UC Irvine

Title

A Web-Based Deep Learning Model for Automated Diagnosis of Otoscopic Images

Permalink

https://escholarship.org/uc/item/1x2850v1

Journal

Otology & Neurotology, 42(9)

ISSN

1531-7129

Authors

Tsutsumi, Kotaro
Goshtasbi, Khodayar
Risbud, Adwight
et al.

Publication Date

2021-10-01

DOI

10.1097/mao.0000000000003210

Peer reviewed

A Web-Based Deep Learning Model for Automated Diagnosis of Otoscopic Images

*Kotaro Tsutsumi, *Khodayar Goshtasbi, *Adwight Risbud, *†Pooya Khosravi, *Jonathan C. Pang, *Harrison W. Lin, *†Hamid R. Djalilian, and *Mehdi Abouzari

*Department of Otolaryngology-Head and Neck Surgery; and †Department of Biomedical Engineering, University of California, Irvine, California

Objectives: To develop a multiclass-classifier deep learning model and website for distinguishing tympanic membrane (TM) pathologies based on otoscopic images.

Methods: An otoscopic image database developed by utilizing publicly available online images and open databases was assessed by convolutional neural network (CNN) models including ResNet-50, Inception-V3, Inception-Resnet-V2, and MobileNetV2. Training and testing were conducted with a 75:25 breakdown. Area under the curve of receiver operating characteristics (AUC-ROC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were used to compare different CNN models' performances in classifying TM images.

Results: Our database included 400 images, organized into normal (n = 196) and abnormal classes (n = 204), including acute otitis media (n = 116), otitis externa (n = 44), chronic suppurative otitis media (n = 23), and cerumen impaction (n = 21). For binary classification between normal versus abnormal TM, the best performing model had average AUC-ROC of 0.902 (MobileNetV2), followed by 0.745 (Inception-Resnet-V2), 0.731 (ResNet-50), and 0.636 (Inception-V3). Accuracy ranged between 0.73–0.77, sensitivity 0.72–0.88, specificity 0.58–0.84, PPV 0.68–0.81, and NPV 0.73–0.83. Macro-AUC-ROC for MobileNetV2 based multiclass-classifier was 0.91, with accuracy of 66%. Binary and multiclass-classifier models based on MobileNetV2 were loaded onto a publicly accessible and user-friendly website (https://head-neckml.com/tympanic). This allows the readership to upload TM images for real-time predictions using the developed algorithms.

Conclusions: Novel CNN algorithms were developed with high AUC-ROCs for differentiating between various TM pathologies. This was further deployed as a proof-of-concept publicly accessible website for real-time predictions. **Key Words:** Classification—Convolutional neural network—Deep learning model—Otoscopic image—Tympanic membrane.

Otol Neurotol **42:**e1382–e1388, 2021.

Otologic diseases are oftentimes initially diagnosed via inspection of patients' tympanic membrane (TM), but the proper use of otoscopes requires training (1) and the diagnoses of otologic disorders can remain prone to error in primary health care (2). For instance, general practitioners and pediatricians correctly diagnosed otologic diseases through inspection of TM images roughly 64% and 50% of the time, respectively, while extensively trained otolaryngologists provided accurate diagnoses 73% of the time (3). This is especially important to consider given how low and middle-income countries can have a limited number of trained otolaryngologists (4), despite their high prevalence of otologic disorders (4–6). Consequently, there is a great need to develop automated diagnostic technologies that allow for early and effective diagnosis of such diseases especially in regions with limited resources.

Developments in artificial intelligence (AI) and increases in available annotated medical data have allowed for successful application of these technologies to many fields of medicine (7–9). A subset of AI, deep learning, is particularly suited for image classification and segmentation tasks and has recently been applied for making diagnoses based on TM images taken via otoscopes (10,11). Such technology has significant potential for supporting proper patient care in regions lacking in medical resources by outputting automatic diagnoses. This is especially pertinent due to recent developments in smartphone otoscopes, as smartphone-enabled medical devices have been shown to be effective at mediating telemedicine (12–15).

© 2021, Otology & Neurotology, Inc.

Given such context, we aimed to develop a proof-of concept, mobile-compatible website that classifies otoscopic images as one of five different classes (normal, acute otitis media, otitis externa, chronic suppurative otitis media, and cerumen impaction). We constructed deep learning models, specifically convolutional neural networks (CNNs), based on four pre-existing networks (ResNet-50, Inception-V3, Inception-Resnet-V2, and MobileNetV2) and trained them on an original image database consisting of publicly available online images and open databases. Such technology could serve as a useful diagnostic tool, especially in regions lacking in otology resources and medical personnel.

METHODS

Image Database

This study did not require approval from the Institutional Review Board's biomedical research committee because of the use of publicly available databases. A database of TM images was developed by utilizing publicly available online images and open databases. A total of 400 high-quality TM images from the Van Akdamar Hospital eardrum database (16) and Google Images (17) were queried using the terms ''tympanic membrane,'' ''otoscopic image,'' ''eardrum,'' ''cerumen impaction," ''acute otitis media,'' ''otitis externa,'' and ''chronic suppurative otitis media.'' The Van Akdamar Hospital eardrum database consisted of images with a dimension of 500×500 pixels. These were taken from a cohort of 282 patients who volunteered for the study and were evaluated by three otolaryngologists, and represented a wide range of TM presentations. Images were also acquired from Google Images based on hits that appeared postquery and may constitute of images that are more prototypical of each pathology while varying in resolution and size. The database was bipartitely organized into normal (196) and abnormal (204) classes. The abnormal class further contained images representing four different pathologies: acute otitis media (AOM; 116), otitis externa (OE; 44), chronic suppurative otitis media (CSOM; 23), and cerumen impaction (21) (Fig. 1).

Convolutional Neural Network

Since our dataset was relatively small for training a deep learning model from scratch, we employed a technique referred to as transfer learning, in which pretrained networks are fine-tuned on new datasets. For this study, we utilized some of the many publicly available models pretrained on the ImageNet database (http://www.image-net.org) including the ResNet-50, Inception-V3, Inception-Resnet-V2, and MobileNetV2

networks (18–21). The pretrained networks were loaded through the open sourced Keras library written in the Python programming language.

To develop and test the algorithms, 60% of the database was used for training, 15% for validation, and 25% for testing of the model. Two different sets of models were trained and compared. We first trained our model solely for detection of abnormal versus normal TM (binary classification). Next, to further expand classification capacities, we trained our model as a multiclass classifier that identified the five individual classification classes. All layers of the loaded models were frozen excluding BatchNormalization layers. The following five layers were added at the ends of the models: GlobalAveragePooling2D, Dense (256, activation = ''relu''), Dropout (0.25), and BatchNormalization. A fully connected layer with two output nodes with a SoftMax activation function was added as the last layer. This final layer was modified to contain five output nodes for the five-class classification task. Hyperparameters for the training process were as follow: batch size 32, number of epochs 20, learning rate 0.001, optimizer root mean square propagation (RMSprop). All images were resized to $224 \times 224 \times 3$ pixel images containing three color channels of red, green, and blue. We performed data augmentation with rotation range 180, sheering range 0.3, zoom range 0.6, random brightness change between ranges of $0.2-2.7$, and random horizontal and vertical flips. The study was conducted via Google Collaboratory notebook ran on its GPU.

Website Structure

A web-based platform was developed to demonstrate how these algorithms work in real-time by allowing any TM image to be uploaded for an immediate prediction with the associated probability. Our MobileNetV2 model was selected for this deployment due to its small size and high performance, allowing for classification with low-latency and low-power. Using TensorFlow.js (22), a JavaScript interface for machine learning algorithms, our model runs entirely in the browser session of a user's device to ensure privacy and security of the data. This allows the model to predict results without the need to send information to a server for inference. Additionally, the web interface is also available on mobile devices.

RESULTS

Algorithm Performance

We tested our models using 100 images that were never seen by the models during training. The algorithms were primarily evaluated according to their classification accuracy and area under the curve of the receiver operating characteristic curve (AUC-ROC). Accuracy was
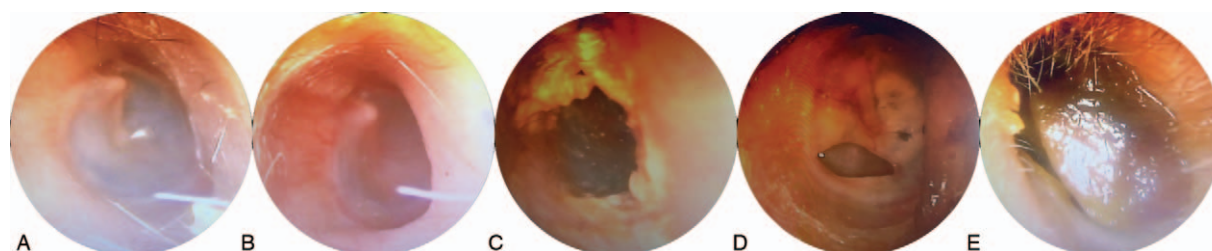

FIG. 1. Examples of our tympanic membrane image classifications representing *A*, normal, *B*, acute otitis media, *C*, otitis externa, *D*, chronic suppurative otitis media, and *E*, cerumen impaction.

TABLE 1. *Performance of models for binary classification (classification threshold = 0.5)*

Model	AUC-ROC	Accuracy	Sensitivity	Specificity	PPV	NPV	Network Size (MB)
Inception-V3	0.590	0.73	0.720	0.740	0.735	0.725	92
ResNet-50	0.718	0.74	0.800	0.680	0.714	0.773	98
Inception-Resnet-V2	0.745	0.73	0.880	0.580	0.677	0.829	215
MobileNetV2	0.902	0.77	0.700	0.840	0.814	0.737	14

AUC-ROC, area under the curve of the receiver operating characteristic; NPV: negative predictive value; PPV, positive predictive value.

calculated as the percentage of correct classifications made in classifying the test dataset with a classification threshold of 0.5. AUC-ROC represents the area under the curve created by plotting true positive rates against false positive rates of the test dataset classification.

We first conducted a binary classification task through which the models classified the input images into either normal or abnormal classes. The highest AUC-ROC and accuracy were yielded by the Mobile-NetV2 (AUC-ROC 0.902, accuracy 77%), followed by Inception-Resnet-V2 network (AUC-ROC 0.745, accuracy 73%), ResNet-50 (AUC-ROC 0.731, accuracy 71%), then Inception-V3 (AUC-ROC 0.636, accuracy 72%) (Table 1). The AUC-ROC curves of all models at various discrimination thresholds when classifying the test dataset are demonstrated in Figure 2. MobileNetV2 outperformed other models while also being a much smaller network, being suited for effective deployment via a web application. We then proceeded to test the model for classifying the input images into five individual diagnostic categories. Given the results of the previous binary classification task, we employed the MobileNetV2 model. The macro-AUC-ROC was 0.91 and all the individual AUC-ROC per each classification are demonstrated in Figure 3. Additionally, the sensitivity and positive predictive value (PPV) of the model for predicting each classification are shown in Table 2.
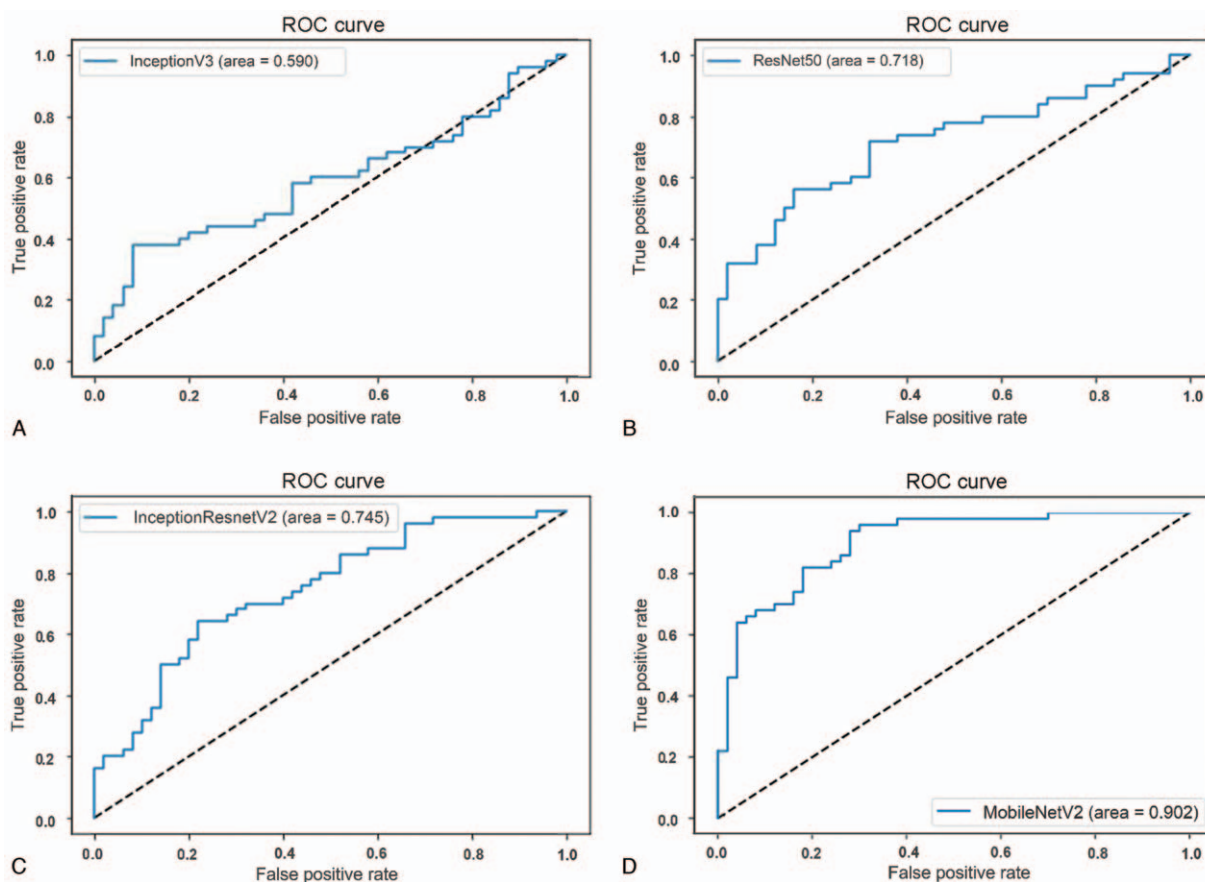

FIG. 2. AUC-ROC curves for *A*, Inception-V3, *B*, ResNet-50, *C*, Inception-Resnet-V2, and *D*, MobileNetV2 networks for binary classification. AUC-ROC, area under the curve of receiver operating characteristics.
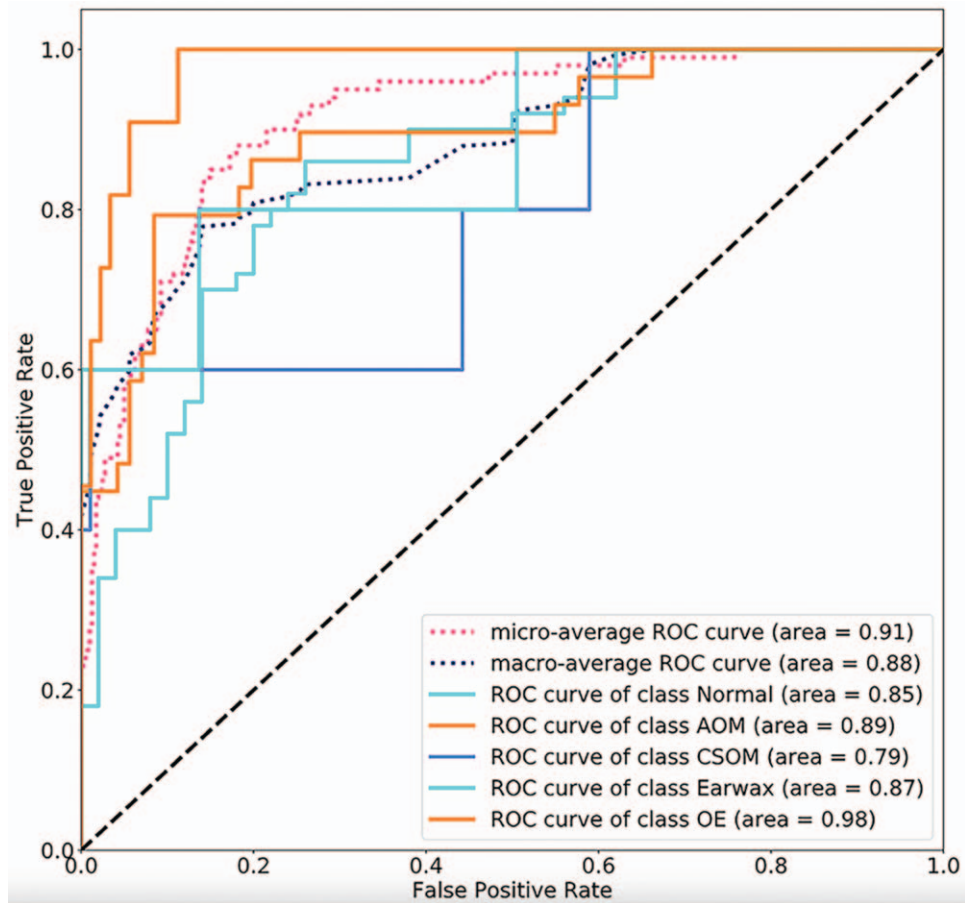
FIG. 3. AUC-ROC curve for MobileNetV2 network multiclass classification. AUC-ROC, area under the curve of receiver operating characteristics.

Web Application

A web-based platform was created to allow users to upload images and use our MobileNetV2-based model for their classification (Fig. 4). Each input image is first resized into a 224 × 224 × 3 image. The image is then fed into our MobileNetV2-based model and the output predictions are shown to the user with the associated probabilities. To enhance user experience, we organized the website such that the users are able to intuitively drag and drop one or many images into a box. Both binary and multiclass classification results are presented and organized based on image name, thumbnail, and its classification. The website is publicly accessible at the following link: https://headneckml.com/tympanic.html.

DISCUSSION

In this study, we developed novel deep learning algorithms and a proof-of-concept website that allows users to classify TM images belonging to five different classifications. We first created an original otoscopic image dataset by collecting images from various public sources.

TABLE 2. *Performance of MobileNetV2 for multiclass classification (classification threshold = 0.5)*

Classification	N: Training and Validation (Test)	AUC-ROC	Sensitivity	PPV
Normal	146 (50)	0.85	0.62	0.82
Acute otitis media	87 (29)	0.89	0.90	0.50
Chronic suppurative otitis media	18 (5)	0.79	0.40	0.67
Cerumen	16 (5)	0.87	0.40	1.00
Otitis externa	33 (11)	0.98	0.45	1.00

AUC-ROC, area under the curve of the receiver operating characteristic; PPV, positive predictive value.
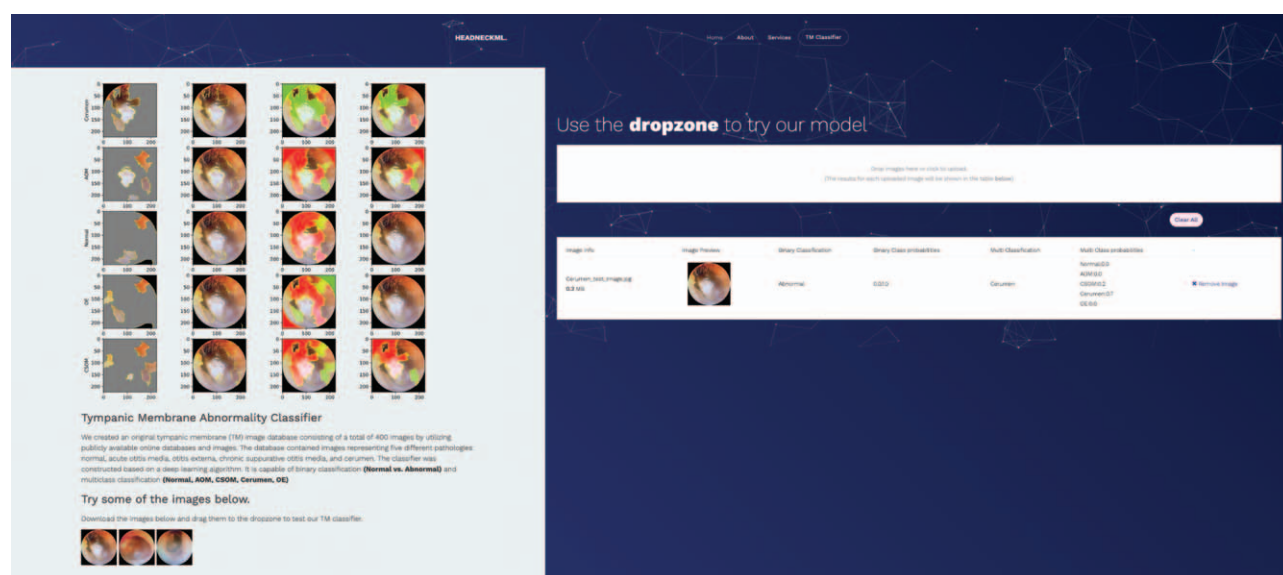
FIG. 4. User interface for the web deployment: https://headneckml.com/tympanic.html.

These images were then used for training deep learning algorithms based on multiple different pre-existing models, after which the models were tested using a fraction of the original dataset and the models were compared regarding their performance. The main novelty of this study is the proof-of-concept website that allows the readership to upload pictures for predicting whether an uploaded TM image is normal or abnormal, along with further prediction of possible pathologies with their associated probabilities. To the best of our knowledge, this is the first study that illustrates the viability of a user-friendly website that classifies TM images based on a deep learning model.

Our deep learning models were constructed based on a method referred to as transfer learning (23). This is founded on the concept of fine-tuning pretrained networks with new datasets rather than training a deep learning model from scratch. Since these networks have already learned general features present among many types of images, consisting of millions of labelled images belonging to around 1000 different categories, training on top of these knowledge allows for classification of images via small datasets. Hence, transfer learning was the ideal technique for this study, given our relatively small image dataset. Since transfer learning builds on top of pre-existing networks, it was crucial to determine the most ideal model for the specific task at hand. To do so, we developed different algorithms based on four distinct models that have been widely utilized in previous literature, including the ResNet-50, Inception-V3, Inception-Resnet-V2, and MobileNetV2 networks.

The algorithm that was transfer-learned on Mobile-NetV2 performed the best regarding AUC-ROC, accuracy, and specificity, while the Inception-ResNet-V2 based model performed the best in terms of sensitivity. Given that classification technology will most likely become implemented in the form of devices that aid physicians in making diagnoses, it is important to heavily weight both the sensitivity and specificity in evaluating these technologies. This will put pressure on these devices to detect abnormalities with a higher reliability and to have the capability of differentiating between various otologic pathologies, which is necessary given the disease-specific treatments. Together, this will allow physicians to make final diagnoses on top of the output results. In this regard, it is important to note Inception-Resnet-V2-based model's high sensitivity that outperformed the counterpart models.

Given our goal of deploying a functional website based on these deep learning models, we must also consider their speed and size. Although recent trends in computer vision work towards achieving higher accuracies by making networks deeper and more complicated, this approach comes with tradeoffs in speed and size. In this context, MobileNet and MobileNetV2 were designed as lightweight models for solving computer vision tasks on computationally limited platforms such as mobile devices (20). MobileNets achieve this reduction in computation and model size by first applying a single depthwise convolution filter to each color channel. This is then combined with a 1×1 pointwise convolution in separate layers, in contrast to standard convolutional strategies in which filters and inputs are both combined in one step. As a result, our model transfer-learned on MobileNetV2 achieved AUC-ROC, accuracy, and specificity higher than that of other models while also being the most lightweight. Taking these into account, this particular model may be the most suited for the aim of this study. Furthermore, we proceeded to conduct multiclass classification using MobileNetV2, which yielded high AUC-ROCs for each class that were comparable to those from the binary classification task. Importantly, a previous

study has shown that pediatricians and otolaryngologists provided accurate diagnoses based on TM images roughly 50% and 73% of the time, respectively (3). Other studies showed that general practitioners diagnose AOM correctly roughly 64% of the time based on both TM images and symptoms (2). Though underlying processes behind algorithm-based diagnoses and physician interpretation of patient presentation contrasts significantly, our CNN-based algorithm performed favorably to the average performance of physicians when solely considering accuracy; however, since the number of images belonging to each class was limited, it is important to interpret these results with caution.

Machine learning approaches are often limited by the size of available samples. This was no different for the present study, which solely relied on publicly available datasets and images. While we utilized techniques designed to overcome this challenge such as image augmentation and modification of hyperparameters, a larger image database would likely have increased the performances of different models. However, given the diversity of the images contained within our dataset collected through various sources, a high AUC-ROC could indicate that our model may be capable of classifying a diverse range of target images that will most likely be encountered in real clinical settings. Nevertheless, our future study will involve incorporating institutional images to continue building up on the current model.

It is further important to note that this study is particularly significant in the deployment of a functional website that allows users to classify otoscopic images into multiple classes. Proper use of otoscopes requires many years of training and experience, presenting with a moderate learning curve for many (1). Aiming to mitigate this issue, recent studies have highlighted the potential and the utility of smartphone-enabled otoscopes as a telemedicine tool, through which medical personnel without sufficient otoscope training can relay information to otolaryngologists for remote evaluation (13). The present study effectively takes this one step further by automating the final diagnostic step, or it can be used as educational purposes for the initial evaluators. The ideal implementation workflow in the future would be for users to capture images via a smartphone-enabled otoscope and evaluate it on the same device through our diagnostic website. This workflow could be applied for varying scenarios, may it be for large hospitals that need extra diagnostic capacity or for health care workers located in under-served areas in need of expert specialists. With such use in mind, our website was designed to have a user-friendly platform. While operating on computers, the user can simply drag images to the dropzone for analysis. The website is also mobile-friendly, allowing for analysis of photos stored on mobile phones or those directly captured through the device. Furthermore, since medical images are highly sensitive, our deep learning model runs locally on users' devices to ensure security and privacy. The data is never sent to a server or stored for inference.

Several limitations exist despite the accomplishments of this study. As previously mentioned, performance of machine learning models is often limited by the size of its training database. Image augmentation techniques were implemented to overcome this issue, though, utilization of a larger image database would most likely have improved the performance and applicability of our algorithms. Second, machine learning algorithms are often considered a ''black-box,'' providing users with little to no information on how the algorithm classified the input image. Techniques such as Grad-CAM exist to highlight particular areas of an image considered important through the classification process, though this still does not provide an explanation on how each aspect of the image precisely influences the prediction (24). Machine learning explainability and interpretation is a topic that is under extensive investigation and will be extremely important as such systems become more widely implemented in the medical field in the future.

With regards to its current clinical utility, there exist certain limitations inherent to our study methods and image selection. It is important to note that the four ear pathologies analyzed with our model are often straight-forward for clinicians to discriminate between; the greater challenge lies in distinguishing normal ears from effusion and more subtle acute otitis media, which unfortunately, was not evaluated using our algorithms. Moreover, our classifier was likely trained/tested using many prototypical images of each pathology, introducing another source of bias and possibly resulting in artificially high performance outcomes. In clinical practice, specific otoscopic features combined with additional signs and symptoms aid in distinguishing the four pathologies from one another. For acute otitis media, there must be evidence of moderate to severe bulging of the TM (or mild bulging with recent onset otalgia or erythema), and new onset of otorrhea not due to otitis externa (25). For otitis externa, signs of ear canal inflammation must be present, including ear canal and TM erythema/edema (26). The diagnosis of chronic suppurative otitis media is confirmed by visualization of TM perforation, with findings of thickened granular middle ear mucosa and polyps typically present (27). Lastly, the diagnosis of cerumen requires otoscopic confirmation of cerumen accumulation within the ear canal (28). Though derivation of clinical criteria for diagnosing different conditions is imperative for clinical application of novel technologies, it is currently not possible to derive the exact formula deep learning uses in making its decisions (29). This issue of machine learning being a black box is being heavily researched and will most likely be crucial in implementing such technology to real-life clinical settings in the future. Lastly, deep learning performance is inevitably correlated to similarity in quality of the training images and the input images. Hence, a significant variance in the quality of input images may render the algorithm less effective in making accurate classifications. In real life, quality of input images would depend on the healthcare practitioners

using the otoscope, and this may potentially result in variability in the performance of our algorithm depending on the user.

CONCLUSION

We developed otoscopic image classifiers trained on publicly available images, of which MobileNetV2 had the best relative performance with high AUC-ROC and accuracy. An additional novelty of this work included the development of a proof-of-concept website which uses our best-performing novel algorithm to predict whether an uploaded TM image is normal/abnormal and which of five imaging classifications it belongs to. Improvements of this platform and its effective application to clinic practices could prove to be useful in bolstering the shortage of diagnostic capacity of otologic diseases in various scenarios.

REFERENCES

1. Davies J, Djelic L, Campisi P, Forte V, Chiodo A. Otoscopy simulation training in a classroom setting: a novel approach to teaching otoscopy to medical students. *Laryngoscope* 2014;124:2594–7.
2. Blomgren K, Pitkäranta A. Is it possible to diagnose acute otitis media accurately in primary health care? *Fam Pract* 2003;20:524–7.
3. Pichichero ME, Poole MD. Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. *Arch Pediatr Adolesc Med* 2001;155:1137–42.
4. Fagan JJ, Jacobs M. Survey of ENT services in Africa: need for a comprehensive intervention. *Glob Health Action* 2009;2. doi:10.3402/gha.v2i0.1932.
5. Monasta L, Ronfani L, Marchetti F, et al. Burden of disease caused by otitis media: systematic review and global estimates. *PLoS One* 2012;7:e36226.
6. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:743–800.
7. Yamamoto Y, Tsuzuki T, Akatsuka J, et al. Automated acquisition of explainable knowledge from unannotated histopathology images. *Nat Commun* 2019;10:5642.
8. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
9. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24:1342–50.
10. Livingstone D, Talai AS, Chau J, Forkert ND. Building an Otoscopic screening prototype tool using deep learning. *J Otolaryngol Head Neck Surg* 2019;48:66.
11. Cha D, Pae C, Seong SB, Choi JY, Park HJ. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. *EBioMedicine* 2019;45:606–14.
12. Richards JR, Gaylor KA, Pilgrim AJ. Comparison of traditional otoscope to iPhone otoscope in the pediatric ED. *Am J Emerg Med* 2015;33:1089–92.
13. Moshtaghi O, Sahyouni R, Haidar YM, et al. Smartphone-enabled otoscopy in neurotology/otology. *Otolaryngol Head Neck Surg* 2017;156:554–8.
14. Coulibaly JT, Ouattara M, D'Ambrosio MV, et al. Accuracy of mobile phone and handheld light microscopy for the diagnosis of schistosomiasis and intestinal protozoa infections in Côte d'Ivoire. *PLoS Negl Trop Dis* 2016;10:e0004768.
15. Chang J, Arbeláez P, Switz N, et al. Automated tuberculosis diagnosis using fluorescence images from a mobile microscope. *Med Image Comput Comput Assist Interv* 2012;15 (Pt 3):345–52.
16. Başaran E, Cömert Z, Çelik Y. Convolutional neural network approach for automatic tympanic membrane detection and classification. *Biomed Signal Process Control* 2020;56:101734.
17. Google. Google Images. 2020. Available at: https://www.google.com/imghp?hl=EN. Accessed July 2020.
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016; 770–8.
19. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016; 2818–26.
20. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018; 4510–20.
21. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv* 2016. 160207261.
22. Abadi M, Agarwal A, Barham P, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. *arXiv* 2016;160304467.
23. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *arXiv* 2014;1411:1792.
24. Zhang Z, Beck MW, Winkler DA, et al. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Ann Transl Med* 2018;6:216.
25. Harmes KM, Blackwood RA, Burrows HL, Cooke JM, Harrison RV, Passamani PP. Otitis media: diagnosis and treatment. *Am Fam Physician* 2013;88:435–40.
26. Rosenfeld RM, Schwartz SR, Cannon CR, et al. Clinical practice guideline: acute otitis externa. *Otolaryngol Head Neck Surg* 2014;150 (1 Suppl):S1–24.
27. Rosenfeld RM, Shin JJ, Schwartz SR, et al. Clinical practice guideline: Otitis media with effusion (update). *Otolaryngol Head Neck Surg* 2016;154 (1 Suppl):S1–41.
28. Schwartz SR, Magit AE, Rosenfeld RM, et al. Clinical practice guideline (update): Earwax (Cerumen Impaction). *Otolaryngol Head Neck Surg* 2017;156 (1 Suppl):S1–29.
29. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Cent Rep* 2019;49:15–21.