A Three-Dimensional and Explainable Artificial Intelligence Model for Evaluation of Chronic Otitis Media Based on Temporal Bone Computed Tomography: Model Development, Validation, and Clinical Application

Binjun Chen, Yike Li, Yu Sun, Haojie Sun, Yanmei Wang, Jihan Lyu, Jiajie Guo, Yushu Cheng, Xun Niu, Lian Yang, Jianghong Xu, Juanmei Yang, Yibo Huang, Fanglu Chi, Bo Liang, Dongdong Ren

Submitted to: Journal of Medical Internet Research
on: August 09, 2023

Table of Contents

A Three-Dimensional and Explainable Artificial Intelligence Model for Evaluation of Chronic Otitis Media Based on Temporal Bone Computed Tomography: Model Development, Validation, and Clinical Application

Binjun Chen[1*] MD; Yike Li[2*] MD, PhD; Yu Sun[3*] MD, PhD; Haojie Sun[1] MD; Yanmei Wang[1] MD; Jihan Lyu[1] MD; Jiajie Guo[4] PhD; Yushu Cheng[5] MD; Xun Niu[3] MD; Lian Yang[6] MD; Jianghong Xu[1] MD, PhD; Juanmei Yang[1] MD, PhD; Yibo Huang[1] MD, PhD; Fanglu Chi[1] MD, PhD; Bo Liang[6] MD; Dongdong Ren[1] MD, PhD

[1]Department of Otorhinolaryngology Eye, Ear, Nose and Throat Hospital Fudan University Shanghai CN
[2]Department of Otolaryngology-Head and Neck Surgery Vanderbilt University Medical Center Nashville US
[3]Department of Otorhinolargnology Union Hospital, Tongji Medical College Huazhong University of Science and Technology Wuhan CN
[4]State Key Laboratory of Digital Manufacturing Equipment and Technology School of Mechanical Science and Engineering Huazhong University of Science and Technology Wuhan CN
[5]Department of Radiology Eye, Ear, Nose and Throat Hospital Fudan University Shanghai CN
[6]Department of Radiology Union Hospital, Tongji Medical College Huazhong University of Science and Technology Wuhan CN
[*]these authors contributed equally

Corresponding Author:
Yike Li MD, PhD
Department of Otolaryngology-Head and Neck Surgery
Vanderbilt University Medical Center
1215 21st Avenue South
Nashville
US

Abstract

Background: Computed tomography (CT) of the temporal bone has become a critical diagnostic approach to chronic otitis media (COM), but it requires training and experience for interpretation. Artificial intelligence may assist clinicians in evaluating COM using CT with efficiency and reliability, but the logic for decision-making can be incomprehensible and there is currently no model that makes full use of the multidimensional diagnostic information.

Objective: This study was to develop an explainable and three-dimensional (3D) deep learning framework for detection and differential diagnosis of COM based on CT.

Methods: Temporal bone CT scans were retrospectively obtained from patients receiving surgeries for COM between December 2015 and July 2021 at two independent institutes. The region of interest containing the middle ear was automatically segmented, followed by 3D convolutional neural networks trained to identify pathological ears and cholesteatoma. Gradient-weighted class activation mapping was used to generate heatmaps highlighting the critical regions for decision-making. Model performance was evaluated over five rounds of cross-validation and external validation and benchmarked against clinical experts.

Results: The internal and the external datasets contained 1,661 patients (number of eligible ears[n]= 3,153) and 108 patients (n=211), respectively. The deep learning model achieved decent and comparable area under the receiver operating characteristic curve (AUROC) scores ([mean ± SD]: 0.96±0.01 and 0.93±0.01) and accuracies (87.8±1.7% and 84.3±1.5%) in detection of pathological ears on two datasets. Similar outcomes were also observed in identifying cholesteatoma, with AUROC of 0.85±0.03 and 0.83±0.05, and accuracies of 78.3±4.0% and 81.3±3.3%, respectively. The model exhibited equivalent or superior performance and a much higher consistency compared to experts' averages in both tasks. The heatmaps properly highlighted the middle ear and mastoid regions, consistent with human knowledge in interpreting temporal bone CT.

Conclusions: This study suggests the feasibility of a 3D deep learning framework in automatic evaluation of COM using CT. This model demonstrates decent performance, generalizability, and transparency, making it a useful tool to assist clinicians in assessment of COM.

(JMIR Preprints 09/08/2023:51706)

DOI: https://doi.org/10.2196/preprints.51706

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

A Three-Dimensional and Explainable Artificial Intelligence Model for Evaluation of Chronic Otitis Media Based on Temporal Bone Computed Tomography: Model Development, Validation, and Clinical Application

Binjun Chen, MD[1,†], **Yike Li**, MD, PhD[2,†,#], **Yu Sun**, MD, PhD[3,†], **Haojie Sun**, MD[1], **Yanmei Wang**, MD, PhD[1], **Jihan Lyu**, MD[1], **Jiajie Guo**, PhD[4], **Shunxing Bao**, PhD[5], **Yushu Cheng**, MD[6], **Xun Niu**, MD[3], **Lian Yang**, MD[7], **Jianghong Xu**, MD, PhD[1], **Juanmei Yang**, MD, PhD[1], **Yibo Huang**, MD, PhD[1], **Fanglu Chi**, MD, PhD[1], **Bo Liang**, MD[7,#], **Dongdong Ren**, MD, PhD[1,#]

1. Department of Otorhinolaryngology, Eye, Ear, Nose and Throat Hospital, Fudan University, Shanghai, China; **2.** Department of Otolaryngology-Head and Neck Surgery, Bill Wilkerson Center, Vanderbilt University Medical Center, Nashville, TN, USA; **3.** Department of Otorhinolaryngology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China; **4.** State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China; **5.** Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA; **6.** Department of Radiology, Eye, Ear, Nose and Throat Hospital, Fudan University, Shanghai, China; **7.** Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

† These authors contributed equally to this work.

Corresponding authors: **Yike Li**, Department of Otolaryngology-Head and Neck Surgery, Vanderbilt University Medical Center, 1215 21st Avenue South, Rm. 10410, Medical Center East, Nashville, TN 37232, USA. E-mail: yike.li.1@vumc.org. ORCID: 0000-0001-8465-130X; **Bo Liang**, Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Jiefang Avenue #1277, Wuhan, Hubei, 430022, China. E-mail: xiehelb@163.com; **DongDong Ren**, Department of Otorhinolaryngology, Eye, Ear, Nose and Throat Hospital, 83 Fenyang Road, Shanghai, 200031, China. E-mail: dongdongren@fudan.edu.cn.

Abstracts

Background: Temporal bone computed tomography (CT) plays a crucial role in the diagnosis of chronic otitis media (COM). However, its interpretation requires training and expertise. Artificial intelligence (AI) has the potential to assist clinicians in evaluating COM through CT scans, but existing models lack transparency and may not fully leverage multidimensional diagnostic information.

Objective: This study aimed to develop an explainable AI system based on three-dimensional (3D) convolutional neural networks (CNNs) for the automatic evaluation of COM via CT scans.

Methods: Temporal bone CT scans were retrospectively obtained from patients receiving surgeries for COM between December 2015 and July 2021 at two independent institutes. A region of interest encompassing the middle ear was automatically segmented, and 3D CNNs were subsequently trained to identify pathological ears and cholesteatoma. An ablation study was performed to refine model architecture. Benchmark tests were conducted against a baseline two-dimensional (2D) model and seven clinical experts. Model performance was measured through cross-validation and external validation. Heatmaps, generated using gradient-weighted class activation mapping, were employed to highlight critical decision-making regions. Finally, the AI system was assessed with a prospective cohort to aid clinicians in preoperative COM assessment.

Results: The internal and the external datasets contained 1,661 patients (number of eligible ears[n]= 3,153) and 108 patients (n=211), respectively. The 3D model exhibited decent performance with area under the receiver operating characteristic curve (AUROC) scores (mean ± SD) of 0.96±0.01 and 0.93±0.01, and accuracies of 87.8±1.7% and 84.3±1.5%, respectively, for detecting pathological ears on the two datasets. Similar outcomes were observed for cholesteatoma identification (AUROC: 0.85±0.03 and 0.83±0.05; accuracies: 78.3±4.0% and 81.3±3.3%). The proposed 3D model achieved a commendable balance between performance and network size relative to alternative models. It

significantly outperformed the 2D approach in detecting COM (P<=.05) and exhibited a substantial gain in identifying cholesteatoma (P<.001). The model also demonstrated superior diagnostic capabilities over resident fellows and the attending otologist (P<.05), rivaling all senior clinicians in both tasks. The generated heatmaps properly highlighted the middle ear and mastoid regions, aligning with human knowledge in interpreting temporal bone CT. The resulting AI system achieved an accuracy of 81.8% in generating preoperative diagnoses for 121 patients and contributed to clinical decision-making in 90.1% cases.

Conclusions: This study presents a 3D CNN model trained to detect pathological changes and identify cholesteatoma via temporal bone CT scans. In both tasks, this model significantly outperforms the baseline 2D approach, achieving levels comparable to or surpassing those of human experts. The model also exhibits decent generalizability and enhanced comprehensibility. This AI system facilitates automatic assessment of COM and shows promising viability in real-world clinical settings. These findings underscore the potential of AI as a valuable aid for clinicians in COM evaluation.

Keywords

Artificial intelligence; Cholesteatoma; Deep learning; Otitis media; Tomography, Xray computed

Introduction

Chronic otitis media (COM) represents a recurrent inflammatory condition inside the tympanic cavity [1]. COM encompasses various forms, including chronic suppurative otitis media (CSOM) and cholesteatoma, each with unique histological characteristics. CSOM involves the accumulation and discharge of purulent fluid, affecting an estimated 330 million people worldwide, with approximately half experiencing hearing loss [2]. Cholesteatoma is characterized by the build-up of keratinized squamous epithelium, which has the potential to erode auditory structures and exhibits a notable tendency for relapse. Accurate identification and differentiation of COM types are crucial for effective disease management and surgical planning [3]. Mastoidectomy, which involves the removal of part of the temporal bone, is the conventional surgical approach for COM. However, less invasive techniques like endoscopic tympanoplasty are gaining favor for treating CSOM and other non-cholesteatoma conditions due to their potential for reduced structural damage and faster recovery [4-9].

Temporal bone computed tomography (CT) is vital for assessing COM and aiding in surgical planning, especially when initial otoscopic examinations have restricted views and yield inconclusive findings [10]. Offering a cost-effective alternative to magnetic resonance imaging (MRI), CT is instrumental in distinguishing cholesteatoma from CSOM by detecting osseous erosion in the tympanum. Although studies have shown that clinicians are capable of diagnosing COM based on CT alone [11-17], distinguishing between COM subtypes poses greater challenges to the human eye. Moreover, interpreting temporal bone CT scans requires specialized training and experience, which may not be universally available across otolaryngologists.

Artificial intelligence (AI) is making remarkable advancements in healthcare. Deep learning (DL) models, particularly convolutional neural networks (CNNs), have demonstrated enhanced efficiency and reduced errors in disease diagnoses and prediction of clinical outcomes [18-21]. While

a few recent papers have reported CNN models in evaluating COM with accuracy scores ranging from 0.77 to 0.85, these studies primarily relied on otoscopic or single-layer CT images [22,23]. These two-dimensional (2D) representations may not be optimal for revealing pathological changes in concealed or peripheral anatomical structures, such as the attic space and the mastoid air cells. Additionally, the inherent “black box” nature of DL models, where decision-making strategies are challenging to understand, has been a common criticism [24,25]. This lack of comprehensibility hinders the widespread adoption of AI models in clinical practice.

In light of these challenges, this study aimed to create an explainable, three-dimensional (3D) CNN model for the automatic interpretation of temporal bone CT scans. The model was designed to pinpoint the region of interest (ROI) and identify pathological and cholesteatomatous conditions in a 3D fashion. Comprehensive benchmarks against baseline methods and human experts on distinct datasets were conducted to demonstrate the robustness and generalizability of this model. Additionally, heatmap generation was employed to highlight potential pathological changes in CT scans and elucidate the model’s rationale for making predictions. These features were integrated into an AI system for the automatic, end-to-end evaluation of COM, which was subsequently assessed in clinical settings. The overarching goal of this system is to support clinicians in making informed decisions for common otologic conditions, thereby enhancing efficiency, reliability, and transparency.

Methods

Ethical considerations

This study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was granted by the Institutional Review Boards at Vanderbilt University Medical Center (IRB number: 191804) and the Eye, Ear, Nose and Throat (EENT) Hospital of Fudan University (IRB number: 2019076). Informed consent was waived as all data were de-identified. The

observational study, which aimed to assess the model's viability in aiding preoperative assessment, was registered with the Chinese Clinical Trial Register (ChiCTR: 2000036300). No compensation was provided to any study participants.

Participants

Data were retrospectively obtained from patients admitted for middle ear surgeries from December 2015 to July 2021 at EENT Hospital. Patients diagnosed with acute otitis media, any inner or external ear diseases, or those with missing temporal bone CT scan were excluded, resulting in 1,661 eligible for model development. An extra dataset containing 108 patients with COM was collected from Wuhan Union (WU) Hospital for external validation (Figure 1).

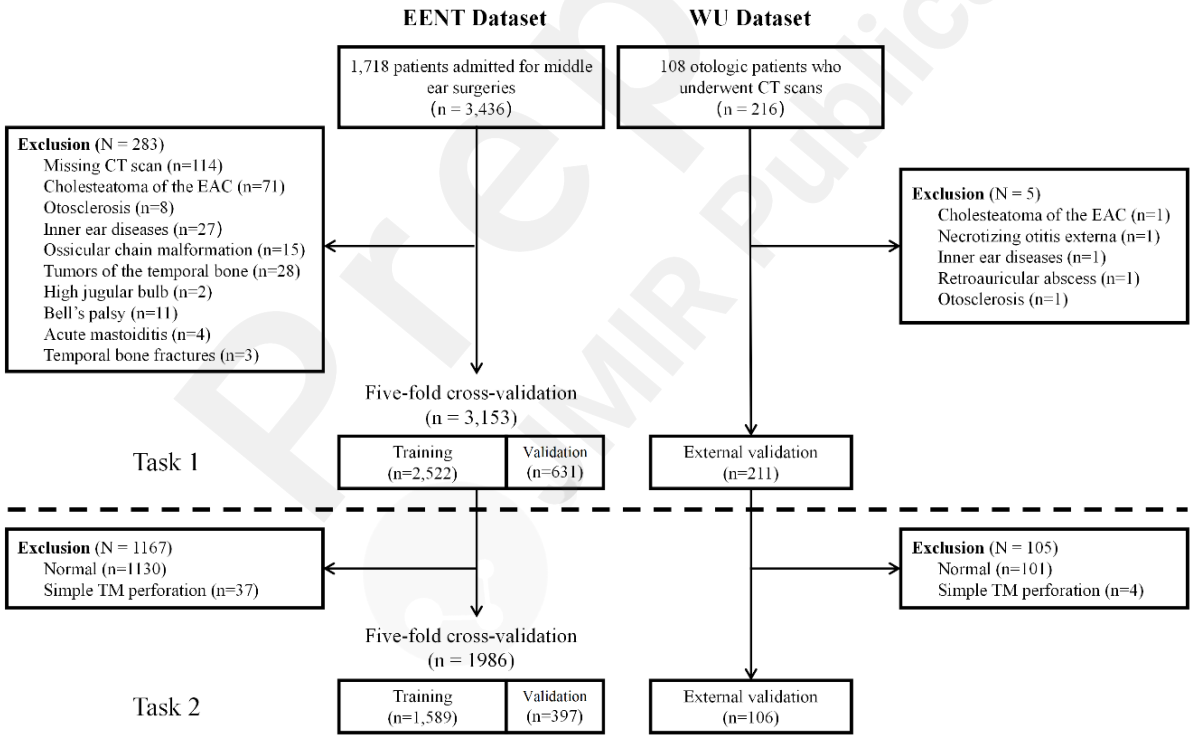


Figure 1. Flow chart of data retrieval. EAC: external auditory canal; TM: tympanic membrane.

Temporal bone CT scans

As part of the routine preoperative assessment, each patient underwent at least one temporal bone CT scan, conducted from the lower margin of the external auditory meatus to the top margin of the petrous bone using a SOMATOM Sensation 10 CT scanner (Siemens Inc., Munich, Germany) at the EENT Hospital. The scanning parameters were as follows: matrix (512 x 512), field of view (220mm x 220mm), tube voltage (140kV), tube current (100mAs), section thickness (0.6 - 0.75mm), window width (4000HU), and window level (700HU). CT scans from the WU Hospital were obtained using a SOMATOM Plus 4 model (Siemens Inc., Munich, Germany) with different settings for field of view (100mm), voltage (120kV), and thickness (0.75mm). All images were saved in the DICOM format.

Label assignment

All eligible ears were treated as independent cases and assigned ground truth labels based on their diagnoses (Table 1). Each label was verified according to intraoperative findings and pathology reports for operated ears, and employing a combination of history, ear examination, audiogram results, and imaging findings for unoperated ears. In cases of unoperated ears, a “normal” label was assigned when there was an absence of ear symptoms, hearing loss, or signs of inflammation. A diagnosis of CSOM was assigned when chronic purulent discharge, conductive hearing loss, and the presence of a perforated tympanic membrane or soft tissue shadow in the tympanic cavity were observed. Cholesteatoma was considered if keratin debris was identified, or if there were signs of osseous damage along with retraction or perforation of the pars flaccida [22]. Two otolaryngology residents with full access to patients’ medical records independently reviewed these labels as unblinded annotators. Any discrepancies were addressed with senior specialists until a consensus was reached. All data were de-identified and stored on password-protected computers.

Table 1. Summary of patient characteristics and label assignment.

Characteristics	EENT dataset (N = 1,661; Number of ears [n] = 3,153)	WU dataset (N=108; n=211)
Patient age (year), mean ± SD	41.1 ± 16.6	39.8 ± 14.0
Patient sex, male: female (%)	832 (50.1): 829 (49.9)	49 (45.4): 59 (54.6)
Diagnosis per ear, n (%)		
Normal	1130 (35.8)	101 (47.9)
Cholesteatoma	728 (23.1)	30 (14.2)
CSOM	1011 (32.1)	69 (32.7)
Tympanosclerosis	142 (4.5)	2 (0.1)
Cholesterol granuloma	72 (2.3)	1 (0.05)
OME	41 (1.3)	7 (3.3)
Adhesive otitis media	29 (0.1)	1 (0.05)
Task 1 labels, n (%)		
Normal	1130 (35.8)	101 (47.9)
Pathological	2023 (64.2)	110 (52.1)
Task 2 labels, n (%)		
Cholesteatoma	728 (36.7)	28 (26.4)
Non-cholesteatoma	1258 (63.3)	78 (73.6)

SD: standard deviation; OME: otitis media with effusion

Model architecture

The framework consists of two functionally distinct units: a region proposal network for 3D segmentation of ROI, and a classification network for generating predictions. Both networks are established based on CNN models.

Region proposal network

This network is designed to extract the middle ear on each side from a full set of temporal bone CT scan (Figure 2A). It contains a YOLO (You Only Look Once, v5) model that is trained to detect and locate two auditory structures, including the internal auditory canal and the horizontal semicircular canal, in a series of 2D axial CT images [26]. These landmarks, positioned at or around the central level of the middle ear, possess unique graphical appearances recognizable by the object

detection model. In our recent study, this model demonstrated a 100% success rate in identifying the middle ear region from temporal bone CT scans [22]. Subsequently, a 3D data matrix (150x150x32) of the ROI is extracted based on the center coordinates of these two structures on each side.

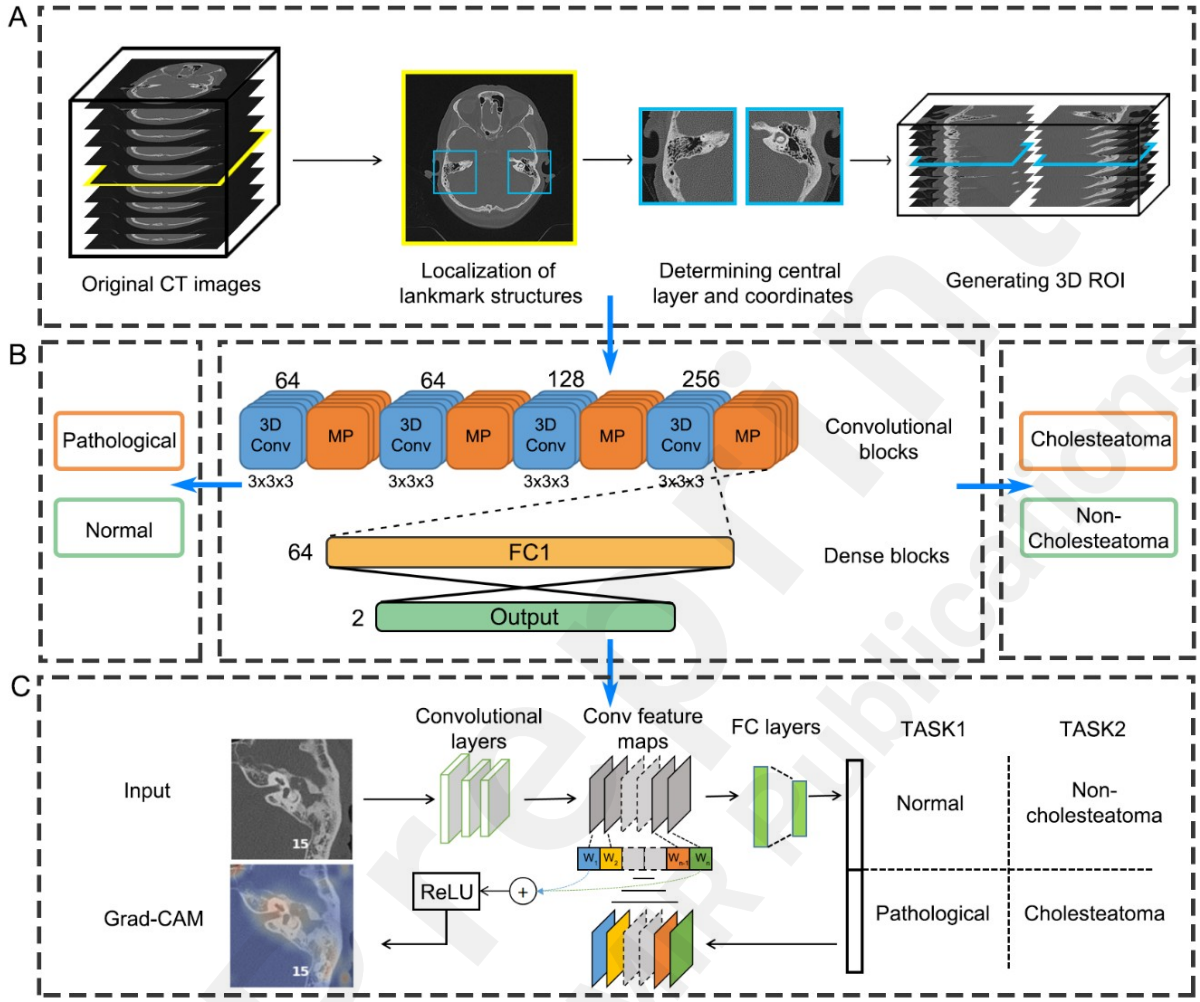


Figure 2. An overview of the AI framework. (A) The region proposal network used to locate landmark structures and segment the 3D ROI from the original CT images. (B) The classification network based on a 3D CNN architecture and trained to perform two classification tasks. (C) The gradient heatmaps generated to highlight the critical regions for decision-making. Conv: Convolution; FC: Fully connected; ReLu: Rectified Linear unit.

Classification network

A 3D CNN model is built to interpret the extracted ROI and classify different types of conditions (Figure 2B). This model features four convolution blocks and two dense blocks (Table 2). Each convolution block consists of a 3D convolutional layer to summarize graphical features along all

axes of the input image, followed by a max-pooling layer for downsampling these features and another layer for batch normalization. These high-level features are then pooled and passed to the fully connected layers of the dense blocks, where the diagnosis is predicted based on the calculated probability of each class by a softmax function. A dropout layer is applied to prevent overfitting [27].

Table 2. Architecture of the 3D CNN model.

Block	Kernel	Settings
Input	Input	
Convolution 1	Conv3D	(3,3,3,64)
	MaxPooling3D	(2,2,2)
	BatchNormalization	
Convolution 2	Conv3D	(3,3,3,64)
	MaxPooling3D	(2,2,2)
	BatchNormalization	
Convolution 3	Conv3D	(3,3,3,128)
	MaxPooling3D	(2,2,2)
	BatchNormalization	
Convolution 4	Conv3D	(3,3,3,256)
	MaxPooling3D	(2,2,2)
	BatchNormalization	
	GlobalAveragePooling3D	
Dense 1	Fully connected	64
	Dropout	0.3
Output	Fully connected	2

Conv3D: three-dimensional convolutional layer; MaxPooling3D: three-dimensional max pooling layer; BatchNormalization: batch normalization layer; GlobalAveragePooling3D: layer performing global average pooling for three-dimensional data.

Model training and testing

Task 1 – Detection of COM

The first classification model was trained in a binary task distinguishing between normal and pathological ears in all cases (n=3,153). The training and testing procedures involved five-fold cross-validation on the internal (EENT) dataset. Specifically, the dataset was evenly partitioned into five non-overlapping subsets in a random, stratified fashion. In each iteration, one subset was reserved for

testing (n=631), while the remaining four were used for training (n=2,522). Model performance metrics were averaged over five iterations of this process. During each training session, a random 20% of training images (n=504) were allocated for validation. Training was set for 1000 epochs with an initial learning rate of 0.0001, and the Adam optimizer was employed to dynamically adjust the algorithm's learning capability and minimize errors [28]. Early termination was implemented if no further decrease in validation loss was observed for a consecutive 10 epochs. These hyperparameters were determined based on the resultant model performance and training efficiency shown in a preliminary study. The trained model was also evaluated on the external dataset (n=211) in each round.

Task 2 – Identification of cholesteatoma

The second classification model was trained to specifically identify cholesteatoma on selected CT images that displayed signs of inflammation in the middle ears. This task was designed to simulate a common clinical scenario where clinicians need to differentiate cholesteatoma from other types of COM in patients with positive imaging findings. The aim was to provide a preoperative assessment of the risk of cholesteatoma, assisting clinicians in surgical planning [3,29]. For this task, a subset of CT scans with visible soft tissue density or increased opacification in the middle ear or mastoid was selected from both the internal (n=1,986) and external sets (n=106). The remaining methods, including extraction of ROI, network architecture, and the training and testing procedures, were consistent with those used in the first task.

Ablation study

To refine model selection and gain a better understanding of the network’s behavior, an ablation study was performed to compare the proposed classification network with three alternative models,

each incorporating modifications to certain features. Specifically, the number of convolutional blocks was decreased and increased by 1 in alternative Model 1 and Model 2, respectively, and a different size of filter was applied in Model 3 (Table S1-S3 in Multimedia Appendix 1). To ensure adequate statistical power for detecting differences across models, experiments were conducted on the main dataset using the same methodology as outlined in the preceding sections.

Benchmarking against the 2D approach

To investigate whether the use of 3D CT images may enhance diagnostic performance, a benchmark study was designed to compare the proposed system with a baseline model utilizing 2D images. This baseline model, previously established by our team, uses transfer learning on a pretrained Inception-V3 (Google LLC, Mountain View, CA) model [22]. In the current study, the base model of Inception-V3 was retained, and the final classification layer was customized with a binary output. Training and validation were conducted in the same manner as the 3D model, except that only a single CT image at the central layer of the ROI was used as the input for the 2D model. All image preprocessing techniques and hyperparameter settings remained consistent with those outlined in the previous study [22].

Benchmarking against human experts

Another benchmark test was performed against human experts to provide an additional unbiased evaluation of the proposed system. Seven human specialists with a broad range of qualifications were recruited to perform both tasks based on the same image data. The participants included 2 senior otologists, each with 12 years of clinical experience, 1 senior head and neck radiologist with 21 years of experience, 1 attending otologist with 7 years of experience, and 3 otolaryngology residents with 3, 3, and 2 years of experience, respectively. Each expert was provided only with the

CT images and instructed to make a task-specific diagnosis to each ear (task 1: normal or pathological; task 2: cholesteatoma or non-cholesteatoma). The test data for clinicians comprised a random selection of 244 ears from the EENT set and all eligible ears from the WU set. To assess intra-rater reliability, a random replication of 10% of test cases (n=48) was mixed with these data. All test cases (N=502) had not been previously seen by any experts. They were anonymized, shuffled and stored on a password-protected computer along with spreadsheets to record each expert's diagnoses for these cases.

Generation of heatmaps

Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized to visualize model's rationale for decision-making (Figure 2C). In essence, this approach leverages the gradients of the target class flowing into the final convolutional layer to produce a coarse localization heatmap, highlighting the critical regions in the image [30]. In this study, heatmaps were generated in a 3D fashion and rescaled to match the original images using TensorFlow 2.11 in Python 3.91 [31].

Clinical applications

The validated model was integrated into a Python program, enabling the automated assessment of COM from raw CT inputs to the generation of explainable diagnoses in an end-to-end fashion (See Data availability statements and Multimedia Appendix 2). To evaluate its viability in assisting otologists in clinical settings, this system was employed with a prospective cohort of patients undergoing middle ear surgeries at EENT hospital from November 2023 to January 2024 in a single-arm observational study. Preoperative model predictions, along with routine assessments, were provided to two senior otologists, who were given autonomy to determine surgical strategies based on their discretion. Surgeons were surveyed regarding the utility of model-generated

information in their decision-making processes for these cases. Model predictions were utilized to analyze the selection of surgical approaches and to measure model performance against pathological findings. Hearing gain was assessed by comparing the air conduction threshold at 2 weeks postoperatively with the baseline.

Statistical analysis

Descriptive statistics were applied as appropriate. The overall predictability of a model was evaluated by the area under the receiver operating characteristic (AUROC) curve. The optimal cutoff threshold on the curve was determined at the point with minimal distance to the upper left corner on the validation set and subsequently applied to the test set. The numbers of correctly and incorrectly classified cases were displayed in a confusion matrix, and these were used to calculate the performance metrics, including accuracy, recall, specificity, precision and F1 score. These metrics offer comprehensive insights into the model's performance, covering overall correctness in identifying both positives and negatives (accuracy), sensitivity in detecting positive cases (recall), capability in ruling in patients (specificity), propensity for preventing false alarms (precision), and effectiveness in identifying positive cases while minimizing false positives and false negatives (F1 score). They were derived by:

$$\text{Accuracy} = (\text{True Positive} + \text{True Negative}) / \text{Total Sample Size}$$

$$\text{Recall} = \text{True Positive} / (\text{True Positive} + \text{False Negative})$$

$$\text{Specificity} = \text{True Negative} / (\text{True Negative} + \text{False Positive})$$

$$\text{Precision} = \text{True Negative} / (\text{True Negative} + \text{False Negative})$$

$$\text{F1} = 2 \times \text{True Positive} / (2 \times \text{True Positive} + \text{False Positive} + \text{False Negative}).$$

Results are averaged over 5 iterations of cross-validation or external validation and presented as mean ± standard deviation. Intra-rater consistency was evaluated using Cohen's kappa. Significance

was determined through pairwise t-test for difference in performance between models and via one-way analysis of variance between the proposed model and human experts. The alpha level was set at 0.05. Statistical analyses were conducted using Python 3.91 [31].

Results

ROI extraction

The region proposal network successfully extracted the 3D ROI containing the critical anatomies on each side, including the tympanic cavity and sinus tympani (Figure 3). This has been confirmed by manual inspection of the generated images in all cases from both datasets.

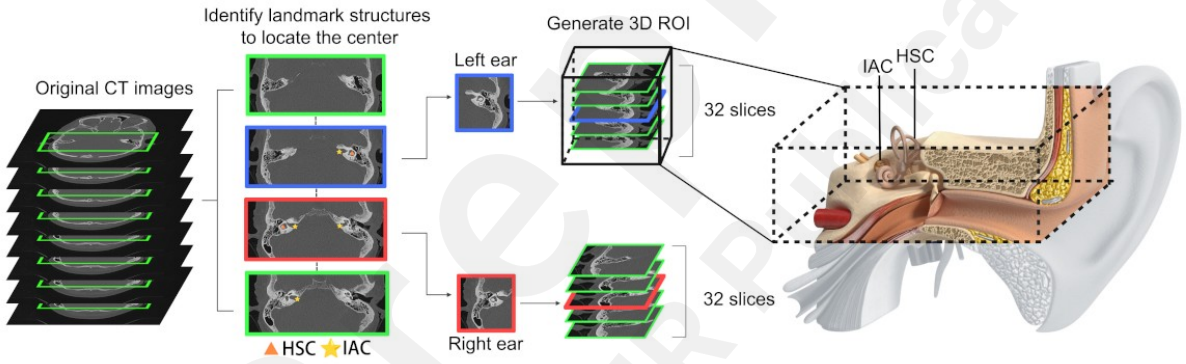


Figure 3. Generation of the 3D ROI. The region proposal network identifies landmark structures in each of the full-sized sequential CT slices and determines the center of the middle ear on each side. A 3D image comprising 32 stacks of axial slices in 150x150 pixels is subsequently segmented. This ROI encompasses an extensive range of critical anatomies within the temporal bone for the evaluation of COM. HSC: horizontal semicircular canal; IAC: internal auditory canal.

Task 1

Our model exhibited decent performance in identifying pathological changes in the middle ear, achieving a mean accuracy of 87.8%, recall of 85.3%, specificity of 91.3%, and precision of 93.3% on the internal dataset (Table 3). It also demonstrated a near-perfect AUROC score of 0.96. These performance metrics remained generally consistent on the external dataset, with a comparable

AUROC score of 0.93, indicating reasonable generalizability (Figure 4).

Table 3. Performance of the baseline 2D and the proposed 3D models.

Task	Model	Size (MB)	Dataset	Accuracy	Recall	Specificity	Precision	F1 Score	AUROC	*P*
1	3D	14.2	EENT	87.8 ± 1.7%	85.3 ± 3.2%	91.3 ± 6.7%	93.3 ± 4.5%	89.0 ± 1.2%	0.959 ± 0.011	.003
	2D	274		86.1 ± 1.9%	84.5 ± 2.8%	88.3 ± 5.2%	90.9 ± 3.6%	87.5 ± 1.6%	0.939 ± 0.013	
	3D	14.2	WU	84.3 ± 1.5%	75.6 ± 4.7%	93.4 ± 2.1%	92.4 ± 1.8%	83.0 ± 2.2%	0.933 ± 0.010	.05
	2D	274		82.1 ± 2.3%	74.4 ± 7.8%	90.1 ± 4.6%	89.1 ± 3.9%	80.8 ± 3.6%	0.918 ± 0.012	
2	3D	14.2	EENT	78.3 ± 4.0%	80.8 ± 2.5%	77.0 ± 5.4%	65.2 ± 6.0%	72.1 ± 4.2%	0.853 ± 0.030	<.001
	2D	274		67.0 ± 3.7%	71.6 ± 14.4%	64.6 ± 11.9%	52.3 ± 4.4%	59.6 ± 3.6%	0.744 ± 0.025	
	3D	14.2	WU	81.2 ± 3.3%	61.4 ± 8.5%	87.8 ± 3.1%	62.6 ± 7.8%	61.8 ± 6.9%	0.826 ± 0.055	<.001
	2D	274		67.6 ± 10.3%	47.9 ± 22.4%	74.1 ± 18.5%	41.0 ± 8.6%	41.1 ± 9.6%	0.714 ± 0.049	

Task 2

This model also demonstrated satisfactory predictive capabilities in differentiating between cholesteatoma and non-cholesteatomatous cases. On both datasets, the model managed to correctly identify whether a case involved cholesteatoma in approximately four out of five instances (with accuracies of 78.3% and 81.3%). Generalizability was further supported by the comparable AUROC scores of 0.85 and 0.83 on the internal and the external dataset, respectively (Table 3).

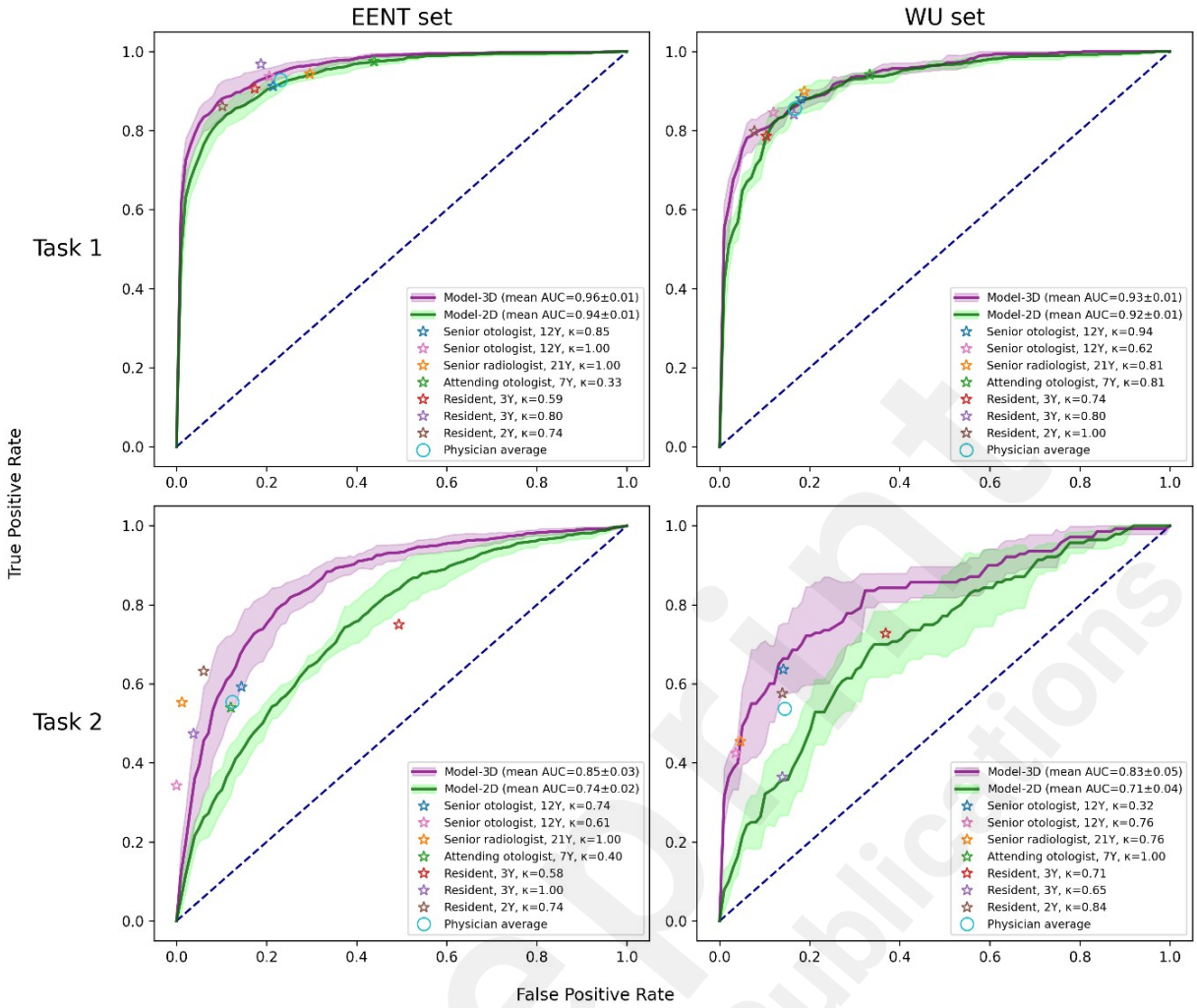


Figure 4. ROC plots for the benchmark tests. The curve and the shaded area indicate the mean and ±1 standard deviation of a model, respectively. Clinical experts are marked by colored asterisks for individual performance and by an open circle for averaged performance. The dotted diagonal line represents a random classifier.

Ablation study

This model exhibited a reasonable balance between predictability and efficiency (Table 4). Compared to Models 1 and 3, it achieved significantly better performance in both tasks (P<.01). Additionally, despite having approximately 60% fewer parameters, the proposed model demonstrated equivalent performance to Model 2 in both tasks (P =.26 and .91, respectively), indicating its enhanced computational efficiency.

Table 4. Ablation study on the 3D classification network.

Task	Model	Size (MB)	Accuracy	Recall	Specificity	Precision	F1 Score	AUROC	*P*
1	Proposed	14.2	87.8 ± 1.7%	85.3 ± 3.2%	91.3 ± 6.7%	93.3 ± 4.5%	89.0 ± 1.2%	0.959 ± 0.011	
	Model 1	4.0	85.8 ± 3.0%	82.7 ± 4.6%	90.1 ± 5.8%	92.1 ± 4.3%	87.0 ± 2.8%	0.947 ± 0.019	<.001
	Model 2	34.5	88.4 ± 1.4%	86.2 ± 2.1%	91.4 ± 4.1%	93.3 ± 3.0%	89.5 ± 1.2%	0.961 ± 0.009	.26
	Model 3	64.8	86.4 ± 2.2%	85.1 ± 6.2%	88.7 ± 7.4%	91.4 ± 5.3%	87.8 ± 1.9%	0.950 ± 0.019	.003
2	Proposed	14.2	78.3 ± 4.0%	80.8 ± 2.5%	77.0 ± 5.4%	65.2 ± 6.0%	72.1 ± 4.2%	0.853 ± 0.030	
	Model 1	4.0	75.8 ± 4.8%	71.2 ± 11.8%	78.3 ± 6.5%	63.6 ± 6.4%	66.8 ± 7.5%	0.817 ± 0.060	.006
	Model 2	34.5	78.2 ± 3.6%	79.5 ± 7.1%	77.5 ± 7.4%	65.9 ± 7.1%	71.6 ± 3.2%	0.862 ± 0.031	.91
	Model 3	64.8	75.6 ± 5.6%	76.0 ± 5.9%	75.4 ± 10.9%	63.4 ± 8.8%	68.5 ± 3.7%	0.826 ± 0.047	.003

Benchmarks

Compared to the 2D approach, the 3D network demonstrated significantly superior performance in both tasks across datasets (*P*<=.05). In particular, the proposed model exhibited a substantial performance gain in differentiating between cholesteatoma and non-cholesteatomata, with an increase of over 10% in all outcome metrics on both datasets (Table 3).

This model also matched or even surpassed the diagnostic capabilities of human experts in both tasks (Figure 4). It exhibited marginally superior performance compared to human eyes in the first task (*P*=.05) and significantly outperformed them in the visually challenging task 2 (*P*<.001). Post-hoc pairwise comparisons revealed that the model excelled over the attending otologist in task 1 and two resident fellows in task 2, rivaling all senior clinicians (Table 5). Similar results were shown across the breakdown of data sources, with a notable finding that the model outperformed a senior otologist in task 2 on the EENT subset (Table S4 in Multimedia Appendix 1). Moreover, the proposed model demonstrated perfect consistency, surpassing all human experts who exhibited higher standard deviations in all outcome metrics and lower scores of intra-rater reliability.

Table 5. Benchmark performance against human experts.

Task	Rater	Accuracy	Recall	Specificity	Precision	F1 Score	Kappa	*P*
1	The 3D model, mean ± SD	87.8 ± 1.7%	85.3 ± 3.2%	91.3 ± 6.7%	93.3 ± 4.5%	89.0 ± 1.2%	1.00 ± 0.00	
	Expert average, mean ± SD	85.7 ± 2.2%	89.8 ± 4.2%	80.4 ± 9.4%	86.0 ± 5.0%	87.6 ± 1.3%	0.82 ± 0.09	.05
	Senior otologist A - 12Y	87.3%	89.8%	84.1%	87.9%	88.8%	0.75	.79
	Senior otologist B - 12Y	85.7%	89.9%	80.4%	85.6%	87.7%	0.92	.49
	Senior radiologist - 21Y	85.4%	92.4%	76.3%	83.4%	87.7%	0.87	.37
	Attending otologist - 7Y	81.1%	96.0%	61.9%	76.5%	85.2%	0.73	.002
	Resident A - 3Y	85.9%	85.5%	86.4%	89.1%	87.2%	0.71	.56
	Resident B - 3Y	87.4%	91.2%	82.5%	87.1%	89.1%	0.81	.74
	Resident C - 2Y	86.8%	83.5%	91.2%	92.4%	87.7%	0.92	.96
2	The 3D model, mean ± SD	84.3 ± 1.5%	75.6 ± 4.7%	93.4 ± 2.1%	92.4 ± 1.8%	83.0 ± 2.2%	1.00 ± 0.00	
	Expert average, mean ± SD	74.1 ± 5.2%	54.9 ± 12.3%	86.5 ± 13.9%	77.2 ± 13.5%	62.2 ± 6.1%	0.72 ± 0.12	<.001
	Senior otologist A - 12Y	73.8%	36.7%	98.2%	93.0%	52.6%	0.70	.07
	Senior otologist B - 12Y	75.8%	60.6%	85.7%	73.3%	66.3%	0.47	.25
	Senior radiologist - 21Y	79.5%	52.3%	97.0%	91.9%	66.7%	0.86	.82
	Attending otologist - 7Y	74.5%	55.0%	87.0%	73.2%	62.8%	0.74	.11
	Resident A - 3Y	63.8%	74.3%	56.9%	52.9%	61.8%	0.67	<.001
	Resident B - 3Y	72.3%	44.0%	90.9%	76.2%	55.8%	0.77	.02
	Resident C - 2Y	78.8%	61.5%	89.9%	79.8%	69.4%	0.80	.96

Visual assessment of heatmaps

Heatmaps from both models consistently highlighted the tympanic cavity and mastoid that manifested pathological findings characteristic of the target condition (Figure 5). Specifically, the first model generated a hot signal indicative of soft tissue density in an affected middle ear (Figure 5A), while the signal remained subdued in a normal ear (Figure 5B). Similarly, the second model

revealed a distinct hotspot in a cholesteatomatous ear exhibiting the classic patterns of tympanum widening and ossicular destruction [17,32,33] (Figure 5C). In contrast, a case of CSOM showing intact ossicles surrounded by soft tissue shadows in a normal-sized tympanic cavity did not exhibit a corresponding hotspot (Figure 5D). These observations reflect that the AI's decision-making strategy aligns reasonably well with established human knowledge for both tasks.

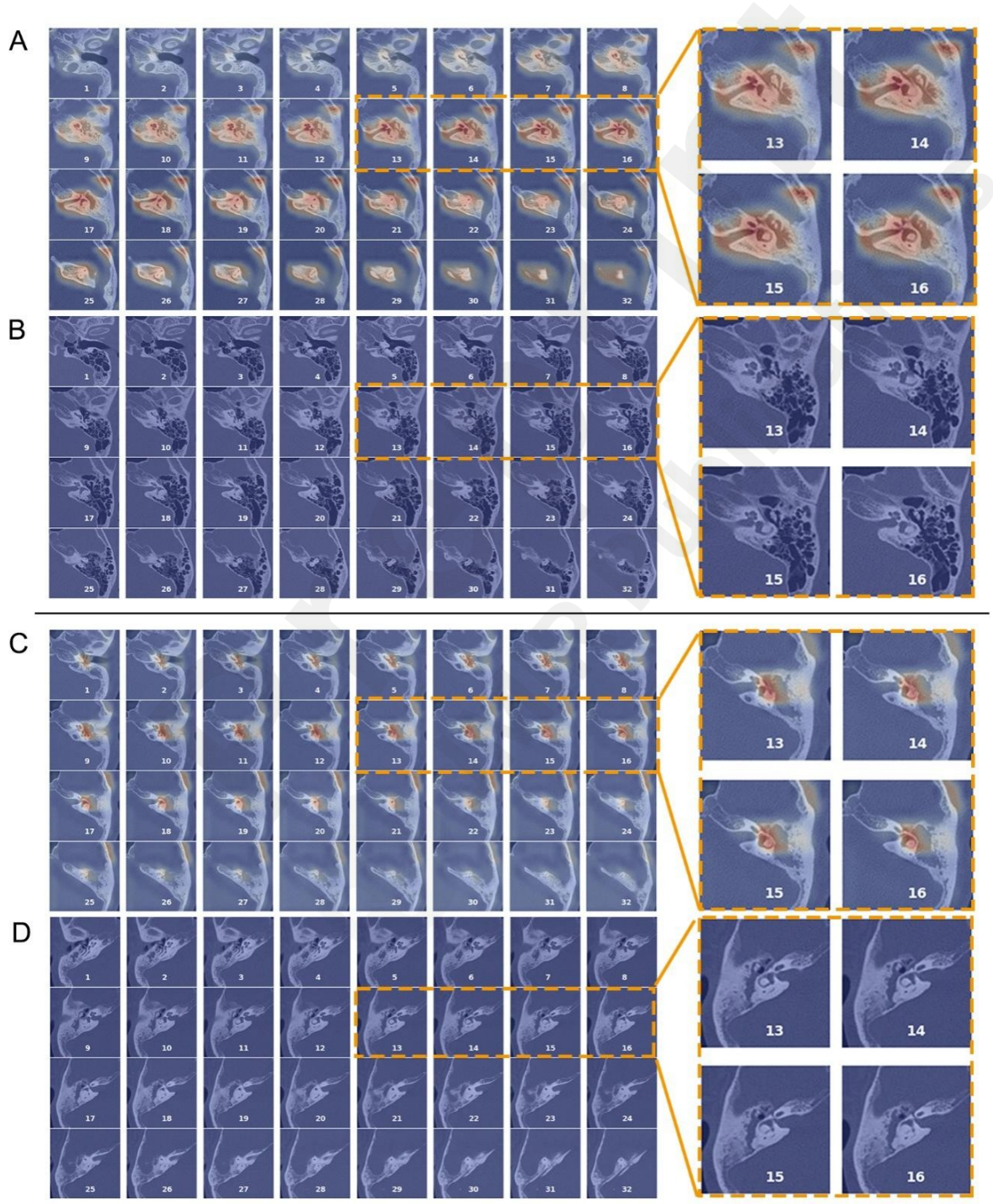


Figure 5. Examples of heatmaps. The heatmaps, generated in 3D fashion, are superimposed on the original CT and flattened to a series of 2D images for demonstration purpose. (A-B) A pathological and a normal ear, respectively. (C-D) A cholesteatoma and a non-cholesteatoma case, respectively. Area marked by hot signals indicate the presence of graphic patterns contributing to a “positive” prediction (i.e., a pathological ear in task 1 and a cholesteatoma in task 2).

Clinical utilization

The automatic evaluation system, incorporating the validated 3D model and the heatmap visualization technique, was evaluated for its viability in aiding preoperative assessment in 121 patients with COM (mean age 46.8 ± 16.1 years, 40.5% male). This system achieved an overall accuracy of 81.8% in distinguishing between cholesteatoma and non-cholesteatoma cases. Sixty-nine ears were identified as free of cholesteatoma by the model, all of which received minimally invasive tympanoplasty under endoscopy. During the procedure, 9 ears (13.0%) revealed signs of cholesteatoma, and 5 of them required additional bone grinding technique for complete removal of the mass. Cholesteatoma was initially predicted in 52 ears, with 37 (71.2%) of them undergoing canal-wall-down mastoidectomy. In the remaining 15 ears, the treating surgeons opted for endoscopic tympanoplasty, overriding the conventional technique for the model's predicted diagnosis. Clinicians reported that the model predictions aligned with their initial judgement or helped with their decision-making in 109 cases (90.1%). Postoperative hearing results were obtained in 106 patients (87.6%) who maintained follow-up. Both groups of ears showed normal recovery, with a mean hearing gain of 8.5 ± 15.6 and 5.5 ± 18.1 dB, respectively (Figure 6).

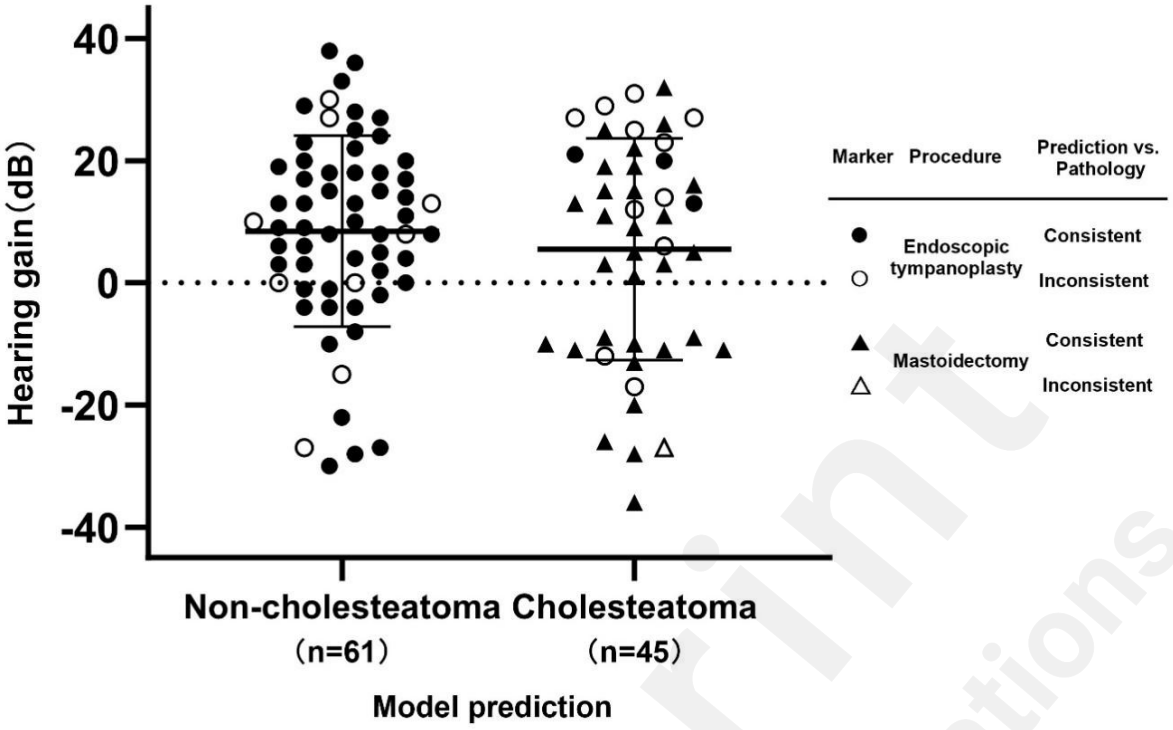


Figure 6. Postoperative hearing gain for the operated ears with available audiometry outcomes (n=106). Data are categorized according to model predictions. Predictions that agree with the pathological results are denoted by close symbols, while open symbols indicate disagreements. Circles and triangles represent the treatment of endoscopic tympanoplasty and mastoidectomy, respectively. The error bars indicate ±1 standard deviation from the mean.

Discussion

Principal results

This study demonstrates the robustness and generalizability of an AI model based on 3D CNN for the detection and differential diagnosis of COM using temporal bone CT scans. This model leverages multidimensional diagnostic information from the middle ear, resulting in a significant performance improvement compared to the traditional 2D approach. The framework exhibits comparable or even superior performances to human experts in otologic tasks with clinical significance and visual challenges, especially for classifying between cholesteatoma and non-cholesteatomatous cases. Additionally, the novel heatmap technique allows inspection of the AI's logic for decision-making, thereby enhancing the transparency of this model. The resulting AI

system serves to automate summarization of critical radiologic findings and enables efficient evaluation of COM with minimum manual input. It provides tangible benefit in assisting otologists during preoperative assessment and results in favorable clinical outcomes that are comparable to historical results [34-37]. These findings further support the clinical viability and advantages of AI technology, which is expected to improve efficiency, reduce errors, and facilitate precision medicine in healthcare in the new era of big data.

Comparison with prior work

A few AI models have recently been developed to classify common middle ear conditions, such as CSOM, otitis media with effusion, and cholesteatoma [38-41]. However, these models were primarily based on traditional otoscopic images, which are potentially limited by a narrow field of view and insufficient diagnostic information. Temporal bone CT scans, which are increasingly used in otologic workup by virtue of its accessibility, rich amount of anatomical information and adequate sensitivity in revealing pathological changes, have also been explored in a limited number of studies [22,42-45]. Although these AI models demonstrated decent AUROC scores (e.g., >0.9) in common classification tasks, they were all trained to generate predictions based on 2D single-layer CT images. A potential drawback is the increased likelihood of missing small or peripheral pathological changes (e.g., an attic cholesteatoma) and the resultant false negatives.

Efforts were made in this study to establish a 3D approach to take full advantage of all available anatomical information and achieve a better coverage of the tympanum and the mastoid. Inspection of the extracted ROI suggests all critical anatomies are visible. Results from the benchmark test indicate that the proposed 3D model outperforms the state-of-the-art 2D approach by a modest performance gain in the detection of COM, and by a much larger extent in differentiating between cholesteatoma and non-cholesteatoma. This finding has several implications. First, both models are

generally adequate in identifying common abnormal patterns from the CT, which are graphically characterized by increased opacification or soft tissue shadows in the middle ear cavity and indicative of pathological conditions in general. This is a relatively simple visual task, during which diagnostic information obtained from a single 2D CT slice is likely sufficient for the purpose and extra findings from other layers only provide minimal contribution to the decision-making. Second, the 3D model has huge advantage over the 2D approach in differentiating cholesteatoma from other types of COM. This task is known to be more visually challenging for humans, often requiring detection of subtle osseous erosions from multiple CT slices, as quite a few pathological changes caused by cholesteatoma are peripheral or non-characteristic [32,33] . A substantial increase in each outcome measure justifies the advantage of the current 3D model for this task. Moreover, this 3D model only has a simple network structure with a small size (14.5 MB) as opposed to a complex and large-sized 2D network (274 MB), suggesting both higher computational efficiency and performance of the 3D approach. Finally, the AUROC of 0.92-0.94 and accuracy scores of 82.1-86.1% achieved by the 2D network in this study in detecting COM were equivalent to historical results (0.92 and 86%, respectively) in our previous study [22], further indicating the reliability of these findings and potentially the intrinsic limit of using single-layer CT image for this task. To the best of our knowledge, this is the first study showing quantitative evidence to support the advantage of a 3D CNN model in two common otologic tasks based on temporal bone CT scans. It also advances beyond prior retrospective research by showcasing the practicality and benefits of the model in a clinical environment.

Clinical implications

Cholesteatoma exhibits distinct histology marked by local invasiveness and a propensity for recurrence. The imperative for successful outcomes necessitates complete removal of the mass,

particularly because recurrent cholesteatoma complicates revision surgery [46]. Suspected cases often require a canal-wall-down mastoidectomy to expose the tympanum, resulting in an open cavity and a permanently altered sound conduction pathway [46]. Accumulating evidence suggests non-cholesteatoma may spare from mastoidectomy and benefit from minimally invasive procedures like endoscopic tympanoplasty [47,48]. Therefore, the current AI system holds potential value for otologists in surgical planning. Ears with a low risk of cholesteatoma, as identified by the model, could potentially be treated by less invasive procedures that retain the integrity of canal wall, leading to reduced procedural time and enhanced recovery [6,7,9,49,50]. This clinical merit is supported by the superior benchmark performance in identifying cholesteatoma and the favorable outcomes observed in the prospective study.

While detecting COM in task 1 involves spotting any pathological patterns on CT, which may not fully capture the differences between models in diagnostic capabilities, the increased visual challenges in identifying cholesteatoma substantiate the advantages of the proposed 3D approach for this task. In this study, the 3D model outperformed junior clinicians and demonstrated equivalent or superior performance to senior experts in identifying cholesteatoma based on CT. Notably, the 3D model achieved outcomes that were on par with or better than those based on human interpretation of MRI, which, despite its higher sensitivity, is a more expensive diagnostic method [22,43,45,51-53]. These findings underscore the 3D model's potential as a reliable and cost-effective alternative, offering sufficient COM evaluation with CT imaging alone, thereby reducing the need for the pricier MRI.

The findings from the prospective study indicate that the model is efficacious in clinical environments, especially in distinguishing cholesteatoma from non-cholesteatoma. Feedback from our clinical team highlights that the system serves as a reliable and streamlined source for a second opinion. Before surgery, the treating physician can rapidly identify essential details like the lesion's location and properties, using the model's diagnostic output and heatmaps. Concordance between the

model's predictions and the physician's initial assessment bolsters confidence in surgical planning, thereby streamlining the diagnostic and therapeutic process. In contrast, discrepancies between the model's results and the physician's judgment prompt a detailed case reassessment or team consultation, aiding in the validation of a suitable treatment plan or preparing for intraoperative modifications. This process provides timely advisory support for complex cases, encouraging meticulous evaluation by the physician, minimizing errors, and keeping the clinician's cognitive load in check without compromising their autonomy in decision-making.

It should be noted that even for seasoned otologists and radiologists, who are adept at quickly and accurately reading temporal bone CT scans, a second opinion can add an extra layer of confidence to their assessments. For novice clinicians, who may find the diagnostic process more challenging and time-intensive (Table 5), the model may offer substantial improvements in both the accuracy and speed of diagnosing and managing COM. This is particularly beneficial for physicians in smaller medical facilities or those early in their careers. Looking ahead, the integration of this model into electronic medical systems or cloud-based servers stands to streamline the provision of immediate second opinions or enable physicians from diverse locations to upload imaging data for dependable diagnostic insights. Such technological progress is poised to advance individualized COM treatments in the big data era, boosting efficiency, reducing costs, and enhancing the quality of patient care.

Research insights

Efforts were undertaken in this study to demystify the criticized non-transparency of DL models, characterized by intricate decision-making strategies within multilayer architectures [30,54]. The nonlinear interactions among these components can yield incomprehensible logic and untraceable predictions vulnerable to bias or errors, posing a significant challenge to the widespread application

of AI in healthcare. To address this issue, heatmaps, and specifically, the Grad-CAM technique, have been employed as a method to inspect AI's strategy and enhance human interpretation in a parsimonious manner [55-57]. In this study, the strategy learned by our models to focus on the middle ear and mastoid regions appeared reasonable and aligned with human knowledge in interpreting CT for COM, reinforcing the reliability of this framework. These informative heatmaps can aid clinicians in understanding and validating AI predictions for specific cases, or serve as educational tools for training medical students or junior residents in reading temporal bone CT scans. Ultimately, this approach presents a viable solution for developing explainable AI models for clinical tasks.

Overfitting is a common concern with DL models, especially when data are limited or sourced from a single institute. It can lead to poor performance on new data despite promising results on the original dataset. Previous DL models were trained on monocentric CT images with participant counts ranging from 61 to 562. Lack of external validation and small sample sizes may raise concern about potential overfitting of these models [22,42,43]. Several approaches were employed in this study to enhance the generalizability of our framework. First, our models underwent cross-validation on a major dataset comprising over 3000 ears, the largest sample size reported to date. Second, these models were evaluated on external data with different patient origins and image properties. Third, several machine learning methods were applied to minimize the risk of models being tuned to the random features, including early termination of training and the use of a dropout function to decrease the interdependency among network nodes [27]. Consistent performance metrics across datasets in both tasks substantiated the generalizability of this framework. Moreover, the region proposal method proved applicable to CT images from both sources, demonstrating adaptability despite differences in CT scanner, scan settings, and image quality.

Limitations

This study has several limitations. First, although an external dataset was obtained from a hospital in a different city, patients in both datasets shared a common racial background. Further validation on data collected from patients with diverse origins may be necessary to ensure the generalizability of these models. Second, the research was constrained to two binary classification tasks relevant to COM. Incorporating additional diagnostic tasks, such as assessing the ossicular chain's integrity and forecasting auditory outcomes, may enrich the diagnostic toolkit. Third, the models were exclusively trained to analyze CT scans, potentially not leveraging AI's full potential in COM evaluation. Comprehensive diagnostics often involve synthesizing information from patient history, clinical symptoms, ear examinations, audiological testing, otoscopy, and various imaging techniques. Over-reliance on CT scans alone may introduce limitations in performance and may not always lead to conclusive diagnoses (Figure 7). Fourth, the ablation study examined a limited array of model alternatives. Despite achieving notable performance through initial model structure refinement, future endeavors should include ongoing optimization of the model architecture and detailed analysis of network component functions to optimize the trade-off between model efficacy and computational demands. Additionally, this study did not place extensive emphasis on exploring common ethical issues, such as patient privacy, data security, and human autonomy, which are critical considerations in the clinical application of AI and warrant ongoing attention. Finally, this study reported initial findings from the clinical application of the AI system in a small, prospective cohort without a control group. Although the main objective was to show that the current model is ready for clinical implementation, a thorough assessment of the model's clinical benefits will be conducted in an upcoming clinical trial with a more rigorous research design.

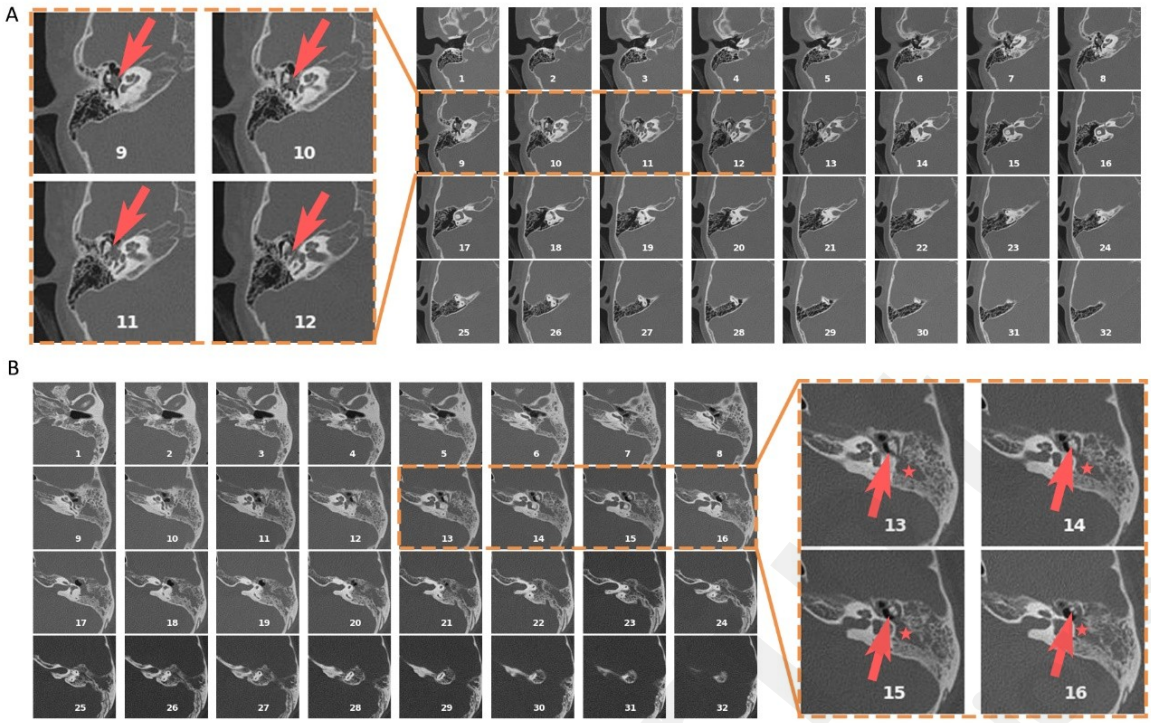


Figure 7. Examples of misclassified cases. (A) A pathological ear showing a small-size soft tissue density near the ossicles (arrows) with no evident sign of osseous erosion or mastoid opacification. (B) A case of cholesteatoma showing soft tissue density (asterisks) but with a visually intact ossicular chain (arrows) and a normal-size tympanic cavity.

Future research

Future studies will focus on leveraging novel techniques to enhance model performance and evaluate the effectiveness in larger-scale controlled trials. For example, new models will be trained to perform additional tasks, including evaluation of ossicular chain and forecasting postoperative hearing, which may enhance features of the current AI framework. A broader dataset will be compiled from hospitals worldwide to assess and refine the generalizability of these models. Moreover, future models will potentially incorporate multiple sources of clinical information with a fusion layer for generating predictions, mimicking human decision-making strategies and potentially enhancing model robustness. Ongoing efforts will also be made to refine model architectures and to address ethical issues associated with the use of AI in healthcare. An active learning framework may be established to integrate feedback loops, allowing clinicians to provide input to the model. This

approach is expected to support ongoing model enhancement and reinforcement learning based on human feedback. In the next stage, multicenter, prospective human trials will be conducted to assess the practical benefits of implementing these AI models in clinical contexts. The ultimate goal of this research line is to establish a robust AI system that can assist clinicians with reliability, efficiency and transparency in the evaluation and management of ear diseases.

Conclusions

This study presents a 3D CNN model trained to detect pathological changes and identify cholesteatoma based on temporal bone CT scans. The model's performance significantly surpasses the baseline 2D approach, reaching a level comparable to or even exceeding that of human experts in both tasks. The model also exhibits decent generalizability and enhanced comprehensibility through the gradient heatmaps. The resulting AI system allows automatic assessment of COM and shows promising viability in real-word clinical settings. These findings imply the potential of AI as a valuable tool for aiding clinicians in the evaluation of COM. Future research will involve enhancing models with additional source of diagnostic information to perform various clinical tasks and evaluating the benefits of AI models in large-scale controlled trials.

Acknowledgements

Author contributions

Yike Li conceptualized and designed the study, reviewed and analyzed the data, performed computer programming, developed and evaluated the AI models, wrote and edited the manuscript, prepared the figures, and supervised the entire project; Binjun Chen retrieved, validated and analyzed the data, drafted the manuscript, and prepared the figures; Yu Sun provided data resources and validated the data. Haojie Sun, Yanmei Wang, and Jihan Lyu retrieved data and evaluated the models. Xun Niu and Lian Yang retrieved and validated the data. Jiajie Guo acquired funding support, validated the model, and edited the manuscript. Shunxing Bao performed model validation and deployment. Yushu Cheng, Jianghong Xu, and Juanmei Yang evaluated the models. Yibo Huang and Bo Liang provided data resources; Fanglu Chi provided data resources and funding support; Dongdong Ren conceptualized the study, provided funding support and data resources, and edited the manuscript. All authors have reviewed, discussed, and approved the manuscript. No generative AI tool was utilized during the preparation of this manuscript.

Funding information

This study is supported by the National Natural Science Foundation of China (Grant Nos. 81970889 to Fanglu Chi and 82271166, 81970880 and 81771017 to Dongdong Ren, U22A20249, 52188102, 52027806 to Jiajie Guo); Natural Science Foundation of Shanghai (Grant No.22ZR1410100 to Dongdong Ren); the “Zhuo-Xue Plan” of Fudan University (Dongdong Ren); Heng-Jie special technical support plan (Dongdong Ren); and the Shanghai Outstanding Young Medical Talent Program (Dongdong Ren). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest

One of the authors (Yike Li) serves as an associate editor for the Journal of Medical Internet Research at the time of manuscript submission. Dr. Li has abstained from participating in any peer-review or editorial decision-making processes related to this article.

Data availability statements

The complete source code and trained models are available in the Multimedia Appendix section as well as in a public repository, which can be accessed at: https://github.com/huntlylee/3D-Otitis-Media. The automatic evaluation system is available for individual use with a detailed instruction manual and a walk-through tutorial. The datasets generated during this study may be obtained from the corresponding authors upon reasonable request.

Abbreviations

AI: Artificial intelligence

AUROC: Area under the receiver operating characteristic

CNN: Convolutional neural network

COM: Chronic otitis media

CSOM: Chronic suppurative otitis media

CT: Computed tomography

DL: Deep learning

Grad-CAM: Gradient-weighted Class Activation Mapping

MRI: Magnetic resonance imaging

ROI: Region of interest

Multimedia Appendices

Multimedia Appendix 1: Supplementary tables

Multimedia Appendix 2: Source code for model development and validation

References

1. Schilder AG, Chonmaitree T, Cripps AW, Rosenfeld RM, Casselbrant ML, Haggard MP, et al. Otitis media. Nat Rev Dis Primers 2016 Sep 8;2(1):16063
2. Organization WH. Chronic suppurative otitis media: burden of illness and management options. 2004
3. Lustig LR, Limb CJ, Baden R, LaSalvia MT. Chronic otitis media, cholesteatoma, and mastoiditis in adults. UpToDate Waltham, MA (citirano 145 2019) 2018
4. Takahashi M, Motegi M, Yamamoto K, Yamamoto Y, Kojima H. Endoscopic tympanoplasty type I using interlay technique. J Otolaryngol Head Neck Surg 2022 Nov 17;51(1):45
5. Ohki M, Kikuchi S, Tanaka S. Endoscopic Type 1 Tympanoplasty in Chronic Otitis Media: Comparative Study with a Postauricular Microscopic Approach. Otolaryngol Head Neck Surg 2019 Aug;161(2):315-323
6. Hsu YC, Kuo CL, Huang TC. A retrospective comparative study of endoscopic and microscopic Tympanoplasty. J Otolaryngol Head Neck Surg 2018 Jul 4;47(1):44
7. Yang Q, Wang B, Zhang J, Liu H, Xu M, Zhang W. Comparison of endoscopic and microscopic tympanoplasty in patients with chronic otitis media. Eur Arch Otorhinolaryngol 2022 Oct;279(10):4801-4807
8. Tsetsos N, Vlachtsis K, Stavrakas M, Fyrmpas G. Endoscopic versus microscopic ossiculoplasty in chronic otitis media: a systematic review of the literature. Eur Arch Otorhinolaryngol 2021 Apr;278(4):917-923
9. Tarabichi M, Ayache S, Nogueira JF, Al Qahtani M, Pothier DD. Endoscopic management of chronic otitis media and tympanoplasty. Otolaryngol Clin North Am 2013 Apr;46(2):155-163
10. Watts S, Flood LM, Clifford K. A systematic approach to interpretation of computed

tomography scans prior to surgery of middle ear cholesteatoma. J Laryngol Otol 2000 Apr;114(4):248-253

11. Selwyn D, Howard J, Cuddihy P. Pre-operative prediction of cholesteatomas from radiology: retrospective cohort study of 106 cases. J Laryngol Otol 2019 Jun;133(6):477-481
12. Songu M, Altay C, Onal K, Arslanoglu S, Balci MK, Ucar M, et al. Correlation of computed tomography, echo-planar diffusion-weighted magnetic resonance imaging and surgical outcomes in middle ear cholesteatoma. Acta Otolaryngol 2015 Aug;135(8):776-780
13. Mahmutoğlu AS, Celebi I, Sahinoğlu S, Cakmakçi E, Sözen E. Reliability of preoperative multidetector computed tomography scan in patients with chronic otitis media. J Craniofac Surg 2013 Jul;24(4):1472-1476
14. Pandey AK, Bapuraj JR, Gupta AK, Khandelwal N. Is there a role for virtual otoscopy in the preoperative assessment of the ossicular chain in chronic suppurative otitis media? Comparison of HRCT and virtual otoscopy with surgical findings. Eur Radiol 2009 Jun;19(6):1408-1416
15. Chee NW, Tan TY. The value of pre-operative high resolution CT scans in cholesteatoma surgery. Singapore Med J 2001 Apr;42(4):155-159
16. Uz Zaman S, Rangankar V, Muralinath K, Shah V, K G, Pawar R. Temporal Bone Cholesteatoma: Typical Findings and Evaluation of Diagnostic Utility on High Resolution Computed Tomography. Cureus 2022 Mar;14(3):e22730
17. Gaurano JL, Joharjy IA. Middle ear cholesteatoma: characteristic CT findings in 64 patients. Ann Saudi Med 2004 Nov-Dec;24(6):442-447
18. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019 Jun;25(6):954-961
19. Mikhael PG, Wohlwend J, Yala A, Karstens L, Xiang J, Takigami AK, et al. Sybil: A

Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography. J Clin Oncol 2023 Jan 12:JCO2201345

20. Yamashita R, Long J, Longacre T, Peng L, Berry G, Martin B, et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. Lancet Oncol 2021 Jan;22(1):132-141
21. Li Y, Guo J, Yang P. Developing an Image-Based Deep Learning Framework for Automatic Scoring of the Pentagon Drawing Test. J Alzheimers Dis 2022 85(1):129-139
22. Wang YM, Li Y, Cheng YS, He ZY, Yang JM, Xu JH, et al. Deep Learning in Automated Region Proposal and Diagnosis of Chronic Otitis Media Based on Computed Tomography. Ear Hear 2020 May/Jun;41(3):669-677
23. Sundgaard JV, Harte J, Bray P, Laugesen S, Kamide Y, Tanaka C, et al. Deep metric learning for otitis media classification. Med Image Anal 2021 Jul;71:102034
24. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: beyond the black box. BMJ 2019 Mar 12;364:l886
25. Castelvecchi D. Can we open the black box of AI. Nature 2016 Oct 6;538(7623):20-23
26. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition 2016 :779-788
27. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research 2014 15(1):1929-1958
28. DP Kingma JB. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 2014
29. Tseng CC, Lai MT, Wu CC, Yuan SP, Ding YF. Comparison of the efficacy of endoscopic

tympanoplasty and microscopic tympanoplasty: A systematic review and meta-analysis. Laryngoscope 2017 Aug;127(8):1890-1896

30. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision 2017 :618-626
31. Team PC. Python: A dynamic, open source programming language ,2022.
32. Baráth K, Huber AM, Stämpfli P, Varga Z, Kollias S. Neuroradiology of cholesteatomas. AJNR Am J Neuroradiol 2011 Feb;32(2):221-229
33. Gulati M, Gupta S, Prakash A, Garg A, Dixit R. HRCT imaging of acquired cholesteatoma: a pictorial review. Insights Imaging 2019 Oct 3;10(1):92
34. Daneshi A, Daneshvar A, Asghari A, Farhadi M, Mohebbi S, Mohseni M, et al. Endoscopic Versus Microscopic Cartilage Myringoplasty in Chronic Otitis Media. Iran J Otorhinolaryngol 2020 Sep;32(112):263-269
35. Prasad SC, La Melia C, Medina M, Vincenti V, Bacciu A, Bacciu S, et al. Long-term surgical and functional outcomes of the intact canal wall technique for middle ear cholesteatoma in the paediatric population. Acta Otorhinolaryngol Ital 2014 Oct;34(5):354-361
36. Wood CB, O'Connell BP, Lowery AC, Bennett ML, Wanna GB. Hearing Outcomes Following Type 3 Tympanoplasty With Stapes Columella Grafting in Canal Wall Down Mastoidectomy. Ann Otol Rhinol Laryngol 2019 Aug;128(8):736-741
37. Chamoli P, Singh CV, Radia S, Shah AK. Functional and Anatomical Outcome of Inside Out Technique For Cholesteatoma Surgery. Am J Otolaryngol 2018 Jul-Aug;39(4):423-430
38. Wu Z, Lin Z, Li L, Pan H, Chen G, Fu Y, et al. Deep Learning for Classification of Pediatric Otitis Media. Laryngoscope 2021 Jul;131(7):E2344-E2351
39. Pichichero ME. Can machine learning and AI replace otoscopy for diagnosis of otitis media. PEDIATRICS 2021 147(4)

40. Tseng CC, Lim V, Jyung RW. Use of artificial intelligence for the diagnosis of cholesteatoma. Laryngoscope Investig Otolaryngol 2023 Feb;8(1):201-211

41. Livingstone D, Chau J. Otoscopic diagnosis using computer vision: An automated machine learning approach. Laryngoscope 2020 Jun;130(6):1408-1413

42. Duan B, Guo Z, Pan L, Xu Z, Chen W. Temporal bone CT-based deep learning models for differential diagnosis of primary ciliary dyskinesia related otitis media and simple otitis media with effusion. Am J Transl Res 2022 14(7):4728-4735

43. Eroğlu O, Eroğlu Y, Yıldırım M, Karlıdag T, Çınar A, Akyiğit A, et al. Is it useful to use computerized tomography image-based artificial intelligence modelling in the differential diagnosis of chronic otitis media with and without cholesteatoma. Am J Otolaryngol 2022 May-Jun;43(3):103395

44. Khosravi M, Jabbari Moghaddam Y, Esmaeili M, Keshtkar A, Jalili J, Tayefi Nasrabadi H. Classification of mastoid air cells by CT scan images using deep learning method. Journal of Big Data 2022 9(1):1-14

45. Wang Z, Song J, Su R, Hou M, Qi M, Zhang J, et al. Structure-aware deep learning for chronic middle ear disease. EXPERT SYSTEMS WITH APPLICATIONS 2022 194:116519

46. Tomlin J, Chang D, McCutcheon B, Harris J. Surgical technique and recurrence in cholesteatoma: a meta-analysis. Audiol Neurootol 2013 18(3):135-142

47. Lee SY, Lee DY, Seo Y, Kim YH. Can Endoscopic Tympanoplasty Be a Good Alternative to Microscopic Tympanoplasty? A Systematic Review and Meta-Analysis. Clin Exp Otorhinolaryngol 2019 May;12(2):145-155

48. Trinidade A, Page JC, Dornhoffer JL. Therapeutic Mastoidectomy in the Management of Noncholesteatomatous Chronic Otitis Media: Literature Review and Cost Analysis. Otolaryngol Head Neck Surg 2016 Dec;155(6):914-922

49. Wu L, Liu Q, Gao B, Huang S, Yang N. Comparison of endoscopic and microscopic

management of attic cholesteatoma: A randomized controlled trial. Am J Otolaryngol 2022 May-Jun;43(3):103378

50. Toulouie S, Block-Wheeler NR, Rivero A. Postoperative Pain After Endoscopic vs Microscopic Otologic Surgery: A Systematic Review and Meta-analysis. Otolaryngol Head Neck Surg 2022 Jul;167(1):25-34

51. Profant M, Sláviková K, Kabátová Z, Slezák P, Waczulíková I. Predictive validity of MRI in detecting and following cholesteatoma. Eur Arch Otorhinolaryngol 2012 Mar;269(3):757-765

52. Lin M, Sha Y, Sheng Y, Chen W. Accuracy of 2D BLADE Turbo Gradient- and Spin-Echo Diffusion Weighted Imaging for the Diagnosis of Primary Middle Ear Cholesteatoma. Otol Neurotol 2022 Jul 1;43(6):e651-e657

53. Sharifian H, Taheri E, Borghei P, Shakiba M, Jalali AH, Roshanfekr M, et al. Diagnostic accuracy of non-echo-planar diffusion-weighted MRI versus other MRI sequences in cholesteatoma. J Med Imaging Radiat Oncol 2012 Aug;56(4):398-408

54. Montavon G, Lapuschkin S, Binder A, Samek W, Müller K. Explaining nonlinear classification decisions with deep taylor decomposition. PATTERN RECOGNITION 2017 65:211-222

55. Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V. A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. CHAOS SOLITONS & FRACTALS 2020 NOV 2020;140

56. Cheng C, Ho T, Lee T, Chang C, Chou C, Chen C, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. EUROPEAN RADIOLOGY 2019 OCT 2019;29(10):5469-5477

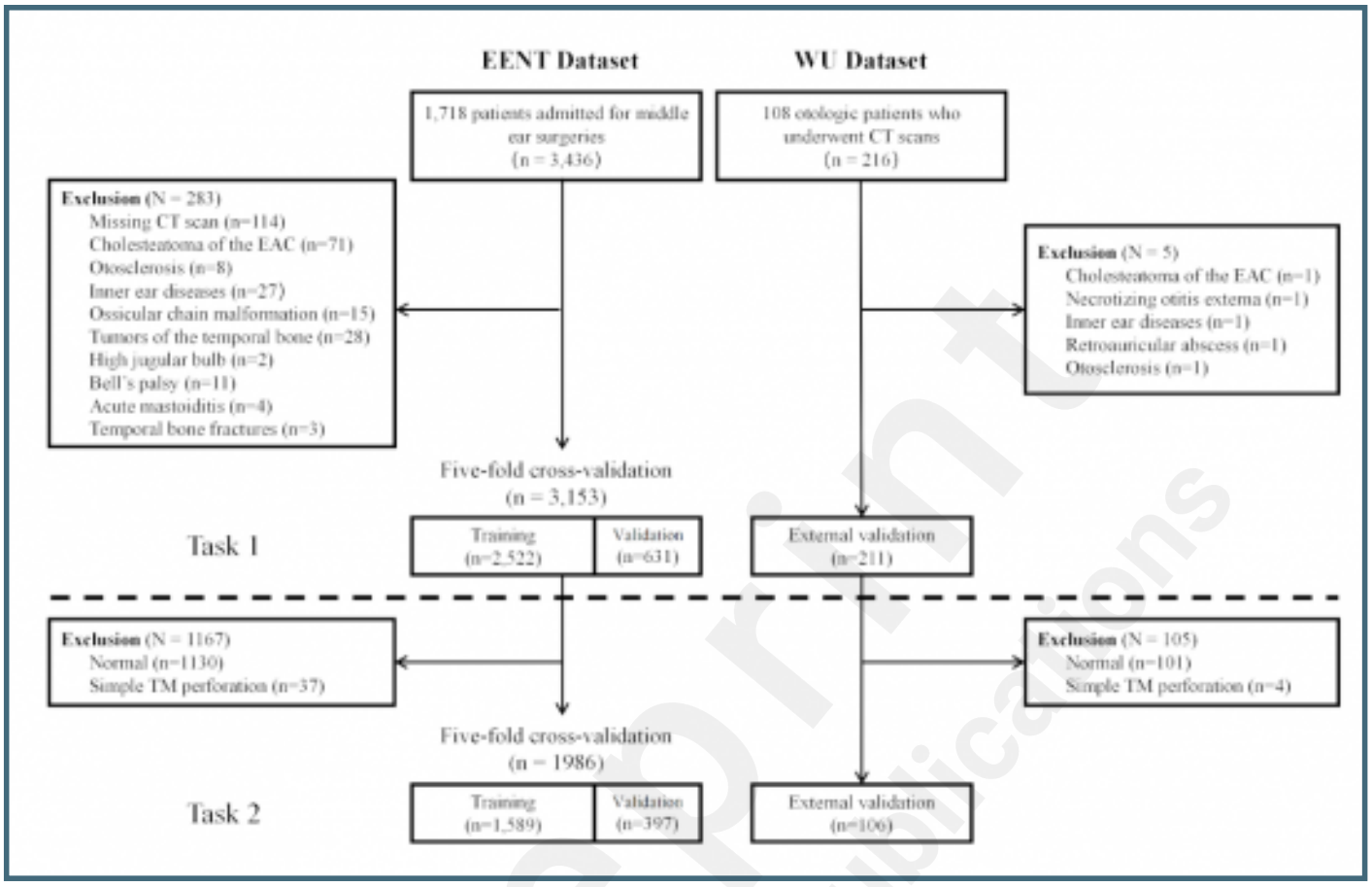
57. He T, Guo J, Chen N, Xu X, Wang Z, Fu K, et al. MediMLP: Using Grad-CAM to Extract Crucial Variables for Lung Cancer Postoperative Complication Prediction. IEEE J Biomed

Health Inform 2020 JUN 2020;24(6):1762-1771

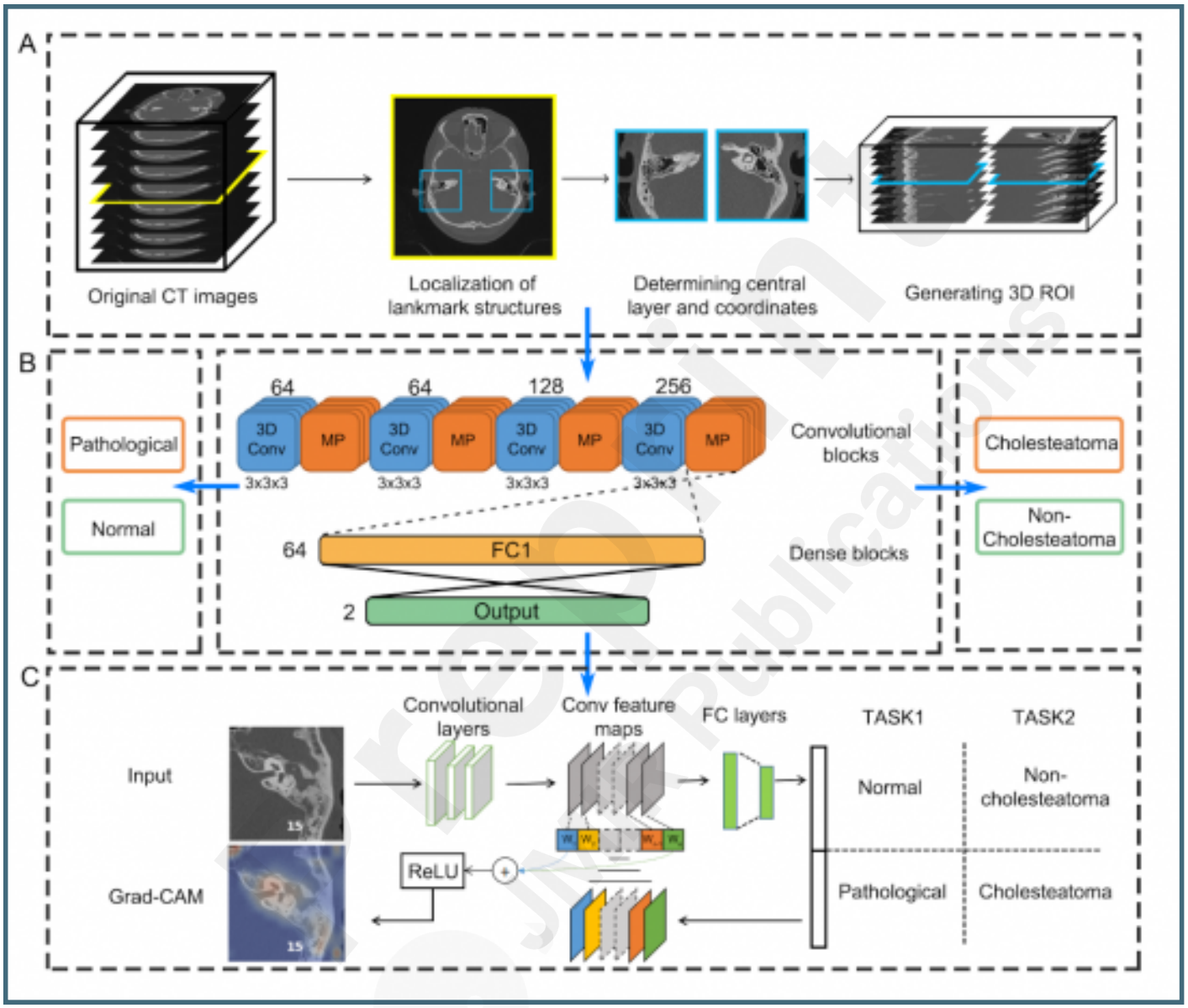
Supplementary Files

Figures

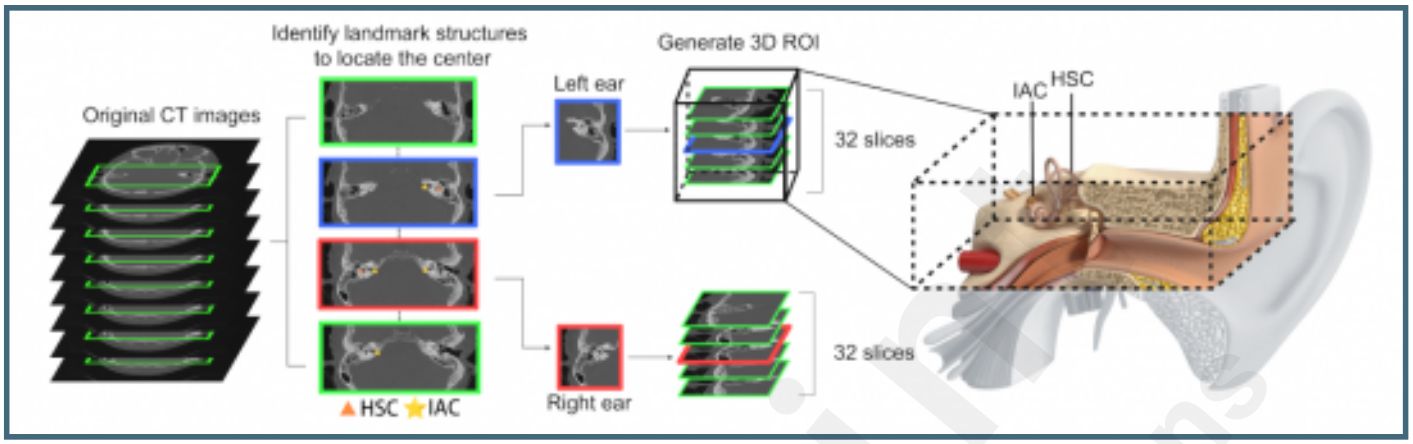
Flow chart of data retrieval. EAC: external auditory canal; TM: tympanic membrane.



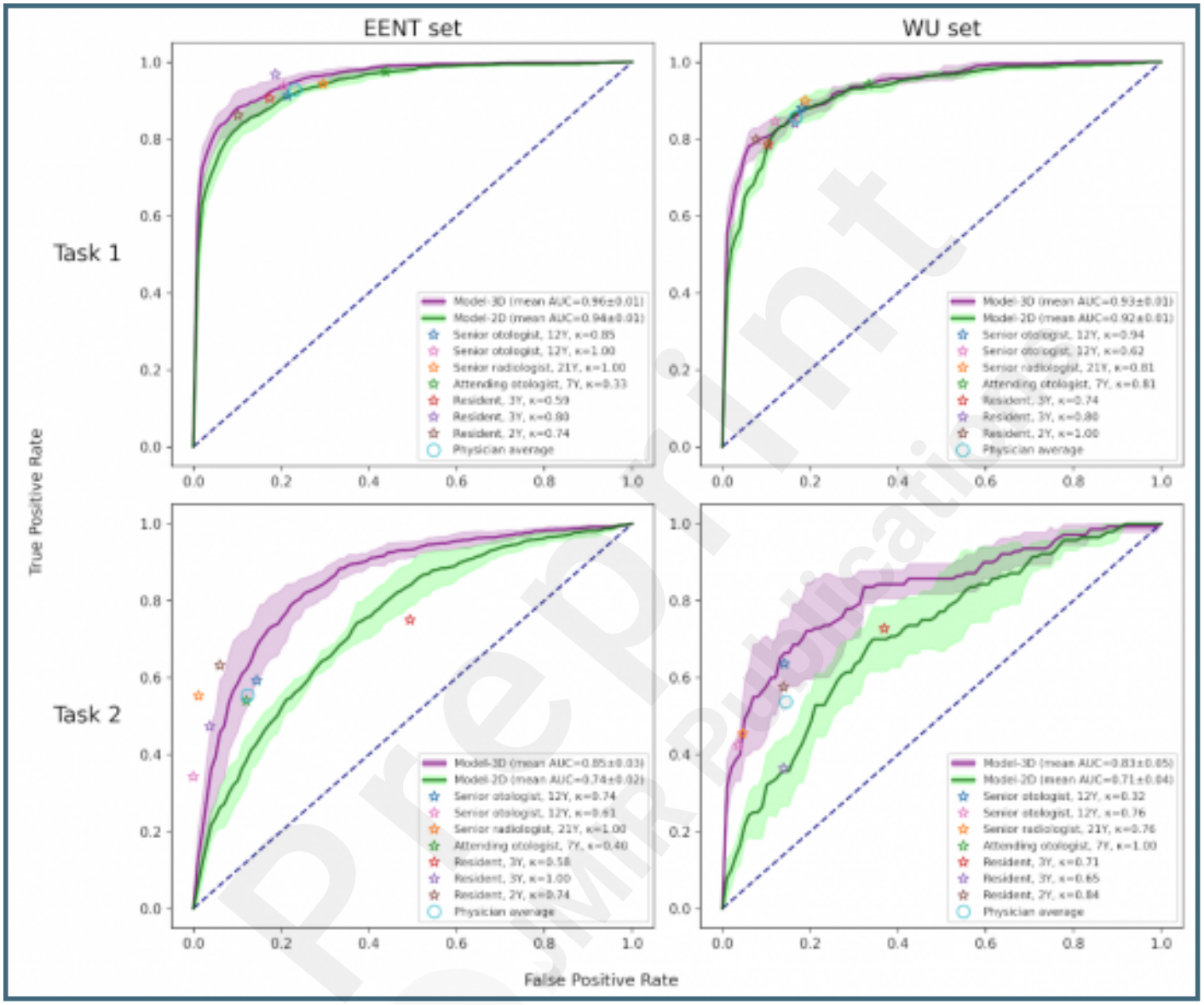
An overview of the AI framework. (A) The region proposal network used to locate landmark structures and segment the 3D ROI from the original CT images. (B) The classification network based on a 3D CNN architecture and trained to perform two classification tasks. (C) The gradient heatmaps generated to highlight the critical regions for decision-making. Conv: Convolution; FC: Fully connected; ReLu: Rectified Linear unit.



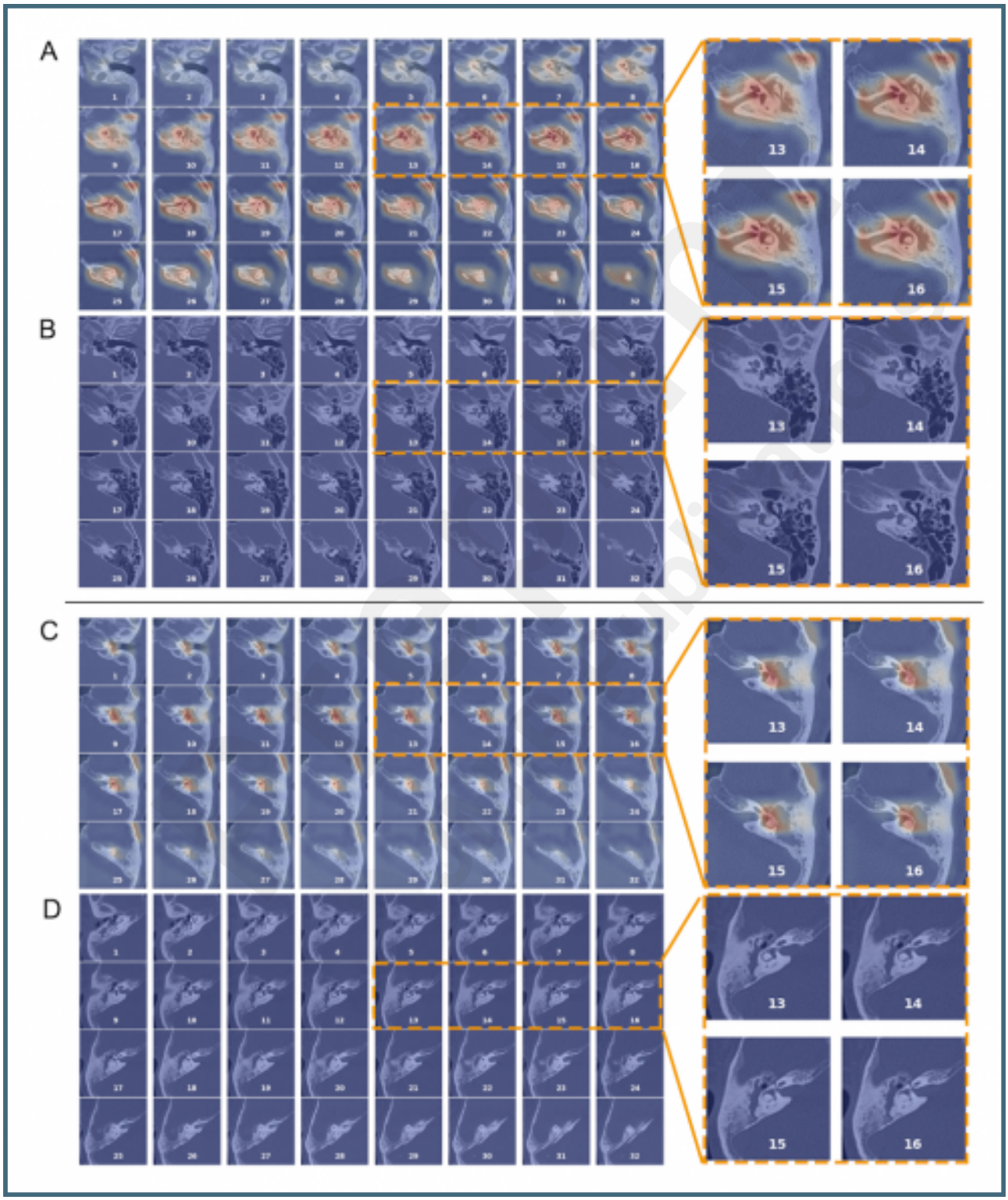
Generation of the 3D ROI. The region proposal network identifies landmark structures in each of the full-sized sequential CT slices and determines the center of the middle ear on each side. A 3D image comprising 32 stacks of axial slices in 150x150 pixels is subsequently segmented. This ROI encompasses an extensive range of critical anatomies within the temporal bone for the evaluation of COM. HSC: horizontal semicircular canal; IAC: internal auditory canal.



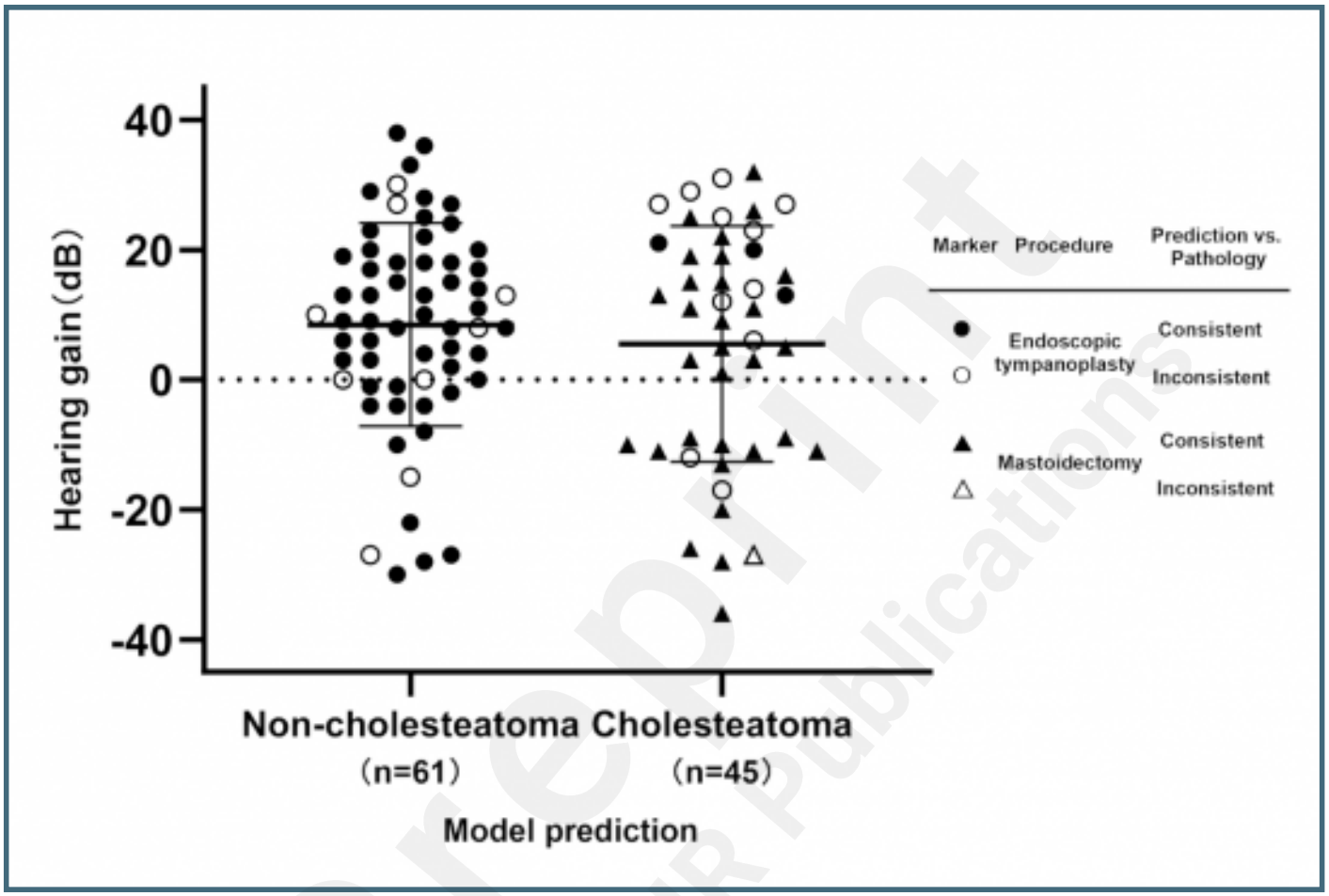
ROC plots for the benchmark tests. The curve and the shaded area indicate the mean and ±1 standard deviation of a model, respectively. Clinical experts are marked by colored asterisks for individual performance and by an open circle for averaged performance. The dotted diagonal line represents a random classifier.



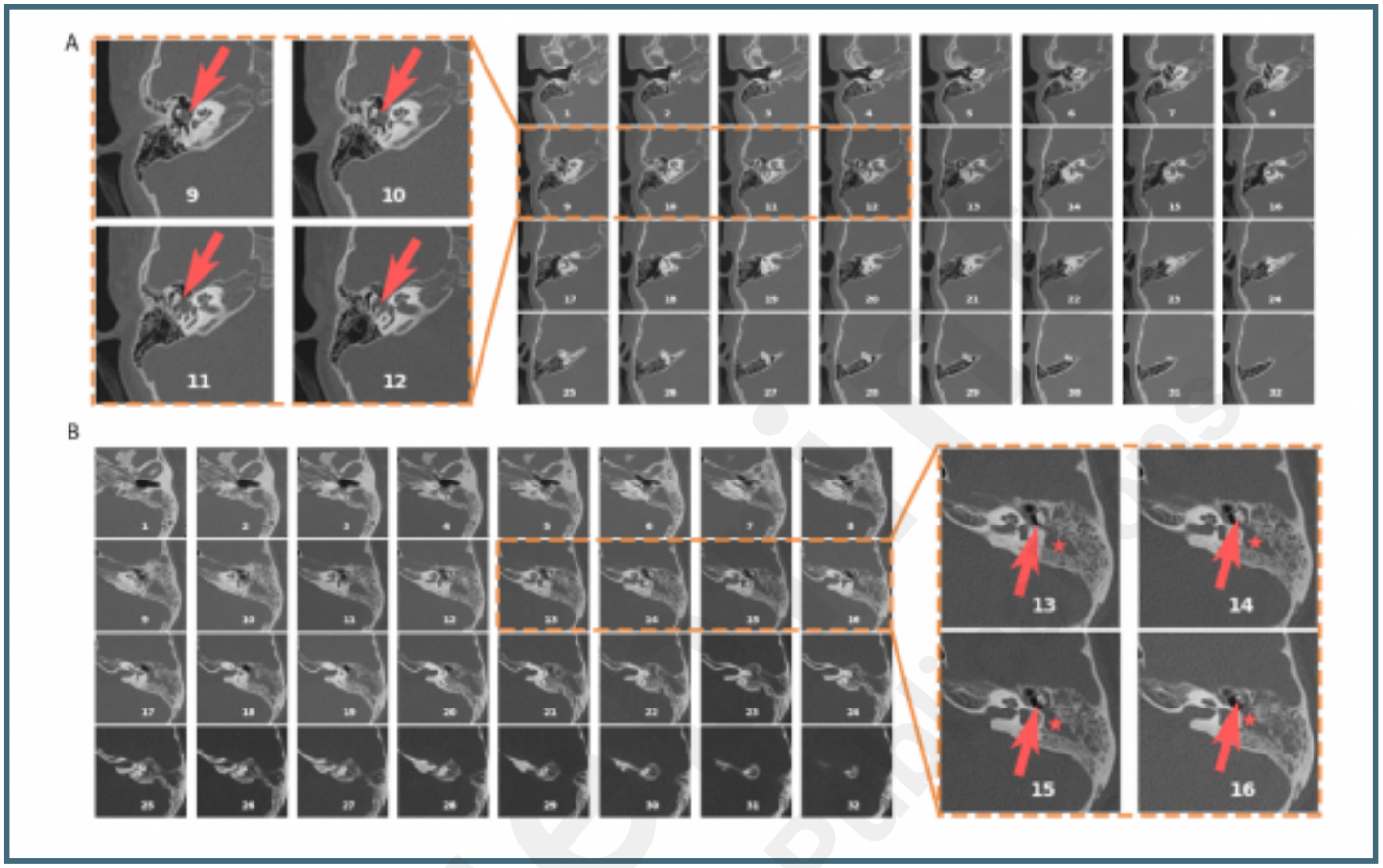
Examples of heatmaps. The heatmaps, generated in 3D fashion, are superimposed on the original CT and flattened to a series of 2D images for demonstration purpose. (A-B) A pathological and a normal ear, respectively. (C-D) A cholesteatoma and a non-cholesteatoma case, respectively. Area marked by hot signals indicate the presence of graphic patterns contributing to a “positive” prediction (i.e., a pathological ear in task 1 and a cholesteatoma in task 2).



Postoperative hearing gain for the operated ears with available audiometry outcomes (n=106). Data are categorized according to model predictions. Predictions that agree with the pathological results are denoted by close symbols, while open symbols indicate disagreements. Circles and triangles represent the treatment of endoscopic tympanoplasty and mastoidectomy, respectively. The error bars indicate ±1 standard deviation from the mean.



Examples of misclassified cases. (A) A pathological ear showing a small-size soft tissue density near the ossicles (arrows) with no evident sign of osseous erosion or mastoid opacification. (B) A case of cholesteatoma showing soft tissue density (asterisks) but with a visually intact ossicular chain (arrows) and a normal-size tympanic cavity.



Multimedia Appendixes

Supplementary tables.
URL: http://asset.jmir.pub/assets/ef022705c966ee6bb610442ed4f8d3c5.docx

Source code for model development and validation.
URL: http://asset.jmir.pub/assets/82eba31cdae38b53af7cbcaf6164eb26.zip