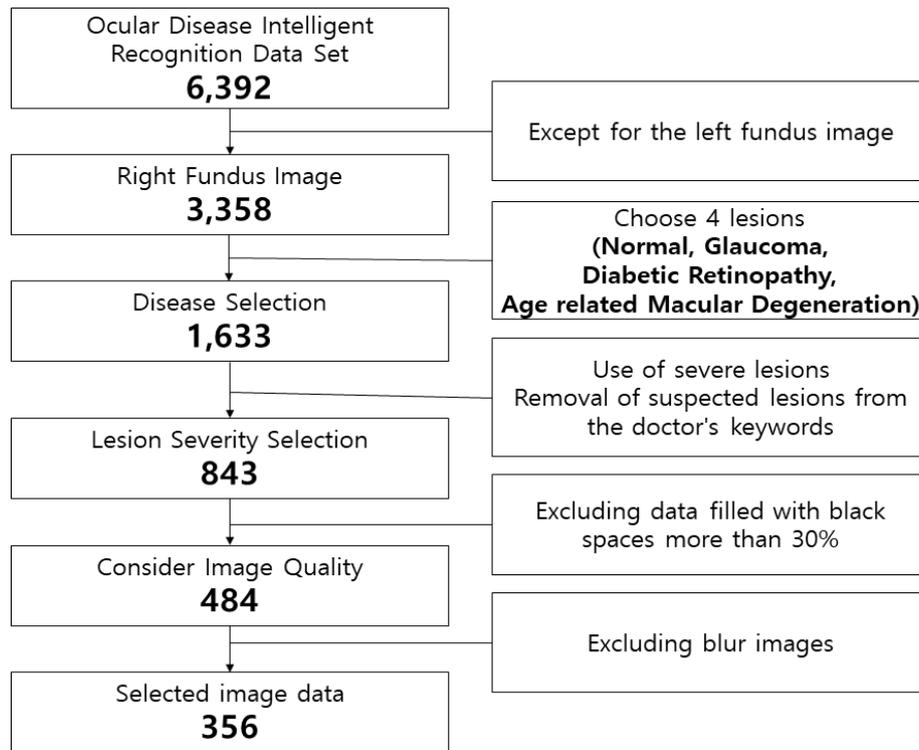
### **3. Materials and Methods**

Materials and methods of this study are composed of two. The first data analysis describes the analysis and classification process of research data. The second fundus image generation model describes Res U-Net used in the study.

#### **3.1. Data Analysis**

In this paper, we use the fundus image dataset, 'Oculative Disease Intelligent Recognition', collected by Kaggle's Shangong Medical Technology Co., Ltd. (Larxel et al., 2020). This data set is collected from hospitals and medical facilities in China. It is photographed with various fundus cameras such as Cannon, Zeiss, and Kowa and has various image resolutions. It was collected in 5,000 patients by age, including fundus images of both eyes and the patient's age and gender. It is also an ophthalmic database consisting of the patient's number and the doctor's diagnostic comment who diagnosed the patient. The dataset is classified into a total of 8 lesions for 6,392 collected fundus images. The classified lesions are normal, diabetic retinopathy, glaucoma, macular degeneration, cataract, high blood pressure, pathological myopia, and other diseases. Among them, this paper uses four types: normal, diabetic retinopathy, glaucoma, and macular degeneration.

Except for normal conditions, three lesions are representative diseases that cause blindness set by The Korean Ophthalmological Society. According to the 2018 National Health and Nutrition Survey conducted by the Korea Disease Control and Prevention Agency and The Korean Ophthalmological Society, the prevalence of these three lesions was 19.6% for diabetic retinopathy, 3.4% for glaucoma, and 13.4% for macular degeneration (Korea Disease Control and Prevention Agency, 2019). Therefore, in this paper, it was judged that research on the fundus image of the lesion would be actively conducted, and four lesions including the normal fundus image were selected as research data.



**Fig. 1; Data Classification Process**

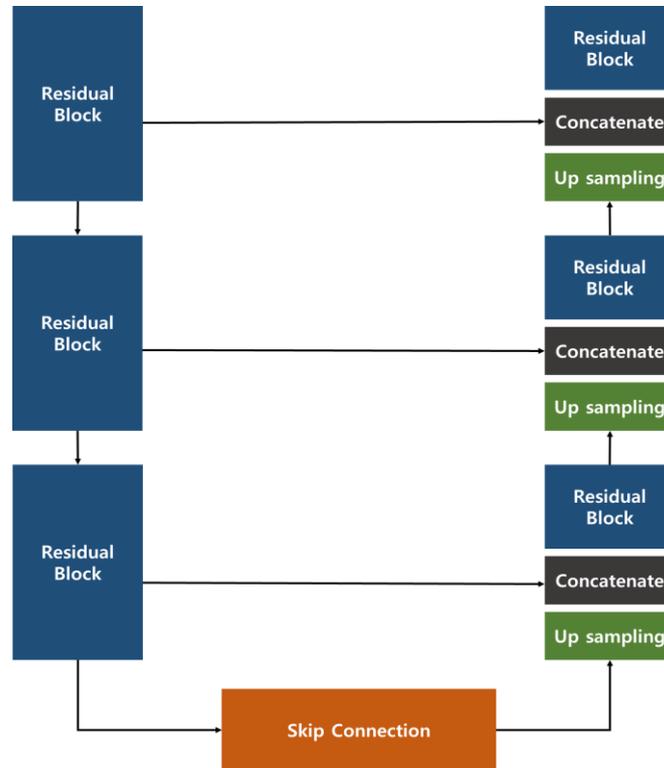
Fig. 1 is the classification process of the dataset used in this paper. Of a total of 6,392 fundus images, we construct a classification process to improve the performance of deep learning models and the quality of generated images. In the first step, only the right fundus image is selected. This considered the consistency of the learning data. In the second step, the lesion to be studied is selected. A total of four lesions, including the normal fundus, were selected. In the third step, lesions of severity or higher were used in consideration of doctor keywords. The fourth step is to exclude data having a space (black area) of 30% or more in the fundus image. In the fifth step, blurred fundus images were excluded in consideration of image quality.

As a result of classification, 356 out of 6,392 data were selected as research data for this paper. There are 114 normal fundus, 67 glaucoma, 78 diabetic retinopathy, and 97 macular degeneration. In the case of normal, the number was larger than that of other lesions, so it was randomly selected out of about 2,000 images. Diabetic retinopathy had a small amount of basic data in the dataset, and glaucoma had poor image quality, so a relatively small amount of data was used as a result of data classification. Subsequently, to solve the problem of ignoring the characteristics of fine blood vessels or lesions in the low-weight fundus image when learning the model, image preprocessing was performed using a sharpening filter, and the size of the input image was adjusted from 512x512 to 256x256.

### 3.2. Fundus Image Generation Model

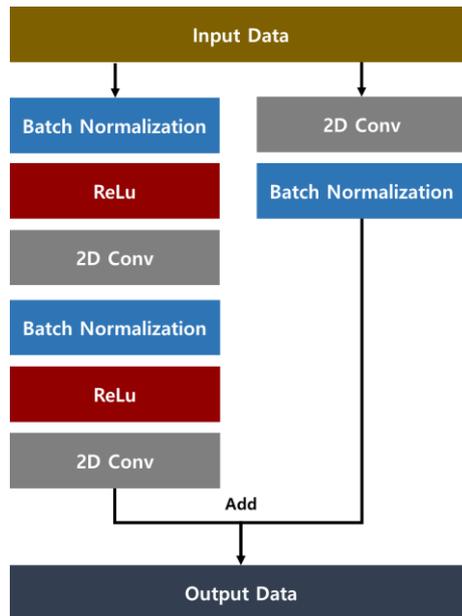
The fundus image generation model of this study consists of Res U-Net. Res U-Net was developed in 2018 for road extraction of high-resolution aerial images (Zhang et al., 2018). Since then, it has been used in various fields, including the medical field (Xiao et al., 2018).

Res U-Net is a combination of Residual Block and U-Net. The overall structure of the model is U-Net to efficiently learn with a small amount of data to extract features of the image. We solve Gradient Vanishing & Exploding using Residual Block in the expansion and contraction processes of U-Net. It also reduces the loss of low-weight information that occurs when the deep learning model structure deepens. This allows for a deeper deep learning network than when using existing U-Net.



**Fig. 2; Res U-Net Model Configuration**

Fig. 2 shows the structure of Res U-Net for generating fundus images. The contraction process on the left consists of three layers of Residual Block. In the contraction process, it serves to extract features of the input image. It then leads to a skip connection that prevents loss of image information. The skip connection has the same structure as the Residual Block. The right side of Figure 2 is the extension process. Like the contraction process, it consists of three layers. Each layer consists of three stages. The first step is the Up Sampling process. The size of the feature map is doubled and the number of channels is halved. Second, the 'Concatenate' operation minimizes the loss of information by using the characteristic information of the contraction process in the expansion process. Third, like the shrinkage process, Residual Block is used. In the last output layer, an image having the same size and channel as the input image is output through the 1x1 convolution layer.



**Fig. 3; Residual Block Configuration**

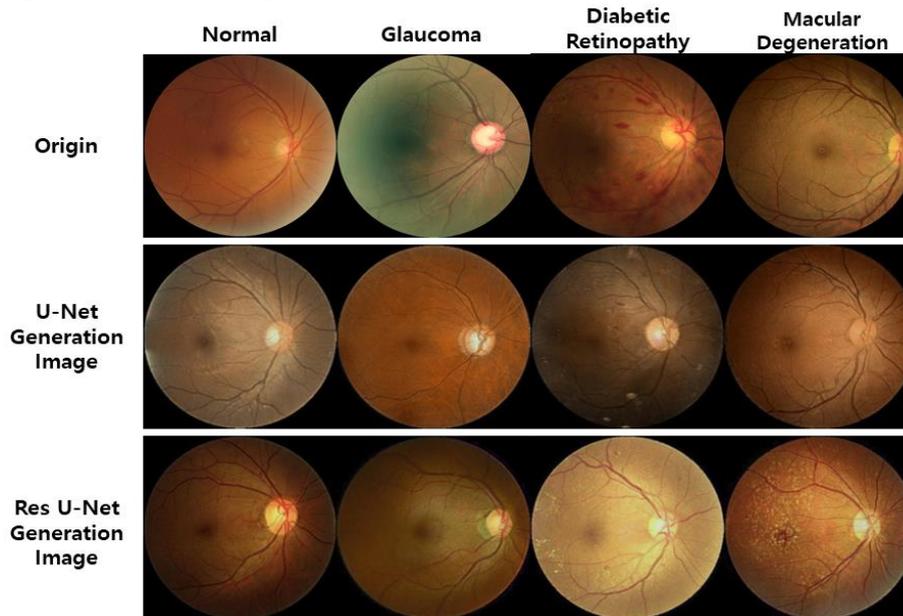
Fig. 3 shows the Residual Block structure of Res U-Net. The Residual Block receives data and passes it to the convolution layer on the left and to the Shortcut Connection on the right. The convolution layer on the left first performs batch normalization. Batch Normalization is a method of accelerating learning speed by stabilizing the learning process of the model. In the second step, the convolution operation is performed using the activation function ReLu (Rectified Linear Unit). The ReLu function is used to accelerate learning speed. The third step proceeds with the 2D convolution operation. The characteristics of the image are identified through convolution operations. After that, proceed with the previous process again. Figure 3 The Shortcut Connection on the right exists to convey the weight of the input data to the computational results of the left convolution layer. The configuration consists of 2D convolution and Batch Normalization. The output data consists of the sum of the calculation results of the convolution layer and the input data transferred from the right Shortcut Connection.

Due to the influence of the sharpening filter used in image preprocessing, the large difference between the black blank part and the pixel value of the fundus image resulted in the emphasis on the outer border part of the fundus image. To solve this problem, blurring is performed by post-processing the image. The Gaussian filter is used, and the kernel size of the kernel is 3 and the element value is 1 kernel. Average value filtering was used by multiplying each element by 1/9, and the sum is configured to be 1.

#### 4. Results and Discussion

In this paper, three methods are used to compare the performance of the fundus image generation model. The first visually compares the original image, the U-Net and Res U-Net generated images. The second is quantitatively compared through three Image Similarity Evaluation. The third is compared through verification by an ophthalmologist.

#### 4.1. Compare Fundus Image Generation Results



**Fig. 4; Compare Generated Fundus Images**

Fig. 4 shows a comparison between the images generated through Res U-Net and U-Net models and the original images. In the previous study, the U-Net generated image had a problem that the overall color of the fundus was similarly adjusted (Kim et al., 2022). When comparing the original image with the U-Net generated image, the retina or blood vessels of the image show a low color. This is a problem caused by the loss of the initial weight value of the blood vessels and retina of the input image as the deep learning network becomes deeper when the U-Net model proceeds with learning. Microvascular vessels with relatively low weights may disappear due to little contrast with the background. As the pixel value of the bright part of the retina also decreases, it becomes difficult to distinguish the color of the background. On the other hand, Res U-Net uses Residual Block to maintain initial weighted values, so the brightness of the blood vessels and retina in the generated image is not different from the original image. In addition, the quality of the image also differs visually compared to the U-Net generated image. The Res U-Net generated image does not look visually different compared to the original image.

#### 4.2. Comparison of Fundus Image Similarity Evaluation Results

In this paper, we quantitatively express the difference in image similarity between the original image and the generated image. Among image similarity evaluation techniques, similarity differences between images are verified using Root Mean Square Error (RMSE), Structural Similarity Index Map (SSIM), and Fréchet inception distance (FID). The values of RMSE and FID become smaller as the original image and the generated image are similar. The SSIM outputs a value close to 1 as the original image and the generated image are similar.

**Table 1. Image Similarity Evaluation**

		RMSE	SSIM	FID
Res U-Net	Normal	15.38	0.94	6.37
	Glaucoma	27.94	0.75	63.48
	Diabetic Retinopathy	20.09	0.90	20.66
	Macular Degeneration	30.4	0.84	38.09
U-Net	Normal	38.58	0.65	254.32
	Glaucoma	39.93	0.65	310.90
	Diabetic Retinopathy	36.76	0.61	284.28
	Macular Degeneration	37.78	0.64	253.49

Table 1 compares the results of the image similarity evaluation index for the generated image. According to the indicators of Res U-Net, RMSE had the lowest normal lesion and the value of macular degeneration was more than twice as high. In SSIM, the value of the normal lesion was the closest to 1, and the value of glaucoma was the lowest at 0.75. FID also had the closest normal lesion value to 0, and glaucoma had the highest value. Looking at the RMSE and FID scores, there was a difference in similarity with the original image on the image similarity index.

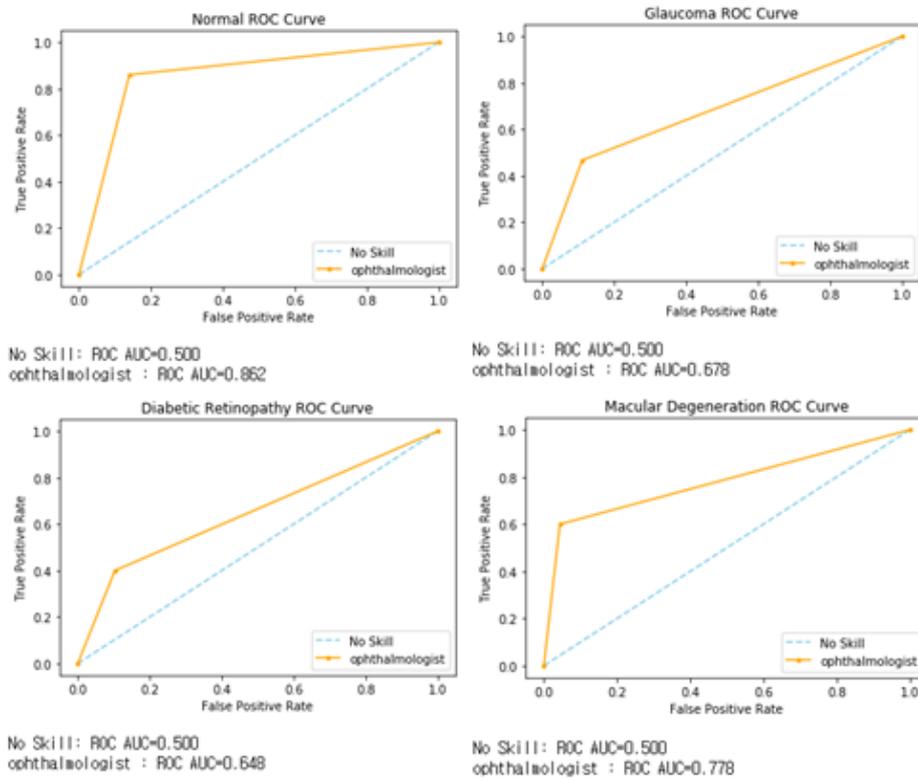
However, looking at the image similarity evaluation index of U-Net, all indicators show better results. Comparing with the mean of the indicators by lesion, RMSE 23.45, SSIM 0.85, FID 32.15 for Res U-Net, and RMSE 38.25, SSIM 0.63, and FID 275.74 for U-Net. RMSE differs by more than 15, SSIM differs by 0.22 and FID by more than 240. All three image similarity indicators show a large difference. This means that the image generated by the Res U-Net model is quantitatively similar to the original image, and that the image generation performance of the Res U-Net model is superior to that of the previous study.

### 4.3. Comparison of Ophthalmologist's Verification Results

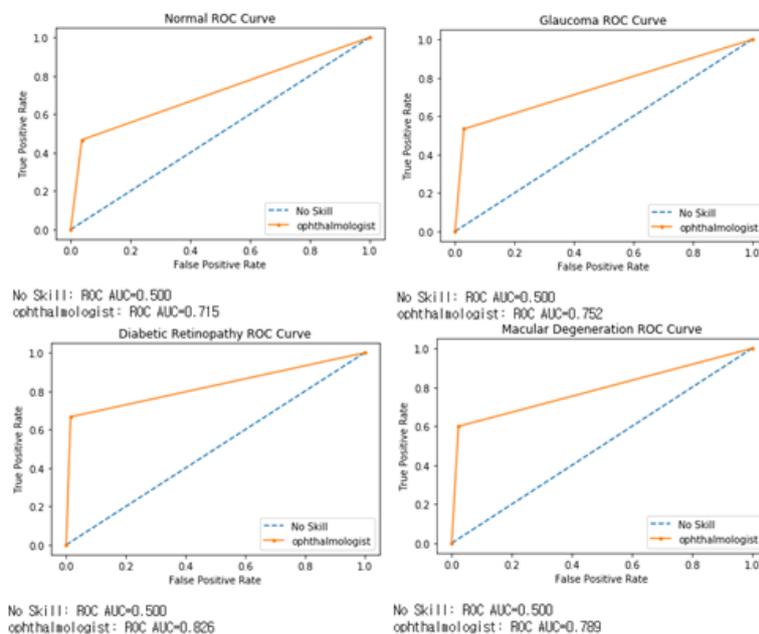
In this paper, data verification was requested by three ophthalmologists to compare and evaluate the clinical efficacy of fundus images generated from two deep learning models. The requested data consists of 45 original fundus images and 5 generated fundus images per lesion, and proceeds in the form of finding the generated image among randomly mixed fundus images. Verification by an ophthalmologist did not provide patient information of the generated fundus image, and the original fundus image proceeds in the same way. This is to minimize the effect of ophthalmologists on fundus image verification due to omission of patient information. Each figure was expressed by averaging the verification scores of three ophthalmologists. This is because it is important to grasp the average reading ability of each ophthalmologist, not to evaluate the reading ability of each ophthalmologist.

The results of verification by ophthalmologists are shown through the Receiver Operation Characteristic (ROC) Curve and the Area Under the Curve (AUC). The ROC Curve is a graph representing the classification performance of artificial intelligence models at all classification thresholds (Fawcett et al., 2006). The portion of the graph where the line breaks is the index point of the classification. The better the classification performance of the model, the closer the ROC Curve is to the top left of the graph. The AUC calculates the area below

the ROC Curve and provides a comprehensive measure of performance for all possible classification thresholds (Huang et al., 2005). The value of AUC is closer to 1 as the classification model performs better, and 0.5 means that the model has no classification performance.



**Fig. 5; Res U-Net Image Verification Results**



**Fig. 6; U-Net Image Verification Results**

Fig. 5 and 6 show the results of ophthalmologist verification of images generated by Res U-Net and U-Net models. As a result of verification by an ophthalmologist, all lesions except for normal generated images show high AUC values. In the normal case, Res U-Net 0.862 and U-Net 0.715 were recorded. For glaucoma, Res U-Net 0.678 and U-Net 0.752 were recorded. In the case of diabetic retinopathy, Res U-Net 0.648 and U-Net 0.826 show the largest difference. In addition, for macular degeneration, Res U-Net 0.778 and U-Net 0.789, showing similar AUC values.

Comparing the overall statistics of ophthalmologist data verification through the average score of AUC, the Res U-Net generated image recorded a score of 0.7415 and the U-Net generated image recorded a score of 0.7705. This means that the Res U-Net generated image is about 3% more difficult to detect than the U-Net generated image. It may be thought that the difference between the two models is not significant. However, when applied to actual clinical sites, it means that the ophthalmologist did not detect 2,585 generated fundus images out of 10,000 fundus images, which is a very large figure. Therefore, it can be seen that the visual difference from the actual fundus image is so small that the image generated by the Res U-Net model in this study is not found by the ophthalmologist about 30%.

## 5. Conclusion

In this paper, a study was conducted to solve the difficulty of obtaining medical image data required for research on deep learning models in the medical field. Data were generated through deep learning models using fundus images, and fundus images generated from existing generation models were compared. Data was classified through the data classification process, and image preprocessing was performed. Subsequently, a fundus image was generated using a Res U-Net-based generation model. As a result of comparing with the existing model through image similarity evaluation, the generation model of this study showed better values for all three image similarity indicators. The ophthalmologist's verification results showed scores of Res U-Net generated image 0.7415 and U-Net generated image 0.7705. Therefore, it can be said that the fundus image generated through the generation model of this paper has no difference from the actual fundus image visually, quantitatively, and clinically.

As a future study, a model for adding and removing lesions and generating patient information will be studied based on the fundus image. Furthermore, it plans to conduct research on the generation of medical data in various ways, such as chest radiation images and 3D medical image images, not limited to fundus images.

## References

1. Dharmo, H., Farshad, A., Laina, I., Navab, N., Hager, G.D., Tombari, F., & Rupperecht, C. (2020). Semantic image manipulation using scene graphs. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5213-5222.
2. Shen, D., Wu, G., & Suk, H.I. (2017). Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, 19(1), 221–248.
3. Kim, Y.J., & Kim, K.K. (2020). Development of an optimized deep learning model for medical images. *Journal of the Korean Society of Radiology*, 81(6), 1274-1289.

4. Dyrbye, L.N., Thomas, M.R., Mechaber, A.J., Eacker, A., Harper, W., Massie, F.S., David, V., & Shanafelt, T.D. (2007). Medical education research and IRB review: an analysis and comparison of the IRB review process at six institutions. *Academic Medicine*, 82(7), 654-660.
5. Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision*, 843-852.
6. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234-241.
7. Esser, P., Sutter, E., & Ommer, B. (2018). A variational u-net for conditional appearance and shape generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8857-8866.
8. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.
9. Larxel. (2020). Ocular Disease Recognition: Right and left eye fundus photographs of 5000 patients. WEB: <https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k/>
10. Korea Disease Control and Prevention Agency. (2019). resource book of the results of the 2018 National Health and Nutrition Survey and Youth Health Behavior Survey. [https://knhanes.kdca.go.kr/knhanes/sub04/sub04\\_04\\_03.do/](https://knhanes.kdca.go.kr/knhanes/sub04/sub04_04_03.do/)
11. Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749-753.
12. Xiao, X., Lian, S., Luo, Z., & Li, S. (2018). Weighted res-unet for high-quality retina vessel segmentation. In *2018 9th international conference on information technology in medicine and education (ITME)*, 327-331.
13. Kim, Y.S., Song, H.J., & Han, J.H. (2022). A Deepfake-Based Deep Learning Algorithm for Medical Data Manipulation Detection. *Journal of System and Management Sciences*, 12(1), 13-26
14. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874.
15. Huang, J., & Ling, C.X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering*, 17(3), 299-310.