

Received 12 May 2023, accepted 27 May 2023, date of publication 31 May 2023, date of current version 9 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3281693

RESEARCH ARTICLE

Development of ResNet152 UNet++-Based Segmentation Algorithm for the Tympanic Membrane and Affected Areas

TAEWAN KIM¹, KYOUNGHO OH², JAEYOUNG KIM^{3,4,5}, YEONJOON LEE¹, AND JUNE CHOI²

¹Major in Bio Artificial Intelligence, Department of Applied Artificial Intelligence, Hanyang University, Ansan 15588, Republic of Korea

²Department of Otorhinolaryngology-Head and Neck Surgery, Ansan Hospital, Korea University College of Medicine, Ansan 15355, Republic of Korea

³Core Research and Development Center, Korea University Ansan Hospital, Ansan 15355, Republic of Korea

⁴Department of Dermatology and Skin Sciences, The University of British Columbia, Vancouver, BC V6T 1Z1, Canada

⁵Cancer Control Research Program, BC Cancer, Vancouver, BC V5Z 1L3, Canada

Corresponding authors: Yeonjoon Lee (yeonjoonlee@hanyang.ac.kr) and June Choi (mednlaw@korea.ac.kr)

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korea Government, Ministry of Science and ICT (MSIT) through the Artificial Intelligence Convergence Innovation Human Resources Development, Hanyang University ERICA, and the Artificial Intelligence Convergence Research Center, Hanyang University ERICA under Grant RS-2022-00155885 and Grant 2020-0-01343; in part by the National Research Foundation of Korea (NRF) funded by the Korea Government (MSIT) under Grant NRF-2022R1F1A1074999; in part by the Korea University Grant and a grant of the Medical Data-Driven Hospital Support Project through the Korea Health Information Service (KHIS), funded by the Ministry of Health and Welfare, Republic of Korea; and in part by the MSIT, South Korea, through the ICT Challenge and Advanced Network of HRD (ICAN) Program supervised by the IITP under Grant IITP-2022-RS-2022-00156439.

ABSTRACT Otitis media (OM) is a common disease in childhood that may have aftereffects such as hearing loss. Therefore, early diagnosis and proper treatment are important. However, the diagnostic accuracies of otolaryngology and pediatrics are low, at 73% and 50%, respectively. Therefore, clinical work that supports the early diagnosis of diseases, such as computer-aided diagnostic (CAD) systems, can be helpful. However, CAD systems for diagnosing ear diseases require an automatic tympanic membrane (TM) segmentation model to assist in diagnosis. This is because it is difficult to detect the TM and affected areas in an endoscopic image of the TM owing to irregular lighting. In this study, we propose a ResNet152 UNet++ image segmentation network. The proposed method applies the ResNet152 layer structure to the encoders in the UNet++ model to detect the location of the TM and affected area with high accuracy. Furthermore, the TM and affected regions can be segmented better than when using the previously proposed UNet and UNet++ models. To the best of our knowledge, this study is the first to use a UNet++-based segmentation model to segment TM areas in endoscopic images of the TM and evaluate its performance. The experiments revealed that ResNet152 UNet++ outperforms conventional methods in terms of segmentation of the TM and affected areas.

INDEX TERMS Convolutional neural network, artificial neural network, segmentation, otitis media, computer-aided diagnosis.

I. INTRODUCTION

Otitis media (OM) is one of the most common childhood diseases [1], collectively term for all inflammatory changes within the middle ear cavity, and involves inflammatory changes in the middle ear mucosa, submucosa, and bone tissue. Failure to receive proper treatment owing to delaying

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Cheng¹.

early diagnosis may result in aftereffects such as hearing loss [2]. In particular, incorrectly treated OM may have serious consequences such as intracranial complications or facial palsy [3]. Therefore, it is important to diagnose and treat OM accurately. However, the average diagnosis rates for otolaryngologists and pediatricians are only 73% and 50%, respectively [4]. The diagnosis of OM is based on the condition of the tympanic membrane (TM), therefore it is very important to clinically identify the TM correctly.

Early identification of the affected areas can help prevent complications associated with untreated or poorly managed middle ear disorders, such as hearing loss, chronic middle ear infections, cholesteatoma, or permanent destruction of the TM [5]. In addition, segmenting the TM and affected areas enables physicians to provide more detailed and accurate diagnoses [6]. As a result, accurate segmentation of the TM and affected areas from endoscopic images, when supported by diagnostic tools like computer-aided diagnostics (CAD), is expected to enhance diagnostic accuracy in diagnosing ear diseases.

CAD systems are important clinical tasks for detecting abnormalities using medical imaging [7]. In the diagnosis of ear diseases, the CAD system includes several steps, including obtaining the original images, preprocessing, and feature extraction [1]. It is important to detect the TM and affected area accurately. However, it is difficult to segment the TM and affected area in the TM image because the video perspective image appears darker than the average color of a specific structure owing to irregular lighting [8]. Therefore, a new method is required for detecting TMs and their affected areas.

With recent developments in deep learning, research on the application of artificial intelligence (AI) to the medical field has been increasing [9]. Among these methods, the convolutional neural network (CNN) is the most commonly used technology in image segmentation [10]. Pham et al. used the UNet-based EAR-UNet model to detect the TMs and affected areas from endoscopic images of the TM, including normal, acute otitis media (AOM), chronic otitis media (COM), and otitis media with effect (OME), with 95.8% accuracy [3]. However, the utility of the detected images in diagnosing ear diseases has not yet been tested. And Basararan et al. used Fast R-CNN to segment TMs and diagnosed six ear diseases with 90.48% accuracy using the VGG16 model [11]. However, the TM segmentation accuracy of the Fast R-CNN was only 79.52% and the improvement in the diagnostic performance was unsatisfactory. Moreover, a study aimed to enhance the performance of the classification model by segmenting the TM in endoscopic images of the TM, utilizing color and geometric structure [12]. However, their approach demonstrated incomplete segmentation outcomes, such as the loss of the retraction feature owing to the inadequate removal of certain boundaries of the malleus, which are challenging to discern visually. In this study, we propose a ResNet152 UNet++ model that uses TM images to detect the TMs and affected areas with high performance. And we evaluated whether the proposed model could accurately detect the TM and affected areas. We verified that images with the segmentation of the TM and affected areas aided OM diagnosis using six CNN models released on ImageNet.

Our paper is organized as follows. Section II briefly reviews the related work and Section III details the experimental methods. Section IV presents and analyzes the experimental results. Finally, the conclusions are outlined in Section V.

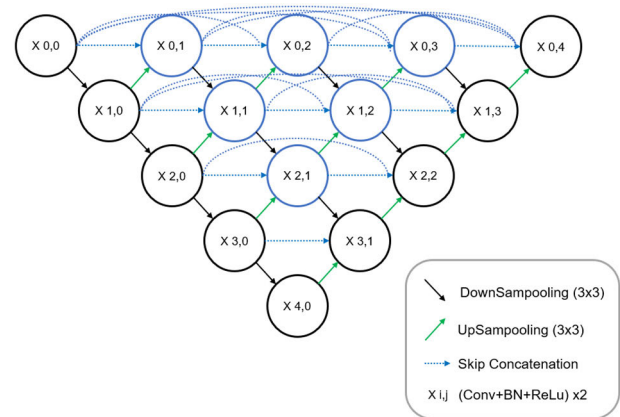


FIGURE 1. UNet++ Backbone.

II. RELATED WORK

A. UNET++

UNet [13] is a segmentation model that has been widely applied to image segmentation in the medical field since its proposal. Many studies have used UNet backbones for medical image segmentation, and there have been various studies that have changed segmentation tasks based on UNet backbones. In Figure 1, UNet++ [14] is a representative UNet-based segmentation model that can achieve more accurate medical image segmentation by applying dense skip connections based on UNet. Each overlapped convolution block is upsampled following downsampling to extract semantic information for multiple convolutions. All of these convolutional layers are connected by dense skip connections to segment medical images with better accuracy.

B. RESNET

Deep convolutional neural networks are an innovative method for classifying images. However, the performance is degraded if the layers are stacked too deeply. ResNet [15] is a CNN model that addresses the degradation problem as the layers deepen by adding residual networks for each layer. The concept of a residual network is illustrated in Figure 2. Identity shortcut connections are connections that skip one or more layers. With the identity shortcut connections, ResNet classifies images with better accuracy, despite their stacking in deep layers, without adding parameters, and with no computational complexity.

III. METHODOLOGY

In this section details the proposed method, including the ResNet152 UNet++ architecture, the type of ImageNet pre-trained CNN model, Data augmentation techniques, and training details. we adopt UNet++'s improved ResNet152 UNet++ architecture to segment the TM and affected areas In the paper. The augmented TM endoscopic image is learned with the proposed segmentation model and evaluated with a published pre-trained CNN model on ImageNet to see if the

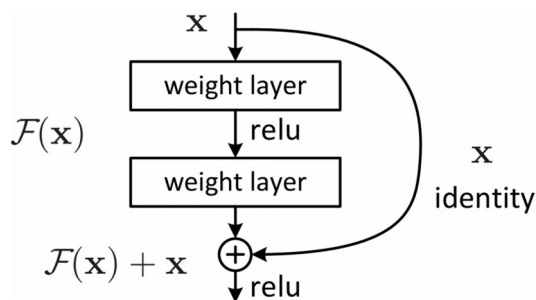


FIGURE 2. Residual learning: a building block [14].

segmented TM and affected area contain information to help diagnose ear disease.

A. RESNET152 UNET++

We designed the ResNet152 UNet++ network to segment endoscopic images of the TM. An overview of the ResNet152 UNet++ is depicted in Figure 3. In $X_{i,j}$, i represents the depth of the layers, and j represents the depth of the convolution layer of the nested block by skip connection.

ResNet UNet++ uses UNet++ as the default network framework. It differs from the existing UNet++ in that the convolutional layer of the encoder, which extracts the image features, uses the ResNet152 architecture. Thus, the image features can be extracted more efficiently. The structure of the ResNet-Bottleneck is shown in Figure 4. The ResNet-Bottleneck layer applies the numbers a and b of the convolution layer and depth c in ResNet152. These configurations segment the TM and affected areas with better accuracy in endoscopic images of the TM.

B. IMAGENET PRE-TRAINED CNN MODEL

We used a published pre-trained CNN model of ImageNet to evaluate whether the segmentation images that were generated by the ResNet152 UNet++ model exhibited better performance in image classification than in the original image. We employed and compared a total of six models: ResNet152 [15], VGG19 [16], GoogleNet [17], DenseNet161 [18], Inception-V3 [19], and Inception-ResNet-v2 [20].

C. DATA AUGMENTATION TECHNIQUES

Data augmentation is a normalization method that is used to prevent overfitting. Image augmentation techniques include flip, rotation, scale, crop, transition, and noise. We used only rotational augmentation techniques to preserve the semantic information of the endoscopic images as far as possible. Figure 5 presents the rotational augmentation results for an endoscopic image sample of the TM. Five augmented images were generated by rotating the original image five times by 60° . Therefore, in addition to the original dataset, this study also used an augmented dataset.

TABLE 1. Learning environments of ResNet152 UNet++ models and CNN models.

Hyperparameters	Segmentation model	CNN model ^a
Batch size	16	16
learning rate	1e-5	1e-4
Optimizer	Adam	Adam
Eps	0.1	-
Loss	Focal Tversky	Cross-Entropy
Epoch	100	100

D. TRAINING DETAILS

The learning environments for both the segmentation and CNN models are presented in Table 1. ResNet152 UNet++ model used a batch size of 16, a learning rate of $5e-2$, the Adam optimizer, and the focal tversky loss function. The Adam optimizer fine-tuned eps to 0.1. Furthermore, the learning environment of the ImageNet pre-trained CNN model used a batch size of 16, a learning rate of $1e-4$, the Adam optimizer, and the cross-entropy loss function. Owing to class-specific data imbalances, a loss weight was applied by calculating the ratio of each amount of data, and both methods were run for 100 epochs. All experiments in this study were conducted on a deep learning server with eight NVIDIA GeForce RTX 3080 12 GB GPUs.

IV. EXPERIMENTS AND RESULTS

A. DATASET SETUP AND PREPROCESSING

We collected and analyzed 1,632 medical images at the Korea University Ansan Hospital and generated 9,792 images using data augmentation. Therefore, endoscopic images were used data from 2,782 normal, 3,626 perforation, 1,866 retraction, and 954 cholesteatoma images. The features of each ear disease are depicted in Figure 6, with blue circles indicating the location of the ear disease feature. TM perforation, retraction, and cholesteatoma are all conditions that may lead to hearing impairment. TM perforation is a condition where a hole forms in the TM and can result from chronic middle ear infections [21]. Perforation of the TM serves as a crucial indicator for chronic middle ear inflammation and significantly influences the decision to perform surgery. Retraction of the TM occurs when a persistent pressure difference between the middle ear and atmospheric pressure arises due to eustachian tube dysfunction, potentially causing patients to experience a sensation of ear fullness [22]. If TM retraction persists, alterations in the middle ear mucosa may develop, ultimately leading to cholesteatoma formation. Therefore, the presence or absence of TM retraction indirectly reflects the patient's eustachian tube function, hinting at potential cholesteatoma development in the middle ear. Cholesteatoma can induce additional symptoms such as hearing loss, headache, and

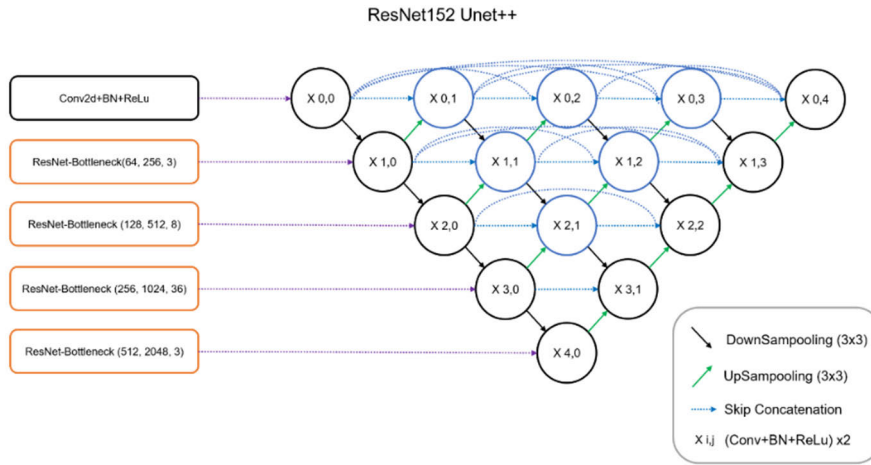


FIGURE 3. Architecture of ResNet152 UNet++.

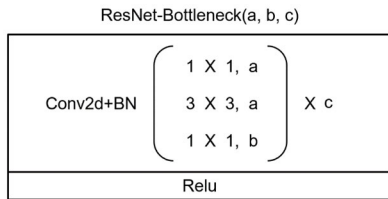


FIGURE 4. Structure of encoder block in the ResNet-Bottleneck.

vertigo. Advanced cholesteatoma can result in serious complications, including damage to middle ear ossicles like the malleus and erosion of the skull base bone, potentially affecting brain function [23]. Accurate assessment of TM findings in cases of cholesteatoma is essential for determining the lesion’s location and size, which ultimately guides the choice of surgical intervention. However, 564 images that could not be recognized owing to severe swelling, bleeding, or shaking were excluded.

Moreover, the TM imaging equipment underwent changes during data collection, resulting in varying image sizes. The majority of the initial images had a resolution of 640×480 pixels, while the remaining images were 1920×1080 pixels. To address the inconsistency in image size, all images were resized to a resolution of 384×384 pixels. The endoscopic images of the TM were labeled on the Computer Vision Annotation Tool (CVAT) website, and the dataset was randomly divided without duplication into 80% for training (7,376) and 20% for testing (1,852). The data of this study were approved by the IRB (2021AS0329) of Korea University Ansan Hospital, and the procedure was followed by the Helsinki declaration in 1975, furthermore informed consent is waived by ethics committee because of a retrospective study.

B. EVALUATION METRICS

We used the pixel accuracy, dice coefficient, and intersection over union (IoU) indicators to evaluate the performance of

segmenting endoscopic images of the TM. Pixel accuracy represents the ratio of correctly predicted pixels to the total number of pixels. The dice coefficient is an evaluation metric that measures the similarity between two sets by considering the overlap of the sets. It evaluates the performance of the model by comparing the predicted regions and ground truth. IoU (Intersection over Union) is a widely used evaluation metric in semantic segmentation that measures the performance by evaluating the overlap between the ground truth and the predicted regions. The equations for the segmentation evaluation metrics are as follows:

$$Pixelaccuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Dicecoefficient = \frac{2 \times TP}{(TP + FP) + (TP + FN)} \tag{2}$$

$$IoU = \frac{TP}{TP + FN + FP} \tag{3}$$

Furthermore, to verify whether the segmentation image helps in diagnosing diseases, we used a published CNN model from ImageNet for a performance comparison with that of the original image. The performance was evaluated using the accuracy and recall indicators. Accuracy is the most commonly used performance metric, representing the ratio of correctly predicted data to the total dataset. Recall is the ratio of correctly predicted data belonging to the true class out of all the datasets in the true class. The formulas for the classification evaluation metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. The higher the value for each metric, the better the segmentation and classification performance.

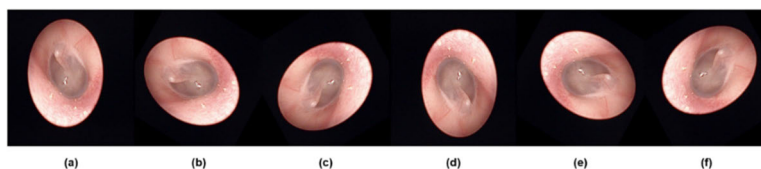


FIGURE 5. Illustration of image augmentation process for tympanic membrane endoscopic image: (a) original image, (b) rotated by 60°, (c) rotated by 120°, (d) rotated by 180°, (e) rotated by 240°, and (f) rotated by 300°.

TABLE 2. Performance comparison of segmentation models.

Model	Dice coefficient	Pixel accuracy	IoU score	Loss
UNet	92.4	97.1	86.3	0.136
EAR-UNet	93.2	97.3	87.5	0.121
UNet++	92.5	97.2	86.4	0.130
VGG19 UNet++	92.7	97.2	86.6	0.129
MobileNet-v2 UNet++	92.8	97.3	86.9	0.128
DenseNet161 UNet++	93.2	97.4	87.5	0.121
Inception-ResNet-v2 UNet++	93.3	97.4	87.7	0.118
ResNet152 UNet++	93.4	97.5	87.8	0.120

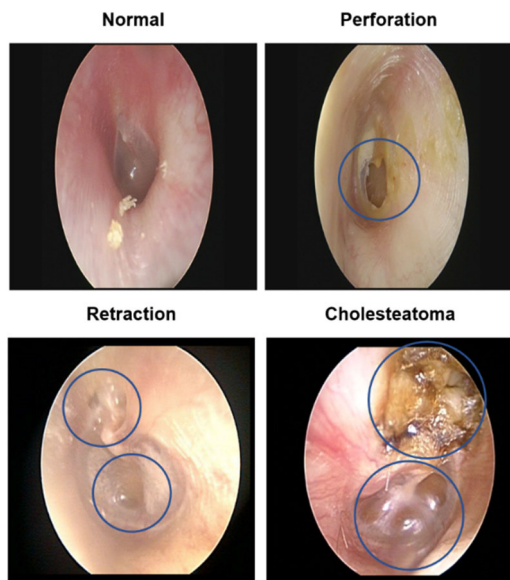


FIGURE 6. The type of ear disease in the data collected. The blue circle represents the location of each disease’s characteristics.

C. SEGMENTATION RESULTS

Table 2 compares UNet, EAR-UNet [3], UNet++, VGG19 UNet++, MobileNet-v2 UNet+, DenseNet161 UNet++,

Inception-ResNet-v2 UNet++, and ResNet152 UNet++ for the segmentation experiments of the TM and affected areas in TM endoscopic images. For the four evaluation metrics, our proposed ResNet152 UNet++ exhibited the best segmentation results, with a dice coefficient of 93.4%, a pixel accuracy of 97.5%, an IoU score of 87.8%, and a loss of 0.120. The existing EAR-UNet model [3], which is based on the UNet framework, enhances its performance compared to the traditional UNet model by incorporating EfficientNet-B4 into the encoder, applying residual blocks to the decoder, and adding attention gates to the skip connections. Conversely, our model is built upon an upgraded UNet++ architecture, featuring re-designed skip pathways that integrate the DenseNet structure into the UNet’s skip connections, as well as a Deep Supervision method that utilizes the average of the up-sampling results from each layer as the final output. Furthermore, we incorporated the ResNet152 structure into the encoder. Consequently, our model exhibited superior performance in comparison to the state-of-the-art EAR-UNet, which segments the TM and affected areas in endoscopic images of the TM, yielding improvements of 0.2% in dice coefficient, 0.2% in pixel accuracy, and 0.3% in IoU score.

Figure 7. presents a comparison of the results of the TM and affected area segmentation between the models. The black line indicates the ground truth and the green line

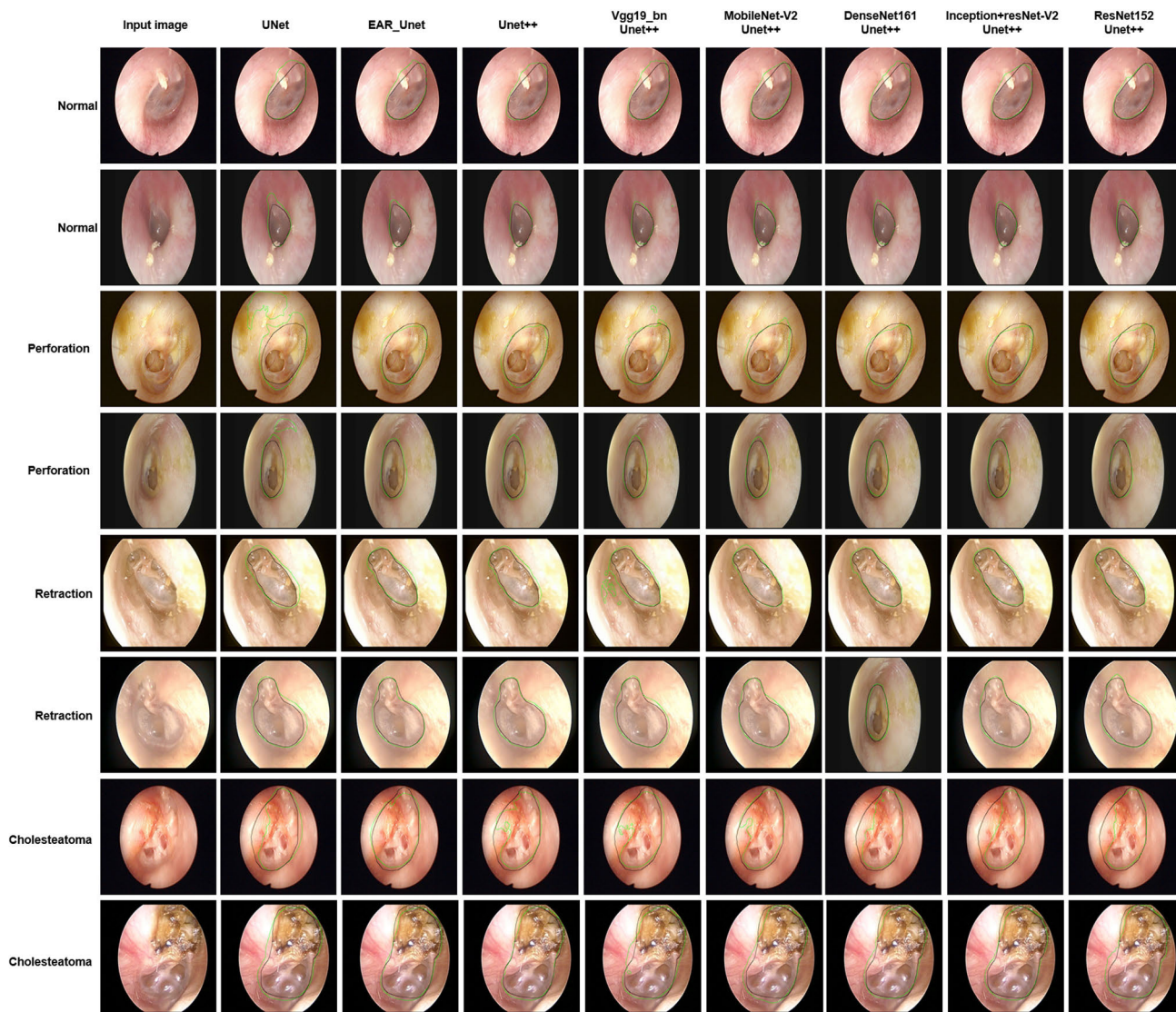


FIGURE 7. Comparison of results of segmentation of tympanic membrane (TM) and affected areas by various segmentation models. The TM and affected areas were segmented from endoscopic images of the TM, including normal and three ear diseases (perforation, retraction, and cholesteatoma). The ground truth is indicated by a black line and the position predicted by the model is indicated by a green line.

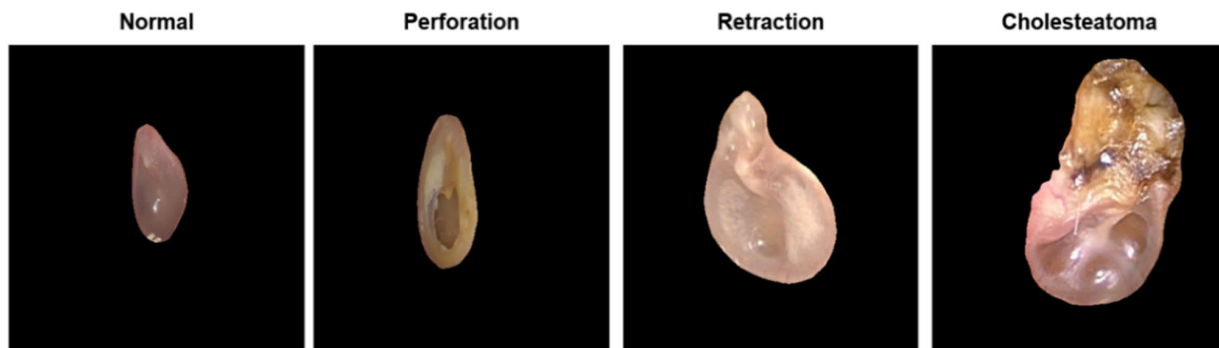


FIGURE 8. Segmentation images generated using the ResNet152 UNet++ model. After removing the out-of-border data at the location predicted by the proposed model, we used only the images of the tympanic membrane (TM) and affected areas, which were important lesions in the endoscopic image of the TM.

represents the location of the TM and affected areas predicted by the proposed model. The results in Figure 7 shows that

ResNet152 UNet++ divided the TM and affected areas better than other models and predicted a more accurate position

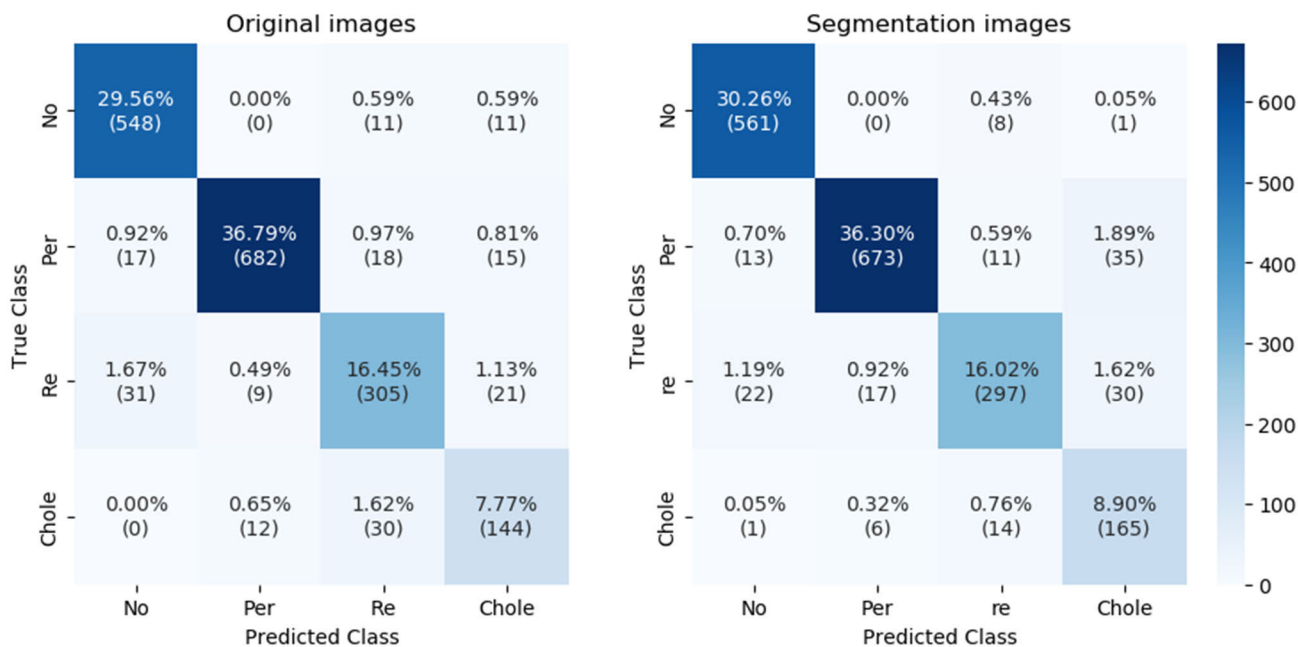


FIGURE 9. Comparison of confusion matrices for DenseNet161 model using original and segmentation images. (No=Normal, Per=Perforation, Re=Retraction, Chole=Cholesteatoma).

TABLE 3. Classification performance comparison of original and segmentation images.

Model	Original image			Segmentation image		
	Accuracy	Recall	Loss	Accuracy	Recall	Loss
VGG19	88.2	84.5	0.721	89.1	87.2	0.784
ResNet152	89.7	87.3	0.698	89.3	86.3	0.467
Inception-ResNet-v2	90.7	86.7	0.588	90.9	88.6	0.481
Inception-v3	90.3	87.3	0.442	90.1	87.2	0.492
GoogleNet	89.4	86.2	0.545	89.5	88.9	0.485
DenseNet161	90.5	87.5	0.447	91.4	90.0	0.536

than the ground truth in the case of cholesteatoma. Therefore, we generated segmentation images based on the prediction results of the proposed model, as shown in Figure 8.

D. IMAGE CLASSIFICATION

We compared the original image with the segmentation image that was generated by the ResNet152 UNet++ model to six CNN models of ImageNet. The comparison results are presented in Table 3. The best performance was obtained when the segmentation images were used on DenseNet161,

with an accuracy and a recall of 91.4% and 90.0%, respectively. Furthermore, a comparison of the confusion matrices (Figure 9) reveals that the number of misdiagnoses increased by 9 for perforations and 8 for retractions when using the segmentation image compared to when using the original image in the DenseNet161 model. However, the number of misdiagnoses for normal and cholesteatoma was reduced by 13 and 21, respectively, and finally, the total number of misdiagnoses was reduced by 17. Thus, the segmentation images that were generated by our proposed ResNet152 UNet++

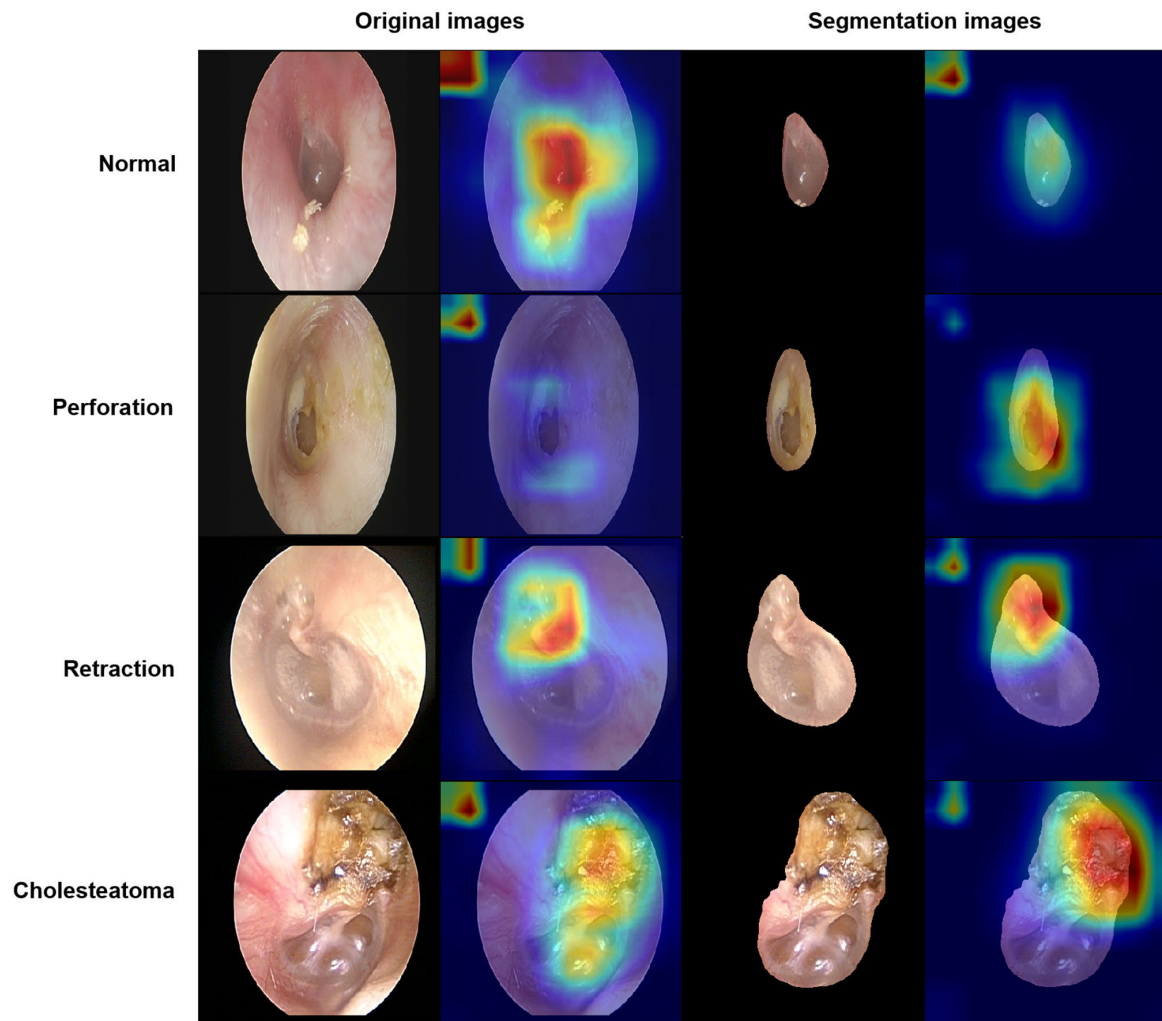


FIGURE 10. Results of disease-specific Grad-CAM comparison of DensenNet161 model using original and segmentation images.

model demonstrated that the positions of the segmented TM and affected areas were useful for diagnosing middle ear diseases.

E. GRAD-CAM

Figure 10 shows the comparison results of Grad-CAM when the original and segmentation images were used for DenseNet161. The comparison indicates that the segmentation image exhibited fewer heat maps in areas other than the TM and affected areas compared to the original image, and more heat maps were displayed in the affected areas. Therefore, when using the segmentation image, we observed the TM and affected area more accurately and confirmed that the disease was classified.

V. DISCUSSION

The UNet++ model employs a substantial number of parameters and consumes significant memory due to its complex connections. In our approach, we further increased the parameter count and memory usage by adopting the ResNet152

architecture, a deeper model, in place of the VGG16 architecture within the UNet++ framework. However, this architectural modification introduces the limitation of increased computational requirements, which should be considered in practical implementation and resource-constrained environments.

In contrast to our study, which segmented the tympanic membrane (TM) as a whole, previous research has focused on segmenting individual parts of the TM and associated lesions, such as perforations, in greater detail [24]. However, our study segmented the tympanic membrane and affected areas as a single object, achieving high accuracy in identifying the TM and affected areas.

However, a more detailed segmentation approach that differentiates between the TM and lesions could potentially offer a foundation for tailored treatment strategies based on the patient's middle ear condition and lesion characteristics in the future. Nonetheless, assessing a patient's condition and establishing precise treatment plans remain constrained by the examination of the TM alone. Clinicians still require

the assistance of other diagnostic tools, such as computed tomography, to obtain a comprehensive understanding of the patient's needs.

Therefore, future research is needed to reduce the parameters and memory used by segmentation models and to investigate whether our model can aid in the automatic diagnosis of ear diseases by individually segmenting the features of the TM in endoscopic images of the TM.

VI. CONCLUSION

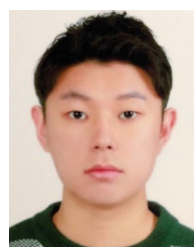
This study proposes a ResNet152 UNet++ model for segmenting the TM and affected areas from endoscopic images of the TM. The combination of endoscope technology and computer algorithms has improved the accuracy of TM diagnosis. Although expert clinicians are still required to interpret the results and provide appropriate treatment, this technology has reduced the chances of making a wrong diagnosis. Our experiments demonstrated the competitive performance of the proposed ResNet152 UNet++ in segmenting TMs and affected areas. This improvement is attributed to the combination of UNet++ and ResNet152 models. The experimental results also confirmed that ResNet152 UNet++ accurately divided the TM and affected areas, except for the external ear area. In addition, the detected TM endoscopic image was learned using a published CNN model of ImageNet, and experiments showed that the detected area was useful for diagnosing ear disease. Therefore, the ResNet152 UNet++ proposed in this study may help detect TMs and affected areas in future remote diagnosis and clinical situations.

ACKNOWLEDGMENT

(Taewan Kim and KyoungHo Oh are co-first authors.)

REFERENCES

- [1] T.-T. Tran, T.-Y. Fang, V.-T. Pham, C. Lin, P.-C. Wang, and M.-T. Lo, "Development of an automatic diagnostic algorithm for pediatric otitis media," *Otol. Neurotol.*, vol. 39, no. 8, pp. 1060–1065, Sep. 2018.
- [2] D. Cha, C. Pae, S.-B. Seong, J. Y. Choi, and H.-J. Park, "Automated diagnosis of ear disease using ensemble deep learning with a big otoscopy image database," *EBioMedicine*, vol. 45, pp. 606–614, Jul. 2019.
- [3] V.-T. Pham, T.-T. Tran, P.-C. Wang, P.-Y. Chen, and M.-T. Lo, "EAR-UNet: A deep learning-based approach for segmentation of tympanic membranes from otoscopic images," *Artif. Intell. Med.*, vol. 115, May 2021, Art. no. 102065.
- [4] M. E. Pichichero and M. D. Poole, "Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media," *Arch. Pediatrics Adolescent Med.*, vol. 155, no. 10, pp. 1137–1142, Oct. 2001.
- [5] J. Patel, R. Aiyer, Y. Gajjar, R. Gupta, J. Raval, and P. Suthar, "Endoscopic tympanoplasty vs microscopic tympanoplasty in tubotympanic CSOM: A comparative study of 44 cases," *Int. J. Res. Med. Sci.*, vol. 3, no. 8, pp. 1953–1957, 2015.
- [6] M. Ashler, H. Özyaz, S. Gürkan, G. Kirkim, and E. A. Güneri, "The effect of tympanic membrane perforation site, size and middle ear, volume on hearing loss," *Turkish Arch. Otorhinolaryngol.*, vol. 57, pp. 86–90, Jun. 2019.
- [7] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Med. Imag. Graph.*, vol. 31, nos. 4–5, pp. 198–211, Jun. 2007.
- [8] E. Comunello, A. von Wangenheim, V. H. Junior, C. Dornelles, and S. S. Costa, "A computational method for the semi-automated quantitative analysis of tympanic membrane perforations and tympanosclerosis," *Comput. Biol. Med.*, vol. 39, no. 10, pp. 889–895, Oct. 2009.
- [9] G. Rong, A. Mendez, E. B. Assi, B. Zhao, and M. Sawan, "Artificial intelligence in healthcare: Review and prediction case studies," *Engineering*, vol. 6, no. 3, pp. 291–301, Mar. 2020.
- [10] S. Sivagami, P. Chitra, G. S. R. Kailash, and S. R. Muralidharan, "UNet architecture based dental panoramic image segmentation," in *Proc. Int. Conf. Wireless Commun. Signal Process. Netw. (WiSPNET)*, Aug. 2020, pp. 187–191.
- [11] E. Başaran, Z. Cömert, and Y. Çelik, "Convolutional neural network approach for automatic tympanic membrane detection and classification," *Biomed. Signal Process. Control*, vol. 56, Feb. 2020, Art. no. 101734.
- [12] C. Shie, H. Chang, F. Fan, C. Chen, T. Fang, and P. Wang, "A hybrid feature-based segmentation and classification system for the computer aided self-diagnosis of otitis media," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 4655–4658.
- [13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2015, pp. 234–241.
- [14] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, 2018, pp. 3–11.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
- [20] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2017, vol. 31, no. 1, pp. 1–15.
- [21] M. Verhoeff, E. L. Van Der Veen, M. M. Rovers, E. A. Sanders, and A. G. Schilder, "Chronic suppurative otitis media: A review," *Int. J. Pediatric Otorhinolaryngol.*, vol. 70, pp. 1–12, Jan. 2006.
- [22] S. Yoshida, S. Seki, T. Sugiyama, S. Kikuchi, N. Yoshida, and Y. Iino, "Comparative study on adhesive otitis media and pars Tensa cholesteatoma in children," *Auris Nasus Larynx*, vol. 49, no. 5, pp. 790–796, Oct. 2022.
- [23] B. A. Jennings, P. Prinsley, C. Philpott, G. Willis, and M. F. Bhutta, "The genetics of cholesteatoma. A systematic review using narrative synthesis," *Clin. Otolaryngol.*, vol. 43, no. 1, pp. 55–67, Feb. 2018.
- [24] J. Seok, J. J. Song, J. W. Koo, H. C. Kim, and B. Y. Choi, "The semantic segmentation approach for normal and pathologic tympanic membrane using deep learning," *BioRxiv*, vol. 8, Jan. 2019, Art. no. 515007.



TAEWAN KIM received the B.S. degree from the Department of Computer Science, Inha Technical College, Republic of Korea. He is currently pursuing the master's degree with the Department of Applied Artificial Intelligence, Hanyang University, Republic of Korea. His research interests include artificial intelligence and medical image processing.



KYOUNGHO OH received the Ph.D. degree from the Department of Otorhinolaryngology-Head and Neck Surgery, Korea University College of Medicine, Republic of Korea. He is currently an Assistant Professor with the Department of Otorhinolaryngology-Head and Neck Surgery, Ansan Hospital, Korea University College of Medicine. His current research interests include machine learning, head and neck cancer, and thyroid cancer.



YEONJOON LEE received the B.S. degree from Hanyang University, in 2012, and the Ph.D. degree in security informatics from Indiana University Bloomington. He is currently an Assistant Professor with the College of Computing, Hanyang University. His research interests include mobile security, the IoT security, network security, as well as artificial intelligence applications in medicine and healthcare.



JAEYOUNG KIM received the Ph.D. degree from the Department of Medicine, Korea University, Republic of Korea. He is currently a Research Professor with the Department of Medicine, Korea University Ansan Hospital, Republic of Korea. His research interests include optical imaging, AI-based clinical imaging analysis, and bioinformatics.



JUNE CHOI received the Ph.D. degree from the Department of Medicine, Korea University, Republic of Korea. He is currently a Professor with the Department of Otorhinolaryngology-Head and Neck Surgery, Korea University Ansan Hospital. His research interests include AI-based new drug development/chemicals or drug toxicity screening systems and virtual reality assisted pre-operative surgery.

...