Heliyon 10 (2024) e29670

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Automated detection of otosclerosis with interpretable deep learning using temporal bone computed tomography images

Zheng Wang [a,c,1], Jian Song [b,d,1], Kaibin Lin [a,c], Wei Hong [a,c], Shuang Mao [b,d], Xuewen Wu [b,d,**], Jianglin Zhang [e,f,g,*]

[a] School of Computer Science, Hunan First Normal University, Changsha, 410205, China
[b] Department of Otorhinolaryngology, Xiangya Hospital Central South University, Changsha, Hunan, China
[c] Key Laboratory of Informalization Technology for Basic Education in Hunan Province, Changsha, 410205, China
[d] Province Key Laboratory of Otolaryngology Critical Diseases, Changsha, Hunan, China
[e] Department of Dermatology, Shenzhen People's Hospital, The Second Clinical Medical College, Jinan University. The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen, 518020, Guangdong, China
[f] Candidate Branch of National Clinical Research Center for Skin Diseases, Shenzhen, 518020, Guangdong, China
[g] Department of Geriatrics, Shenzhen People's Hospital, The Second Clinical Medical College, Jinan University. The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen, 518020, Guangdong, China

ARTICLE INFO

Keywords:
Computed tomography
Deep learning
Area under the receiver operating characteristic curve
Temporal bone computed tomography
Interpretability

ABSTRACT

Objective: This study aimed to develop an automated detection schema for otosclerosis with interpretable deep learning using temporal bone computed tomography images.
Methods: With approval from the institutional review board, we retrospectively analyzed high-resolution computed tomography scans of the temporal bone of 182 participants with otosclerosis (67 male subjects and 115 female subjects; average age, 36.42 years) and 157 participants without otosclerosis (52 male subjects and 102 female subjects; average age, 30.61 years) using deep learning. Transfer learning with the pretrained VGG19, Mask RCNN, and EfficientNet models was used. In addition, 3 clinical experts compared the system's performance by reading the same computed tomography images for a subset of 35 unseen subjects. An area under the receiver operating characteristic curve and a saliency map were used to further evaluate the diagnostic performance.
Results: In prospective unseen test data, the diagnostic performance of the automatically interpretable otosclerosis detection system at the optimal threshold was 0.97 and 0.98 for sensitivity and specificity, respectively. In comparison with the clinical acumen of otolaryngologists at $P < 0.05$, the proposed system was not significantly different. Moreover, the area under the receiver operating characteristic curve for the proposed system was 0.99, indicating satisfactory diagnostic accuracy.
Conclusion: Our research develops and evaluates a deep learning system that detects otosclerosis at a level comparable with clinical otolaryngologists. Our system is an effective schema for the differential diagnosis of otosclerosis in computed tomography examinations.

* Corresponding author
** Corresponding author
E-mail addresses: xwwu840903@hotmail.com (X. Wu), zhangjianglin@szhospital.com (J. Zhang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e29670
Received 25 April 2023; Received in revised form 10 April 2024; Accepted 12 April 2024

Abbreviations

OtoModel	Automatically Interpretable Otosclerosis Detection
CT	Computed Tomography
DL	Deep Learning
AUC	Area under the Receiver Operating Characteristic Curve
TBCT	Temporal Bone Computed Tomography
FaFA	Fissula Ante Fenestra
OAR	Organ at Risk
ACE	Adaptive Cross Entropy
ROI	Regions of Interest
ROC	Receiver Operating Characteristic
ReLU	Rectified Linear Units
XAI	Explainable Artificial Intelligence
BN	Batch Normalization Layer
GAP	Global Average Pooling Layer
FC	Fully Connected Layer
bbox	Bounding Box
CNN	Convolutional Neural Network
CI	Confidence interval
Grad-CAM	Gradient-weighted Class Activation Mapping

1. Introduction

Otosclerosis or otospongiosis is a multifactorial disorder of the temporal bone and stapes that presents with progressive conductive, sensorineural, or mixed hearing loss in humans [1,2]. It results in a sclerotic bone with abnormal osteons and is associated with genetic and environmental factors [3,4]. Otosclerosis is categorized into two subtypes: fenestral (stapedial) and retrofenestral (cochlear). Retrofenestral otosclerosis always occurs with fenestral involvement and is considered to be on a continuum with fenestral otosclerosis [5]. Typically, otosclerosis presents in the second to fourth decade of life and more frequently affects stapes footplate fixation by the foci located anterior to the vestibular window [6,7].

High-resolution computed tomography (CT) of the temporal bone serves as a useful aid to clinicians and remains an effective imaging modality in the diagnosis of otosclerosis [8,9]. Otosclerosis mostly manifests in the common area in the fissula ante fenestram (FaFA) by localizing subtle demineralization and evaluating the oval window, stapes footplate, and round window niche [2]. Otosclerosis findings on CT scan are too subtle and extremely indistinct to be seen, and the margins between normal and abnormal are difficult to delineate [10]. Therefore, an automated, interpretable, and accurate detection of otosclerosis lesions may help otolaryngologists improve diagnostic efficiency and prevent untreated hearing loss and unnecessary costs.

In recent years, deep learning-based diagnostic techniques have been introduced as an aid for radiologists in various fields to improve detection performance [11–16], such as outcome prediction with nonsmall-cell lung in multi-institutional CT image datasets [17], organ at risk delineation in CT images [18], and abnormality classifications from chest radiographs with major thoracic diseases [19]. A fine-tuned deep neural network can be applied for the recognition and classification of CT images.

Fujima et al. [20] used 140 temporal bone CT images to train and assess the utility of deep learning analysis in diagnosing otosclerosis on temporal bone CT images and achieved and 0.915 for the best accuracy and area under the receiver operating characteristic curve (AUC), respectively. Chen et al. [21] introduced W-Net with Adaptive Cross Entropy to detect ultrasmall medical objects on an otosclerosis dataset and achieved 0.954 as the best AUC. Despite their strong predictive power, deep learning models have been criticized for their poor interpretability, and we recognize that these factors represent challenges in the development of useful tools for clinicians [22].

In this study, we present an automatically interpretable deep learning approach (OtoModel) to boost otosclerosis detection on CT images and demonstrate how prediction activation maps learn the relevant features as a complementary means to understand a diagnosis, with the downstream goal of providing reliable and interpretable measures based on the location of otosclerosis. In addition, two experienced otolaryngologists and fellowship-trained radiologists compared the diagnostic performance of the proposed system.

The contributions of this study are as follows:

- The deep neural network models automatically recognized and extracted the Region of Interest (ROI) by detecting contour abnormalities for otosclerosis diagnosis on CT images.
- The diagnostic performance of the proposed system demonstrated good accuracy and was comparable to that of clinical otolaryngologists. The saliency map improves the interpretable ability for diagnosis.
- The interpretable otosclerosis detection system is potentially useful in clinical practice for identifying the presence or absence of otosclerosis on CT images.

2. Materials and methods

2.1. Ethics and consent

This study's ethics and consent were approved by the Institutional Review Board (IRB)of the Xiangya Hospital, Central South University, Changsha, China (IRB #2019121188 in 2019). All procedures performed in the study adhered to the appropriate guidelines and regulations. Written informed consent for participation and for the use of personally identifiable data was obtained from all participants. The study protocol, including the use of computed tomography scans, received IRB approval from Xiangya Hospital, Central South University.

2.2. Data source and patient selection

Fig. 1 depicts the following inclusion criteria: 1) subjects who underwent a CT scan examination; 2) only healthy subjects and subjects with otosclerosis were included; 3) modality with a noncontrast-enhanced CT; and 4) bilateral CT. Subjects were excluded if: 1) the quality of the CT images was poor; 2) the subjects could not be retrospectively identified; and 3) any treatment was performed before the CT scan. Finally, the acquired subjects were randomly divided into the training, validation, and test subsets at a ratio of 8:1:1, split by the number of subjects. The subjects in the three subsets were different from each other. The Xiangya Hospital, Central South University Institutional Review Board approved the study protocol and use of CTs.

CT image datasets (as outlined in Table 1) were acquired from 182 subjects with confirmed otosclerosis, comprising 67 male and 115 female subjects. The average age within this group was 36.4 years, with an age range of 10–58 years. In addition, the dataset encompassed 157 subjects with surgically confirmed healthy otosclerosis, including 52 male and 102 female subjects. In this latter

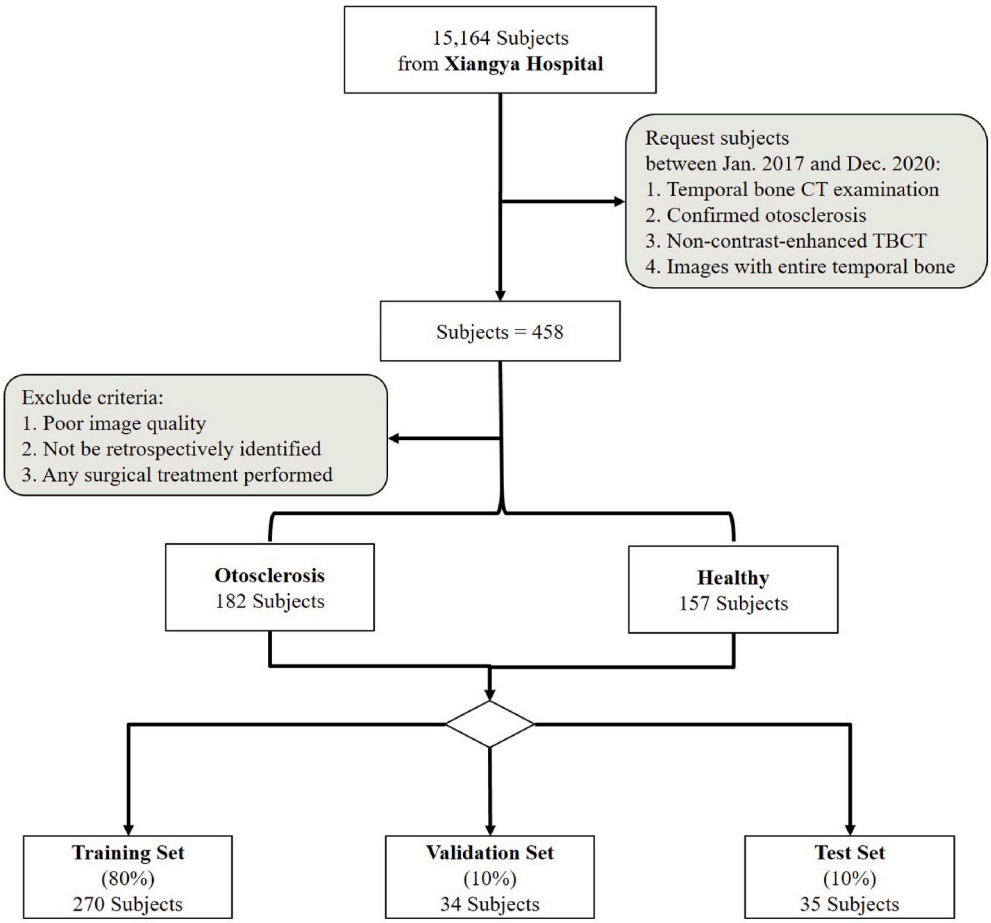

Fig. 1. Acquisition criteria of the temporal bone CT.

Table 1
Demographic distribution and clinical characteristics of the enrolled subjects.

	Healthy	Otosclerosis	All
	n = 157, n(%)	n = 182, n(%)	n = 339, n(%)
Age	30.61 ± 15.34	36.42 ± 19.70	33.52 ± 16
Gender			
Male	52(49)	67(37)	119(35)
Female	102(51)	115(63)	217(65)
Side			
Left	78(50)	97(53)	175(52)
Right	79(50)	85(47)	164(48)
Subtype			
fissula ante fenestraml	x	160(88)	160(88)
cochlear	x	22(12)	22(12)

group, the average age was 30.6 years, ranging from 21 to 48 years.

This study was approved by the medical research and ethics committee of Xiangya Hospital. All subjects underwent CT examination of the temporal bones at Xiangya Hospital between January 01, 2017 and December 31, 2020. The final diagnosis was determined by a board-certified senior otolaryngologist with more than 20 years of experience using temporal bone CT images, audiology examinations, relevant medical history, and an intraoperative confirmation of otosclerosis.

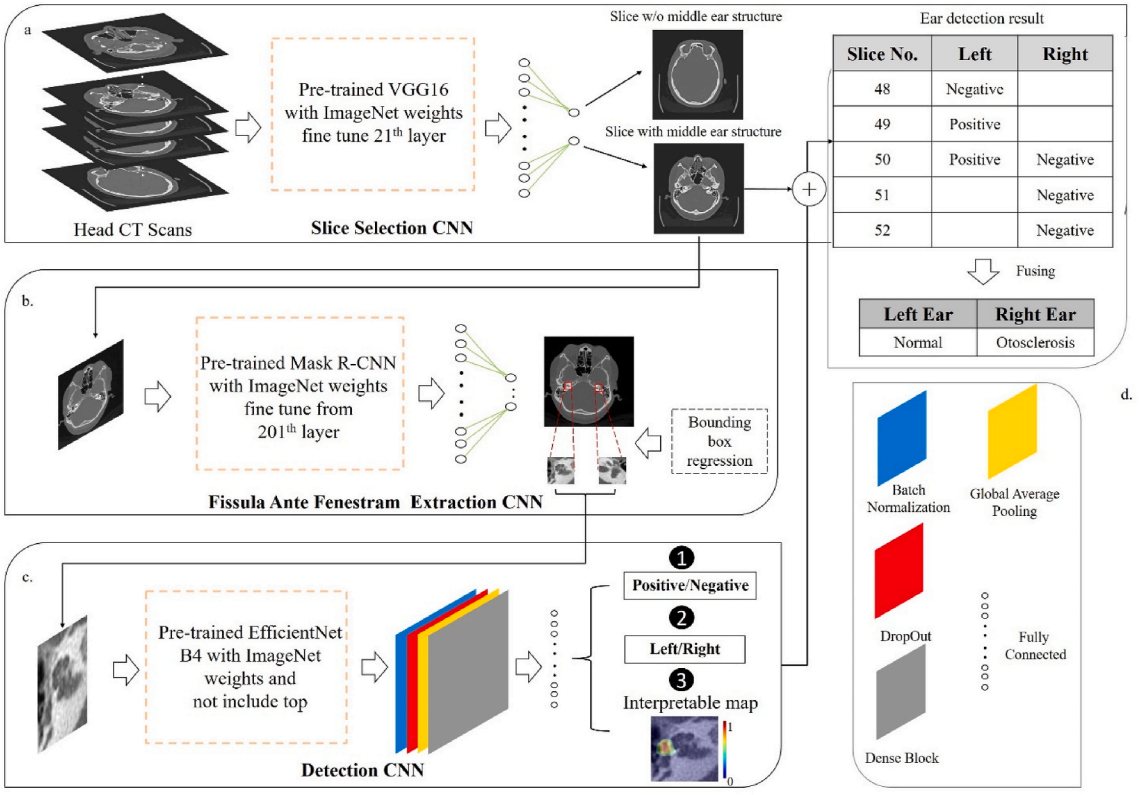

Fig. 2. Overview of the OtoModel system. The system mainly consists of three stages: a) Identify the slices with the middle ear structure and output the final detection result (absence or presence) by heuristically fusing the diagnostic results of a patient. b) Extract useful image features from the selected image slices and predict the coordinates of a rectangular bounding box, which defines the region of the fissula ante fenestram. The extracted patches are downsized to 112 × 112 and are used as input images to the detection CNN. c) Compute the classification probability of otosclerosis on all selected slices containing the fissula ante fenestram. The last three layers advance a gradient backpropagation method to calculate a pixel-by-pixel probability map of the otosclerosis present. d) depicts a simplified diagram of a neural network architecture. CNN: convolutional neural network.

2.3. Data protocol

The temporal bone CT examination was performed using Philips Healthcare CT scanners, which obtained CT images with the following scanning parameters: 1 s, 90–200 mA, 110–130 kV, matrix of 512 × 512, and a bone window width with a high-resolution bone algorithm between 3500 and 4000 Hounsfield centered at 350–650 Hounsfield, in the absence of intravenous contrast. The screening image set was acquired in the axial plane with a slice thickness of 0.5 mm at 0.5-mm intervals and was reconstructed at R 0.3-mm intervals to obtain overlapping slices. A board-certified clinician who specializes in otorhinolaryngology with more than 20 years of clinical experience confirmed the presence or absence of otosclerosis. Otosclerosis confirmation was performed in all subjects. CT images were loaded into the LabelImg software tool (version 1.8.6). They were annotated with rectangular bounding boxes to extract the ROIs containing the entire ear structure by an otolaryngologist and a fellowship-trained radiologist who had more than 10 years and 5 years of clinical experience, respectively.

2.4. System development and training

The proposed system was executed on a Dell xps8930 server (hexa-core 3.20 GHz processor, 16 Gb RAM, and one NVIDIA GeForce GTX 2080 video card), implemented in Python (version 2.7; Python Software Foundation, Wilmington, Del), and coded using the Keras [23] framework with the TensorFlow-GPU 1.15 [24] backend. For a fair comparison, we used 3 × 3 convolutions activated by rectified linear units and trained them using the Adam optimizer. We initially set Adam's learning rate to 1e-3 and then decayed the learning rate by a factor of 5 whenever the validation loss plateaued after an epoch. The optimization process was run for 100 epochs, and a batch size of 16 was selected on the basis of the experiments. To avoid overfitting, we implemented the EarlyStopping function in the Keras framework, which stops training when the monitoring volume stops improving.

The proposed system takes 512 × 512 CT slices as the input, which are normalized to a range of 0 and 1 with respect to the maximum CT signal. The OtoModel system is decomposed into three stages, as illustrated in Fig. 2. The initial stage (shown in Fig. 2a) is dedicated to the precise filtering of CT slices that depict the middle ear structures in a patient during inference. This step is crucial for isolating the ROI pertinent to our analysis. Following this initial phase, a second convolutional neural network (CNN) emerges, as demonstrated in Fig. 2b. Its primary function is to refine the range of information extracted in the first stage. By focusing on the delineated ROI, this network effectively streamlines the data before it is introduced into the disease classification process. Finally, the extracted ROI patch was classified as having the presence or absence of otosclerosis, as depicted in Fig. 2C. Therefore, the result of a slice heuristically fuses all the CT diagnoses, generating the final diagnostic result (normal or otosclerosis) on the series of detection results for a patient at inference. Fig. 2d showcases a flow from batch normalization through dropout and pooling, culminating in a densely connected layer for feature integration.

Explainable artificial intelligence (XAI) addresses the black-box nature of artificial intelligence. One systematic review [25] assessed the state of XAI in healthcare, noting limited research, diverse stakeholder perspectives, and the need for standardized evaluation methods. Zhang et al. [26] explored the growing potential of XAI in medical diagnosis and surgery by examining recent trends, conducting a survey, presenting a breast cancer case study, and highlighting its promising prospects. To assess the potential impact on CT slices [27,28], we used an open-source implementation of guided Gradient-weighted Class Activation Mapping (Grad-CAM) [29] in conjunction with our models. This approach allows us to evaluate the acquired features of the model and generate saliency maps that accentuate the crucial representations related to the target class.

Pretrained VGG16 [30,31], Mask RCNN [32,33], and EfficientNet [34,35] were used to perform transfer learning for the slice selection CNN, fissula ante fenestram extraction CNN, and detection CNN, respectively. Slice selection begins with the radiologist carefully identifying CT slices showing the middle ear through manual screening. These initially selected slices were then fine-tuned using a pretrained VGG16 neural network model. Our proposed system reused the parameters on ImageNet (a large-scale dataset of natural images [36,37]) and enabled a reduction in the number of parameters without degrading the performance of the networks. In this study, we froze the top layers of the networks and fine-tuned them on the CT images. The proposed system executed a 10-fold cross-validation. Data augmentation was used to increase the data size and consistency for robustness. These techniques were subjected to a rotation range of 45°, a shifting width and height range of 0.2, a zooming range of 0.2, and a horizontal flip.

For the first two pretrained networks, we froze a particular layer to preserve the learning representation extracted by these networks. A fully connected (FC) layer was added within the sigmoid activation function to obtain the probability. The last pretrained network added a batch normalization layer, a global average pooling layer, a dropout layer with a threshold of 0.5, and an FC layer to enable very efficient information sharing across the layers. For the outcome, a final probability score for the presence or absence of otosclerosis on the extracted images was predicted from the FC layer. The detection result was calculated using the maximum classification probability of otosclerosis on all selected slices using the ROI bounding box. We used a gradient backpropagation approach [38] to compute a pixel-by-pixel probability map of the otosclerosis present. To compare with other state-of-the-art image classification CNNs, two additional classification CNNs (ResNet-18 [39] and InceptionV3 [40]) were also evaluated.

2.5. Evaluation metrics

We used a well-established measurement criteria, namely sensitivity, specificity, accuracy, precision, and recall, which are defined as Eq. (1):

$$\begin{aligned} \text{Sensitivity} &= \frac{f_{TP}}{f_{TP}+f_{FN}} \\ \text{Specificity} &= \frac{f_{TN}}{f_{TN}+f_{FP}} \\ \text{Accuracy} &= \frac{f_{TP}+f_{TN}}{f_{TP}+f_{FP}+f_{TN}+f_{FN}} \\ \text{Precision} &= \frac{f_{TP}}{f_{TP}+f_{FP}} \\ \text{Recall} &= \frac{f_{TP}}{f_{TP}+f_{FN}} \\ \text{AUC} &= \int_{x=0}^{1} \text{Sensitivity}\left(\frac{1}{\text{1-Specificity}}(x)\right)dx \end{aligned} \tag{1}$$

Here, f_{TP} represents the count of true positives, f_{FP} signifies the count of false positives, f_{FN} denotes the count of false negatives, and f_{TN} stands for the count of true negatives.

All screening CT images of the hold-out test subset were visually evaluated by a clinical expert group comprising a senior otolaryngologist with 20 years of experience, an otolaryngologist with more than 5 years of fellowship training, and a radiologist with 10 years of experience. To facilitate a meaningful comparison of our proposed system's performance, each clinical expert independently reviewed the CT scans within the hold-out test subset, which consisted of 35 subjects, using the same experimental settings. The evaluation used the AUC, which quantifies the entire two-dimensional area beneath the complete receiver operating characteristic curve (ROC) spanning from (0,0) to (1,1). Clinical experts recorded the diagnosis as the presence or absence of otosclerosis. All the clinical experts were blinded to the ground truth and the diagnosis from the proposed system but not to the clinical characteristics.

2.6. Statistical analysis

Statistical analysis was performed using Python (version 2.7, Python Software Foundation, Wilmington, Del), with P values of <0.05, indicating a statistically significant difference. Identifying the presence or absence of otosclerosis on the hold-out test subset of the subjects was evaluated for sensitivity, specificity, and accuracy. These diagnostic results were identified for the OtoModel system using a classification CNN from EfficientNet and alternative otosclerosis detection systems using classification CNNs from ResNet-18 and InceptionV3. Confidence intervals (CIs) of accuracy were calculated for the OtoModel systems. Contingency tables were identified for the otolaryngologist with more than 20 years of experience, the otolaryngologist with 5 years of fellowship training, the radiologist with 10 years of experience, and the OtoModel system with the classification CNN from EfficientNet. In addition, the ROC was used to further analyze the diagnostic probability score of the OtoModel system and clinical experts, with the AUCs compared using the Youden index, which identifies the optimal sensitivity and specificity.

3. Results

The training times for slice selection, ROI extraction, and detection CNNs were 0.5, 1.6, and 1.8 h, respectively. This testing experiment was conducted on an independent subset. Table 2 presents the performance metrics of the CNN developed for extracting the fissula ante fenestram. Table 3 shows the comparison between the sensitivity, specificity, and accuracy for identifying a confirmed diagnosis of the presence or absence of otosclerosis for the proposed system using the classification CNN from EfficientNet and the alternative otosclerosis detection systems using the classification CNNs from ResNet-18 and InceptionV3.

All classification CNNs exhibited strong performance, with accuracy estimates ranging from 0.90 to 0.99. Notably, the proposed system, which uses the classification CNN derived from EfficientNet, demonstrated the highest overall diagnostic efficacy in detecting otosclerosis. Fig. 3 illustrates the ROC curve for the training set and the confusion matrix for the test set, elucidating the diagnostic capabilities of the OtoModel system in identifying the presence or absence of otosclerosis. The AUC for the proposed system was 0.98 (95 % CI: [0.97, 1.00]; $P < 0.005$). In addition, we plotted the sensitivity and 100 - specificity for otolaryngologists with varying levels

Table 2
Performance of the fissula ante fenestram extraction CNN.

	Sensitivity	Specificity	Accuracy	Precision	Recall
	(%) ± Standard Deviation				
Ours	92.1 ± 0.15	93.7 ± 3.35	96.1 + 1.05	92.2 ± 1.10	91.5 ± 1.30

CNN: convolutional neural network.

Table 3
Sensitivity, specificity and accuracy for the proposed system and the alternative otosclerosis detection systems.

Methods	Sensitivity (%)↓	Specificity (%)↓	Accuracy (%) ↓ [95 % CI]	Precision (%)↓	Recall (%)↓	AUC
ResNet-18	93.58	94.04	93.70 [90, 96]	93.2	93.58	93.4
InceptionV3	94.66	95.34	96.04 [93, 98]	93.4	94.66	94.8
Fujima el at [20].	83	93	86 [79, 93]	x	x	91.5
Chen el at [21].	x	x	x	93.58	94.04	95.7
Ours	97.35	98.53	98.06 [95,99]	97.7	97.35	98.0

CI: Confidence interval.

of experience and the radiologist with 10 years of experience.

For comparison, we depicted point assessments of sensitivity and specificity for otolaryngologists with more than 20 years of experience, those with 5 years of fellowship training, and the radiologist with 10 years of experience alongside the ROC curve of the proposed system in Fig. 3-a. Notably, the sensitivity and specificity point assessments of clinical experts fell within the 95 % CIs of the AUC for the OtoModel system. Specifically, the OtoModel exhibited CIs of the presence and absence of otosclerosis between 0.95 and 1.00 (Fig. 3-b). In contrast, clinical experts' diagnoses of the presence of otosclerosis ranged from 0.95 to 0.77, whereas those of the absence of otosclerosis ranged from 0.85 to 1.00. Remarkably, the OtoModel demonstrated no statistically significant differences in diagnostic performance compared with clinical experts.

To investigate the interpretability, the proposed system was visualized by applying Grad-CAM, which produced a coarse localization map highlighting the target. The last convolutional layer of the last res-block was made transparent to predict the presence or absence of otosclerosis. In our study, we applied the "pointing game" method [41,42] to assess how well our explainability map aligns with radiologist-drawn contours. Here, an overlap is counted as a "hit," and no overlap is counted as a "miss." The effectiveness of the explainability map was then measured using the "hit rate" [43], as illustrated in Fig. 4.

The proposed system could render dense probability maps that demonstrate the pixel-by-pixel probability of the presence of otosclerosis on the left ear (shown in Fig. 5a and c), which the OtoModel system explained as positive because of the strong localized activation maps, as illustrated in Fig. 5b and d, at the boundary of the fissula ante fenestram (white arrow).

On an independent test set, the system produced explanations of the presence of otosclerosis in the right ear (shown in Fig. 6a and c), and the OtoModel system was interpreted as positive, as illustrated in Fig. 6b and d. Two explanations of otosclerosis on the left and right ears showed an absence (shown in Fig. 7a and c), and the OtoModel system was explained as negative, as illustrated in Fig. 7b and d. This explains to a certain extent why the model could accurately diagnose otosclerosis. The proposed detection can concentrate on the target area. Thus, it can extract discriminative representations for better detection of otosclerosis on the fissula ante fenestram.

To this end, we roughly calculated the diagnosis time spent by the three clinical experts and our OtoModel system. Specifically, 3 clinical experts typically spend an average of approximately 3–5 min identifying otosclerosis. In contrast, the OtoModel system dramatically reduced the identification time by an average of 0.5 s, resulting in a significant reduction in the diagnosis time of 97.2 %. In addition, it is worth noting that the clinical experts accepted most of the fully automated identification results produced by our OtoModel system without any modification, except for 2 out of 100 CT scans that required extra-human intervention. For example, the localized slice of otosclerosis may not be the best representation. Additional refinements can further improve the reliability of otosclerosis diagnosis and treatment. Our clinical partners have confirmed that such performance is fully acceptable for many clinical and industrial applications, indicating the high clinical utility of the OtoModel system.

4. Discussion

Our study demonstrated the feasibility of using a deep learning system for the automatic interpretable detection of otosclerosis in high-resolution temporal CT scans. The OtoModel system achieved a high diagnostic performance for identifying the presence or absence of otosclerosis, with an AUC of 0.98. Furthermore, there was no statistically significant difference between the clinical experts, who had varied levels of experience in identifying otosclerosis. To the best of our knowledge, only two previous studies [20,21] have reported the utility of deep learning for the diagnosis of otosclerosis on temporal bone CT. In contrast, our proposed system provided automatically interpretable otosclerosis detection in CT volumes to simulate the decision-making of clinicians and outperformed previous studies with a 0.98 accuracy. In addition, our study demonstrated a precise explanation of the presence of otosclerosis.

Our study investigated three different types of pretrained CNNs, including EfficientNet, ResNet-18, and InceptionV3, and found that EfficientNet B4 provided the best diagnostic performance for detecting otosclerosis. EfficientNet [44] used a simple and highly effective compound scaling approach, which easily scaled up a baseline CNN to any target resource and maintained the model's efficiency. EfficientNet consists of eight models from B0 to B7, which refer to variants with more parameters and higher accuracies. Due to clever scaling of the depth, width, and resolution, EfficientNet demonstrated higher accuracy values than the alternative otosclerosis system. The proposed system used the EfficientNet B4 model, as it contains 19 M parameters, which is feasible for our experimental setup, as B5, B6, and B7 include 30 M, 43 M, and 66 M parameters, respectively.

Our interpretability analysis using Grad-CAM confirms the effectiveness of the model by focusing on key areas. Dense probability maps and targeted representation facilitate robust otosclerosis detection, particularly in the anterior fenestram. As shown in the probability map for the left ear in Fig. 4, there is a strong activation (interpreted as positive) at the fenestram-anterior boundary. The

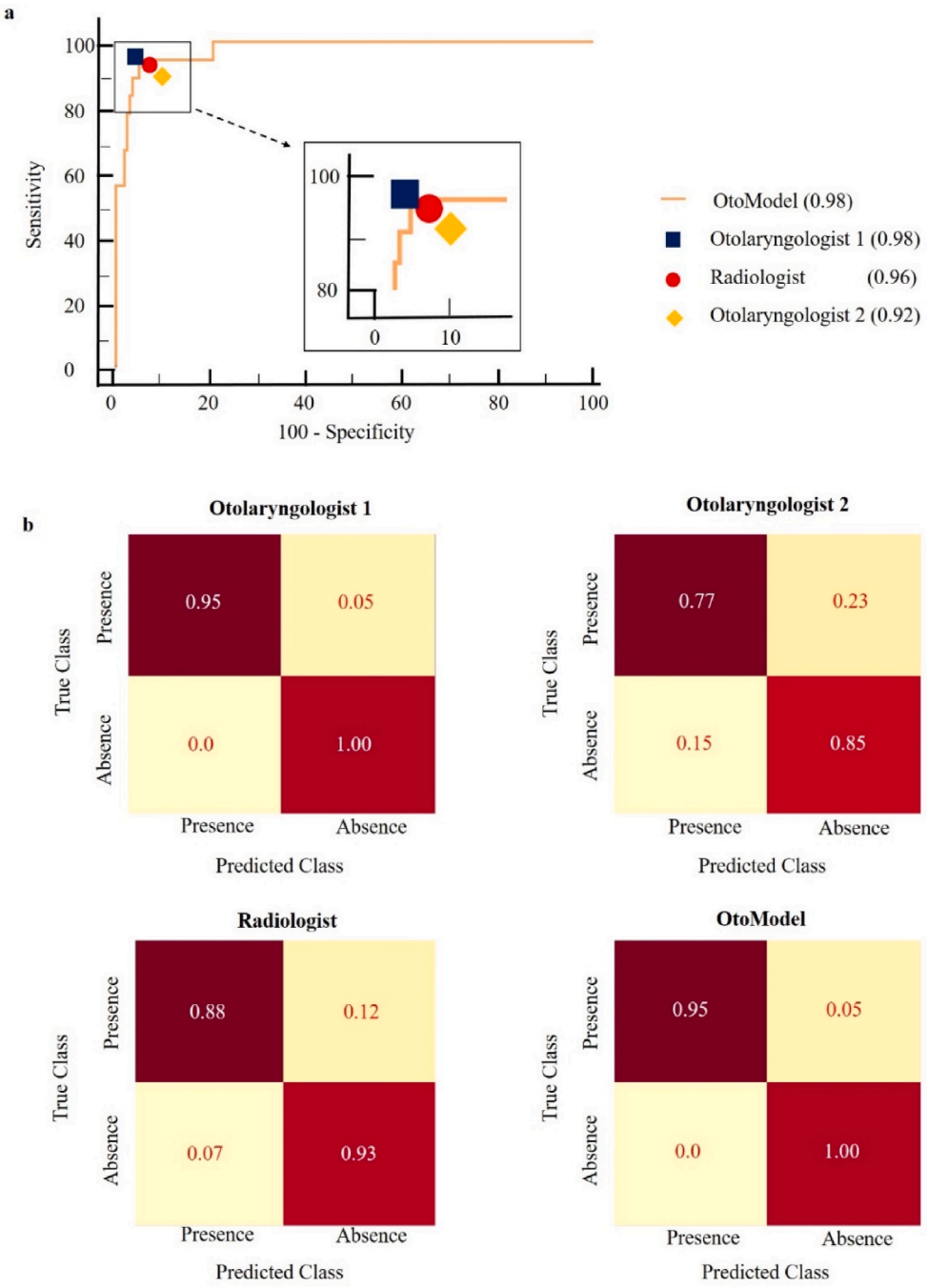

Fig. 3. Comparisons of the diagnostic performance. a) The AUC of the OtoModel was 0.98, indicating a high overall diagnostic accuracy in the test set. Note that the sensitivity and specificity of the clinical experts are closely proximate to the ROC curve of the OtoModel system. **b)** Confusion matrix of the diagnostic accuracy in the test set.

independent test set provides an explanation for the presence of otosclerosis in the right ear (Fig. 5) and absence of otosclerosis in the other two cases (Fig. 6). This targeted detection system increases robustness by focusing on specific areas such as the anterior window cleft for better detection of otosclerosis.

There were certain limitations and challenges encountered during the implementation of our deep learning system. First, our OtoModel system comprises three individual CNNs connected sequentially rather than an end-to-end architecture. This approach may increase the training workload due to the individual training phases. In addition, our study faced limitations in terms of data availability, leading us to rely on transfer learning from pretrained models to optimize training efficiency. Future research with larger

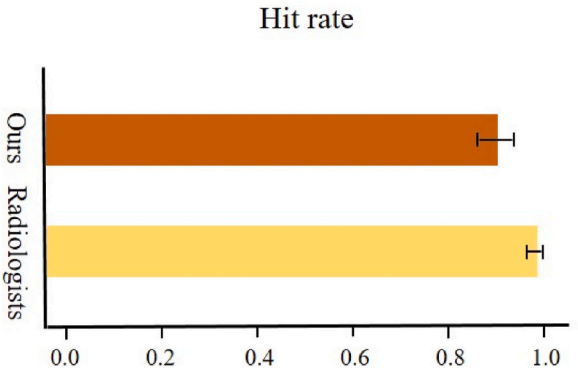

Fig. 4. Comparing with radiologists under the hit rate scheme.

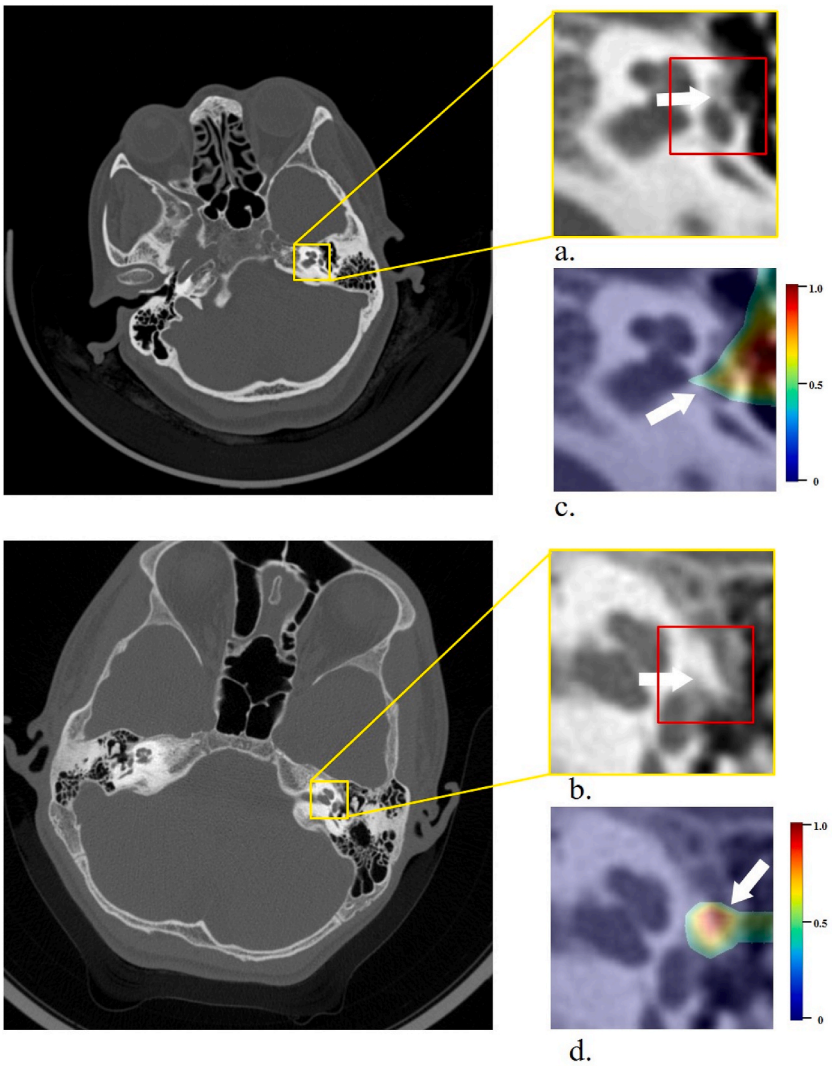

Fig. 5. Demonstration of the confirmed presence of otosclerosis in the left ear. (a) Extracted fissula ante fenestram image patch of a 28-year-old woman and (b) extracted fissula ante fenestram image patch of an 18-year-old woman analyzed by the proposed system as positive with abnormal signals. (c)–(d) Pixel-by-pixel probability map for the extracted image patches showing the high-probability areas in the lesions on which the schema based its explanation of otosclerosis (arrows).

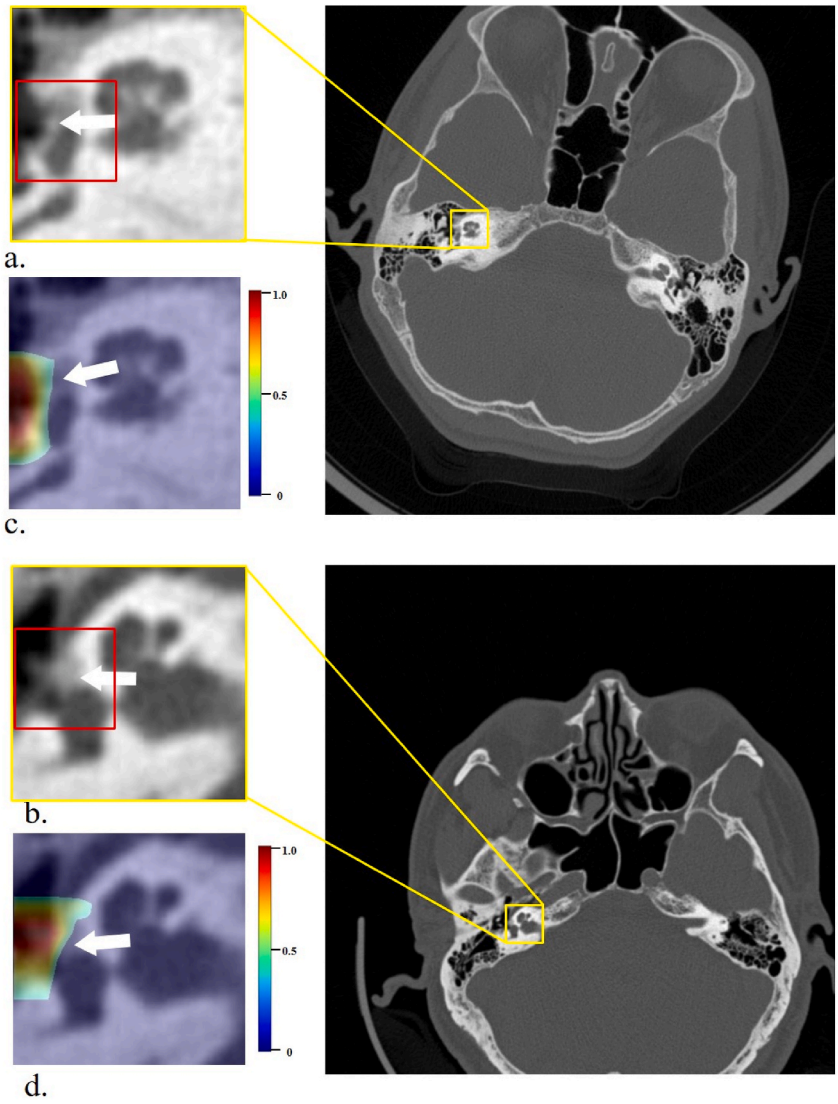

Fig. 6. Demonstration of the confirmed presence of otosclerosis in the right ear.
(a) Extracted fissula ante fenestram image patch of a 23-year-old man and (b) extracted fissula ante fenestram image patch of a 27-year-old woman analyzed by the proposed system as positive with abnormal signals. (c)–(d) Pixel-by-pixel probability map for the extracted image patches showing the high-probability areas in the lesions on which the schema based its explanation of otosclerosis (arrows).

datasets could enhance the diagnostic performance of the proposed system. Furthermore, in our study, all subjects with otosclerosis were evaluated at a single institution using a uniform imaging protocol, primarily focusing on fenestral otosclerosis, with only a few cases of cochlear otosclerosis. Detecting different subtypes of otosclerosis in screening CT examinations is more challenging but holds the potential to aid in developing precise surgical treatment plans and saving valuable diagnosis time for otolaryngologists or radiologists.

In summary, our study has demonstrated the feasibility of implementing a deep learning system to detect otosclerosis on high-resolution temporal bone CT scans and its potential to improve the quality and efficiency of clinical practice. There were no statistically significant differences between the OtoModel system and clinical experts with various levels of experience in identifying the presence or absence of otosclerosis. However, before its implementation in clinical practice, the technical advancements of the OtoModel system must be further evaluated in large prospective studies with multiple institutions that use different CT units and imaging protocols.

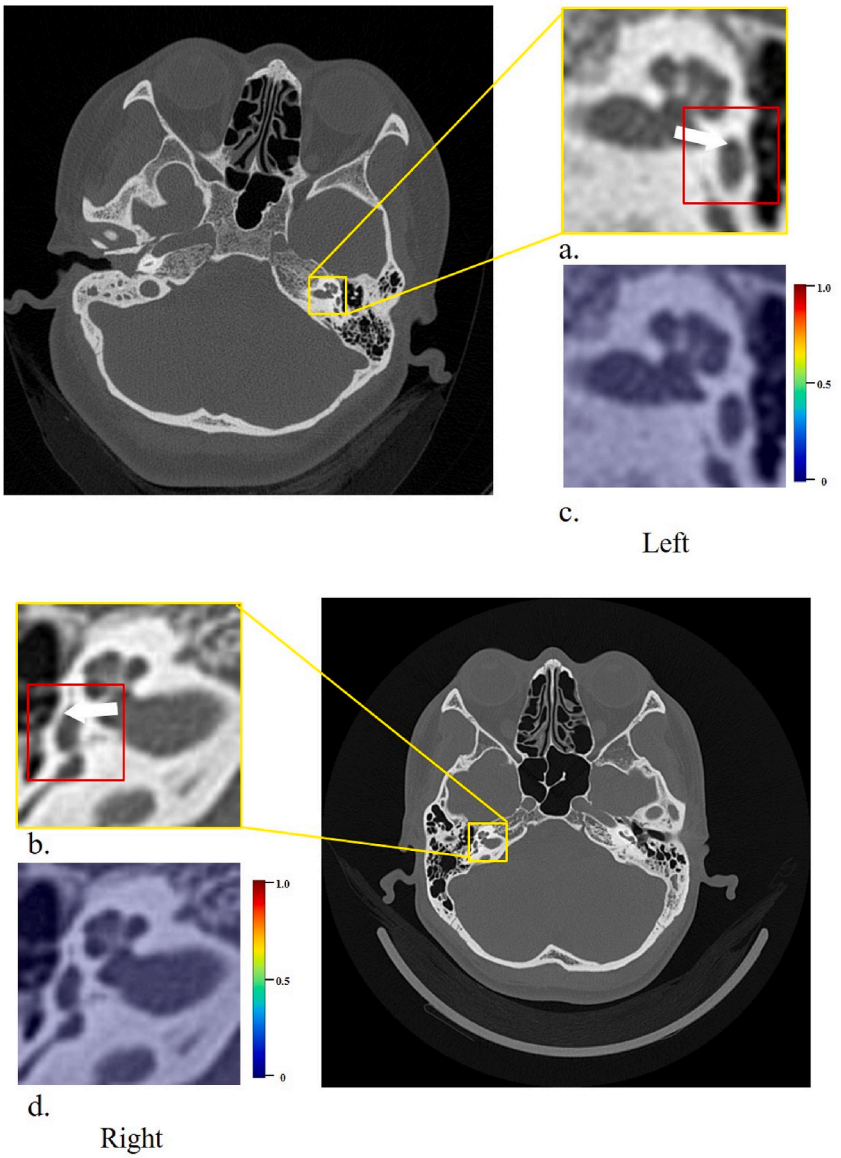

Fig. 7. Demonstration of the confirmed absence of otosclerosis.
(a)Extracted fissula ante fenestram image patch of a 21-year-old woman and (b) extracted fissula ante fenestram image patch of a 20-year-old woman analyzed by the proposed system as negative with normal signals. (c)–(d) Pixel-by-pixel probability map for the extracted image patches showing the low-probability areas in the lesions on which the schema based its explanation of the absence of otosclerosis (arrows).

Data availability statement

All data used in the generation of the results presented in this manuscript will be made available upon reasonable request from the corresponding author.

CRediT authorship contribution statement

Zheng Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Jian Song:** Writing – original draft, Software, Resources, Investigation, Funding acquisition, Data curation, Conceptualization. **Kaibin Lin:** Visualization, Validation, Software, Methodology, Formal analysis. **Wei Hong:** Visualization, Validation, Software, Methodology, Formal analysis. **Shuang Mao:** Resources, Data curation. **Xuewen Wu:** Writing – original draft, Supervision, Project administration, Investigation, Data curation. **Jianglin Zhang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by Hunan Provincial Natural Science Foundation of China (under Grant 2022JJ30189 and 2021JJ31108), and Teaching Reform Research Project of Universities in Hunan Province(under Grant HNJG-2021-1120), and Scientific Research Fund of Hunan Provincial Education Department (grant number 23A0643 and 23C0427), and also funded by the National Natural Science Foundation of China (under Grants 82073018) and Shenzhen Science and Technology Innovation Committee (under Grants JCYJ20210324113001005 and JCYJ20210324114212035).

References

[1] S. Cureoglu, P. Schachern, A. Ferlito, A. Rinaldo, V. Tsuprun, M. Paparella, Otosclerosis: etiopathogenesis and histopathology, Am. J. Otolaryngol. 27 (5) (2006) 334–340, https://doi.org/10.1016/j.amjoto.2005.11.001.
[2] V.C. Andreu-Arasa, E.K. Sung, A. Fujita, N. Saito, O. Sakai, Otosclerosis and dysplasias of the temporal bone, Neuroimaging Clin. 29 (1) (2019) 29–47, https://doi.org/10.1016/j.nic.2018.09.004.
[3] M. Quesnel, R. Ishai, M.J. McKenna, Otosclerosis: temporal bone pathology, Otolaryngol. Clin. 51 (2) (2018) 291–303, https://doi.org/10.1016/j.otc.2017.11.001.
[4] L. Gustave Davis, Pathology of otosclerosis: a review, Am. J. Otolaryngol. 8 (5) (1987) 273–281, https://doi.org/10.1016/S0196-0709(87)80046-7.
[5] B. Purohit, R. Hermans, K. Beeck, Imaging in Otosclerosis: A Pictorial Review, Insights into Imaging 5, 2014, pp. 245–252, https://doi.org/10.1007/s13244-014-0313-9.
[6] H. Schuknecht, W. Barber, Histologic variants in otosclerosis, Laryngoscope 95 (1985) 1307–1317, https://doi.org/10.1288/00005537-198511000-00003.
[7] M. McKenna, Pathophysiology of Otosclerosis, Otology & Neurotology : Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, vol. 22, 2001, pp. 249–257.
[8] S. Lagleyre, T. Sorrentino, M.-N. Calmels, Y.-J. Shin, B. Escude, O. Deguine, B. Fraysse, Reliability of High-Resolution Ct Scan in Diagnosis of Otosclerosis, Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, vol. 30, 2009, pp. 1152–1159, https://doi.org/10.1097/MAO.0b013e3181c2a084.
[9] T. Kanzara, J. Virk, Diagnostic performance of high resolution computed tomography in otosclerosis, World Journal of Clinical Cases 5 (2017) 286–291, 10.12998%2Fwjcc.v5.i7.286.
[10] P. Manning, M. Shroads, J. Bykowski, M. Mafee, Role of radiologic imaging in otosclerosis, Current Otorhinolaryngology Reports 10 (2022) 1–7, https://doi.org/10.1007/s40136-021-00377-z.
[11] S. Soffer, A. Ben-Cohen, O. Shimon, M. Amitai, H. Greenspan, E. Klang, Convolutional neural networks for radiologic images: a radiologists guide, Radiology 290 (2019) 180547, https://doi.org/10.1148/radiol.2018180547.
[12] Rodr'iguez-Ruiz, E. Krupinski, J.-J. Mordang, K. Schilling, S. Heywang-KöbrunnerSechopoulos, R. Mann, Detection of breast cancer with mammography: effect of an artificial intelligence support system, Radiology 290 (2018) 181371, https://doi.org/10.1148/radiol.2018181371.
[13] Z. Wang, Y. Meng, F. Weng, C. Yinghao, F. Lu, X. Liu, M. Hou, Z. Jie, An effective cnn method for fully automated segmenting subcutaneous and visceral adipose tissue on ct scans, Ann. Biomed. Eng. 48 (2019) 312–328, https://doi.org/10.1007/s10439-019-02349-3.
[14] Z. Wang, A. Hounye, J. Zhang, M. Hou, M. Qi, Deep learning for abdominal adipose tissue segmentation with few labelled samples, Int. J. Comput. Assist. Radiol. Surg. 17 (2021) 579–587, https://doi.org/10.1007/s11548-021-02533-8.
[15] Sahayaraj A. Felix, H. Joy Prabu, J. Maniraj, M. Kannan, M. Bharathi, P. Diwahar, J. Salamon, Metal–organic frameworks (MOFs): the next generation of materials for catalysis, gas storage, and separation, J. Inorg. Organomet. Polym. Mater. 11 (2023 May) 1–25, https://doi.org/10.1007/s10904-023-02657-1.
[16] Z. Wang, J. Song, R. Su, M. Hou, M. Qi, J. Zhang, X. Wu, Structure-aware deep learning for chronic middle ear disease, Expert Syst. Appl. 194 (2022 May 15) 116519, https://doi.org/10.1016/j.eswa.2022.116519.
[17] P. Mukherjee, M. Zhou, E. Lee, A. Schicht, Y. Balagurunathan, S. Napel, R. Gillies, S. Wong, A. Thieme, A. Leung, O. Gevaert, A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomography image datasets, Nat. Mach. Intell. 2 (2020) 274–282, https://doi.org/10.1038/s42256-020-0173-6.
[18] H. Tang, X. Chen, Y. Liu, Z. Lu, J. You, M. Yang, S. Yao, G. Zhao, Y. Xu, T. Chen, X. Xie, Clinically applicable deep learning framework for organs at risk delineation in ct images, Nat. Mach. Intell. 1 (2019) 1–12, https://doi.org/10.1038/s42256-019-0099-z.
[19] E.J. Hwang, S. Park, K.N. Jin, J. Im Kim, S.Y. Choi, J.H. Lee, J.M. Goo, J. Aum, J.J. Yim, J.G. Cohen, G.R. Ferretti, Development and validation of a deep learning–based automated detection algorithm for major thoracic diseases on chest radiographs, JAMA Netw. Open 2 (3) (2019 Mar 1) e191095, 10.1001/jamanetworkopen.2019.1095.
[20] N. Fujima, V. Andreu-Arasa, K. Onoue, P. Weber, R. Hubbell, B. Setty, O. Sakai, Utility of deep learning for the diagnosis of otosclerosis on temporal bone ct, Eur. Radiol. 31 (2021) 5206–5211, https://doi.org/10.1007/s00330-020-07568-0.
[21] H. Chen, W. Tan, J. Li, P. Guan, L. Wu, B. Yan, J. Li, Y. Wang, Adaptive cross entropy for ultrasmall object detection in computed tomography with noisy labels, Comput. Biol. Med. 147 (2022) 105763, https://doi.org/10.1016/j.compbiomed.2022.105763.
[22] Y. Lou, R. Caruana, J. Gehrke, Intelligible models for classification and regression. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012, pp. 150–158, https://doi.org/10.1145/2339530.2339556.
[23] N. Kumar Manaswi, Understanding and Working with Keras, 2018, pp. 31–43, https://doi.org/10.1007/978-1-4842-3516-4_2 (Chapter 2)).
[24] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, X. Zheng, Tensorflow : Large-scale machine learning on heterogeneous distributed systems. https://doi.org/10.48550/arXiv.1603.04467.
[25] J. Jung, H. Lee, H. Jung, H. Kim, Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: a systematic review, Heliyon (2023 May 8) e16110, https://doi.org/10.1016/j.heliyon.2023.e16110.
[26] Y. Zhang, Y. Weng, J. Lund, Applications of explainable artificial intelligence in diagnosis and surgery, Diagnostics 12 (2) (2022 Jan 19) 237, https://doi.org/10.3390/diagnostics12020237.
[27] S. Ravichandran, R. Duraiswamy, F.S. Arockiasamy, Tool and formability multi-response optimization for ultimate strength and ductility of AA8011 during axial compression, Adv. Mech. Eng. 14 (10) (2022 Oct) 16878132221131731, https://doi.org/10.1177/16878132221131731.
[28] A.F. Sahayaraj, Revolutionizing energy storage: the rise of silicon-based solutions, Silicon 28 (2023 Apr) 1–7, https://doi.org/10.1007/s12633-023-02417-3.
[29] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, in: InProceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.
[30] B. Talukder Uddin, M.M. Khan, A. Zaguia, Study on convolutional neural network to detect covid-19 from chest x-rays, Math. Probl Eng. 2021 (2021) 1–11, https://doi.org/10.1155/2021/3366057.

[31] J. Yogapriya, V. Chandran, M.G. Sumithra, P. Anitha, P. Jenopaul, C. Suresh Gnana Dhas, Gastrointestinal tract disease classification from wireless endoscopy images using pretrained deep learning model, Comput. Math. Methods Med. 2021 (2021) 5940433, https://doi.org/10.1155/2021/5940433.
[32] G. Kompella, M. Antico, F. Sasazawa, S. Jeevakala, K. Ram, D. Fontanarosa, A.K. Pandey, M. Sivaprakasam, Segmentation of femoral cartilage from knee ultrasound images using mask R-CNN, in: In2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019 Jul 23, pp. 966–969, https://doi.org/10.1109/EMBC.2019.8857645.
[33] U. Rashid, A. Javid, A. Rehman, L. Liu, A. Ahmed, O. Khalid, K. Saleem, S. Meraj, U. Iqbal, R. Nawaz, A hybrid mask rcnn-based tool to localize dental cavities from real-time mixed photographic images, PeerJ Computer Science 8 (2022) e888, https://doi.org/10.7717/peerj-cs.888.
[34] I. David, I. Dinc, Classification of protein crystallization images using efficientnet with data augmentation, in: CSBio2020: the 11th International Conference on Computational Systems-Biology and Bioinformatics, 2020, pp. 54–60, https://doi.org/10.1145/3429210.3429220.
[35] R.N. Lazuardi, N. Abiwinanda, T.H. Suryawan, M. Hanif, A. Handayani, Automatic diabetic retinopathy classification with efficientnet, in: TENCON 2020 - 2020 IEEE REGION 10 CONFERENCE (TENCON), 2020, pp. 756–760, https://doi.org/10.1109/TENCON50793.2020.9293941.
[36] J. Deng Russakovsky, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. KarpathyKhosla, M. Bernstein, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.
[37] M.A. Morid, A. Borjali, G. Del Fiol, A scoping review of transfer learning research on medical image analysis using imagenet, Comput. Biol. Med. 128 (2021) 104115, https://doi.org/10.1016/j.compbiomed.2020.104115.
[38] A. Khosla Zhou, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2016, pp. 2921–2929.
[39] Q.A. Al-Haija, A. Adebanjo, Breast cancer diagnosis in histopathological images using resnet-50 convolutional neural network, in: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2020, pp. 1–7, https://doi.org/10.1109/IEMTRONICS51293.2020.9216455.
[40] T. Aswathi, T.R. Swapna, S. Padmavathi, Transfer learning approach for grading of diabetic retinopathy, J. Phys. Conf. 1767 (1) (2021) 012033.
[41] Adriel Saporta, et al., Deep learning saliency maps do not accurately highlight diagnostically relevant regions for medical image interpretation, medRxiv (2021) p.2021.02.28.21252634.
[42] Hyo-Eun Kim, et al., Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study, The Lancet Digital Health 2 (3) (2020) e138–e148.
[43] Damir Vrabac, et al., DLBCL-Morph: morphological features computed using deep learning for an annotated digital DLBCL image set, Sci. Data 8 (1) (2021) 135.
[44] M. Tan, Q. Le, Efficientnet: rethinking model scaling for convolutional neural networks. InInternational Conference on Machine Learning, PMLR, 2019 May 24, pp. 6105–6114.