# USE OF UNet++ ALGORITHM IN DETERMINATION OF EAR DISEASES

**Atadjanova Nozima Sultan-Muratovna**
Senior Lecturer at Tashkent University of Information Technologies
'Computer engineering' faculty, Department of Computer Systems

***https://doi.org/10.5281/zenodo.10973342***

***Abstract****. 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.*

***Keywords****: ResNet152, UNet++, TM, CNN models, media, ResNet-Bottleneck, receive proper.*

## 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 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.

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.

***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.

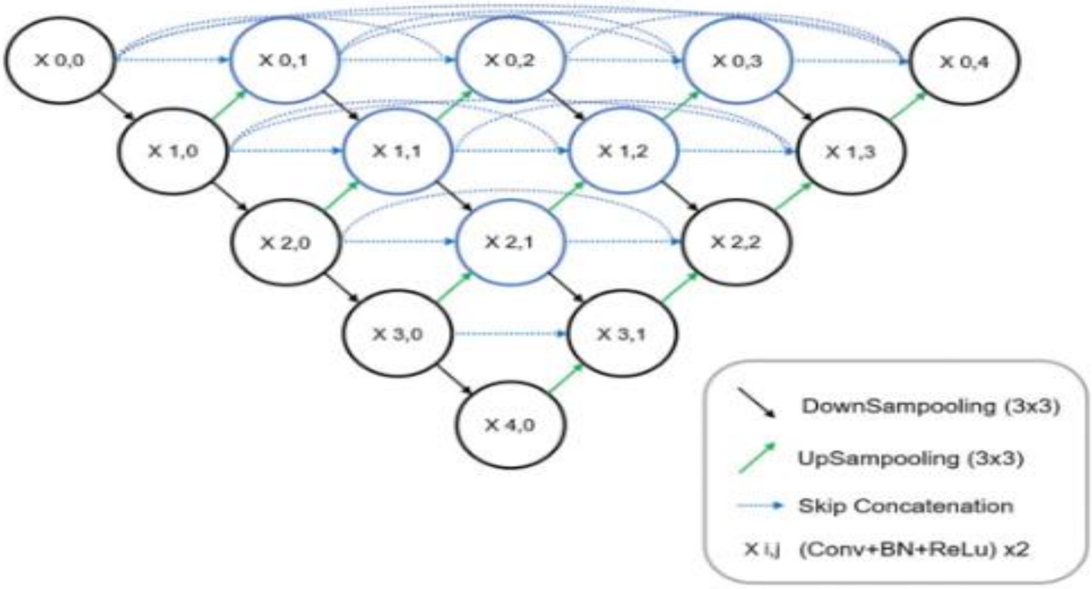

*Figure 1. UNet++ Backbone.*

***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.

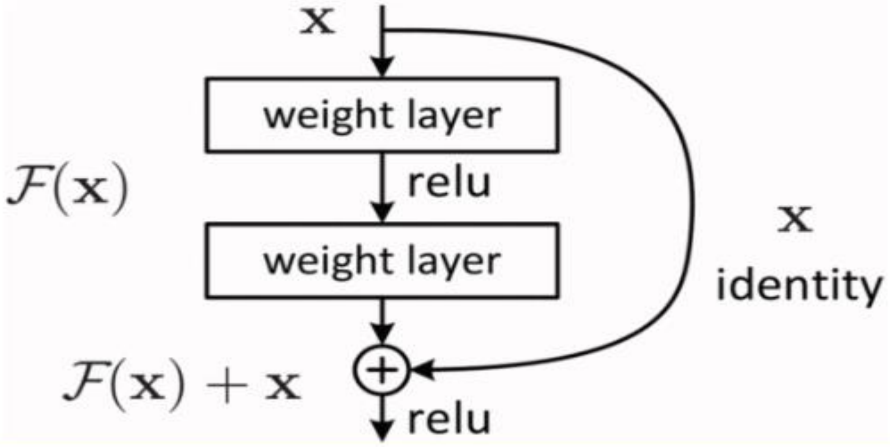
x
weight layer
relu
weight layer
$\mathcal{F}(\mathbf{x})$
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

***Figure 2. Residual learning: a building block***

## II. 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 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.

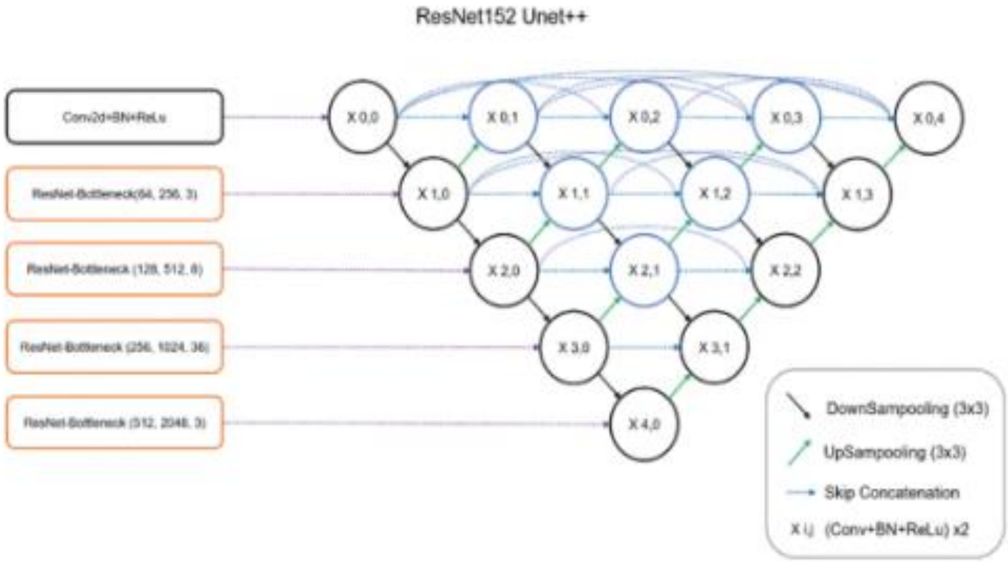

***Figure 3. Architecture of resnet152 unet++***

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.

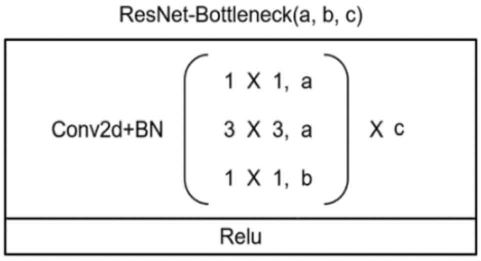

***Figure 4. Structure of encoder block in the resnet-Bottleneck.***

### *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].

### *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.

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

***Table 1. Learning environments of ResNet152 UNet++ models and CNN models***

## III. RESULTS AND DISCUSSION

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 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.

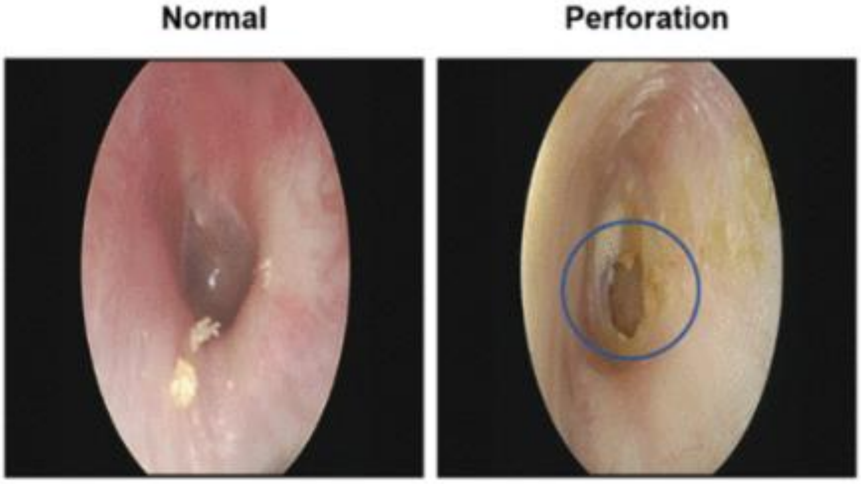

***Figure 6. The type of ear disease in the data collected***

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.

### *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:

$$Pixelaccuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Dicecoefficient = \frac{2 * TP}{(TP + FP) + (TP + FN)} \quad (2)$$

$$IoU = \frac{TP}{TP + FN + FP} \quad (3)$$

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:

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

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

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.

## III. RESULTS AND DISCUSSION

Segmentation results 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.

***Table 2. Performance comparison of segmentation models.***

| | | | | |
|---|---|---|---|---|
| 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,.3 | 87,8 | 0,120 |

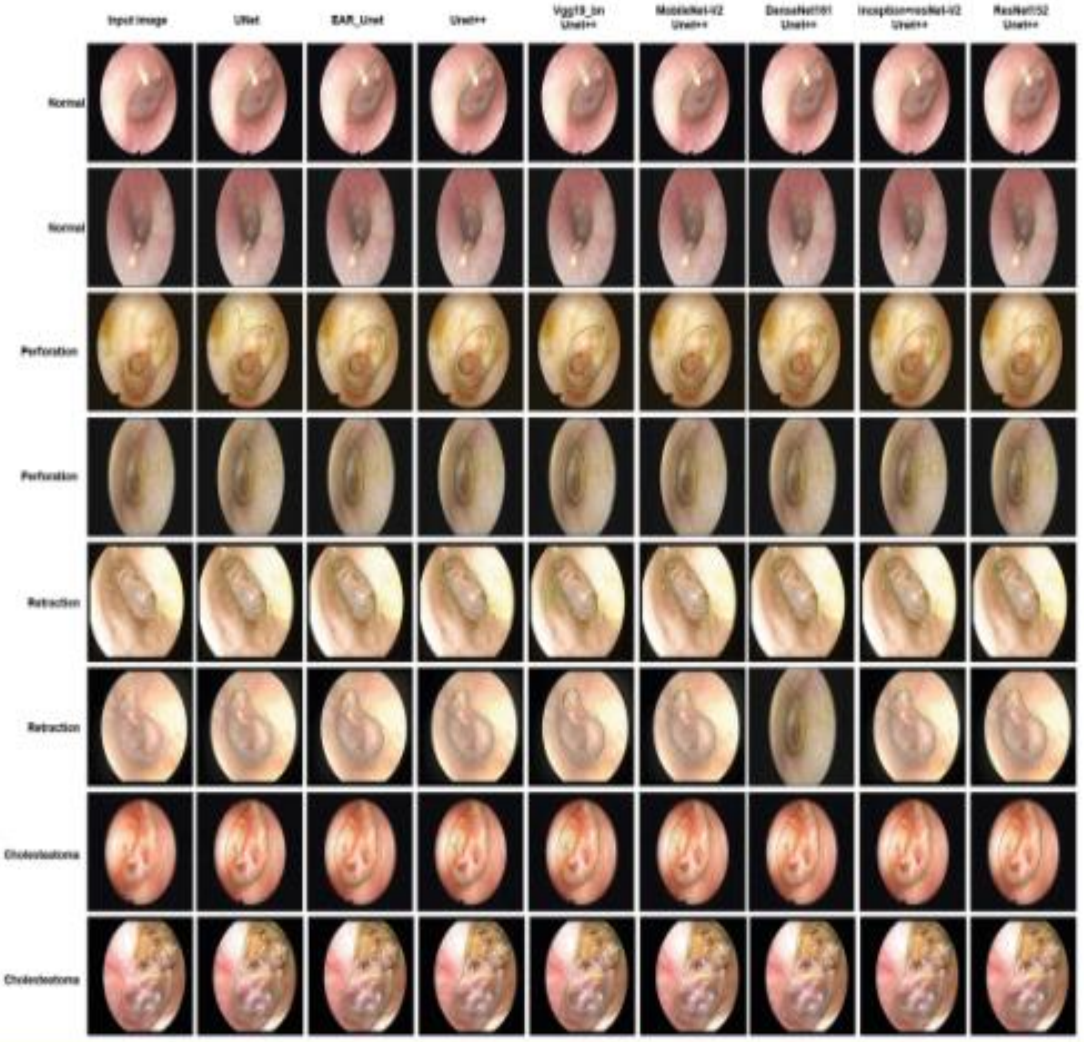

**Figure 7.** Comparison of results of segmentation of tympanic membrane (tm) and affected areas by various segmentation models. the tm and affected areas were segmented from endoscopic images of the tm, including normal and three ear diseases (perforation, retraction, and cholesteatoma). the ground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

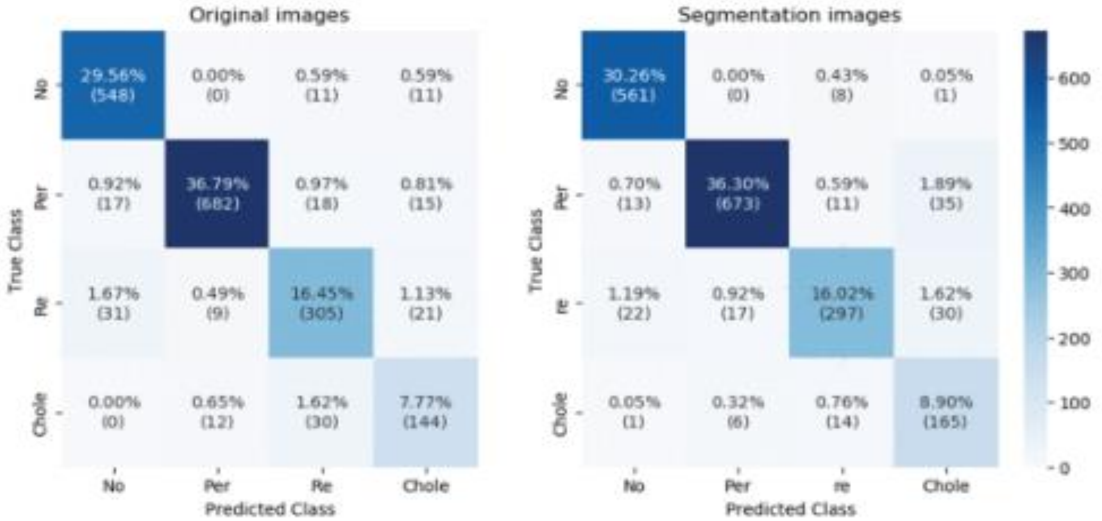

**Figure 8.** Comparison of confusion matrices for densenet161 model using original and segmentation images. (no=normal, per=perforation, re=retraction, chole=cholesteatoma).

## IV. CONCLUSION AND FUTURE WORK

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.

## REFERENCES

1. Jane Y. N. et al. A Vision-Based Approach for the Diagnosis of Digital Asthenopia //2023 4th International Conference on Signal Processing and Communication (ICSPC). – IEEE, 2023. – C. 163-167.
2. Jane, Y. N., Padmanabhan, K., Karthika, S., & Christiana, K. B. (2023, March). A Vision-Based Approach for the Diagnosis of Digital Asthenopia. In *2023 4th International Conference on Signal Processing and Communication (ICSPC)* (pp. 163-167). IEEE.
3. Singh E. et al. Maize Disease Multi-Classification: Leveraging CNN and Random Forest for Accurate Diagnosis //2024 International Conference on Automation and Computation (AUTOCOM). – IEEE, 2024. – C. 75-79.
4. Li D., Chen X., Chen S. Learning and Optimization of Patient–Physician Matching Index in Specialty Care //IEEE Transactions on Automation Science and Engineering. – 2023.
5. Li, Debiao, Xiaoqiang Chen, and Siping Chen. "Learning and Optimization of Patient–Physician Matching Index in Specialty Care." *IEEE Transactions on Automation Science and Engineering* (2023).

6. Başaran, Erdal, et al. "Normal and acute tympanic membrane diagnosis based on gray level co-occurrence matrix and artificial neural networks." *2019 international artificial intelligence and data processing symposium (IDAP)*. Ieee, 2019.
7. Jane Y. N. et al. A Vision-Based Approach for the Diagnosis of Digital Asthenopia //2023 4th International Conference on Signal Processing and Communication (ICSPC). – IEEE, 2023. – C. 163-167.
8. Jane, Y. N., Padmanabhan, K., Karthika, S., & Christiana, K. B. (2023, March). A Vision-Based Approach for the Diagnosis of Digital Asthenopia. In *2023 4th International Conference on Signal Processing and Communication (ICSPC)* (pp. 163-167). IEEE.
9. Bandyopadhyay A., Chaudhuri A., Mondal H. S. IR based intelligent image processing techniques for medical applications //2016 SAI Computing Conference (SAI). – IEEE, 2016. – C. 113-117.
10. Bandyopadhyay, Asok, Amit Chaudhuri, and Himanka Sekhar Mondal. "IR based intelligent image processing techniques for medical applications." *2016 SAI Computing Conference (SAI)*. IEEE, 2016.