

ORIGINAL RESEARCH

Use of artificial intelligence for the diagnosis of cholesteatoma

Christopher C. Tseng MD  | Valerie Lim MD, MBS | Robert W. Jyung MD

Department of Otolaryngology – Head and Neck Surgery, Rutgers New Jersey Medical School, Newark, New Jersey, USA

Correspondence

Robert W. Jyung, Department of Otolaryngology – Head and Neck Surgery, Rutgers New Jersey Medical School, 90 Bergen St., Suite 8100, Newark, NJ 07103, USA.
Email: jyungw@njms.rutgers.edu

Abstract

Objectives: Accurate diagnosis of cholesteatomas is crucial. However, cholesteatomas can easily be missed in routine otoscopic exams. Convolutional neural networks (CNNs) have performed well in medical image classification, so we evaluated their use for detecting cholesteatomas in otoscopic images.

Study Design: Design and evaluation of artificial intelligence driven workflow for cholesteatoma diagnosis.

Methods: Otosopic images collected from the faculty practice of the senior author were deidentified and labeled by the senior author as cholesteatoma, abnormal non-cholesteatoma, or normal. An image classification workflow was developed to automatically differentiate cholesteatomas from other possible tympanic membrane appearances. Eight pretrained CNNs were trained on our otoscopic images, then tested on a withheld subset of images to evaluate their final performance. CNN intermediate activations were also extracted to visualize important image features.

Results: A total of 834 otoscopic images were collected, further categorized into 197 cholesteatoma, 457 abnormal non-cholesteatoma, and 180 normal. Final trained CNNs demonstrated strong performance, achieving accuracies of 83.8%–98.5% for differentiating cholesteatoma from normal, 75.6%–90.1% for differentiating cholesteatoma from abnormal non-cholesteatoma, and 87.0%–90.4% for differentiating cholesteatoma from non-cholesteatoma (abnormal non-cholesteatoma + normal). DenseNet201 (100% sensitivity, 97.1% specificity), NASNetLarge (100% sensitivity, 88.2% specificity), and MobileNetV2 (94.1% sensitivity, 100% specificity) were among the best performing CNNs in distinguishing cholesteatoma versus normal. Visualization of intermediate activations showed robust detection of relevant image features by the CNNs.

Conclusion: While further refinement and more training images are needed to improve performance, artificial intelligence-driven analysis of otoscopic images shows great promise as a diagnostic tool for detecting cholesteatomas.

Level of Evidence: 3.

KEYWORDS

artificial intelligence, cholesteatoma, diagnosis, neural network, otoscopy

Presented as a poster presentation at the AAO-HNSF Virtual Annual Meeting, September 12–October 25, 2020.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Laryngoscope Investigative Otolaryngology* published by Wiley Periodicals LLC on behalf of The Triological Society.

1 | INTRODUCTION

Cholesteatoma is an invasive process of squamous epithelium within the temporal bone characterized by keratin entrapment and bone destruction. It affects about 9.2 per 100,000 adults and about 3 per 100,000 children yearly.¹ Left untreated, cholesteatomas can cause hearing loss, vestibular dysfunction, and even intracranial complications due to progressive bone destruction.²

Currently, the most practical and effective approach for diagnosing cholesteatomas is otoscopic examination by an experienced clinician, usually an otolaryngologist or otologist.³ Widely accepted diagnostic features of congenital cholesteatomas, as described by Levenson et al., include: “a white mass medial to a normal tympanic membrane, normal pars flaccida, and normal pars tensa.” Conversely, retraction pockets and perforations in the tympanic membrane are particularly associated with the development of primary and secondary acquired cholesteatomas respectively, along with accumulated keratin debris.³⁻⁵

However, accurate visual diagnosis can be challenging, as cholesteatomas may be mistaken for cerumen, granulation tissue, otitis externa, serous otitis media, perforation, neoplasm, or other potential pathologies. As a result, early lesions can easily be missed during routine otoscopic exams performed by primary care providers due to multiple factors such as the lesser detail seen on standard otoscope compared to an otomicroscope, used by ear specialists.

Machine learning has the capability to identify important underlying patterns in a given dataset and extrapolate those findings to solving various clinical problems, an approach that has been applied successfully to medical image classification across a range of specialties including radiology, ophthalmology and oncology.^{6,7} One especially high performing machine learning technique, called convolutional neural network (CNN), has been utilized to detect breast cancer on multiple imaging modalities,^{8,9} colon polyps on CT colonography,¹⁰ interstitial lung disease on CT,¹¹ and various pathologies on fundoscopic images.¹²⁻¹⁴

Transfer learning, a technique where a machine learning model trained for a certain classification task is applied to a different but related task, has precedence in medical image classification, successfully employed in detecting diabetic retinopathy on fundoscopy.¹⁵⁻¹⁷ This method has been an especially prominent mainstay of image classification research in otolaryngology, with a number of studies investigating the use of machine learning along with transfer learning to identify a variety of pathologies including otitis media and TM perforations from otoscopic images.¹⁸⁻²² Chen et al. fully trained several standard CNNs to detect 10 common middle ear conditions including acute otitis media and acute myringitis.²³ They then applied transfer learning of the best performing weights to mobile CNNs, achieving an accuracy of 97.6%. Notably, Miwa et al. used transfer learning to train a CNN assisted by digital image enhancement modalities to distinguish cholesteatoma matrix, cholesteatoma debris, and a normal middle ear mucosa.²⁴ Accuracies using the different digital enhancement modalities were compared. In the task of identifying cholesteatoma matrix lesions, they achieved sensitivity of 34.6% and 42.3%, and with a specificity of 81.3% and 87.5%, respectively. The authors attributed the relatively low sensitivity to their small number of training images

as well as artifact in the training images. Along similar lines, transfer learning has also been applied to temporal bone CT imaging analysis. Wang et al. utilized a pretrained CNN to classify normal, chronic suppurative otitis media, and cholesteatoma, with overall model accuracy of 76.7% and cholesteatoma specific accuracy of 76%.²⁵ Toward this end, with our larger dataset of high-quality cholesteatoma images and diverse range of non-cholesteatoma images, we hypothesize that pretrained CNNs can achieve a more robust performance in diagnosing cholesteatomas compared to previous studies.

2 | MATERIALS AND METHODS

This study received approval by the Rutgers New Jersey Medical School Institutional Review Board (protocol ID: Pro20170000936) and every image was deidentified. Otoscopic images of the tympanic membrane (TM) were retrospectively obtained from the records of patients seen at the faculty practice of the senior author. The images were captured using a Karl Storz 0.4 × 6 cm Tele-Otoscope with HOPKINS® Straight Forward Telescope 0° (Karl Storz SE & Co. KG, Tuttlingen, Germany) and Storz IMAGE1 SPIES H3-Z HD camera head (Karl Storz SE & Co. KG, Tuttlingen, Germany) with 1920 × 1080-pixel resolution at 60 frames per second. These images were then manually labeled by the senior author into three categories based on the gross appearance of the TM and external auditory canal on otoscopy, specifically: cholesteatoma, abnormal non-cholesteatoma, and normal. A cholesteatoma image was defined as a TM with visible cholesteatoma on otoscopy. Any images with cholesteatoma and any other coexisting pathology were excluded from the study. All cholesteatoma cases were surgically confirmed with no additional superficial pathologies that could be mistaken for cholesteatoma. An abnormal non-cholesteatoma image was defined as a TM with visible pathology or deformity on otoscopy that is not a cholesteatoma (e.g., serous otitis media, TM perforation, neoplasm). A normal image was defined as a TM with normal anatomical landmarks and no evidence of pathology on otoscopy.

An artificial intelligence (AI) driven workflow was then built using the TensorFlow machine learning platform to automatically categorize these otoscopic images.²⁶ CNN models were trained toward three specific binary classification tasks: cholesteatoma versus normal, cholesteatoma versus abnormal non-cholesteatoma, and cholesteatoma versus non-cholesteatoma (normal images combined with abnormal non-cholesteatoma images). First, images were preprocessed and resized into 224 × 224-pixel input images, then randomly split into separate training, validation, and testing datasets. 80% of images were allocated for CNN training, 10% for validation to iteratively select the top performing model during training, and the remaining 10% retained to evaluate the performance of the final trained model. The next step was data augmentation, a machine learning technique where different transformations such as rotation and horizontal and vertical reflection were applied to the original images, as the addition of these transformed images to the training dataset allows the CNNs to better identify important image features for classification.²⁷ Next, eight distinct CNNs from the Python machine learning library Keras were applied:

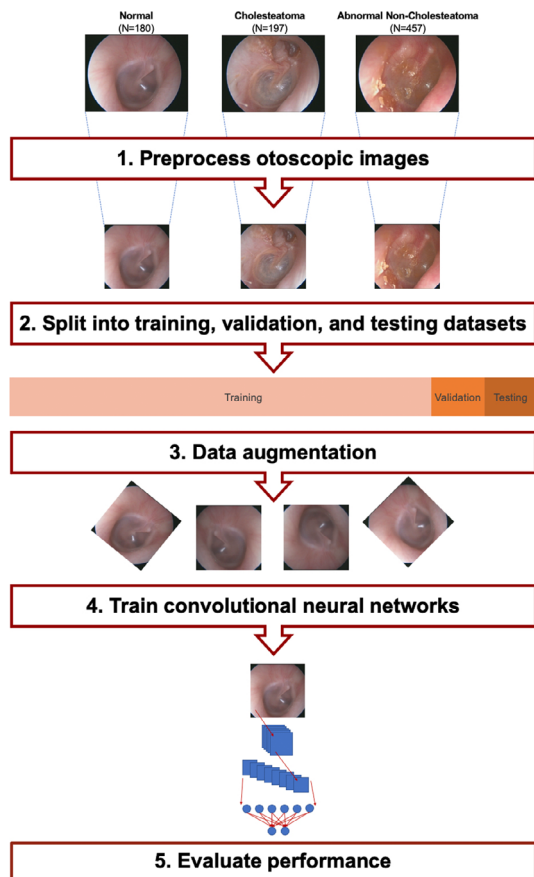


FIGURE 1 Illustrated summary of artificial intelligence driven workflow

VGG19, MobileNetV2, DenseNet201, InceptionV3, ResNet152V2, Xception, InceptionResNetV2, and NASNetLarge.²⁸ The CNNs were optimized during training using the Adam algorithm with a learning rate of 0.0001.²⁹ The final classification layer of each CNN was set as a trainable Dense layer to output the prediction of the neural network. All CNNs were trained and tested independently of each other and for their select classification tasks. Information learned from each CNN was not transferred to other CNNs. The number of training epochs was set at 100 epochs for each model. This CNN workflow was run on a computer with a 2.6 GHz 6-Core Intel Core i7 processor (Santa Clara, CA, USA), Intel UHD Graphics 630 1536 MB graphics card (Santa Clara, CA, USA), and 16 GB 2400 MHz DDR4 memory.

The selected CNNs are known neural network architectures designed and developed for image classification, representative of unique approaches with their own advantages in effectiveness and efficiency. For example, MobileNetV2 is a lightweight neural network capable of running on the limited computing power of a mobile device,³⁰ whereas the more complex NASNetLarge searches for the optimal neural network architecture to classify a given image dataset.³¹ Moreover, these CNNs have been pretrained on ImageNet, a large database containing over 14 million natural images labeled with over 20,000 possible standard categories, used extensively to research and benchmark neural network performance.^{32,33} Thus through transfer learning, CNNs with image feature weights calculated from a fully completed training process

were utilized as the base model, where a single high-level classifier layer for processing outputs from the base model was trained specifically on important features from our dataset. Following training, each CNN was evaluated by measuring their performance in predicting the categories of images reserved in the testing dataset. Evaluation metrics measured included accuracy (proportion of correctly classified images), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area under the receiver operating characteristic curve (AUROC), and overall runtime in minutes. Receiver operating characteristic (ROC) curves were also generated to illustrate classifier performance. This process was repeated for all three classification tasks. The entire workflow is summarized and illustrated in Figure 1.

To counteract overfitting (where CNNs become overly proficient at identifying training image qualities and perform well in classifying training images but are unable to accurately classify unseen testing images), we incorporated dropout and reduced the learning rate to increase classifier generalizability. Dropout is a technique that involves randomly omitting neurons from the network during the training process.³⁴ By temporarily excluding a proportion of neurons and their connections during every training iteration, each neuron should become less overly reliant on other neurons and thus improve its contribution to the given classification task, making the network's overall performance more robust. Toward this end, a dropout layer was implemented with the dropout rate set at 0.2, where 20% of input neurons were randomly selected to be dropped out.

To gain insight into how CNNs interpret images, intermediate activations from the final trained models can be extracted and visualized. An intermediate activation is the transformation of an input image by the activation function of a CNN layer, where activation signals are calculated for the input image based on the specific image features encoded by that layer, thus regions of the input image with greater activation signals had characteristics considered to be more important to the CNN.^{35,36} Generally speaking, initial CNN layers encode broad image features such as brightness and boundaries, whereas deeper layers encode more abstract features such as color and specific angles. As a result, these initial layers are the most amenable for visualization since intermediate activations become more abstract deeper into the CNN. To demonstrate the CNN's capability to identify and delineate potentially important image features, intermediate activations generated and visualized with a heatmap for a given otoscopic image from the initial convolutional layer of the trained InceptionResNetV2 neural network (Figure 2). In this case, InceptionResNetV2 is used as an example; other CNN intermediate activation maps appear similarly.

3 | RESULTS

3.1 | Data characteristics

A total of 834 otoscopic images of the tympanic membranes were collected, which were categorized by the senior author into 197 cholesteatoma images, 457 abnormal non-cholesteatoma images, and 180 normal images.

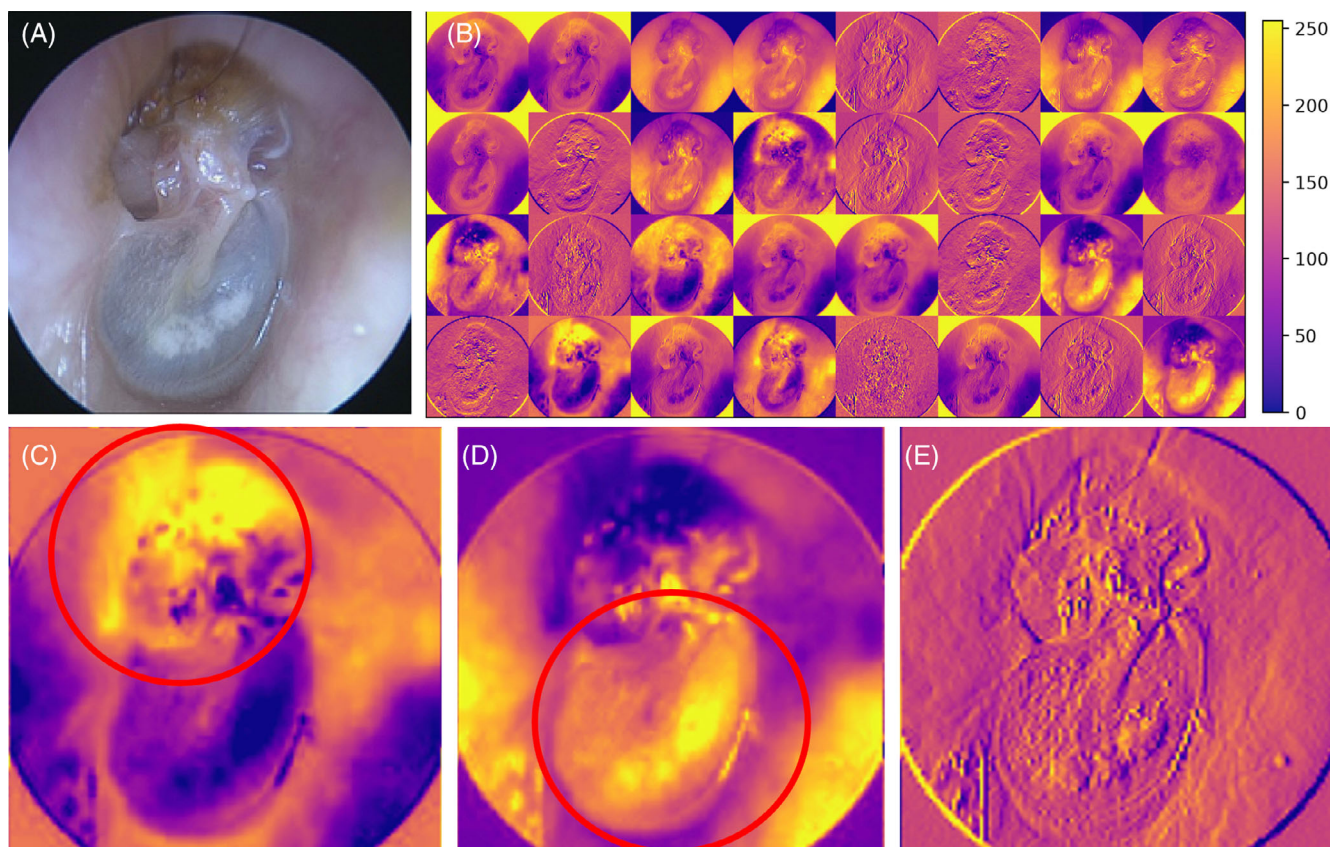


FIGURE 2 Intermediate activations from the trained InceptionResNetV2 neural network for classifying cholesteatoma versus normal otoscopic images. (A) Original otoscopic image of cholesteatoma. (B) Visualization of initial intermediate activations from the neural network with associated color bar, where brighter colors indicate greater activation and darker colors indicate lower activation. (C) Intermediate activation channel appearing to highlight the cholesteatoma (red circle). (D) Intermediate activation channel appearing to highlight the tympanic membrane (red circle). (E) Intermediate activation channel appearing to distinguish the texture of the image

TABLE 1 Cholesteatoma versus normal convolutional neural network evaluation metrics

CNN	Accuracy	Sensitivity	Specificity	PPV	NPV	AUROC	Runtime (min)
VGG19	83.82%	100.00%	67.65%	75.56%	100.00%	0.9922	64.62
MobileNetV2	97.06%	94.12%	100.00%	100.00%	94.44%	0.9991	12.80
DenseNet201	98.53%	100.00%	97.06%	97.14%	100.00%	0.9991	71.53
InceptionV3	95.59%	91.18%	100.00%	100.00%	91.89%	0.9965	24.42
ResNet152V2	92.65%	97.06%	88.24%	89.19%	96.77%	0.9758	67.82
Xception	94.12%	88.24%	100.00%	100.00%	89.47%	0.9983	49.07
InceptionResNetV2	95.59%	91.18%	100.00%	100.00%	91.89%	0.9931	63.85
NASNetLarge	94.12%	100.00%	88.24%	89.47%	100.00%	0.9983	160.18

Abbreviations: AUROC, area under the receiver operating characteristic curve; CNN, convolutional neural network; NPV, negative predictive value; PPV, positive predictive value.

3.2 | Cholesteatoma versus normal

For the cholesteatoma versus normal image classification task, 197 cholesteatoma images and 180 normal images were used for CNN training and testing. The evaluation results are recorded in Table 1, with CNNs listed in ascending order of their accuracy in classifying images from the ImageNet database. Following evaluation of the final CNN models on the testing dataset,

DenseNet201 had the highest accuracy of 98.53%. Three CNNs had the highest sensitivity and NPV of 100%: DenseNet201, VGG19, and NASNetLarge. Four CNNs had the highest specificity and PPV of 100%: MobileNetV2, Xception, InceptionV3, and InceptionResNetV2. DenseNet201 and MobileNetV2 have the highest AUROC score of 0.9991. MobileNetV2 had the overall fastest runtime of 12.80 min. ROC curves for the trained CNNs are shown in Figure 3.

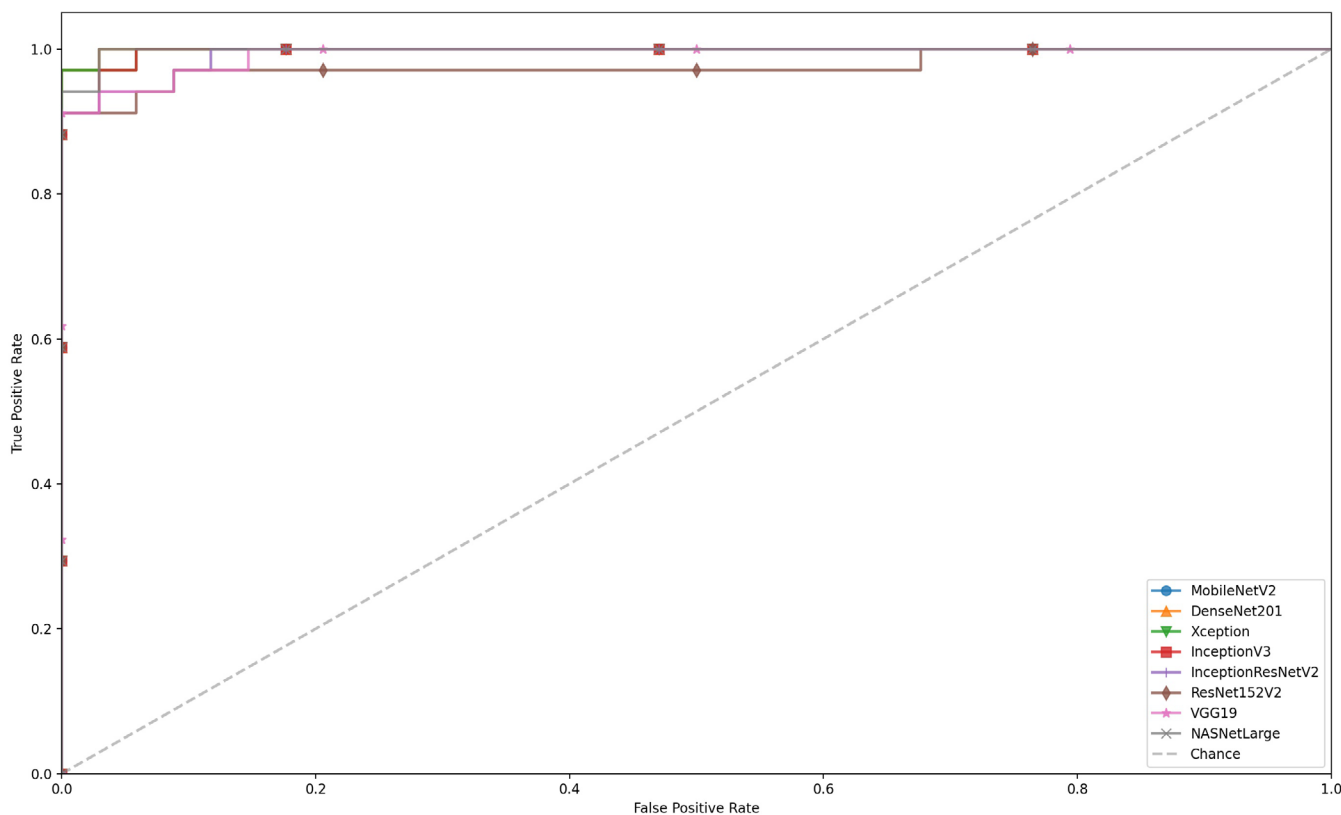


FIGURE 3 Cholesteatoma versus normal receiver operating characteristic (ROC) curves

TABLE 2 Cholesteatoma versus abnormal non-cholesteatoma convolutional neural network evaluation metrics

CNN	Accuracy	Sensitivity	Specificity	PPV	NPV	AUROC	Runtime (min)
VGG19	75.57%	69.44%	77.89%	54.35%	87.06%	0.7906	101.88
MobileNetV2	83.97%	77.78%	86.32%	68.29%	91.11%	0.9032	21.13
DenseNet201	90.08%	88.89%	90.53%	78.05%	95.56%	0.9284	116.28
InceptionV3	82.44%	66.67%	88.42%	68.57%	87.50%	0.8874	39.57
ResNet152V2	86.26%	77.78%	89.47%	73.68%	91.40%	0.9202	100.22
Xception	77.86%	50.00%	88.42%	62.07%	82.35%	0.8751	81.47
InceptionResNetV2	84.73%	66.67%	91.58%	75.00%	87.88%	0.9009	101.73
NASNetLarge	83.21%	83.33%	83.16%	65.22%	92.94%	0.9173	269.52

Abbreviations: AUROC, area under the receiver operating characteristic curve; CNN, convolutional neural network; NPV, negative predictive value; PPV, positive predictive value.

3.3 | Cholesteatoma versus abnormal non-cholesteatoma

For the cholesteatoma versus abnormal non-cholesteatoma image classification task, 197 cholesteatoma images and 457 abnormal non-cholesteatoma images were used for CNN training and testing. The evaluation results are recorded in Table 2, with CNNs listed in ascending order of their accuracy in classifying images from the ImageNet database. Following evaluation of the final CNN models on the testing dataset, DenseNet201 had the highest accuracy of 90.08%, the highest sensitivity of 88.89%, the highest PPV of 78.05%, and the highest

NPV of 95.56%. InceptionResNetV2 had the highest specificity of 91.58%. DenseNet201 had the highest AUROC score of 0.9284. MobileNetV2 had the overall fastest runtime of 21.13 min. ROC curves for the trained CNNs are shown in Figure 4.

3.4 | Cholesteatoma versus non-cholesteatoma

For the cholesteatoma versus non-cholesteatoma (normal + abnormal non-cholesteatoma) image classification task, 197 cholesteatoma images and 637 non-cholesteatoma images were used for CNN

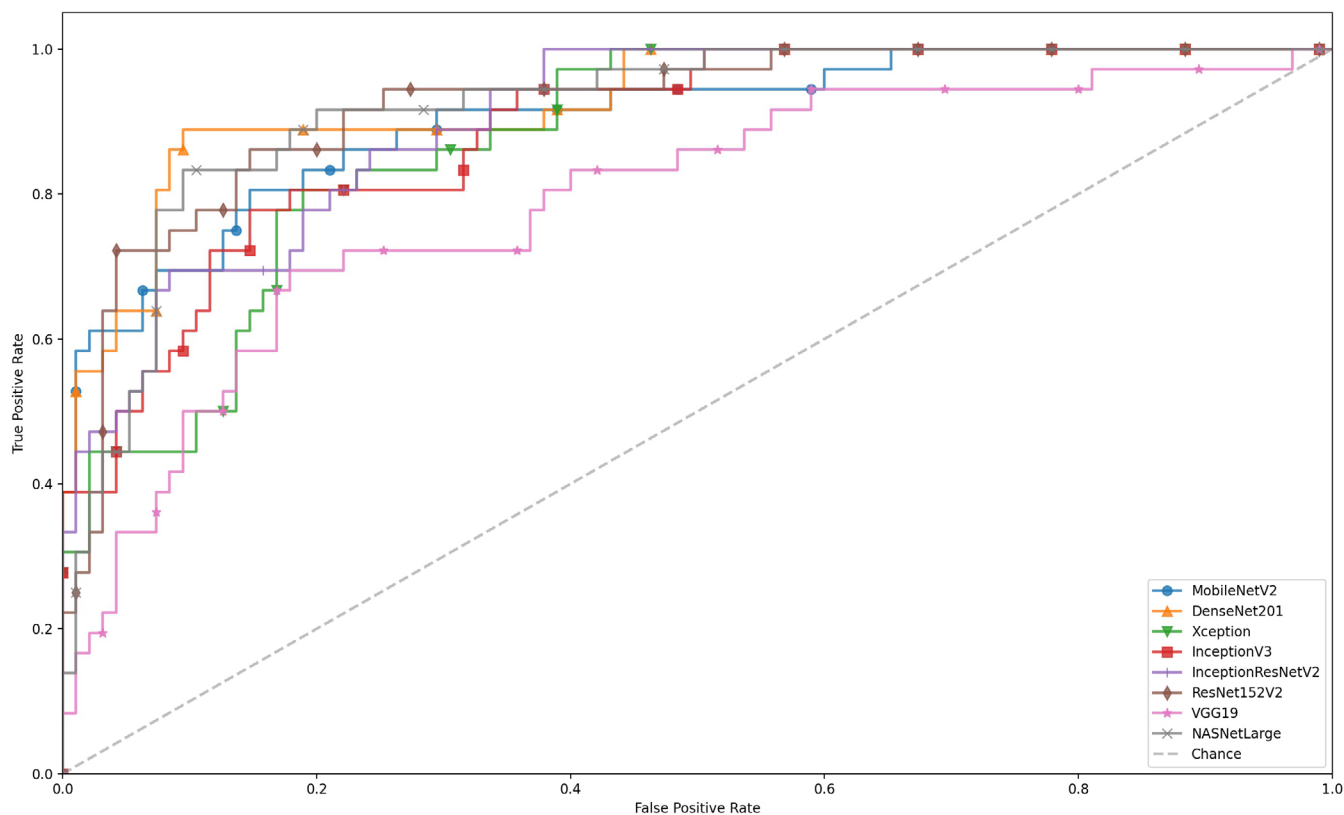


FIGURE 4 Cholesteatoma versus abnormal non-cholesteatoma receiver operating characteristic (ROC) curves

TABLE 3 Cholesteatoma versus non-cholesteatoma convolutional neural network evaluation metrics

CNN	Accuracy	Sensitivity	Specificity	PPV	NPV	AUROC	Runtime (min)
VGG19	87.57%	72.73%	90.97%	64.86%	93.57%	0.9116	129.72
MobileNetV2	88.70%	45.45%	98.61%	88.24%	88.75%	0.9291	30.38
DenseNet201	87.01%	66.67%	91.67%	64.71%	92.31%	0.9465	151.68
InceptionV3	88.70%	54.55%	96.53%	78.26%	90.26%	0.9270	50.00
ResNet152V2	88.70%	72.73%	92.36%	68.57%	93.66%	0.9444	123.37
Xception	90.40%	63.64%	96.53%	80.77%	92.05%	0.9358	102.90
InceptionResNetV2	87.57%	57.58%	94.44%	70.37%	90.67%	0.9343	126.83
NASNetLarge	88.70%	78.79%	90.97%	66.67%	94.93%	0.9609	338.95

Abbreviations: AUROC, area under the receiver operating characteristic curve; CNN, convolutional neural network; NPV, negative predictive value; PPV, positive predictive value.

training and testing. The evaluation results are recorded in Table 3, with CNNs listed in ascending order of their accuracy in classifying images from the ImageNet database. Following evaluation of the final CNN models on the testing dataset, Xception had the highest accuracy of 90.40%. NASNetLarge had the highest sensitivity of 78.79% and the highest NPV of 94.93%. MobileNetV2 had the highest specificity of 98.61% and the highest PPV of 88.24%. NASNetLarge had the highest AUROC score of 0.9609. MobileNetV2 had the overall fastest runtime of 30.38 min. ROC curves for the trained CNNs are shown in Figure 5.

4 | DISCUSSION

In our study, we showed that pretrained CNNs demonstrated strong performance in classifying otoscopic images, specifically in the task of identifying cholesteatomas and differentiating them from other TM appearances. Using pretrained CNNs can be an efficient strategy for machine learning workflows, an approach with precedence in ENT otoscopic image classification. Previous work by Cha et al. used transfer learning to diagnose ear disease into six categories: normal, tumor, perforation, retraction, serious otitis media, otitis externa with

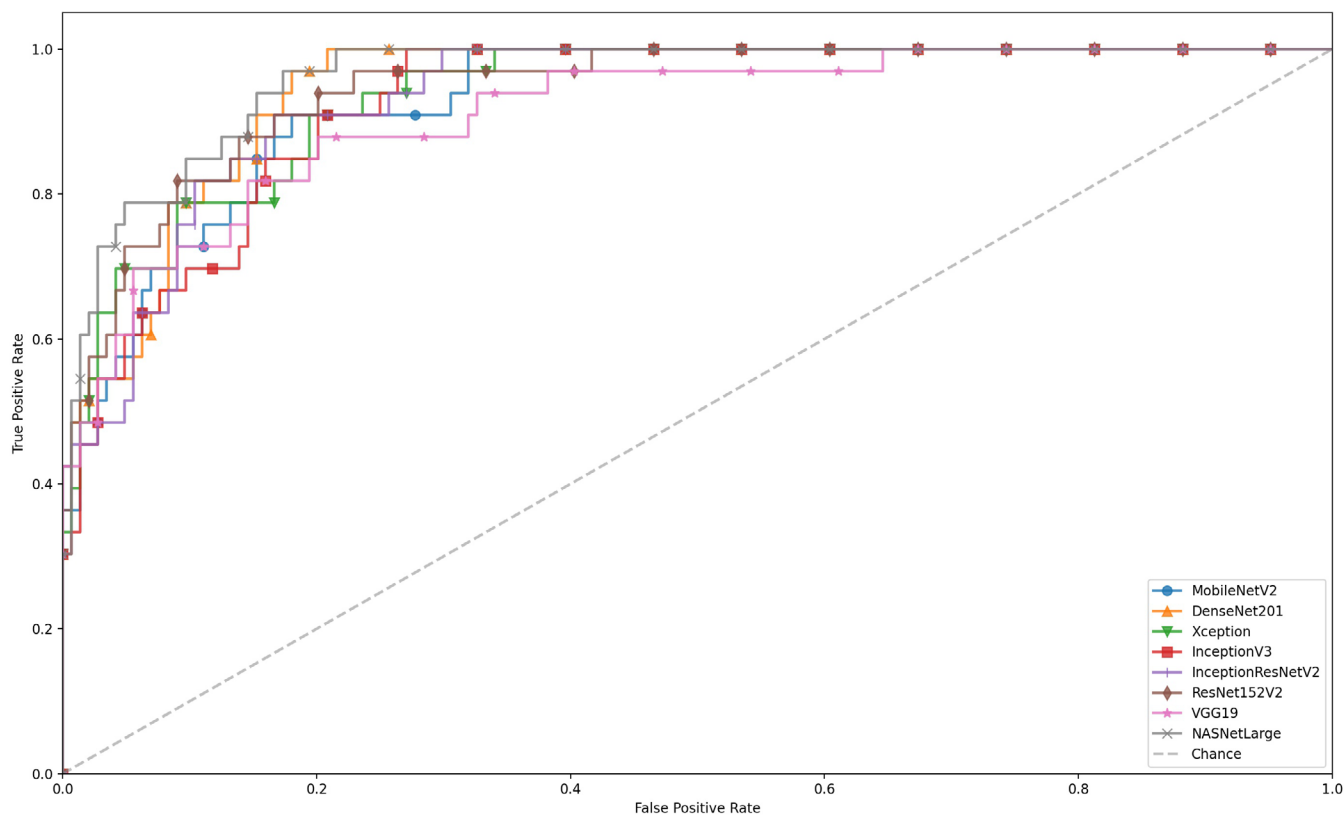


FIGURE 5 Cholesteatoma versus non-cholesteatoma receiver operating characteristic (ROC) curves

myringitis, and otitis externa without myringitis using InceptionResNetV2, InceptionV3, and Resnet101.²⁰ In that study however, cholesteatomas were broadly grouped in a “tumor” category with other TM masses including malignancies. Our study looked specifically at applying transfer learning for the diagnosis of cholesteatoma due to the clinical importance of early diagnosis in preventing severe complications.

Our results showed that the CNNs we trained had accuracies ranging from 83.8% to 98.5% for classifying cholesteatoma versus normal, 75.6%–90.1% for classifying cholesteatoma versus abnormal non-cholesteatoma, and 87.0%–90.4% for classifying cholesteatoma versus non-cholesteatoma. These findings compare favorably to results from previous studies that applied machine learning with transfer learning to classify otoscopic images, with Wang et al. achieving a top accuracy of 90%,¹⁸ Shie et al. achieving a top accuracy of 89.87%,¹⁹ Cha et al. achieving an average accuracy of 93.67%,²⁰ Habib et al. achieving an overall accuracy of 76%,²¹ and Tsutsumi et al. achieving a top accuracy of 77%.²² Byun et al. fully trained several neural networks to detect four classes of TM images: normal, otitis media with effusion, chronic otitis media, and cholesteatoma.³⁷ While their algorithmic approach reached an accuracy of 97.2%, it was evaluated on a small testing set of 71 images, and only had to differentiate between images from four distinct categories. Multiclass classifiers can be artificially limited to the specific categories they are trained to recognize, further constrained to categories which have a sufficient number of images in the dataset to adequately train and test

the models to detect each possible class. Comparably, Livingstone and Chau trained a machine learning classifier to identify 14 pathologies from otoscopic images, including cholesteatomas specifically.³⁸ While their trained neural network had an overall accuracy of 88.7% in classifying various pathologies, it only achieved 50% accuracy in diagnosing cholesteatomas. The authors attributed this fairly poor performance to only having 19 cholesteatoma images available to train the neural network, concluding that a greater number of higher quality training images should improve machine learning classification accuracy. While previous studies have applied a multiclass approach to automatically classify otoscopic images into several possible categories, our study sought to address the observed limitation among these classifiers by devolving the problem into several binary tasks differentiating cholesteatomas from other TM appearances. A binary classification task directed toward differentiating cholesteatomas from other potential TM appearances allows for training based on images unrestricted by a predefined set of specific classes. Though there is the tradeoff of being unable to specify multiple possible diagnoses like multiclass classifiers, a trained binary classifier with the primary objective of determining if an image shows a cholesteatoma or not may be more readily applicable to assess the wide variety of potential TM appearances encountered from otoscopy in a real-world clinical setting. Our results are very promising considering each CNN has its own specifically designed network architecture and were originally trained to classify ImageNet pictures, not necessarily all well-suited for this particular task. ImageNet images are pictures of

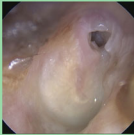
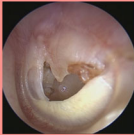

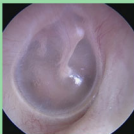
	Predicted Cholesteatoma	Predicted Non-Cholesteatoma
Actual Cholesteatoma		
Actual Non-Cholesteatoma		

FIGURE 6 Examples of correctly and incorrectly classified otoscopic images by the trained Xception neural network for classifying cholesteatoma versus non-cholesteatoma otoscopic images

everyday subjects like pencils and dogs, with more obvious, generally recognizable distinctions between different categories, whereas our study's otoscopic images can appear similar in shape and lighting to the untrained eye and require expert clinical experience to make a diagnosis. Due to the high performance level of these pretrained CNNs, our findings indicate that image features like color, shade, and edges which are used to classify everyday objects can also, to a certain extent, be effectively generalized to distinguish between cholesteatomas and non-cholesteatomas in otoscopic images.

Our findings show that CNNs can be trained to utilize the same basic rules humans use to discriminate between different types of otoscopic images for cholesteatomas versus other potential TM appearances. This is observed in the intermediate activations extracted from a trained CNN, whose visualizations showed increased activation signals which delineated key regions such as the TM and the retraction pocket in the cholesteatoma image. Additionally, there are varying degrees of activation based on range of light and shadow as well as texture present in the image, which are also highlighted by the algorithm. Therefore, it is evident that CNNs are capable of detecting basic image features that are important for cholesteatoma identification, outputting the respective activation signals to the rest of the CNN architecture, and synthesizing this information to classify the original image with a high level of accuracy. To further illustrate this, images correctly and incorrectly classified as cholesteatoma or non-cholesteatoma cases by the fully trained Xception neural network are shown in Figure 6. Correctly classified images are fairly characteristic, such as the grossly normal TM correctly predicted as non-cholesteatoma, and a notable retraction pocket medial to the TM correctly predicted as cholesteatoma. On the other hand, incorrectly classified images bear challenging features that may have potentially misled the classifier. The image incorrectly predicted as a cholesteatoma has a slightly shadowy area in the upper right hand corner due to the nature of the view that may perhaps be vaguely mistaken as a retraction pocket. Moreover, the image

incorrectly predicted as a non-cholesteatoma shows a significant TM perforation, an infrequently encountered primary otoscopic finding for a cholesteatoma compared to other characteristics that may present more frequently in our dataset. Overall, while the fully trained models performed well, these instances of misclassification point to the need for a large, diverse array of training otoscopic images to better improve performance.

The trained CNNs demonstrated the best performance in differentiating a cholesteatoma from a normal TM, compared to an abnormal non-cholesteatoma TM or a non-cholesteatoma TM, where all but one CNN achieved over 90% classification accuracy, with multiple neural networks including DenseNet201, NASNetLarge, InceptionV3, and MobileNetV2 scoring highly across all evaluation metrics. It is obvious that the greatest visual difference in otoscopic image appearance exists between a TM with a cholesteatoma and a TM without any abnormalities, likely due to the image noise which various abnormal TM pathologies can introduce to complicate the classification process. With this in mind, one potential workflow would be to process images sequentially, where one CNN first differentiates abnormal (cholesteatoma + abnormal non-cholesteatoma) images from normal images, and a subsequent CNN differentiates cholesteatoma images from abnormal non-cholesteatoma images. While this appears intuitive, our study results indicate that these CNNs perform better when differentiating cholesteatoma from non-cholesteatoma (normal + abnormal non-cholesteatoma) at the start. Of note, CNN sensitivities and PPVs were on average distinctly lower compared to their respective specificities and NPVs when differentiating cholesteatomas from abnormal TMs and non-cholesteatoma TMs. This finding shows that CNNs trained to recognize cholesteatomas may be particularly well suited for ruling out cholesteatomas from otoscopic images when applied as a clinical diagnostic tool.

It is particularly noteworthy that MobileNetV2, the smallest and fastest CNN utilized in this study, demonstrated surprisingly strong performance, ranking in the top four best performing CNNs by accuracy in all three classification tasks. As MobileNetV2 was developed to run efficiently on mobile devices and other platforms with limited computational resources, this CNN may play a potentially key role in the development of mobile or web-based machine learning applications for assisting physicians in identifying cholesteatomas and for training and teaching the diagnostic process.

There still exists room for improving CNN performance in otoscopic image classification. To preserve the pretrained layers of the neural network, the base neural network's image feature weights were frozen so only high-level layers were trained on our study's images. Additional fine tuning of the base model's layers would be needed to further focus CNN training to otoscopic images specifically. There is also the risk of overfitting, though considering our final trained CNNs achieved high accuracy even on the testing dataset, the effect of overfitting was likely minimized, though this may be attributable to the small size of the testing dataset which simplifies the final evaluation. Another approach to address this issue would be the implementation of cross-validation during the training phase. Cross-validation involves partitioning the full dataset into a predetermined number of

groups then evaluating every possible training–testing split of these groups, where each group is the testing set in a split and the remaining groups utilized for training.³⁹ While more computationally intensive, being able to sample the entire dataset for training and testing the classifiers should help further improve our workflow by reducing model overfitting, leading to more robust assessment of model performance.

AI-driven detection of cholesteatoma has many potential benefits such as standardizing, streamlining, and enhancing cholesteatoma screening worldwide. In the US, it can improve diagnosis especially in regions known to be under-resourced with otolaryngologists.⁴⁰ Under this model, an otoscopic image can be acquired by trained staff, put through our trained CNN, and reviewed by doctors of other specialties (e.g. family medicine doctors, pediatricians) alongside clinical findings. Cha et. al points out that multiple studies have shown an average diagnosis rate of <80% for ear disease²⁰ with specific studies showing lower averages in pediatricians and general practitioners (about 50% and 46% respectively) than in general otolaryngologists overall (about 73–74%).^{41,42} These studies demonstrate the variability and challenges of diagnosing ear disease. With further development, our method has the potential to enhance early detection and address healthcare gaps in the diagnosis of cholesteatoma. Moreover, this technology has also been utilized to recognize specific image characteristics that are more prone to diagnostic error,⁴³ and thus could potentially be applied to reduce the rate of cholesteatoma misdiagnosis. Lastly, this technology can be applied to enhance medical education in otolaryngology. A study by Ping et al. demonstrated that radiology residents trained with computer-aided diagnostic tools showed improved performance in evaluating mammograms⁴⁴; this can be similarly applied to teaching otolaryngology residents. Even with its many potential benefits, however, these tools should not replace the current standards of human image interpretation, rather they should enhance it.⁴⁵ Continued awareness of this technology's drawbacks remains prudent.

Limitations to our study include implementing every model with minimal parameter tuning, which may lead to suboptimal CNN performance. Additional methodological considerations including fine tuning the number of training epochs for each individual model and early termination to improve efficiency. Moreover, our preprocessing step included resizing, which reduced the resolution of the original image. Though consistent with current CNN standards, this step carries the risk of diminishing details important for diagnosis. Additionally, we used data augmentation to increase the number of images for the training step. While vertical flips of an isolated mass, for example, cholesteatoma, can still result in a realistic mass,²⁴ we used images that included the entire tympanic membrane and other anatomical features which would not realistically be vertically flipped, though such data augmentation steps should contribute toward improving CNN performance. Furthermore, we are limited by the number of images and representation of categories available to our study. Because machine learning performs better with more data and better representation of each possible class of image used in training, increasing the number of images can further improve classification accuracy and generalizability.

Increasing the testing dataset size by splitting the images into 60% training, 20% validation, and 20% testing sets could also be considered to better verify CNN performance in classifying cholesteatomas. Such a split would reduce the number of images available for training though, thus likely reducing classification performance compared to the 80% training, 10% validation, 10% testing dataset split utilized in this study. Finally, the inner workings of neural networks can be notoriously difficult to unravel, where the specific features considered by any given CNN layer to be the most important for classification remain obscure.³⁶ However, our study's visualization of intermediate activations provides a glimpse into how the algorithms perceive the image, which ultimately contributes to the CNN's classification decision. Additional CNN visualization tools including saliency and class activation maps can be explored in further research. From a clinical application standpoint, it should also be kept in mind that the images in this study were taken using an endoscope. Non-otolaryngologists who perform otoscopy regularly including primary care physicians and pediatricians do not have access to this tool and may not be able to obtain similar high-quality images for analysis by these classifiers.

Furthermore, future studies can utilize more powerful computational resources to more thoroughly investigate CNN capabilities by fully training these algorithms for the definite purpose of identifying cholesteatomas, reducing our current reliance on pretraining. Moreover, this image classification approach can be applied to other TM pathologies and expand AI-driven otoscopic image analysis. While interpretation of the results of this study are currently limited by lack of external validation, further direction would include investigating the effectiveness of a trained neural network-driven classifier for detecting cholesteatomas by analyzing images directly from in-office otoscopy. Moreover, while our proposed sequential approach would allow for immediate demonstrable utility of the trained binary classifiers developed through this work, a multiclass model classifying cholesteatoma versus non-cholesteatoma pathologies versus normal TM may demonstrate similar accuracy while being more computationally efficient. Thus, implementing the aforementioned combined multiclass approach with a larger dataset of abnormal non-cholesteatoma images in conjunction with our cholesteatoma dataset should be explored in future work.

5 | CONCLUSION

With modern technological advances in image processing and machine learning, the field of AI-driven medical image classification has rapidly expanded to a growing number of specialties, providing physicians with enhanced capabilities in disease diagnosis. Otoscopic images are particularly well suited for this analytical approach due to varying TM pathologies which warrant prompt recognition and evaluation. In our study, we demonstrated that the use of pretrained CNNs for otoscopic image analysis shows considerable capacity to detect important image features and differentiate cholesteatomas from other potential TM appearances, though fine tuning and larger image datasets would be needed to further improve and validate classifier

performance. Our results are an encouraging step forward toward practical application of machine learning in the analysis of otoscopic images, for the goal of optimal diagnosis of cholesteatoma.

CONFLICT OF INTEREST

None.

ORCID

Christopher C. Tseng  <https://orcid.org/0000-0002-2168-7935>

REFERENCES

- Olszewska E, Wagner M, Bernal-Sprekelsen M, et al. Etiopathogenesis of cholesteatoma. *Eur Arch Oto-Rhino-Laryngol Head Neck*. 2004; 261(1):6-24.
- Kuo C-L, Shiao A-S, Yung M, et al. Updates and knowledge gaps in cholesteatoma research. *Biomed Res Int*. 2015;2015:1-17.
- Kim S, Chang P. Cholesteatoma-diagnosing the unsafe ear. *Aust Fam Physician*. 2008;37(8):631.
- Rutkowska J, Özgirgin N, Olszewska E. Cholesteatoma definition and classification: a literature review. *J Int Adv Otol*. 2017;13(2):266-271.
- Tono T, Sakagami M, Kojima H, et al. Staging and classification criteria for middle ear cholesteatoma proposed by the Japan Otological Society. *Auris Nasus Larynx*. 2017;44(2):135-140.
- Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500-510.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6):84-90.
- Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys*. 2017;44(10):5162-5171.
- Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J Med Imaging*. 2016;3(3):034501.
- Roth HR, Lu L, Liu J, et al. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Trans Med Imaging*. 2015;35(5):1170-1181.
- Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging*. 2016;35(5):1207-1216.
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410.
- Choi JY, Yoo TK, Seo JG, Kwak J, Um TT, Rim TH. Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database. *PLoS One*. 2017;12(11):e0187336.
- Perdomo O, Arevalo J, González FA. Convolutional network to detect exudates in eye fundus images of diabetic subjects. Paper presented at: *12th International Symposium on Medical Information Processing and Analysis* 2017.
- Li X, Pang T, Xiong B, Liu W, Liang P, Wang T. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. Paper presented at: *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* 2017.
- Ishtiaq U, Abdul Kareem S, Abdullah ERMF, Mujtaba G, Jahangir R, Ghafoor HY. Diabetic retinopathy detection through artificial intelligent techniques: a review and open issues. *Multimed Tools Appl*. 2020;79(21):15209-15252.
- Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2009;22(10):1345-1359.
- Wang X, Valdez TA, Bi J. Detecting tympanostomy tubes from otoscopic images via offline and online training. *Comput Biol Med*. 2015; 61:107-118.
- Shie C-K, Chuang C-H, Chou C-N, Wu M-H, Chang EY. Transfer representation learning for medical image analysis. Paper presented at: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 2015.
- Cha D, Pae C, Seong S-B, Choi JY, Park H-J. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. *EBioMedicine*. 2019;45:606-614.
- Habib AR, Wong E, Sacks R, Singh N. Artificial intelligence to detect tympanic membrane perforations. *J Laryngol Otol*. 2020;134(4): 311-315.
- Tsutsumi K, Goshtasbi K, Risbud A, et al. A web-based deep learning model for automated diagnosis of otoscopic images. *Otol Neurotol*. 2021;42(9):e1382-e1388.
- Chen YC, Chu YC, Huang CY, et al. Smartphone-based artificial intelligence using a transfer learning algorithm for the detection and diagnosis of middle ear diseases: a retrospective deep learning study. *EClinicalMedicine*. 2022;51:101543.
- Miwa T, Minoda R, Yamaguchi T, et al. Application of artificial intelligence using a convolutional neural network for detecting cholesteatoma in endoscopic enhanced images. *Auris Nasus Larynx*. 2022;49(1): 11-17.
- Wang YM, Li Y, Cheng YS, et al. Deep learning in automated region proposal and diagnosis of chronic otitis media based on computed tomography. *Ear Hear*. 2020;41(3):669-677.
- Abadi M, Barham P, Chen J, et al. Tensorflow: a system for large-scale machine learning. Paper presented at: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* 2016.
- Hussain Z, Gimenez F, Yi D, Rubin D. Differential data augmentation techniques for medical imaging classification tasks. Paper presented at: *AMIA Annual Symposium Proceedings* 2017.
- Chollet F. Keras documentation. *Keras io*. 2015.
- Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.
- Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. Mobilenetv2: inverted residuals and linear bottlenecks. Paper presented at: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2018.
- Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. Paper presented at: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2018.
- Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. Paper presented at: *2009 IEEE Conference on Computer Vision and Pattern Recognition* 2009.
- Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115(3):211-252.
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929-1958.
- Chollet F. *Deep Learning with Python*. Manning Publications Company; 2017.
- Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. Paper presented at: *European Conference on Computer Vision* 2014.
- Byun H, Yu S, Oh J, et al. An assistive role of a machine learning network in diagnosis of middle ear diseases. *J Clin Med*. 2021;10(15): 3198.
- Livingstone D, Chau J. Otosopic diagnosis using computer vision: an automated machine learning approach. *Laryngoscope*. 2020;130(6): 1408-1413.
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825-2830.

40. Lango MN, Handorf E, Arjmand E. The geographic distribution of the otolaryngology workforce in the United States. *Laryngoscope*. 2017; 127(1):95-101.
41. Pichichero ME, Poole MD. Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. *Arch Pediatr Adolesc Med*. 2001;155(10):1137-1142.
42. Pichichero ME, Poole MD. Comparison of performance by otolaryngologists, pediatricians, and general practitioners on an otoendoscopic diagnostic video examination. *Int J Pediatr Otorhinolaryngol*. 2005;69(3):361-366.
43. Rodríguez JH, Fraile FJC, Conde MJR, Llorente PLG. Computer aided detection and diagnosis in medical imaging: a review of clinical and educational applications. *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, Salamanca, Spain 2016.
44. Luo P, Qian W, Romilly P. CAD-aided mammogram training. *Acad Radiol*. 2005;12(8):1039-1048.
45. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? *Radiology*. 2009;253(1):17-22.

How to cite this article: Tseng CC, Lim V, Jyung RW. Use of artificial intelligence for the diagnosis of cholesteatoma. *Laryngoscope Investigative Otolaryngology*. 2023;8(1):201-211. doi:[10.1002/lio2.1008](https://doi.org/10.1002/lio2.1008)