Deep Learning in Automated Region Proposal and Diagnosis of Chronic Otitis Media Based on Computed Tomography

Yan-Mei Wang,1,2,5 Yike Li,3,5 Yu-Shu Cheng,4 Zi-Yu He,1,2 Juan-Mei Yang,1,2 Jiang-Hong Xu,1,2 Zhang-Cai Chi,1,2 Fang-Lu Chi,1,2 and Dong-Dong Ren1,2

OBJECTIVES: The purpose of this study was to develop a deep-learning framework for the diagnosis of chronic otitis media (COM) based on temporal bone computed tomography (CT) scans.

DESIGN: A total of 562 COM patients with 672 temporal bone CT scans of both ears were included. The final dataset consisted of 1147 ears, and each of them was assigned with a ground truth label from one of the 3 conditions: normal, chronic suppurative otitis media, and cholesteatoma. A random selection of 85% dataset (n = 975) was used for training and validation. The framework contained two deep-learning networks with distinct functions: a region proposal network for extracting regions of interest from 2-dimensional CT slices; and a classification network for diagnosis of COM based on the extracted regions. The performance of this framework was evaluated on the remaining 15% dataset (n = 172) and compared with that of 6 clinical experts who read the same CT images only. The panel included 2 otologists, 3 otolaryngologists, and 1 radiologist.

RESULTS: The area under the receiver operating characteristic curve of the artificial intelligence model in classifying COM versus normal was 0.92, with sensitivity (83.3%) and specificity (91.4%) exceeding the averages of clinical experts (81.1% and 88.8%, respectively). In a 3-class classification task, this network had higher overall accuracy (76.7% versus 73.8%), higher recall rates in identifying chronic suppurative otitis media (75% versus 70%) and cholesteatoma (76% versus 53%) cases, and superior consistency in duplicated cases (100% versus 81%) compared with clinical experts.

CONCLUSIONS: This article presented a deep-learning framework that automatically extracted the region of interest from two-dimensional temporal bone CT slices and made diagnosis of COM. The performance of this model was comparable and, in some cases, superior to that of clinical experts. These results implied a promising prospect for clinical application of artificial intelligence in the diagnosis of COM based on CT images.

KEY WORDS: Artificial intelligence, Cholesteatoma, Deep learning, Otitis media, Tomography, X-ray computed.

(Ear & Hearing 2020;41:669–677)

INTRODUCTION

Chronic otitis media (COM) is characterized by the presence of persistent inflammation in the middle ear and/or the mastoid air cells. It is one of the most common global health issues, affecting up to 0.3 billion individuals worldwide and accounting for a disease burden of over 2 million disability-adjusted life years (World Health Organization 2004). This disease is the major cause of preventable hearing loss, especially in developing countries (World Health Organization 2004; Park et al. 2015; Kaspar et al. 2018; Libwea et al. 2018). Clinical classification of COM is typically based on the content of inflammatory buildup inside the middle ear. In particular, chronic suppurative otitis media (CSOM, characterized by purulent fluid buildup and discharge) and cholesteatoma (characterized by keratinized, desquamated epithelial collection) represent two common, yet distinct types of COM. Differential diagnosis between them is essential, as each of these diseases has different etiologies, clinical features, and management guidelines (Lustig et al. 2019).

An initial ear exam is not always conclusive in detection or differential diagnosis of COM; therefore, it is frequently supplemented by imaging techniques. A high-resolution computed tomography (CT) scanning of the middle ear and the temporal bone is currently regarded as the first-line diagnostic imaging modality in the case of COM. It assists clinicians in evaluating the extent of COM for refinement of surgical plans and improving patient outcomes (Kusak et al. 2018). At a lower cost compared with magnetic resonance imaging, a CT scan achieves adequate sensitivity in the detection of osseous erosion, which is characteristic of cholesteatoma (Gaurano & Joharji 2004; Molteni et al. 2019).

Artificial intelligence (AI) has made remarkable progress in recent years. Development of multilayered network architectures that compile mathematical functions with millions of parameters allows machines to “think” deeply and to interpret complex data in a highly precise manner. Such methods in machine learning are referred to as deep learning. Computer vision has benefited greatly from application of deep-learning techniques. For instance, with the evolution of convolutional neural networks (CNNs), AI performance in image classification, and object detection and localization have significantly improved from what has been achieved using traditional machine-learning algorithms. Recent attempts to implement deep learning in classification of medical images have demonstrated its tremendous potential in diagnostic decision support, prognostic prediction, and hospital triage (Arbabshirani et al. 2018; Coudray et al. 2018; De Fauw et al. 2018; Yasaka et al. 2018b; Chowdhury et al. 2019; Kim et al. 2019; Liang et al. 2019).
A limited number of previous studies were conducted on detection of otitis media using machine-learning techniques (Kuruvilla et al. 2013; Kasher 2018). However, these studies were based on otoscopic images, and none of them attempted to distinguish between different types of COM. Compared with otoscopy, which only allows inspection of the tympanic membrane, a CT scan presents a broader view of auditory structures including the entire middle ear and the mastoid air cells. In addition, its digital nature connects to machine learning seamlessly and therefore is more practical in developing a clinically applicable AI.

The objective of this study was to develop a deep-learning framework for automated detection and differential diagnosis of COM based on patients’ temporal bone CT scans. To achieve this goal, two networks with different functions were built and trained (Fig. 1). The first network was used to extract the region of interest (i.e., region proposal) from sequential temporal bone CT slices. The second network was designed to analyze the images extracted by the first network and to predict the most likely diagnoses. Finally, the performance of this model was tested on a separate dataset and compared with that of clinical experts, including otologist, otolaryngologist, and radiologist.

MATERIALS AND METHODS

Study Subjects
This study was approved by the institutional review board. Query of the clinical database based on procedural terms identified 737 adults (≥18 years old) who underwent middle ear surgeries between May 2013 and October 2018. Medical records were then reviewed to exclude any patient diagnosed with acute otitis media or any inner ear disease, as well as any patient missing temporal bone CT scan. A final cohort of 562 patients was eligible for retrospective data collection (Table 1).

CT Imaging
Each enrolled patient had received one or more temporal bone CT scans, resulting in a total of 672 scans available for analysis. These scans were obtained by a 128-channel multidetector SOMATOM Definition Edge CT scanner (Siemens Inc., Munich, Germany). Per our standard protocol, scanning was carried out from the lower margin of the external auditory meatus to the top margin of the petrous bone in each patient. Axial sections of 0.6 mm thick were obtained at the following settings: collimation of 128 × 0.6 mm, field of view of 220 × 220 mm, pitch of 0.8 mm, matrix size of 512 × 512, voltage of 120 kV, and current of 240 mAs. The number of CT slices in the axial plane ranged from 30 to 50 in each scan. All images were downloaded from a physician workstation and saved in 512 × 512 pixels JPEG format for analysis (Fig. 1A).

Clinical Labeling
The axial view of a temporal bone CT scan allowed visualization of both ears, resulting in a total of 1344 ears regarded as independent cases. Clinical label for each ear was first assigned based on the diagnosis of patient’s medical notes. Two experienced otolaryngologists with full access to all patients’ medical records independently reviewed these labels and addressed any discrepancy until consensus was reached. These clinical labels represented the ground truths that were used in the training and testing sessions of the classification network (Table 2). Further inspection of these labels revealed that “CSOM,” “normal,” and “cholesteatoma” represented three classes with the most numbers of cases. To achieve an adequate performance by deep learning, the network was trained to classify these 3 conditions only (n = 1147).

The protocol for assignment of clinical labels is summarized later. The label (diagnosis) of an operated ear was assigned based on its pathology. For an unoperated ear, the label was assigned according to a combination of history, physical examination, and imaging findings. The final classification was made by an experienced otologist.

TABLE 1. Baseline characteristics of enrolled patients

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>254</td>
<td>308</td>
<td>562</td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
<td>41.26±14.22</td>
<td>43.88±13.24</td>
<td>42.70±13.74</td>
</tr>
<tr>
<td>Number of CT scans</td>
<td>309</td>
<td>363</td>
<td>672</td>
</tr>
<tr>
<td>Labels assigned to ears</td>
<td>618</td>
<td>726</td>
<td>1344</td>
</tr>
</tbody>
</table>

CT, computed tomography.
A faster region-based convolutional neural network (Faster R-CNN) was applied for detection and localization of these landmark structures (Girshick 2015; Ren et al. 2017; Liu & Stathaki 2018; Li et al. 2019). The base model of this network was a pretrained CNN (Inception-V2, Google LLC, Mountain View, CA) serving as an image feature extractor (Szegedy et al. 2016). A CNN is a deep-learning algorithm that mimics the biological connectivity patterns between neurons in the visual cortex and is extremely useful for pattern-recognition tasks. In simple words, this base model extracted basic graphical features across an image and collectively formed patterns with increasing complexity. These features were then pooled within each of multiple fix-sized bounding boxes (as known as anchors) distributed across the image. Trained with those annotated images, this network learned to classify the contents in these anchors and to resize the box to fit the two landmark structures. Training was conducted using Python (Python Core Team 2019). A momentum optimizer was applied and set at a learning rate of 0.0002 for the first 900,000 steps and 0.00002 for the last 1,200,000 steps.

After this network was trained, each auditory structure was predicted with a probability of its presence and location in each slice of a temporal bone CT scan. The image with the highest aggregating probability of the two structures on each side was used for region proposal of the ear on the same side. The region of interest was cropped into a 150 × 150 pixels JPEG image, using the same center coordinates as those of the horizontal semicircular canal box (Fig. 1C).

Data Augmentation Techniques for the Classification Tasks

To improve the efficacy of deep learning, all these cropped images were horizontally flipped (mirrored), lightly rotated, sheared, zoomed, or shifted in side length to increase the amount of training data for the classification tasks (Fig. 1D). Such data augmentation was performed by the classification network itself in a real-time fashion. All images were finally resized to 299 × 299 pixels using bicubic interpolation.

Classification Tasks Using a Classification Network

The classification network was built upon the base model of a pretrained CNN (Inception-V3, Google LLC, Mountain View, CA) (Szegedy et al. 2016). Similarly, this base model served to extract graphical features that higher-level layers use to classify images. The final classification layer was a softmax function with an output of a probability distribution for the three classes: “normal,” “CSOM,” and “cholesteatoma” (Fig. 1E).

Training was performed in Python (Python Core Team 2019) using TensorFlow (Google LLC, Mountain View, CA). A random selection of 85% of the dataset (n = 975) was used for training and validation. A stratified fivefold cross-validation method was applied to obtain an optimal model. For each of the 5 training sessions, a distinct set of 20% training and validation data (n = 195) was selected for validation while the other 80% (n = 780) were used for training. These training sessions yielded five models with different sets of parameters. Each training session was set for 1000 epochs with an initial learning rate of 0.0001. The Adam optimizer (Kingma & Ba 2014) was applied to adjust the learning ability of the algorithm and to reduce errors. In each epoch, the model’s performance was evaluated by the accuracy and the number of
errors (as known as losses) using both training and validation datasets. The training session of a model would be terminated if there was no progress made for a consecutive five epochs. In addition, a dropout function was applied to the algorithm to prevent overfitting of the trained model. During the training session, each class was given a weight inversely proportional to its sample size to balance the influence of each class to the algorithm. The model with the highest validation accuracy at its best epoch was selected as the classification network and used in the subsequent testing session.

The remaining 15% of the data (n = 172) was kept as a clean test dataset on which to assess the performance of the model after training was completed. This same dataset was also used to evaluate the performance by clinical experts.

**Evaluation of Expert Performance**

Six clinical experts were recruited to evaluate their diagnostic performance on the test dataset. Participants included 2 certified otologists with extensive clinical experience (both with 15 years of experience), 3 otolaryngologists (12, 8, and 6 years of experience), and 1 radiologist with specialist training in temporal bone CT interpretation and ear diseases (16 years of experience). Each expert was instructed to independently provide a diagnosis to each case from one of the three categories by reading the CT images only. Any patient information or other assessment was kept inaccessible to the panel. A random replication of 10% cases (n = 18) was mixed with the test dataset to assess the intrarater reliability.

**Statistical Analysis**

The performance of the deep-learning algorithm and clinical experts on the test dataset was summarized in confusion matrices. To evaluate the ability of the AI model in the diagnosis of COM, all labels were binarized into normal (“normal”) and COM (either “CSOM” or “cholesteatoma”), and a receiver operating characteristic curve was plotted. The area under the receiver operating characteristic curve (AUC) of the deep-learning algorithm was measured using the predicted probability, and its 95% confidence interval was estimated using DeLong’s method (DeLong et al. 1988). The optimal cutoff threshold of the curve was determined at the point with minimal distance to the upper left corner. The true positive rate was plotted against the false-positive rate for each of the six clinical experts on the same graph. Inter-rater reliability was measured using Fleiss’ kappa. This measure calculated the level of agreement between the AI model and the clinical experts in a classification task over that which would be expected by chance. All statistical analyses were performed using Python (Python Core Team 2019) and Excel (Microsoft Corporation, Redmond, WA). The alpha level was set at 0.05.

**RESULTS**

**The Region Proposal Network**

Examples of this network’s workflow and outcomes are shown in Figure 2. The trained deep-learning algorithm precisely detected and localized each of the two landmark structures. As a result, the region of interest for each side of ear was properly generated in all (100%) cases, confirmed by manual inspection of all cropped images.

**Performance of the Classification Network and Clinical Experts**

A fivefold cross-validation yielded the best model with a training accuracy of 78.9% and a validation accuracy of 76.2% at its most balanced epoch. This model was applied in the classification network. For a binarized classification task, its performance on classifying the presence of COM (either CSOM or cholesteatoma) versus normal condition on the test dataset was comparable to that of clinical experts. The AUC of this network was 0.92 with a 95% confidence interval between 0.88 and 0.96 (Fig. 3A). At the optimal cutoff probability determined by the curve, the AI model achieved a sensitivity of 83.3% and a specificity of 91.4%. This result even exceeded the overall performance of 6 clinical experts, with an averaged sensitivity of 81.1% and specificity of 88.8%. The model’s performance at the default cutoff probability was also assessed, resulting in a higher sensitivity of 89.5%, a lower specificity of 79.3%, and an accuracy of 86%. According to the rule of thumb (Landis & Koch 1977), the inter-rater agreement between this network and the clinical experts in detection of COM was substantial (Fleiss’ kappa = 0.72).

A more complete display of their performance is provided in Figure 3B, C. For a 3-class classification task, the overall accuracy of the AI model was 76.7%, versus a lower averaged 73.8% of clinical experts. Particularly, comparison of these numbers revealed that this network had higher recall rates in both CSOM (75% versus 70%) and cholesteatoma (76% versus 53%) cases. In addition, all diagnoses made by this network were completely repeatable (100%), and clinical experts only showed an averaged diagnostic consistency in 81% cases (ranging from 61 to 94%).

**DISCUSSION**

Owing to rapid advances in machine-learning techniques and big data available for training, AI has been making significant strides in healthcare. In particular, use of deep-learning methods in those specialties relying on medical images, such as radiology, pathology, and ophthalmology, resulted in satisfactory performance which matched or even surpassed that of humans in various tasks. For instance, deep learning has been applied in X-rays for detection of hip fractures (Gale et al. 2017) and tuberculosis (Lakhani & Sundaram 2017); in CT scans for detection of spine compression fractures (Bar et al. 2017), pancreatic cancer (Liu et al. 2018), intracranial hemorrhage (Arbabshirani et al. 2018), and critical head abnormalities (Chilamkurthy et al. 2018), classification of liver masses (Yasaka et al. 2018a), and staging of liver fibrosis (Yasaka et al. 2018b); in magnetic resonance imaging for automated cardiac segmentation (Lieman-Sifry et al. 2017), classification of brain tumor (Deepak & Ameen 2019), and prognosis prediction of amyotrophic lateral sclerosis (van der Burgh et al. 2017); in optical coherence tomography for diagnosis and referral of retinal diseases (De Fauw et al. 2017); in digital pathology slides for classifying lung cancers and predicting specific genomic mutations (Couderay et al. 2018); in dermoscopic images for detection of skin cancers (Esteva et al. 2017; Haenssle et al. 2018); in electrocardiograms for detection of arrhythmia (Rajpurkar et al. 2017; Hannun et al. 2019) and echocardiograms for classification of cardiomyopathy (Zhang et al. 2018b). The AUCs of some models even reached a near-perfect level of 0.99 (Gale et
al., 2017; Coudray et al., 2018; De Fauw et al., 2018). In addition, deep-learning algorithms have been shown to reliably predict a variety of important clinical outcomes, ranging from readmission to mortality, from patients’ electronic health data (Elfiky et al., 2018; Rajkomar et al., 2018). Several diagnostic devices and wearables implemented with deep-learning algorithms have been approved by Food and Drug Administration for detection of atrial fibrillation (Apple Inc., 2018), monitoring of blood glucose (US Food and Drug Administration, 2018b) or diagnosis of diabetic retinopathy (US Food and Drug Administration, 2018a). In the fields of neurotology and otology, a few pioneer works using basic machine-learning techniques to predict audiologic outcomes in patients with sudden sensorineural hearing loss (Bing et al., 2018) or to predict auditory and speech perception outcomes in cochlear implant children (Feng et al., 2018) have been reported.

This article represented the first study to apply deep learning in the diagnosis of COM based on CT images. This AI model has achieved a satisfactory performance. Specifically, its accuracy (86%) in classifying the presence of COM from normal
condition was equivalent or superior to those reported in previous studies based on otoscopic images, ranging from 80 to 85.6% (Kuruvilla et al. 2013; Kasher 2018). In this study, its sensitivity and specificity matched and, in some cases, surpassed those of clinical experts. In addition, its higher recall rates in identifying CSOM and cholesteatoma cases may potentially assist clinicians in differential diagnosis between these two common types of COM. All these advantages, in addition to AI’s absolute consistency and superior efficiency in decision making, implied a promising prospect for clinical application of AI in diagnosis of COM based on CT images.

To our best knowledge, this study was also the first to apply a deep-learning method to extract regions of interest from temporal bone CT scans. In previous studies, regions of the ear in CT images were commonly segmented using an atlas-based approach (Ferreira et al. 2014; Powell et al. 2017; Senaras et al. 2018). The main idea of this method is to build an atlas by compiling several known information of the auditory structures and use this atlas for region proposal of the same structures in new CT images. This method, however, required copious anatomical knowledge and lacked generalizability (Ferreira et al. 2014). In a recent study, Zhang et al. (2018a) applied a deep-learning method for automated detection of the inner ear, but only aimed at counting the number of ears in a CT image. The present study went beyond these prior studies by applying deep learning in detection and localization of auditory structures, extracting key regions of the auditory system, and bridging them to a classification network. Moreover, because the temporal bone and inner, middle, and outer regions of the ear are all visible in the images extracted by this model, the potential application of this network is not only in diagnosis of otitis media but also expandable to other ear diseases.

A classic problem of machine learning is overfitting, especially in cases with extensive training steps or limited sample sizes. Consequently, the overfitting model is likely to perform worse on an unseen dataset. To prevent overfitting and improve the generalizability of this model, several approaches were applied. First, early termination of the training process was executed if no progress was made for a consecutive five epochs. As a result, all training sessions were finished by 15 epochs, so that the risk of the model tuned very specific to random features of the training data was minimized. Second, a dropout function was used in the classification network. The term “dropout” refers to randomly ignoring a number of units, resembling neurons, in a neural network during the training phase. A simple interpretation of dropout is to decrease the interdependency among these units, forcing each of them to take more individual responsibility in generating the final output. It has been recognized as an effective method to prevent overfitting (Srivastava et al. 2014). Third, the technique of data augmentation was applied to largely increase the number of images available for training. With all these approaches, the accuracy of this model was 78.9%, 76.2%, and 76.7% in a 3-class classification task on the training, validation, and test dataset, respectively, indicating an appropriate fitting pattern. In addition, the method of cross-validation and assignment of weights to different classes contributed to improving the validity of this model.

A major barrier against the widespread implementation of deep learning in medicine is difficulty in understanding the rationale from which the algorithm draws a conclusion. In a
recent study (De Fauw et al. 2018), researchers built a framework with a segmentation network to generate tissue maps from digital optical coherence tomography scans, followed by a classification network for referral suggestion and diagnosis. Such workflow divided the decision-making process into two distinct steps: interpretation of the original scan and generating judgments based on findings from the first step. In the present study, the AI framework was designed in a similar fashion, separating the extraction of essential regions from the diagnosis of COM. This strategy resembled the decision-making process of a clinician and allowed inspection of the extracted regions of interest by an observer. Such method may provide a solution to understand the mysterious “black box” behind an AI model.

Further inspection of some misclassified cases provided insight into the network’s logic. Examples are shown in Figure 4. Both the AI model and the clinical experts managed to classify most of the typical, less challenging cases, as represented in Figure 4A. However, there were still some cases which this model and/or the clinical experts were not able to correctly classify (Fig. 4B–D). These challenging cases usually presented with atypical or subtle patterns in the CT scans. Although the exact strategy that the AI model adopted for classification task was still beyond comprehension, the higher recall rates in classifying CSOM and cholesteatoma cases reflected AI’s greater sensitivity in reading abnormal patterns in the CT images. This feature is valuable for providing initial diagnostic screening by minimizing the risk of missing COM cases. Clinical experts, however, demonstrated higher specificity by correctly identifying greater percentage of normal ears. Therefore, they may serve to rule out those false-positive cases misclassified by the AI model. Such utilization of AI technology in clinical settings is expected to increase efficiency, reduce errors, and accomplish goals with minimal manual input.

Last but not the least, it is worth mentioning that each diagnosis made by the classification network of this AI model was in fact based on only a single, two-dimensional, cropped image, while clinical experts had access to all of the full-sized CT images. Not only could they obtain diagnostic information from multiple slices but they were also able to compare the opacification of mastoid air cells between two ears. Both advantages in reading temporal bone CT images are very helpful for diagnosis of COM. Despite these disadvantages, the AI model still managed to achieve a comparable, in some cases superior performance to that of clinical experts. In this sense,
any improvement in the algorithm by incorporating additional information obtained from the adjacent CT slices or from the contralateral side is expected to promote this model’s performance to a greater level.

Several limitations should be noted in this study. First, only a moderately sized dataset was obtained, and the sample sizes for several categories of COM were either too small to establish a reliable classifier or mildly imbalanced. Although this limitation was partially alleviated by removing small-sized categories from training and by applying several machine-learning methods, such as data augmentation and assignment of class weights, it might still have affected the performance of this model to a certain extent. As a result, it was only trained to classify a narrow spectrum of COM. Second, application of this model was only limited to a single two-dimensional CT image based on the current availability of machine-learning techniques. Obviously, this is less optimal compared with using three-dimensional CT images, which provide an entire coverage of the middle ear region and the temporal bone. Therefore, a validated model that takes three-dimensional images is expected to improve the diagnostic performance of this model. Such an algorithm also matches the way in which a clinician interprets a CT scan. Third, the accuracy of ground truth labels assigned to the unoperated ears remained unproven. In fact, the gold standard for determining the type of COM is pathology. In this study, however, diagnoses to the unoperated ears could only be made based on a combination of history, ear exams, audiograms, and temporal bone CT scans. Although this was a widely accepted practical guideline for diagnosis of COM, it was still not as accurate as pathology. With this regard, a potential improvement in the method is to train this model using only cases with confirmed pathology and to incorporate other assessments with the CT scans. The ultimate goal is to establish a comprehensive AI model that reliably assists clinicians in diagnosis of COM as well as other ear diseases.

CONCLUSIONS

This article presents a deep-learning framework that automatically extracted the region of interest from two-dimensional temporal bone CT slices and made diagnosis of COM. The performance of this model was comparable and, in some cases, superior to that of clinical experts. These results implied a promising prospect for clinical application of AI in diagnosis of COM based on CT images. Future study is to extend this approach using three-dimensional CT images to a larger scale of ear diseases.

ACKNOWLEDGMENTS

We appreciate the suggestions provided to improve our methodology by Dayi Bian, Shuxing Bao, Yiyuan Zhao from Vanderbilt University, and Chenghua Tao from Indiana University at Bloomington. We acknowledge Maria Powell from Vanderbilt University Medical Center for her invaluable opinions in manuscript writing.

Supported by the National Key Research and Development Program of China (2016YFC0905200, 2016YFC0905202) to F.-L.C.; the National Natural Science Foundation of China (NSFC) (Grant Nos. 8142018010 to F.-L.C. and 81771017 and 81570920 to D.-D.R.); the “Zhao-Xue Plan” of Fudan University to D.-D.R.; the Shanghai Outstanding Young Medical Talent Program to D.-D.R.. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Y.L. conceptualized and designed the study, reviewed and analyzed the data, performed computer programming, evaluated the AI model, wrote and edited the manuscript; Y.-M.W. retrieved and validated the data; Y.-S.C. retrieved the data and evaluated the AI model; Z.-Y.H. retrieved and validated the data; J.-M.Y., J.-H.X., and Z.-C.C. evaluated the AI model; F.-I.C. provided funding support and the data resources; D.-D.R. conceptualized the study, provided funding support and data resources, administered the project and edited the manuscript. All authors have reviewed, discussed, and approved the manuscript.

The authors have no conflicts of interest to disclose.

Address for correspondence: Yike Li, Department of Otolaryngology, Vanderbilt University Medical Center, 1313 21st Avenue South, 602 Oxford House, Nashville, TN 37232, USA. E-mail: yike.li1@vumc.org; Dong-Dong Ren, Department of Otorhinolaryngology, Eye and ENT Hospital, 83 Fenyang Road, Shanghai, 200031, China. E-mail: dongdongren@fudan.edu.cn

Received May 28, 2019; accepted July 22, 2019.

REFERENCES


