



OPEN ACCESS

EDITED BY

Guo-Wei Wei,
Michigan State University, United States

REVIEWED BY

Nenad Filipovic,
University of Kragujevac, Serbia
Ali Kouhi,
Tehran University of Medical Sciences, Iran
Menglun Wang,
United States Food and Drug Administration,
United States

*CORRESPONDENCE

Qiyuan Li,
✉ qiyuan.li@xmu.edu.cn
Chengfu Cai,
✉ ysc96@126.com

†These authors share first authorship

RECEIVED 30 June 2023

ACCEPTED 27 December 2023

PUBLISHED 21 March 2024

CITATION

Qi F, You Z, Guo J, Hong Y, Wu X, Zhang D, Li Q and Cai C (2024), An automatic diagnosis model of otitis media with high accuracy rate using transfer learning.
Front. Mol. Biosci. 10:1250596.
doi: 10.3389/fmolb.2023.1250596

COPYRIGHT

© 2024 Qi, You, Guo, Hong, Wu, Zhang, Li and Cai. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An automatic diagnosis model of otitis media with high accuracy rate using transfer learning

Fangyu Qi^{1,2,3†}, Zhiyu You^{2,3†}, Jiayang Guo^{2,3}, Yongjun Hong⁴, Xiaolong Wu^{2,3}, Dongdong Zhang², Qiyuan Li^{2,3*} and Chengfu Cai^{5*}

¹Department of Anesthesiology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China, ²School of Medicine, Xiamen University, Xiamen, China, ³National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China, ⁴Zhongshan Hospital Affiliated to Xiamen University, Xiamen, China, ⁵College of Otorhinolaryngology Head and Neck Surgery, Xiamen Haicang Hospital, Xiamen, China

Introduction: Chronic Suppurative Otitis Media (CSOM) and Middle Ear Cholesteatoma are two common chronic otitis media diseases that often cause confusion among physicians due to their similar location and shape in clinical CT images of the internal auditory canal. In this study, we utilized the transfer learning method combined with CT scans of the internal auditory canal to achieve accurate lesion segmentation and automatic diagnosis for patients with CSOM and middle ear cholesteatoma.

Methods: We collected 1019 CT scan images and utilized the nnUnet skeleton model along with coarse grained focal segmentation labeling to pre-train on the above CT images for focal segmentation. We then fine-tuned the pre-training model for the downstream three-classification diagnosis task.

Results: Our proposed algorithm model achieved a classification accuracy of 92.33% for CSOM and middle ear cholesteatoma, which is approximately 5% higher than the benchmark model. Moreover, our upstream segmentation task training resulted in a mean Intersection of Union (mIoU) of 0.569.

Discussion: Our results demonstrate that using coarse-grained contour boundary labeling can significantly enhance the accuracy of downstream classification tasks. The combination of deep learning and automatic diagnosis of CSOM and internal auditory canal CT images of middle ear cholesteatoma exhibits high sensitivity and specificity.

KEYWORDS

chronic suppurative otitis media (CSOM), middle ear cholesteatoma, CT images, computer-aided diagnosis (CAD), transfer learning (TL)

Introduction

Otitis media is a prevalent ear disease that affects a significant portion of the global population, with an estimated 65 to 350 million individuals affected worldwide ([World Health Organization, 2004](https://www.who.int/news-room/fact-sheets/detail/otitis-media)). In developing countries, the prevalence of Chronic Suppurative Otitis Media (CSOM) ranges from 0.4% to 33.3% ([Kaur et al., 2017](https://doi.org/10.1155/2017/1234567)). Our study mainly focuses on non-invasive temporal bone CT images, in order to help clinicians quickly get a relatively accurate preliminary diagnosis and lay the foundation for further judgment of whether patients need surgical treatment.

Otitis media is classified into three categories: acute otitis media, chronic otitis media (COM), and middle ear cholesteatoma. Chronic otitis media typically exhibits more pronounced pathological changes on CT images due to its protracted course, whereas acute otitis media does not usually display this characteristic. Consequently, it is often recommended that patients with chronic otitis media undergo internal auditory canal CT scan to assess their condition. Chronic otitis media is further divided into two subcategories: chronic non-suppurative otitis media and chronic suppurative otitis media (CSOM) (Schilder et al., 2017). CSOM and Middle Ear Cholesteatoma are two typical otitis media diseases that are diagnosed primarily through temporal bone CT scans (Fukudome et al., 2013; Lustig et al., 2018). CSOM typically occurs following improper treatment of acute otitis media, often resulting in tympanic membrane perforation and persistent middle ear purulence (Ahmad et al., 2022). Middle ear cholesteatoma, on the other hand, is the pathological outcome of abnormal accumulation of keratin squamous epithelium, primarily composed of keratinized, exfoliated epithelium. It often accumulates in the middle ear, with a tendency to erode the ossicular chain, tympanic wall, and/or mastoid area (Sun et al., 2011; Jang et al., 2014).

Clinicians usually identify CSOM and middle ear cholesteatoma through CT scanning of the internal auditory canal. However, CT reports of these two types both show erosion and/or loss of the ossicular chain with diffuse abnormal soft tissue shadow (Madabhushi and Lee, 2016). Theoretically, the two types differ in bone erosion margins and soft tissue shadow contours: The soft tissue shadow of cholesteatoma has a smooth, clear outline, while that of CSOM lacks a clear outline and is often accompanied by pus accumulation. In addition, the edge of the bone erosion caused by CSOM is serrated, while bone destruction caused by cholesteatoma is frequently surrounded by a ring of sclerosis. Therefore, our group proposed using deep learning to differentiate between CSOM and cholesteatoma to achieve more accurate clinical diagnoses based on the theoretical differences between these two diseases (Kemppainen et al., 1999; Yorgancilar et al., 2013).

Deep learning techniques have seen widespread use in the medical field in recent years, enabling the extraction of key features from patients to facilitate predictive modeling (Elfiky et al., 2018). Transfer learning is a deep learning technique that involves leveraging knowledge gained from solving one problem to address another related problem. It is particularly useful when the amount of labeled data for the target task is limited. By leveraging transfer learning, a pre-trained model developed for one task can be fine-tuned and adapted for another task with different data but similar features. This approach enables the model to benefit from the knowledge learned by the pre-trained model on a larger dataset and adapt it to the new task by making only minor adjustments to the model's architecture or parameters. A bunch of researches have uncovered that superiority of transfer learning over traditional strategy. S. Deepak implement brain tumor classification using deep CNN features via transfer learning (Deepak and Ameer, 2019). For tuberculosis detection, a VGGNet based model had been proposed combining transfer learning (Ahsan et al., 2019). Dube S presented an automatic content-based image retrieval system for brain tumors on contrast-enhanced MRI (Dube et al., 2006).

Given the small morphological differences between various types of otitis media, the challenge of manual identification, and the unclear contour of lesions, we sought to establish a transfer learning framework by integrating coarse-grained labeled contour information as pre-trained data and employing the CNN model

skeleton to extract high-level features from internal auditory canal CT images. Specifically, the deep representation of images obtained through pre-training was utilized to accurately classify CSOM, middle ear cholesteatoma, and normal samples in downstream tasks.

In conclusion, this paper's key contributions are:

1. We propose a transfer learning-based framework that utilizes coarsely annotated segmentation data as input for pretraining the model. The proposed model effectively extracts implicit information from the data and can subsequently be used for classification prediction.
2. We propose an end-to-end learning model that can effectively improve the accuracy of middle ear infection classification prediction. Our proposed model outperforms non-pretrained models in all metrics.
3. In the field of deep learning combined with medicine, we look forward to replacing the heavy and repetitive manual labeling task with more mature machine automated labeling.
4. Our combination of otological diseases and computer learning can increase the coverage of related research and provide more precise and diversified help for clinicians in diagnosis and treatment.

Materials and methods

Data acquisition

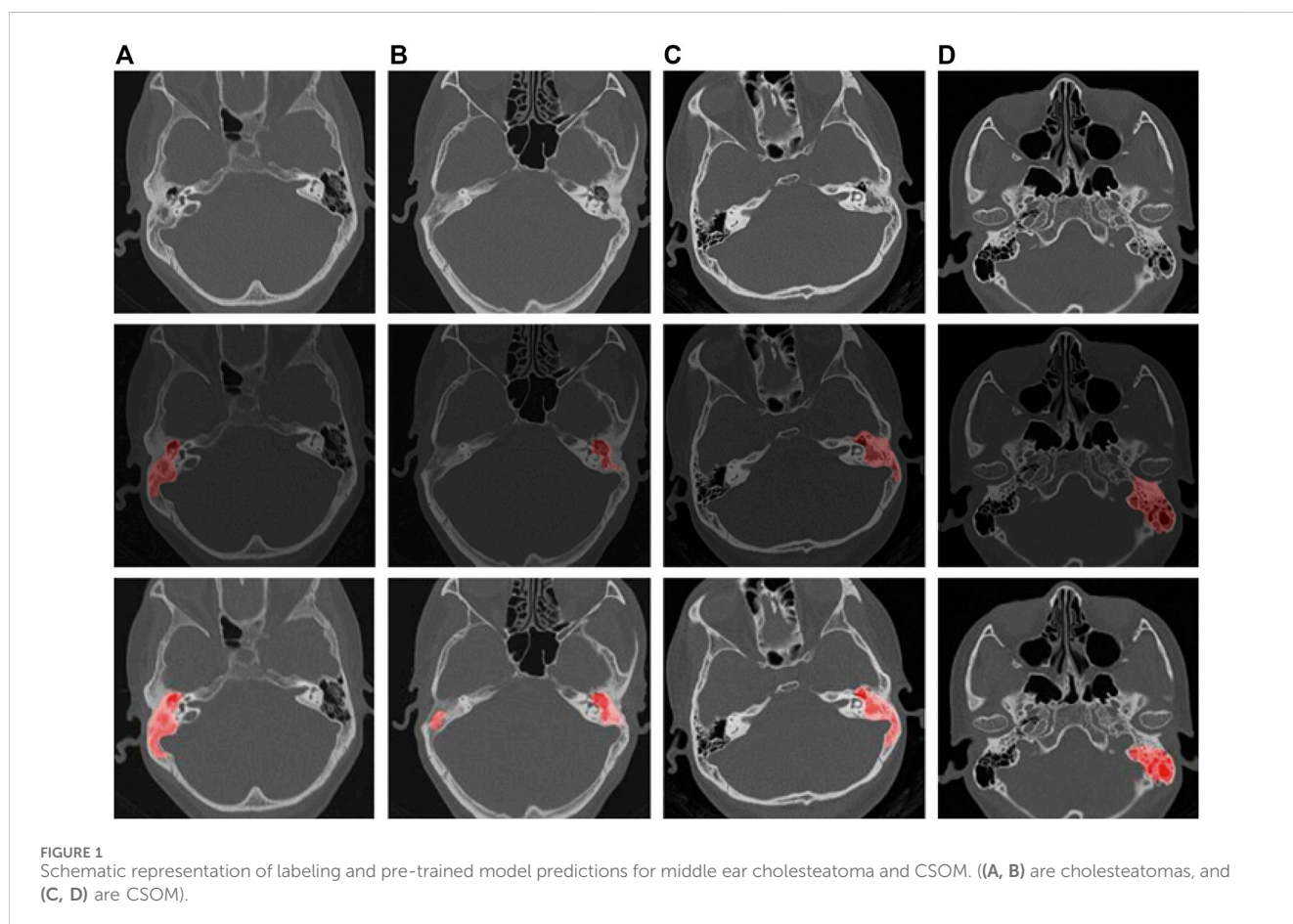
We conducted a retrospective study at Zhongshan Hospital Affiliated to Xiamen University to investigate patients diagnosed with otitis media and middle ear cholesteatoma from 2012 to 2021. This study was approved by the ethics committee and informed consent was waived due to the retrospective nature of the study. The inclusion criteria for the study were based on pathology or medical history, ear examination, audiogram, and imaging examination of the surgical side of the ear. We referred to the previous medical records of the hospital to obtain specific diagnosis results. A total of 6,967 axial high-resolution CT images of temporal bone were collected from 180 patients, including 27 female patients with middle ear cholesteatoma, 31 male patients with middle ear cholesteatoma; 50 female patients and 51 male patients with chronic otitis media, including 40 females and 39 males with CSOM, and Chronic otitis media with effusion, including 10 females and 12 males. Besides, there were 15 normal controls (6 females and 9 males), and 2 children with middle ear cholesteatoma, 2 children with CSOM (less than 10 years old), and 2 normal controls. Except for children, the patients collected were between 30 and 80 years of age. All the patients information has been shown in Table 1.

The CT features of chronic otitis media with effusion (COME) are often very similar to those of chronic suppurative otitis media (CSOM), with both conditions typically presenting with varying degrees of effusion in mastoid cells. As such, we excluded a total of 412 axial CT data from 22 patients diagnosed with COME. For each patient, we selected approximately 10–20 CT images that showed well-defined lesions. Ultimately, we used 410 CSOM CT images, 398 middle ear cholesteatoma CT images, and 211 normal CT control images for our analysis.

TABLE 1 The collected patients information.

	Female (30–80 years old)	Male (30–80 years old)	Pediatric patients (<10 years old)
Middle Ear Cholesteatoma	27	31	2
CSOM	40	39	2
Chronic Otitis Media with Effusion	10	12	0
Normal	6	9	2
Total	83	91	6

Data sources: Zhongshan Hospital Affiliated to Xiamen University.
 Note: numbers represent the quantity of patients.



CT scanner settings

Temporal bone CT is derived from GE LightSpeed 64-row volume CT. Its detector is the core technology of multi-slice spiral CT. The detector arrangement adopts 64×0.625 mm detector unit to ensure the maximum coverage of 40 mm/circle at present, and at the same time, it can also perform sub-millimeter thick scanning in any mode. The isotropic resolution is up to 0.30 mm, which ensures a large range of volume acquisition and high resolution acquisition. The CT imaging parameters used were as follows: CT collimator 128×1.0 mm, field of view 220×220 mm, matrix size $1,024 \times 1,024$, voltage 120 kV, current 240 mAs, and axial CT slice number 30–50 per scan.

Data marking

The CT findings of chronic suppurative otitis media are often difficult to distinguish from middle ear cholesteatoma. To accurately identify middle ear cholesteatomas, we marked local or isolated cholesteatomas in the erosion area of the incudostapedial joint or hammer-incus joint in the rotation plane of the middle tip of the cochlea. We also highlighted areas of bone destruction within the tympanic sinus, epitympanic region, or mastoid process in other levels. Additionally, we marked any irregular soft tissue shadows with smooth edges on any plane. In contrast, when marking CT images of suppurative otitis media, we identified the soft tissue shadow around the auricle in the tympanic cavity, the sclerotic

hyperplasia part of the mastoid, and the bone with uniform density and serrated edge in the tympanic sinus at the vestibular level. These characteristics helped distinguish them from the sclerosing ring that is formed by the compression of a cholesteatoma. At the apical spiral layer of the cochlea, we marked the hammer-incus joint of the ossified epitympanic, the “ice cream cone-like structure,” and the serrated bone around the mastoid cavity and sinus. At the bottom spiral level of the cochlea, we marked the thickened mucosa on the promontory surface. Finally, at the mastoid level, we marked the erosion of the mastoid bone and the thickening of the mucous membrane caused by suppurative effusion (Gomaa et al., 2013; Zelikovich, 2004). We provide visual examples of typical lesion markers and predictions for these two diseases in Figure 1.

The original data was stored in the Dicom format, which we converted into PNG image data using MicroDicom software. This step allowed us to separate the patient’s personal information from the image, thereby ensuring patient privacy.

To mark the lesion area on each image, we enlisted the help of a team consisting of five professional otolaryngologists and two radiologists. They used polygonal markers on LabelMe software to eliminate background interference and generate unified coordinates for each lesion area.

Data pre-processing

To improve the robustness of our model, we applied various image data augmentation and processing techniques during the training process. Specifically, we randomly transformed the input images by performing horizontal and vertical translations, flipping, rotation, slight scaling, and adjustments to hue, contrast, and numerical values.

In order to balance the size of our training model and the time required for training, we scaled all images to a uniform size of 224×224 using bilinear interpolation. This allowed us to efficiently process and train on a large dataset of images while still maintaining a high level of accuracy and performance in our final model.

Model architecture and training strategy

We utilized the nnUnet (Isensee et al., 2021) architecture as the foundation for our deep learning model to extract critical features from CT images. This model has demonstrated exceptional performance in various medical image segmentation tasks. Our model consists of two branches: coarse-grained segmentation task and exact classification task.

In our workflow, we first pre-trained a model for lesion segmentation, which includes the nnUnet skeleton and a pixel-level prediction head that outputs three classification results for each pixel: CSOM, middle ear cholesteatoma, or normal samples. On the back of the above-mentioned process, our goal was to acquire a well-trained backbone that could extract underlying information containing pixel-level features, which would then be fine-tuned for picture-level classification. We trained this model using the gradient descent algorithm until convergence was achieved.

The nnUnet model is composed of an encoder and a decoder. The encoder reduces the image size layer by layer while capturing features of varying granularity from different images. It consists of seven layers, each containing {1, 3, 4, 6, 6, 6, 6} blocks, with each block containing two convolutional layers, two activation layers, and two normalization layers. Successive layers are directly connected with a pooling layer, which reduces the image size by half. The first layer of the encoder contains 32 features, and the number of features in each subsequent layer doubles but does not exceed the maximum number of features, which is 512.

The decoder has six layers, each consisting of {2, 2, 2, 2, 2, 2} blocks. These layers use linear interpolation upsampling to increase the image size. The encoder and decoder are connected using residual layer hopping. The decoder outputs the hidden variables of the image as inputs to both the pixel-level projection head and the image-level projection head for further processing.

Once the model was successfully trained, we extracted the hidden variables before the pixel-level prediction head of the image as inputs for the downstream classification prediction head. This allowed us to efficiently classify images with high accuracy by leveraging the previously extracted features.

We employed a five-fold cross-validation approach to train and evaluate our model (see Figure 2). Given that a series of adjacent CT images from the same subject tend to exhibit strong similarities, we took care to avoid overfitting due to data leakage. Specifically, during the training process, we randomly divided the dataset into a training set and validation set at a ratio of 4:1, based on the subject’s name. This ensured that CT images from the same subject were assigned to either the training or validation set, but not both.

Our models were compiled using Python 3.8, trained with PyTorch version 1.10, and accelerated with Nvidia A100 high-performance GPUs. During the training process, we set the maximum training epoch to 500 epochs, with a training batch size of 8 samples. We used Adam as the model optimizer, with an initial learning rate of 0.001. The dynamic learning rate decreased gradually with each increase in training batch until it reached $10e-5$. For pre-training optimization, we utilized the cross-entropy of each pixel classification of the image. Downstream training utilized the cross-entropy of the image classification as the loss function. Our specific workflow is depicted in Figure 3 for further visualization.

(The pre-training phase uses the pixel-level labels of route A to train the CNN, and the pixel predictor is responsible for output the category of each pixel. When performing the downstream picture classification task, according to route B, the pre-trained model is used to fine-tune the neural network through the category classifier, and the final picture prediction result is output.)

Result

In the upstream segmentation task, our deep learning model achieved a mean Intersection of Union (mIoU) index of 0.5376, indicating excellent performance in accurately removing background noise. Subsequently, we employed this well-performing model for the downstream fine-tuning step, where we aimed to classify otitis media into three distinct categories. Our model achieved a micro-f1 index of 92.33%, a significant improvement of 4.83% compared to the benchmark model.

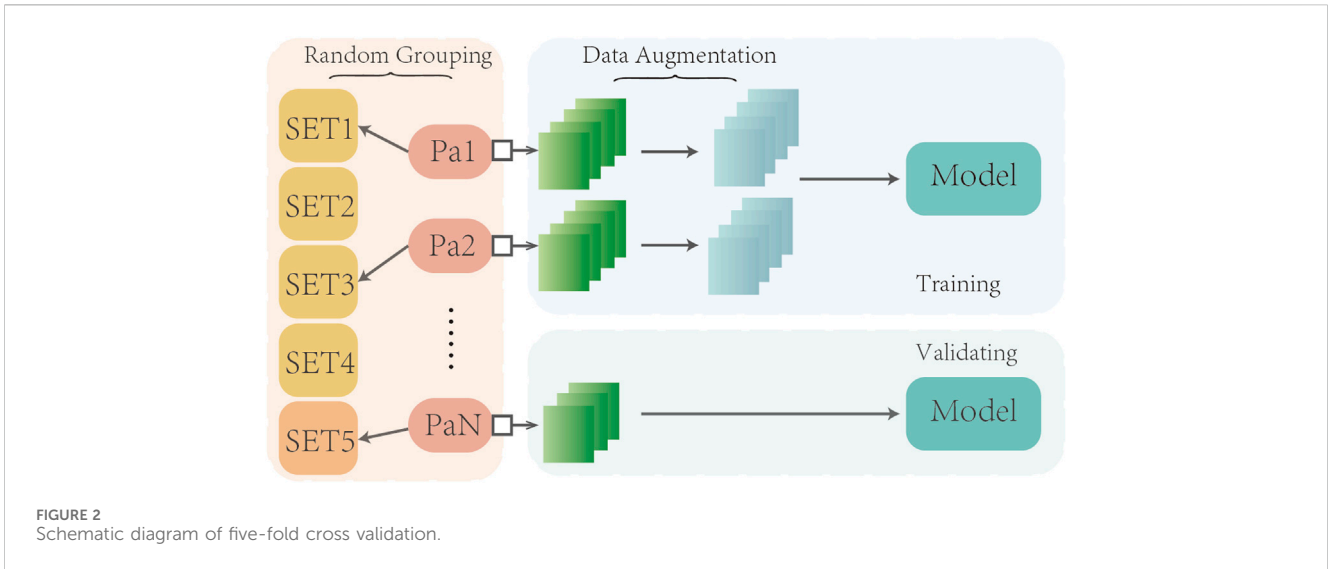
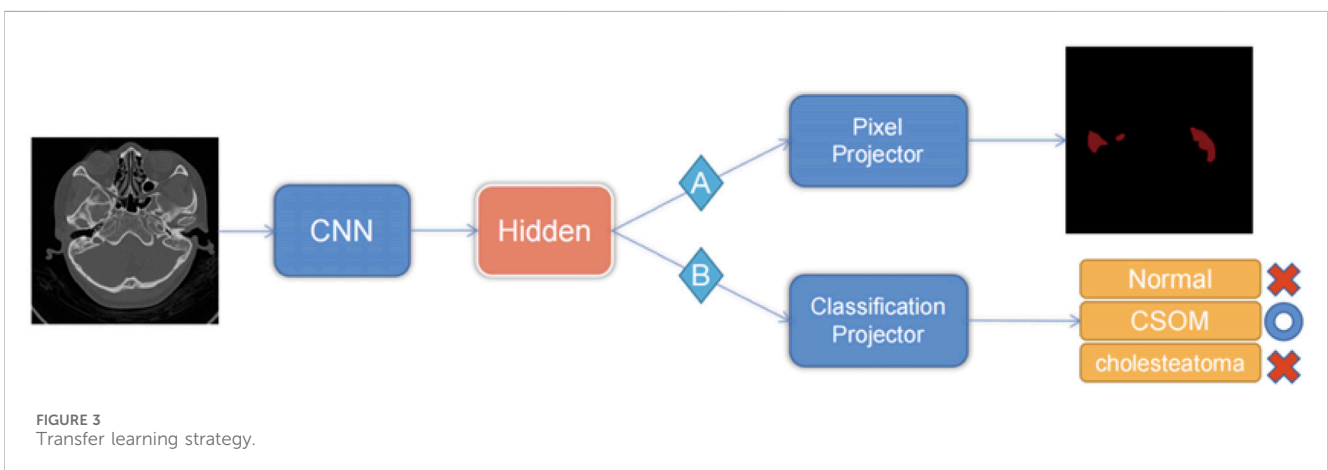


TABLE 2 Comparison of results.

		Accuracy	Auc
nnUnet	Normal vs. the others	0.8732 ± 0.1694	0.9558 ± 0.0454
	CSOM vs. the others	0.8619 ± 0.0776	0.9695 ± 0.0212
	Cholesteatoma vs. the others	0.9013 ± 0.0394	0.9555 ± 0.0242
p_nnUnet	Normal vs. the others	0.8873 ± 0.1593	0.9530 ± 0.0526
	CSOM vs. the others	0.9434 ± 0.0343	0.9916 ± 0.0069
	Cholesteatoma vs. the others	0.9290 ± 0.0170	0.9622 ± 0.0316

p represents the fine-tuning results after using the pre-trained model. The above results are the means after five-fold cross-validation. The mIoU index is used to describe the average ratio of intersection and union of all pixel categories in the image segmentation task. In this experiment, the background normal tissue categories were removed to obtain more accurate prediction results. mIoU is described as follows.



On the other hand, the pre-trained model exhibits an overall area under the receiver curve of 0.9689, which is slightly higher than that of the benchmark model which reach 0.9603. As is depicted in Figure 4, it can be observed that the performance in distinguishing chronic suppurative otitis media (CSOM) is the best, by a margin of 8.15%, as is showed in Table 2. These results

indicate that the pre-trained model has a superior ability to accurately classify CSOM cases compared to the benchmark model. These results highlight the potential of deep learning technology in medical image analysis and its ability to significantly improve diagnostic accuracy and treatment outcomes.

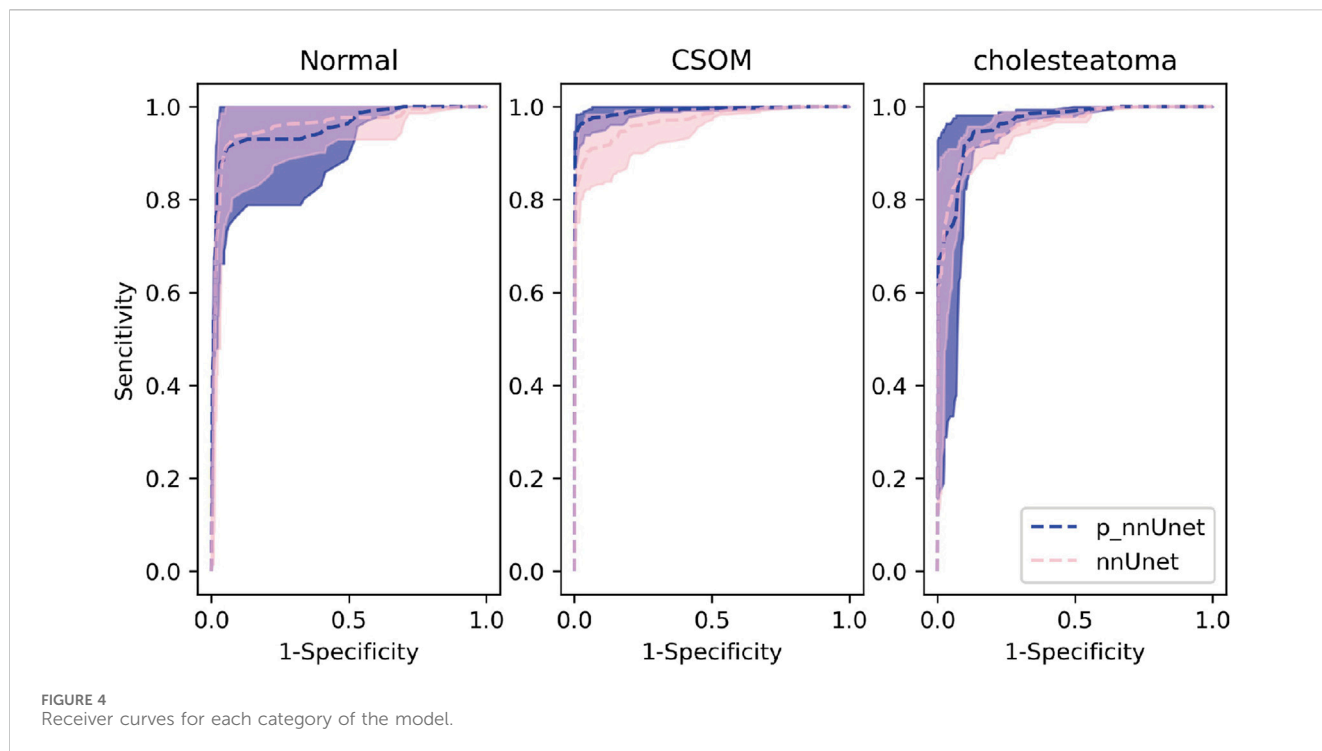


TABLE 3 Diagnostic accuracy results of manual diagnostic group.

The clinical experts (n = 3)/Real diagnose (n = 3,039)	COSM (n = 1,182)	Cholesteatoma (n = 1,224)	Normal (n = 633)
CSOM	948 (80.20%)	489 (39.95%)	48 (7.58%)
Cholesteatoma	222 (18.78%)	708 (57.84%)	9 (1.42%)
Normal	12 (1.01%)	27 (2.20%)	576 (90.99%)

The horizontal represents the actual diagnosis results, while the vertical represents the manual diagnosis results.

$$mIoU = \frac{1}{|D|} \sum_{i \in D} \frac{TP_i}{TN_i + FP_i + FP_i}$$

Note: In the above equation, FN is the false negative class, FP is the false positive class, TP is the true class, and D represents the CSOM and cholesteatoma set.

On the other hand, we demonstrate the disparity in accuracy between AI models and manual diagnosis by clinicians. We intentionally randomized and combined a total of 1013 CT images of middle ear cholesteatoma, CSOM, and normal control images, while effectively concealing the actual diagnostic labels. These images were then distributed to both the manual diagnosis group and the model diagnosis group for a double-blind evaluation of diagnostic accuracy. The comprehensive test results are shown in Tables 3, 4, and it can be observed that CSOM and cholesteatoma exhibit a high misdiagnosis rate.

Our experimental results show that in the otitis media classification task, the use of contour boundary labeling can well improve the accuracy of downstream classification tasks, and the area under the receiver operating characteristic curve is better than that of the non-pre-trained model shown in Figure 4. These results indicate that the predictive power of our model on this task has the possibility of real-world application.

(p-represents the fine-tuning results after using the pre-trained model. In the multi-classification ROC curve, the positive samples belong to a particular category while the negative samples belong to all other categories combined. Based on this distinction, the true positive rate and false positive rate have been accurately calculated).

Discussion

Otitis media is characterized by a prolonged course of illness, high incidence, easy recurrence, conductive deafness, and potentially fatal intracranial infection (Otten and Grote, 1990; Hutz et al., 2018). A case analysis conducted in a public hospital in the United States revealed that the incidence of postoperative complications associated with complex chronic otitis media with middle ear cholesteatoma was similar to that observed in developing regions (Greenberg and Manolidis, 2001). As a result, early diagnosis, intervention measures, and clinical management of this disease are especially crucial, regardless of whether one resides in developed or developing regions. In our study, we utilized CT images of CSOM and middle ear cholesteatoma labeled by

TABLE 4 Diagnostic accuracy results of model diagnosis group.

The AI model/Real diagnose (<i>n</i> = 1,013)	COSM (<i>n</i> = 394)	Cholesteatoma (<i>n</i> = 408)	Normal (<i>n</i> = 211)
COSM	336 (85.28%)	16 (3.92%)	7 (9.52%)
Cholesteatoma	44 (11.17%)	336 (82.35%)	2 (0.95%)
Normal	14 (3.55%)	20 (4.90%)	188 (89.52%)

The horizontal represents the actual diagnosis results, while the vertical depicts the diagnostic outcomes of the deep learning model.

medical experts as the training set for our algorithmic model. Our algorithm model accurately predicted unlabeled CT images with a high degree of precision, achieving excellent agreement between predicted lesion types and actual clinical findings.

So far, CT scan and Endoscopy of ear, as the classical methods for the diagnosis of various types of otitis media, are still the latest diagnostic methods (Gomaa et al., 2013; Zelikovich, 2004). The golden standard for the diagnosis of CSOM and middle ear cholesteatoma is intraoperative histopathological examination. However, it takes a long time to make a preliminary diagnosis of one single patient. Despite efforts to reduce the prevalence of Chronic Suppurative Otitis Media (CSOM) in underdeveloped areas, clinical diagnosis, treatment, and prognosis of the disease remain suboptimal. Recent epidemiological investigations have shown that CSOM has shifted to a population dominated by adults, despite a decrease in overall prevalence during the past 2 decades (Orji et al., 2016). Additionally, an Australian survey highlighted the high incidence of CSOM and middle ear cholesteatoma among impoverished individuals and the need for early diagnosis (Benson and Mwanri, 2012).

Deep learning has emerged as a valuable tool in various medical fields. A substantial amount of research on deep learning applied to clinical datasets, using high-quality medical examination images, has showcased its efficacy in defining patient categories, identifying and locating lesions, and other relevant tasks (Wang et al., 2020). With our transfer learning model, medical researchers can avoid the time-consuming and resource-intensive process of training models from scratch, while also benefiting from the wealth of knowledge captured in existing non-medical datasets. In the field of computational vision, pre-trained models have become a commonly used tool in many applications, particularly in addressing medical imaging challenges. These challenges can arise from imaging modalities such as X-ray, Magnetic Resonance Imaging (MRI), CT scan, and Ultrasound data. Many works have demonstrated the potential of pre-trained models to improve diagnostic accuracy, reduce processing time, and assist in the development of automated diagnosis systems. Our transfer learning model could also be used in other diseases which need CT scan or endoscope or any examinations that take images as the method to diagnose. Once the medical examination images are too similar to find the differences, our model could give several suggestions in differential diagnosis based on the previous history image labels.

Among the different types of otitis media, there are varying methods for diagnosis and treatment. However, the CT image features tend to be similar across these types, which can pose challenges for clinicians in terms of differential diagnosis. Such challenges can lead to delays in proper treatment, and potentially result in errors or overmedication. Moreover, the COVID-19

patients were found to have relationship with Otitis media (Choi et al., 2022), they demand to be diagnosed earlier than before, as otitis media always intend to recurrence and even cause Sensorineural-hearing-loss (Xia et al., 2022). As a result, achieving rapid differential diagnosis for otitis media is crucial to ensure optimal patient outcomes, in this situation, an efficient diagnosis can be given using our transfer learning model.

However, our transfer learning models have shown some limitations in classifying certain ear diseases. For instance, when differentiating between secretory otitis media and suppurative otitis media, deep learning models tend to confuse the two because their CT scans are very similar. This could be attributed to inadequate data sets. As such, there is a need for more comprehensive and diverse medical data to improve the accuracy of diagnostic models used to differentiate between various ear diseases. Moreover, our model showed significant differences in lesion information extraction. For instance, some predicted lesions would perform fewer or more lesions compared to those marked by medical experts. Also, in images with unclear lesions, there were discrepancies in identifying the lesion. For example, the images of the mastoid layer of chronic suppurative otitis media often have varying degrees of mucosal thickening due to chronic inflammation, while the images of the mastoid layer of chronic secretory otitis media show fluid levels caused by chronic effusion. These conditions are quite similar, with only slight differences in the contour of the mucosal within the mastoid bone. In most cases, our transfer learning networks could detect and label prominent lesions such as large soft tissue shadow of middle ear cholesteatoma, eroded bone structure surrounded by soft tissue shadow, and eroded bone structure of chronic suppurative otitis media. Unfortunately, the CT images of chronic otitis media with effusion (COME) do not have the typical erosive features of CSOM and middle ear cholesteatoma. Therefore, our research group excluded the CT images of chronic secretory otitis media and focused solely on collecting CT images of middle ear cholesteatoma and chronic suppurative otitis media as the objects of our study. In the future, when the amount of data collection is large enough, we will continue to promote the application of new migration models to this type of classification project.

What's more, due to the limited clinical applications of deep learning and the laborious, time-consuming nature of acquiring supervised data such as lesion regions, our research group aims to identify alternative weakly supervised signals for model transfer learning pre-training. "Human-in-the-loop" is an effective interactive mode between doctors and models, which can provide weakly supervised signals and ensure continuous learning of the model. This approach also represents a practical scenario for the clinical application of the model. This can help reduce manual labeling costs and improve overall prediction performance. In the

future, we hope to replace the repetitive, cumbersome task of manual labeling with more advanced machine automated labeling techniques in the field of deep learning combined with medicine.

Regarding the types of otological diseases combined with deep learning models, there are few studies on using detection results of acoustic immittance, acoustic reflex, and pure tone hearing threshold to achieve accurate predictions, prognosis, and treatment. Additionally, while otitis media and vertigo have received significant research attention, other diseases such as otosclerosis, ear tumors, and sudden neurotropic hearing loss remain understudied. Future research on combining otological diseases with computer learning may increase the coverage of relevant studies and provide clinicians with more precise and diverse tools for diagnosis and treatment.

Data availability statement

The datasets presented in this article are not readily available because The data is not publicly available. Requests to access the datasets should be directed to moon dancer122@163.com.

Ethics statement

The studies involving humans were approved by the Scientific Research Sub-Committee of Medical Ethics Committee of Zhongshan Hospital Affiliated to Xiamen University. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because The data collected were CT images of historical medical records, which were non-invasive medical examination images. Written informed consent was not obtained from the minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article because The data collected were historical medical record CT images, which were

References

- Ahmad, R. U., Ashraf, M. F., Qureshi, M. A., Shehryar, M., Tareen, H. K., and Ashraf, M. A. (2022). Chronic Suppurative Otitis Media leading to cerebellar brain abscess, still a problem in 21st century: a case report. *Ann. Med. Surg. (Lond)*. 80, 104256. doi:10.1016/j.amsu.2022.104256
- Ahsan, M., Gomes, R., and Denton, A. (2019). "Application of a Convolutional Neural Network using transfer learning for tuberculosis detection," in 2019 IEEE International Conference on Electro Information Technology (EIT), USA, 20-22 May 2019 (IEEE). doi:10.1109/EIT.2019.8833768
- Benson, J., and Mwanri, L. (2012). Chronic suppurative otitis media and cholesteatoma in Australia's refugee population. *Aust. Fam. Physician* 41 (12), 978–980. doi:10.3316/informit.998418740667479
- Choi, S. Y., Yon, D. K., Choi, Y. S., Lee, J., Park, K. H., Lee, Y. J., et al. (2022). The impact of the COVID-19 pandemic on otitis media. *Viruses* 14 (11), 2457. doi:10.3390/v14112457
- Deepak, S., and Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Comput. Biol. Med.* 111, 103345. doi:10.1016/j.combiomed.2019.103345
- Dube, S., Elsadon, S., Cloughesy, T. F., and Sinha, U. Content based image retrieval for MR image studies of brain tumors[J]. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 2006, 1(1): 3337–3340. doi:10.1109/IEMBS.2006.260262
- Elfiky, A. A., Pany, M. J., Parikh, R. B., and Obermeyer, Z. (2018). Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy. *JAMA Netw. open* 1 (3), e180926. doi:10.1001/jamanetworkopen.2018.0926
- Fukudome, S., Wang, C., Hamajima, Y., Ye, S., Zheng, Y., Narita, N., et al. (2013). Regulation of the angiogenesis of acquired middle ear cholesteatomas by inhibitor of DNA binding transcription factor. *JAMA Otolaryngology-Head Neck Surg.* 139 (3), 273–278. doi:10.1001/jamaoto.2013.1750
- Gomaa, M. A., Abdel Karim, A. R., Abdel Ghany, H. S., Elhiny, A. A., and Sadek, A. A. (2013). Evaluation of temporal bone cholesteatoma and the correlation between high resolution computed tomography and surgical finding. *Clin. Med. Insights Ear Nose Throat* 6, 21–28. Published 2013 Jul 23. doi:10.4137/CMENT.S10681
- Greenberg, J. S., and Manolidis, S. (2001). High incidence of complications encountered in chronic otitis media surgery in a U.S. metropolitan public hospital. *Otolaryngol. Head. Neck Surg.* 125 (6), 623–627. doi:10.1067/mhn.2001.120230
- Hutz, M. J., Moore, D. M., and Hotaling, A. J. Neurological complications of acute and chronic otitis media. *Curr. Neurology Neurosci. Rep.*, 2018, 18(3): 11–17. doi:10.1007/s11910-018-0817-7

non-invasive medical examination images. All patients had signed waivers of informed consent with the ethics committee of the hospital, and the identity information of all patients was concealed.

Author contributions

QF performed the major jobs including literature search and data marking. YZ provided model training. All authors contributed to the article and approved the submitted version.

Funding

The project was supported by the Natural Science Foundation of Fujian Province of China (No. 2022J05006), Natural Science Foundation of Fujian Science and Technology (No. 2020J02060), Key Medical and Health Project of Xiamen Science and Technology Bureau (No. 3502Z20204009), Key Research and Development (Digital Twin) Program of Ningbo City (No. 2023Z219) and the Fundamental Research Funds for the Chinese Central Universities (No. 0070ZK1096 to JG).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. methods* 18 (2), 203–211. doi:10.1038/s41592-020-01008-z
- Jang, C. H., Choi, Y. H., Jeon, E. S., Yang, H. C., and Cho, Y. B. (2014). Extradural granulation complicated by chronic suppurative otitis media with cholesteatoma. *vivo* 28 (4), 651–655.
- Kaur, I., Goyal, J. P., and Singh, D. (2017). Prevalence of chronic suppurative otitis media in school going children of patiala district of Punjab, India. *J. Evol. Med. Dent. Sci.* 75, 5402–5407. doi:10.14260/jemds/2017/1171
- Kemppainen, H. O., Puhakka, H. J., Laippala, P. J., Sipilä, M. M., Manninen, M. P., and Karma, P. H. (1999). Epidemiology and aetiology of middle ear cholesteatoma. *Acta oto-laryngologica* 119 (5), 568–572. doi:10.1080/00016489950180801
- Lustig, L. R., Limb, C. J., Baden, R., et al. (2018). *Chronic otitis media, cholesteatoma, and mastoiditis in adults*. MA: UpToDate Waltham. (citirano 145 2019).
- Madabhushi, A., and Lee, G. (2016). Image analysis and machine learning in digital pathology: challenges and opportunities. *Med. image Anal.* 33, 170–175. doi:10.1016/j.media.2016.06.037
- Orji, F. T., Ukaegbe, O., Alex-Okoro, J., Ofoegbu, V. C., and Okorafor, I. J. (2016). The changing epidemiological and complications profile of chronic suppurative otitis media in a developing country after two decades. *Eur. Arch. Otorhinolaryngol.* 273 (9), 2461–2466. Epub 2015 Nov 26. PMID: 26611685. doi:10.1007/s00405-015-3840-1
- Otten, F. W. A., and Grote, J. J. (1990). Otitis media with effusion and chronic upper respiratory tract infection in children: a randomized, placebo-controlled clinical study. *Laryngoscope* 100 (6), 627–633. doi:10.1288/00005537-199006000-00014
- Schilder, A. G. M., Marom, T., Bhutta, M. F., Casselbrant, M. L., Coates, H., Gisselsson-Solén, M., et al. (2017). Panel 7: otitis media: treatment and complications. *Otolaryngology-Head Neck Surg.* 156 (4_Suppl.1), S88–S105. doi:10.1177/0194599816633697
- Sun, X. W., Zhang, J. J., Ding, Y. P., Dou, F. f., Zhang, H. b., Gong, K. b., et al. (2011). Efficacy of high-resolution CT in differential diagnosis of chronic suppurative otitis media and cholesteatoma otitis media by soft-tissue shadows. *Zhonghua er bi yan hou tou jing wai ke za zhi= Chin. J. Otorhinolaryngology Head Neck Surg.* 46 (5), 388–392. (Language: Chinese).
- Wang, Y. M., Li, Y., Cheng, Y. S., He, Z. Y., Yang, J. M., Xu, J. H., et al. Deep learning in automated region proposal and diagnosis of chronic otitis media based on computed tomography. *Ear Hear.*, 2020, 41(3): 669–677. doi:10.1097/AUD.0000000000000794
- World Health Organization (2004). *Chronic suppurative otitis media: burden of illness and management options*.
- Xia, A., Thai, A., Cao, Z., Chen, X., Chen, J., Bacacao, B., et al. (2022). Chronic suppurative otitis media causes macrophage-associated sensorineural hearing loss. *J. Neuroinflammation* 19 (1), 224. doi:10.1186/s12974-022-02585-w
- Yorgancılar, E., Yıldırım, M., Gun, R., Bakir, S., Tekin, R., Gocmez, C., et al. (2013). Complications of chronic suppurative otitis media: a retrospective review. *Eur. Archives Oto-rhino-laryngology* 270 (1), 69–76. doi:10.1007/s00405-012-1924-8
- Zelikovich, E. I., Vozmozhnosti, K. T., and Visochnoi Kosti, V. (2004). diagnostike khronicheskogo gnoinogo srednego otita i ego oslozhnenii [Potentialities of temporal bone CT in the diagnosis of chronic purulent otitis media and its complications]. *Vestn. Rentgenol. Radiol.* 1, 15–22. (Language: Russian).