Deep learning techniques for ear diseases based on segmentations of normal tympanic membrane

Yong Soon Park*, Jun Ho Jeon*, Tae Hoon Kong**, Tae Yun Chung*, Young Joon Seo**

*Gang-won Research Institute of ICT Convergence, GangNeung Won Ju National University, GangNeung, South Korea
**Department of Otorhinolaryngology, Yonsei University Wonju College of Medicine, 20 Ilsan-ro, Wonju, Gangwon-do 26426, South Korea
***Research Institute of Hearing Enhancement, Yonsei University Wonju College of Medicine, Wonju, South Korea

Acknowledgment

This work was supported by the Technology Innovation Program (20010587, Development and Dissemination on National Standard Reference Data) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and by the project Joint Demand Technology R&D of Regional SMEs funded by the Korea Ministry of SMEs and Startups in 2020 (Project No. S3035723).

Declarations of Interest: none

Funding: None

Correspondence

Young Joon Seo, MD, PhD
Department of Otorhinolaryngology, Yonsei University Wonju College of Medicine, 20 Ilsan-ro, Wonju, Gangwon-do 26426, South Korea.
E-mail: okas2000@hanmail.net; Tel.: +82-33-741-0644; Fax: +82-33-732-8287
ORCID ID https://orcid.org/0000-0002-839-4676
Conceptualization: YJS, Data curation: YSP, JHJ, THK. Formal analysis: YSP, JHJ, YJS. Methodology: YJS. Project administration: YJS. Resources: YJS. Software: JHJ. Supervision: YJS. Validation: YSP. Visualization: YJS. Writing - original draft: YJS. Writing - review & editing: YSP, YJS.
Abstract

Background: Otitis media is a common infection affecting people worldwide. Owing to the limited number of ear specialists and rapid development in telemedicine, several trials have been conducted for developing novel diagnostic strategies to improve the diagnostic accuracy and screening of patients afflicted with otologic diseases based on abnormal otoscopic findings. Although these strategies have demonstrated a high diagnostic accuracy for the tympanic membrane (TM), the insufficient explainability of such techniques limits their deployment in clinical practices.

Methods: Herein, we used a deep convolutional neural network (CNN) model based on the segmentation of a normal TM into five substructures (malleus, umbo, cone of light, pars flaccida, and annulus) to identify abnormalities in otoscopic ear images. The Mask R-CNN algorithm learned the labeled images. Subsequently, the values were combined according to the five substructures using a three-layer fully connected neural network to determine whether an ear disease was present.

Results: We obtained the receiver operating characteristic (ROC) curves of the optimal condition for the presence or absence of eardrum diseases according to each substructure or a combination of substructures. The results indicated that the highest area under the curve ranged from a 0.911 true-positive rate in the ROC curve combined with malleus, cone of light, and umbo, compared with the corresponding range of 0.737–0.873 in each substructure. Thus, an algorithm using these five important normal anatomical structures could prove to be explainable and effective in screening abnormal TMs.

Conclusion: This automated algorithm can improve the diagnostic accuracy to discriminate between normal and abnormal TMs and facilitate appropriate and timely referral consultations to improve the quality of life in public healthcare.
Keywords Tympanic membrane, Deep learning, Mask R-CNN, Otitis media, Otoendoscopy
1. Introduction

Otitis media has one of the highest disease prevalence rates in the world [1,2]. However, otoscopic screening is highly subspecialized, creating diagnostic difficulties for primary care providers who are relatively inaccurate in providing otologic diagnoses. The average diagnostic accuracy for acute otitis media (AOM) and otitis media with serous (SOM) using video otoscopy is only 51% and 46%, respectively. By contrast, the diagnoses of otolaryngologists are more accurate, although very far from perfect, with diagnostic accuracies of approximately 74%. Hence, there is a need for a new diagnostic strategy to improve the diagnostic accuracy and an improved method for screening patients with otologic diseases based on abnormal otoscopic findings. Because of the limited number of otolaryngologists covering all regions and the rapid development of telemedicine worldwide, tele-otoscopy may be beneficial in optimizing diagnoses and treatments for otitis media [3].

Artificial intelligence (AI)-based techniques, particularly disease screening tools to support clinicians’ decisions, have impacted and improved traditional healthcare circumstances. Deep learning is the leading AI method for a wide range of tasks, including medical imaging problems. Recently, deep neural networks have been applied with great success in otologic diagnosis. Moreover, many trials in the analysis of the tympanic membrane (TM) have shown the usefulness of using a deep learning model in the early detection and treatments of ear diseases [4,5]. However, although they showed
high-accuracy diagnoses with the TM, such techniques have limitations in being successfully deployed into clinical practice. These limitations are due to the underlying unexplainable 'black-box nature' of deep learning algorithms. For clinical use, studies should be conducted using an understandable algorithm in implementation among different algorithms. Moreover, an AI system to complement medical professionals should have a certain amount of explainability and allow human experts to retrace decisions and use their judgment [6]. Medical image segmentation has been an emerging biomedical image processing technology in deep learning algorithms used for the explainability of decisions [7].

The normal TM has several properties that makes it distinguishable from an infected TM. Therefore, the objective of this study was to develop and evaluate a software prototype using a deep convolutional neural network (CNN) model based on the segmentation of a normal TM into five substructures (malleus, umbo, cone of light, pars flaccida, and annulus) for identifying abnormalities in otoscopic ear images. The combination of thresholds among the five substructures would help improve the accuracy of discriminating differences between normal and abnormal TMs. This would also facilitate the appropriate triaging of middle ear diseases for primary clinicians and facilitate timely access of the transfer of findings to otologic specialists.

2. Methods
2.1. Ear imagery database

Medical images of eardrums from patients who visited the outpatient clinic in the Department of Otorhinolaryngology, S Hospital, from 2015 to 2020 were used retrospectively. This study was approved by the Institutional Ethics Review Committee (CR19081) and adhered to the principles of the Declaration of Helsinki. Eardrum images were taken by otolaryngology residents and professors using 2.7 and 100 mm Insight 0-degree telescopes (MIone, Korea) equipped with a full HD camera system with a 21.5'' LED monitor. The resolution of the images was 1920 × 1080, 60 fps with an illuminance of 30 and 000 lx. A total of 12,444 endoscope photos were reviewed and screened by two ENT surgeons. Mostly subjective and reliable images were selected by the specialized otologists for defined images in each disease after removing the images which had vague boundaries of substructures on tympanic membrane, 2,597 photos of which were deemed appropriate and hence included for examination. We classified 1,370 photos in the normal TM group and 1,227 in the abnormal TM group. We included nine ear diseases (acute suppurative otitis media (AOM), SOM, otitis media with mucoid (MOM), chronic otitis media without perforations (COM w/o P), chronic otitis media with perforations (COM w P), traumatic drum perforation (traumatic TM), tympanosclerosis (sclerosis TM), tympanostomy tube insertions (Tube), and congenital cholesteatoma (Chole)), as shown in Fig. 1. This retrospective study was approved by the Hospital
Institutional Review Boards (Yonsei IRB-CR319081). Among the 2,597 images used for this study, the training set was composed of 2,358 images (90%), and the validation set was composed of 239 images (20%) chosen randomly.

2.2. Data preprocessing

Eardrum photos were labeled as 10 categories, including normal TM and nine diseases (acute otitis media; AOM, otitis media with serous; SOM, otitis media with mucous; MOM, chronic otitis media without perforation; COM w/o P, chronic otitis media with perforation; COM w P, traumatic drum perforation; traumatic TM, tympanosclerosis; Sclerosis TM, tympanostomy tube inserted status; Tube, and congenital cholesteatoma; Chole). We followed the guideline of the American Academy of Otolaryngology–Head and Neck Surgery (2016) [8] to define the diseases with the retrospective clinical records of patients. AOM refers to the rapid onset of signs and symptoms of inflammation of the middle ear. Otitis media with effusion (OME) is the presence of fluid in the middle ear without signs or symptoms of acute ear infections. COM is an OME persisting for 3 months from the date of onset (if known) or from the date of diagnosis (if onset is unknown). Furthermore, we added the tympanosclerosis group for comparison to a normal eardrum because plaques could shade the visualization of normal substructures, such as the cone of light, umbo, or malleus. Images of a normal TM should not have any surgical history of ear and chronic otitis media, and they also
should include the five substructures (malleus with lateral process and handle, whole annulus, pars flaccida, umbo, and cone of light) [9]. Along with the unique pearl gray or white color in the pars tensa of the TM, a normal TM has a translucent and concave-shaped membrane (umbo, center of the TM) with malleus bone visibility. Light reflection, also known as the cone of light, is another important feature. Earwax should be below 10% of the whole eardrum in the image. The number of images used and samples representing each classification is shown in Fig. 1.

2.3. Labeling segmentation extraction from the normal TM

An in-house graphic user interface software implemented on MATLAB2019a® (MathWorks, Inc., Natick, Massachusetts, United States) was used for manual labeling. As shown in Fig. 2, filmed otoscopic images that contained both ears were divided into two segmented images, i.e., the right and left parts. The images were manually labeled with the contours of the five substructures (malleus with lateral process and handle, whole annulus, pars flaccida, umbo, and cone of light) by two specialized otologist. When all five substructures in one image were confirmed by the otologists, it was used as the training data. We used “LabelMe,” a database and an online annotation tool that allows the sharing of images and annotations. This online tool provides functionalities, such as drawing polygons, querying images, and browsing the database [10]. The labeling results were converted into JSON files.
2.4. Deep learning models to discriminate the normal TM and abnormal TM

The learning process shown in Fig. 2A was performed. The mask R-CNN by Matterport Inc. (Mountain View, CA, USA) (https://github.com/matterport/Mask_RCNN) with ResNet-50 was used to detect and segment the contours of predicted substructures. The classification result of the mask R-CNN was passed through a three-layer fully connected neural network to detect the presence of an ear disease. To extract features from the eardrum image, Matterport’s mask R-CNN model was used, and ResNet-101 was used as the backbone [11]. Mask R-CNN was trained with a batch size = 4, learning momentum = 0.9, and weight decay = $10^{-4}$. Up to epoch 100, only the head part was trained with a learning rate of $10^{-2}$; up to 340, layers of stage 4 or higher were trained with a learning rate of $10^{-3}$; and finally up to epoch 400, the entire layer was trained with a learning rate of $10^{-4}$. In the model training process, random flips were performed in the horizontal direction of the inputted images to augment the data, and the training datasets were inputted into a deep neural network to extract the features of the eardrum image sample. The k-fold cross-validation was used as a tool to evaluate machine learning models. Then, we observed the performance of the training model until the values were stabilized. In addition, based on the weight obtained through the above learning process, we extracted the classification probability value of the input data of the three-layer fully connected neural network. Through this neural network, the presence or absence
of an eardrum disease is judged. In the neural network for judging the presence of an ear disease, two fully connected layers of size 32 were used. To prevent overfitting, L2 regularization of $10^{-4}$ and dropout of 0.5 were applied, and the epoch value was set to 1000 for learning. While processing the data, we extracted the mask R-CNN classification probability value by entering the input of the above process. Then, we identified the missing value for the class that was not detected. This study was performed using the TensorFlow deep learning framework on multiple GPU (Tesla V100, NVIDIA) to increase the training speed of neural network. Supple.fig.1 shows the training and validation loss graphs according to the epoch of the mask R-CNN classification, and the results was obtained by inferring the test data based on the weights obtained through transfer learning.

2.5. Statistical analysis

The metrics used for evaluating the performance of the final model were calculated using a one-versus-rest strategy applied separately for each class. The F1 score provides a harmonic mean of the sensitivity (recall) and positive predictive value (precision). Receiver operating characteristic (ROC) analyses, including the area under the curve (AUC), were performed separately for each class by varying the cut-off applied to the probabilistic output of the neural network for the class in question using the SPSS 23.0 statistical software. The micro-averaged ROC AUC was calculated using Scikit-learn.
3. Results

3.1. Accuracy of TM segmentation

Supple.fig.1 shows the results trained with several epochs to obtain good results in training and validation loss. An example of the classification results with five substructures using the mask R-CNN model is shown in Fig. 3. We analyzed the results by dividing the images into individual structures, and we evaluated the detection rate and segmentation accuracy using the intersection over union (IoU) score. The average IoU scores in the normal group were $0.9 \pm 0.14$ in the malleus, $0.99 \pm 0.05$ in the annulus, $0.88 \pm 0.15$ in the cone of light, $0.84 \pm 0.17$ in the umbo, and $0.89 \pm 0.13$ in the pars flaccida (Fig. 3). The most significant differences were found between the normal group and other disease groups in the malleus and pars flaccida. With a cut-off value of 0.8, we could distinguish a normal TM from abnormal TMs with each substructure of the malleus, cone of light, umbo, or pars flaccida.

3.2. Accuracy of discriminating between Normal and Abnormal TM through deep learning
Because the missing values below 0.5 of the IoU values were applied, the optimal missing value could be found and replaced in each substructure. To make the best distinction between the normal and abnormal groups, the inputted missing values for each class were as follows: malleus = 0.2, cone of light = 0.7, pars flaccida = 0.2, annulus = 0.8, and umbo = 0.3.

We achieved the accuracy of Mark R-CNN according to the fine tunings of the learning rate with 0.01, 0.001, 0.0001, 0.00001, and scheduled, and we had a better result in 0.001 learning rate (Fig. 4A). We also have tested the fine tuning layers with stage 1(Network heads), stage 2(over Resnet Stage4), and stage 3(all layers), and we have chosen the layer of stage 2 with lowest validation loss and lowest computation power (Fig. 4B). The model with the best performance on the validation data was evaluated using the test set.

The ROC curves for each subgroup are shown in Fig. 4C. The highest AUC was 0.873 in the umbo, 0.826 in pars flaccida, 0.797 in cone of light, and 0.794 in malleus. The annulus (AUC=0.737) did not have a suitable structure for discrimination between normal and abnormal TMs. To increase sensitivity and specificity, we combined several segmentations to classify the TM in diseases from the normal TM (Fig. 4D). Based on the three-layer fully connected neural network, we obtained the ROC curves of the optimal condition for the presence or absence of eardrum diseases. Except for the annulus substructure, we could obtain good prediction results with combinations of the other four substructures (20 cases of combinations). We could also diagnose abnormal TMs with malleus,
cone of light, and umbo compared with the normal TM, which showed a satisfactory result (AUC = 0.911). This value was higher than those with each substructure. The scores of precision and recall in the combination of malleus, cone of light, and umbo was the best compared to each substructure or other combinations (Fig. 4E and F).

**Accuracy of discriminating between Normal and each disease TM**

We finally have applied the deep learning model of gradient boosting classifier (gbc) in this study, because it had highest values in accuracy, precision, recall, and f1 (supple. fig. 2). We have compared the precision, recall, f1, and support values between the normal TM and each disease TM groups (AOM, SOM, MOM, COM w/o P, COM w P, traumatic, sclerosis, tube, and chole). We had the results that there were significant values over 0.911 of AUC in SOM, MOM, COM w/o P, COM w P, Traumatic TM, Tube, and Chole). The specificity of disease was lack in AOM and Sclerosis TM group). The combination group of SOM, COM w P, traumatic TM had the most significant value (precision; 0.950, recall; 0.960) in compared to the normal TM group (Fig. 5).

**4. Discussion**

This deep neural network framework is the first prototype implementation of a CNN for the
substructure-based classification of the TM with diseases as compared to a completely healthy ear. The classification accuracy of the current model reached 91.1% by the combination of malleus, cone of light, and umbo substructures which is higher than those of deep learning models based on a single substructure. We could increase the true-positive rate of precision up to 0.950 and the false-negative rate of recall up to 0.960 to discriminate TMs between the normal and the certain diseases.

A medical diagnosis needs to be transparent, understandable, and explainable to gain the trust of clinicians. Hence, we believe that this ear image segmentation would be an important first step in numerous applications. It segmented images into five anatomically meaningful regions (malleus, umbo, cone of light, annulus, and pars flaccida), based on which substructures could be extracted. Until now, other recent deep learning studies on ear diseases have classified several diseases using whole images of the TM. Cha et al. proposed an image classification model based on transfer learning with a deep CNN, which classified middle ear and EAC pathologies into six categories with a mean accuracy of 93.73% [4]. Another study reported that otoendoscopic images of the eardrum and external auditory canal were classified into eight categories. The classification accuracy of the current model reached 95.59% [5]. Even though they showed high accuracies with their models, they are not sufficient to show a certain amount of explainability and allow human experts to retrace decisions and use their judgment for disease diagnoses. The analysis of medical images with segmentations delineating the anatomical or pathological structures from medical images can help explain deep learning models in medical imaging diagnoses. Typical medical image segmentation tasks include brain and tumor segmentation, cardiac segmentation, liver and tumor segmentation,
and cell and subcellular structures [12]. The AUC values ranged between 0.864 and 0.937 for diagnosing lung nodules or lung cancer on the chest X-ray or CT scan. For breast imaging, the AUC values ranged between 0.868 and 0.909 for diagnosing breast cancer on mammogram, ultrasound, MRI, and digital breast tomosynthesis. Compared with these values, our deep learning algorithm with combinations of five substructures had high AUC values ranging from 0.905 to 0.932 in the ROC curve. Hence, using the five important normal anatomic structures would be an explainable and effective algorithm for screening abnormal TMs.

We included nine diseases (AOM, SOM, MOM, COM w/o P, COM w P, traumatic TM, sclerosis TM, tube, and chole) in the group with abnormal TMs. However, we did discriminate the abnormal TMs compared with the normal TMs based on five segmentations; hence, we did not classify all diseases into subgroups. Only distinguishing between normal and abnormal TMs would be meaningful in the primary medical care system. This deep learning model would support the primary screening for ear diseases before seeing a specialist. If a person using it is diagnosed with an abnormal TM, then he or she must consult an ENT specialist for further detailed diagnosis with additional tests, including hearing test or computed tomography of the temporal bones. The diagnosis of various ear diseases with deep learning algorithms should not overly predict diseases just by using a single otoendoscopic image.

Mask R-CNN has enormous importance in medical imaging analysis. He et al. [11] proposed a mask R-CNN model in 2017. Mask R-CNN is a multitask network and can simultaneously implement
detection and segmentation. It also detects small objects effectively, such as ear substructures, because of the introduction of the feature pyramid network mechanism. The cascaded CNN was designed with multiple layers of anisotropic and dilated convolution filters for automatic segmentation for brain tumor [14]. Mask R-CNN is an important AI-based scheme that has been used in automatic nucleus segmentation [15], lung nodule detection and segmentation [16], liver segmentation [17], automated blood cell counting, and multiorgan segmentation [18]. In ENT research, a CNN with transfer learning was used for the classification of dental diseases [19]. Mask R-CNN can also be used in specialized areas, such as oral pathology [20]. Because the substructures of the TM can overlap one another and be distorted with disease conditions, such as perforations or ear wax, mask R-CNN was chosen in our study as the best deep learning algorithm to enable each substructure to be detected separately on otoendoscopic images. It could effectively perform object detection and instance segmentation for five substructures of the TM.

Nonetheless, this study has a limitation. Considering diseases, other considerable characteristics should be considered by clinicians except for the five substructures (malleus, umbo, cone of light, annulus, and pars flaccida). The red color of the TM can be noticed as the bulged shape in the TM of AOM. The bulged shape of the TM in otitis media changes the central concavity of the membrane with loss or relocations of the cone of light [21]. OME shows small bubbles of fluid behind the TM with several fluid characteristics. However, light reflection and malleus bone can also be noted. The discrimination between normal and abnormal TMs was analyzed with only five substructures as normal TM components, not including the color, perforation of the TM, and otorrhea from the
middle ear. We did not include further parameters because we did not try to classify the diseases into more detailed subgroups. The absence of perforations of the membrane in otitis media with serous will affect the judgement of the deep learning model for normal TMs. Although the colors of the TM are important to decide whether there are fluids behind the membrane, the lights through otoendoscopy may cause biases in images. Based on this algorithm, we will add other parameters to discriminate different diagnoses among various diseases in future studies. A collaborative, multi-institutional approach to obtaining high-quality images would be necessary.

Considering many previous studies regarding the diagnosis of ear diseases with considerable accuracy rates [4,22], this study demonstrates the usefulness of applying multiple-object detections and segmentations of five substructures using mask R-CNN in otoendoscopic images to discriminate ear diseases and the normal TM. It would help clinicians sufficiently to show a certain amount of explainability and allow human experts to retrace decisions and use their judgment for abnormal eardrum. The proposed model may achieve a high accuracy similar to those of primary clinicians who judge whether patients must meet a specialist or not. We hope that this automated algorithm would improve the diagnostic accuracy for abnormal TMs and facilitate appropriate and timely clinical referral consultations to improve the quality of life in public healthcare.
References


Figure legends

Figure 1. Normal anatomic substructures of tympanic membrane (A and B). (C) Otoendoscopy image and two diagnostic classes of normal and abnormal tympanic membranes including 9 diseases subgroups. acute otitis media; AOM, otitis media with serous; SOM, otitis media with mucous; MOM, chronic otitis media without perforation; COM w/o P, chronic otitis media with perforation; COM w P, traumatic drum perforation; traumatic, tympanosclerosis; sclerosis, tympanostomy tube inserted status; tube, and congenital cholesteatoma; chole.
Figure 2. Pre-processing with ‘Lebelme’. (A) A schematic flow of image analysis (B) Labeling with contours of five substructures (malleus with lateral process and handle, whole annulus, pars flaccida, umbo, and cone of light) have been done manually by specialized otologist. (C) Sample images which showed the delineation of five substructures on normal tympanic membranes (D) Example of result for 5 substructures analyzed with mask R-CNN

Figure 3. The comparisons of IoUs in the subgroups according to 5 substructures (malleus, annulus, cone of light, umbo, and pars flaccida). Y-axis shows the values of IoUs.

Figure 4. Results of Mark R-CNN. (A) Fine tunings according to the learning rate with 0.01, 0.001, 0.0001, 0.00001, and scheduled. (B) Fine tunings according to the layers with stage 1(Network heads), stage 2(over Resnet Stage4), and stage 3(all layers). The layer of stage 2 showed the lowest validation loss and lowest computation power. (C) ROC curves of the 3-layer fully connected neural network algorithm according to each substructure. (D) ROC curve according to combined substructures. (E) Precision and Recall curves on each substructure. (F) Precision and Recall curves on combined substructures. We could obtain good prediction results with combinations of the other four substructures. We could also diagnose abnormal TMs with malleus, cone of light, and umbo compared with the normal TM, which showed a satisfactory result (AUC = 0.911).
Figure 5. The Matrix of precision, recall, f1, and support values between the normal TM and the combination group of SOM, COM w P, Traumatic TM. (A) Matrix of raw cases sorted between True and Predicted classes. (B) Matrix of proportion for precision, recall, f1, and support between the normal and the combination groups. The combination group of SOM, COM w P, Traumatic TM had the most significant value (precision; 0.950, recall; 0.960) in compared to the normal TM group.

Supplement Figure 1. Accuracy and loss on the training and validation datasets with TensorBoard between normal and abnormal TM across epochs. (A) Accuracy in the training and validation both improve until the model converges at around 1000 epochs (B) Loss curve in the training and validation. The training (blue line) and validation (red line)

Supplement Figure 2. Models of deep learning to discriminate the normal and the abnormal TMs.
Figure 1

A

[Diagram showing anatomical structures of the ear with labels such as Long crus of incus, Posterior malleolar fold and chorda tympani nerve, etc.]

B

[Image showing a close-up view of the ear with labeled structures like Lateral process, Handle, Incus visible through drum, etc.]

C

<table>
<thead>
<tr>
<th>Normal (n=1370)</th>
<th>Diseases (n=1227)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Normal Image" /></td>
<td><img src="image2" alt="Diseases Images" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AOM</th>
<th>SOM</th>
<th>MOM</th>
<th>COM w/o P</th>
<th>COM w P</th>
<th>Traumatic</th>
<th>Sclerosis</th>
<th>Tube</th>
<th>Chole</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3" alt="AOM Image" /></td>
<td><img src="image4" alt="SOM Image" /></td>
<td><img src="image5" alt="MOM Image" /></td>
<td><img src="image6" alt="COM w/o P Image" /></td>
<td><img src="image7" alt="COM w P Image" /></td>
<td><img src="image8" alt="Traumatic Image" /></td>
<td><img src="image9" alt="Sclerosis Image" /></td>
<td><img src="image10" alt="Tube Image" /></td>
<td><img src="image11" alt="Chole Image" /></td>
</tr>
</tbody>
</table>
Figure 2

A

Input data → Mask-RCNN → Classification result

Prediction ← 3-layer fully connected neural network ←

B


C

D

annulus 0 593
more Precise 0 548

Corneal 0 709
Corneal 0 709
Figure 4

A. Validation loss

B. Validation loss (learning rate: 0.001)

C. ROC Curve

D. ROC Curve

E. Precision vs. Recall Curve

F. Precision vs. Recall Curve
Figure 5

A. GradientBoostingClassifier Confusion Matrix

B. GradientBoostingClassifier Classification Report