



Structure-constrained deep feature fusion for chronic otitis media and cholesteatoma identification

Cong Cao¹ · Jian Song^{2,3,4} · Ri Su¹ · Xuewen Wu^{2,3,4} · Zheng Wang⁵ · Muzhou Hou¹ 

Received: 18 November 2022 / Revised: 19 March 2023 / Accepted: 18 April 2023 /

Published online: 4 May 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Chronic suppurative otitis media (CSOM) and middle ear cholesteatoma (MEC) were two most common chronic middle ear disease(MED) clinically. Accurate differential diagnosis between these two diseases is of high clinical importance given the difference in etiologies, lesion manifestations and treatments. The high-resolution computed tomography (CT) scanning of the temporal bone presents a better view of auditory structures, which is currently regarded as the first-line diagnostic imaging modality in the case of MED. In this paper, we first used a region-of-interest (ROI) network to find the area of the middle ear in the entire temporal bone CT image and segment it to a size of 100*100 pixels. Then, we used a structure-constrained deep feature fusion algorithm to convert different characteristic features of the middle ear in three groups as suppurative otitis media (CSOM), middle ear cholesteatoma (MEC) and normal patches. To fuse structure information, we introduced a graph isomorphism network that implements a feature vector from neighbourhoods and the coordinate distance between vertices. Finally, we construct a classifier named the “otitis media, cholesteatoma and normal identification classifier” (OMCNIC). The experimental results achieved by the graph isomorphism network revealed a 96.36% accuracy in all CSOM and MEC classifications. The experimental results indicate that our structure-constrained deep feature fusion algorithm can quickly and effectively classify CSOM and MEC. It will help otologist in the selection of the most appropriate treatment, and the complications can also be reduced.

Keywords Structure-constrained deep feature fusion · Middle ear structure · Computed tomography · Chronic suppurative otitis media · Cholesteatoma and normal (CSOMCN)

Cong Cao and Jian Song contributed equally to this work.

✉ Zheng Wang
w8614@hotmail.com

✉ Muzhou Hou
houmuzhou@sina.com

Extended author information available on the last page of the article

1 Introduction

The ear is a susceptible organ of the human body. One of its primary functions is to use hearing by detecting, transmitting, and converting sounds. Another significant function of the ear is to maintain a sense of balance in our bodies. Like vision, hearing is one of our most important sources of biological information [25, 49].

Anatomically, the ear can be divided into three discriminable parts: the inner, middle, and outer ear. It includes a chain by the malleus, incus, and stapes, collectively called the auditory ossicles. It is a complex system of fluid-filled channels and cavities located deep in the hard, rocky part of the temporal bone [3, 28].

Middle ear diseases (MEDs), with the high incidence, described as lesions in the middle ear, which play a vital role in otorhinolaryngology clinical practice [10, 42]. It is the primary reason for hearing loss in many developing countries around the world [18, 34], and chronic suppurative otitis media (CSOM) and middle ear cholesteatoma (MEC) are two common types seen in clinical practice. Both of them seriously affect the patients' quality of life and increase the burden of life [9, 29, 47]. Early diagnosis and effective treatment are the clinical goals of Otolaryngologists at present.

Otitis media (OM) is an inflammatory disease that affects all or part of the middle ear. That can also occur in the middle ear but does not usually damage the surrounding bony structures, presenting a filling phenomenon in CT scans [41]. In this paper, we mainly focus on chronic suppurative otitis media (CSOM), the most common type of otitis media, characterized by persistent repeated inflammation and complications of the tympanum or mastoid cavity [2, 8, 27]. In the CT images, the middle ear structure of patients with CSOM is relatively complete, but there are still some fillers can be found in the tympanic and mastoid cavity [30, 45].

Middle ear cholesteatoma (MEC), an uncontrolled expanding growth of the epithelium tissues surrounded by an inflammatory reaction, is another common type of middle ear disease. Although these lesions are benign, they can erode the normal bone structure of temporal bone and cause various complications. For example, damage to the auditory ossicle chain can lead to conducted hearing loss (HL). In addition, the erosion of semicircular fistulas and facial neural tubes may lead to disorders of the balance function and facial paralysis, respectively [13, 14, 19, 38]. High-resolution CT provides excellent detail in detecting cholesteatoma in the middle ear [23, 26]. From CT images, the middle ear bone structure of patients with cholesteatoma is usually partially eroded, and the surface of eroded bone is relatively smooth [40].

In recent studies, some researchers have conducted retrospective studies of patients at the BC Children's Hospital Microtia Clinic from Jan.1, 1990 to Apr.17, 2017 to examine existing imaging and clinical records to determine the presence or nonexistence of middle ear cholesteatoma [31]. Recently, there has developed a new consensus statement on the definition and classification of middle ear cholesteatoma [5].

In clinical practice, the symptoms and results of the objective examination characteristics of MEC and CSOM are similar. MEC is often not detected or even misdiagnosed at the early stage, especially when combined with inflammation. Due to differences in pathological changes, lesion manifestations, and complications between these two diseases, achieving accurate diagnostic discrimination is crucial in clinical practice. Therapy for MEC is usually surgical intervention as early as possible. In contrast, CSOM can be treated conservatively in most cases [32], which may avoid unnecessary injuries and improve recovery efficiency to a great extent.

Previous studies have used deep learning to classify and diagnose MEDs, or other otolaryngology diseases. The most recent research indicates that using CNNs in endoscopic examinations to automate the diagnosis of ear disease and detect of tympanic membrane perforations and middle ear infection can achieve excellent precision [6, 15, 20, 52, 53, 57]. Yan-Mei Wang et al. proposed a deep learning framework for the diagnosing of CSOM based on CT scans, and the model's performance was shown to be superior to that of clinical experts in some cases [51]. Wang et al. presented a deep-learning method for the diagnosis of CSOM and MEC, which chose Visual Geometry Group 16 (VGG-16) as the model's backbone [55]. Moreover, Wang et al. fused individual features from both CNN and GCN to assist radiologists in rapidly detecting COVID-19 from chest CT images [54]. Parvaze et al. extracting crafted features to analyze and identification the pathologies features of peritumoral vasogenic edema [35, 36].

Comparatively, graph neural networks (GNNs) [7, 16, 22, 48] is a promising technology and emerging network architecture, which can efficiently deal with graph structure data by modeling relations between sample nodes (or vertices). In these variants of the GNN model, the graph isomorphism network (GIN) [58] has the greatest ability to represent from different graph structures and has quantifiable generalization ability, which quickly attracted wide attention from the GNN community. This special graph structural information shows some meaningful structural patterns. Because the symptoms and results of objective features of MEC and CSOM are very similar, and MEC is often not found or even misdiagnosed at an early stage, therefore, accurate diagnosis and differential diagnosis of these two diseases in clinical practice are of great significance. However, most of the previous studies used deep learning to classify and diagnose meds, but the structural features of the middle ear were not analyzed. In this work, our goal is to enhance the GIN toward better structural modeling and middle ear disease property identification via using special structural information. Moreover, our OMCNIC achieved a better classification result than [51, 55], and we analyze the effects with different parts of the middle ear structure on CSOM and MEC. We provide several middle ear examples of CSOM, MEC and normal structure in Fig. 1.

To reduce the workload of radiologists and improve their work efficiency, in this paper, we proposed the first work that uses a graph isomorphism network method to evaluate the impact of structure-constrained deep feature fusion with the middle ear on chronic suppurative otitis media and cholesteatoma with CT images. The experimental results indicate the validity of our algorithm. The major contributions of this work are summarized as follows:

- We first automatically crop the region of interest (ROI) from CT images, i.e., ME patches.
- Based on the image patches, we use structure-constrained deep feature fusion to represent the middle ear structure and convert the middle ear structure image into a graph.
- We use the graph isomorphism network to identify ME disease and evaluate the impact of different structures with the middle ear diseases efficiently.
- By analyzing the effects of different structures in the middle ear on cholesteatoma and chronic suppurative otitis media, this approach can provide a new direction for preoperative and postoperative care for cholesteatoma surgery.

The rest of the contents of this paper are as follows: Part II gives the data description and preprocessing details. Part III will present our GIN model to classify chronic suppurative

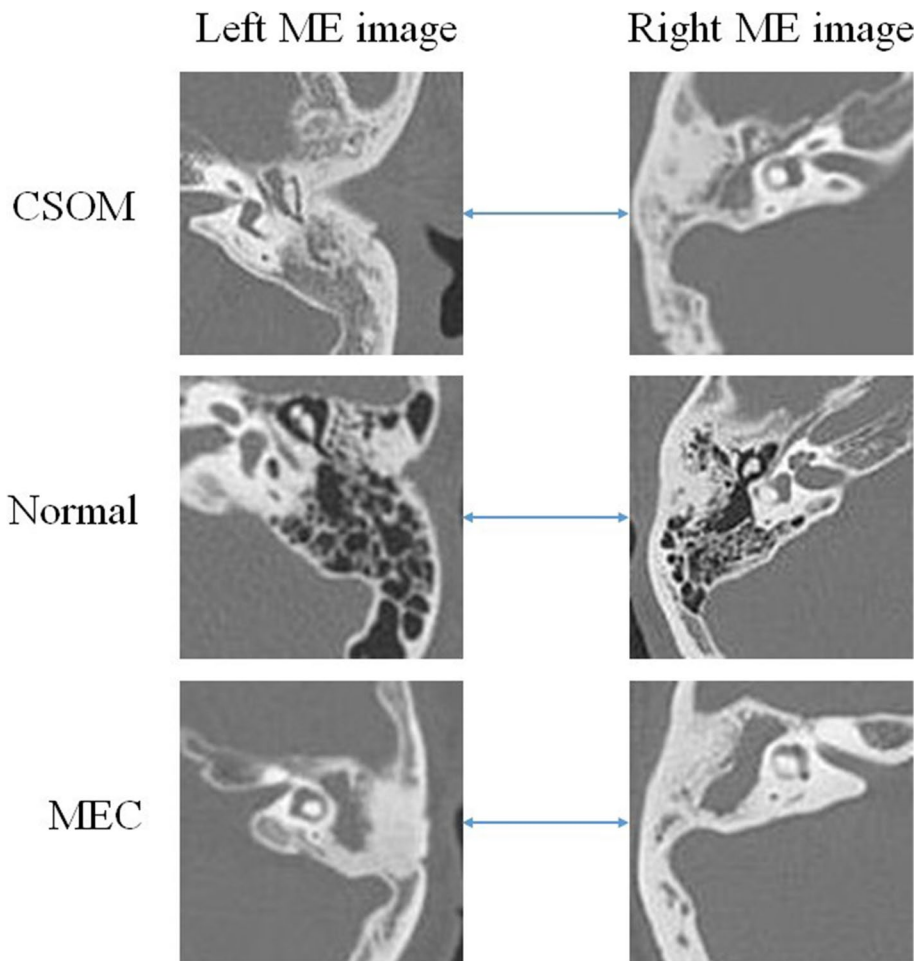


Fig. 1 Illustration of 3 classes of ME patch examples

otitis media, cholesteatoma and normal (CSOMCN) diseases. Part IV gives our experimental results. Part V is the discussion, and Part VI is the conclusion.

2 Data description and preprocessing

The medical research and ethics committee of Xiangya Hospital, Central South University approved this study. The researchers collected these data from 573 patients who underwent middle ear surgeries in the Department of Otorhinolaryngology, Xiangya Hospital, from Jan. 2018 to Oct. 2020, the age range of patients was 5–72 years, with a mean $\pm$ SD of 38.75 ± 14.38 years. They then reviewed medical records to exclude any patient diagnosed with a congenital malformation or any postoperative situation.

Each enrolled patient at least had received one temporal bone CT scans, resulting in a total of more than 573 scans available for this study. These scans were obtained by a

256-channel multidetector Revolution CT scanner (GE Healthcare). The parameters of CT scanning were: tube voltage of 100 kV, tube current of 325mAs, pitch coefficient of 0.6875, matrix size of 512×512 , field of view of 220×220 mm, thickness of layer of 0.625 mm. Body position was the standard cranial-anteriorly, scanning mode was spiral scanning, window width was 3000~4000 HU and window position was 300~500 HU.

Based on cooperation with Xiangya Hospital, the dataset that we finally adopted consists of 499 patients. Our final dataset consists of 998 unilateral ears, including 108 cases of middle ear cholesteatoma (MEC) (46 in the left ear and 62 in the right ear), 622 cases of CSOM (314 in the left and 308 in the right) and 268 cases of normal ME (139 in the left and 129 in the right), and each of cases was classified and labeled by professionals. All data were obtained with informed consent signed by the subjects. The training labels of the ROI network were annotated by an otolaryngologist. Inspired by the main ideas of Neural Style Transfer proposed by the researchers Garg [11, 46], by combining with the specificity and scarcity of our middle ear MEC data, we reused MEC data, and adding inverted left ear MEC case data to the training of the right ear pathology classifier.

Our dataset is divided into a training set, a verification set and a test set in proportions of 80%, 10% and 10%. By working together with the hospital on artificial intelligence and medical data, we obtained the hospital's medical data and needed to preprocess the CT scans. To achieve better training results, we systematically labeled the original data under the guidance of a specialist in otolaryngology. We chose to cover the middle ear structure and crop the middle ear structure and then used it to create the ROI label to better train our classification network and to analyze the correlation between different middle ear structures. Figure 2 is an example of data preprocessing of CSOMCN on the left and right ears.

2.1 Region of interest search net

In our experiment, we wanted our GIN algorithm to focus on the ME region in the CT image by ignoring as much other noise as possible, so we tried to build a network of regions of interest (ROIs) to extract important parts of the ME. Only in this way can we verify the accuracy and effectiveness of our OMCNIC. In addition, in order to improve accuracy and increase

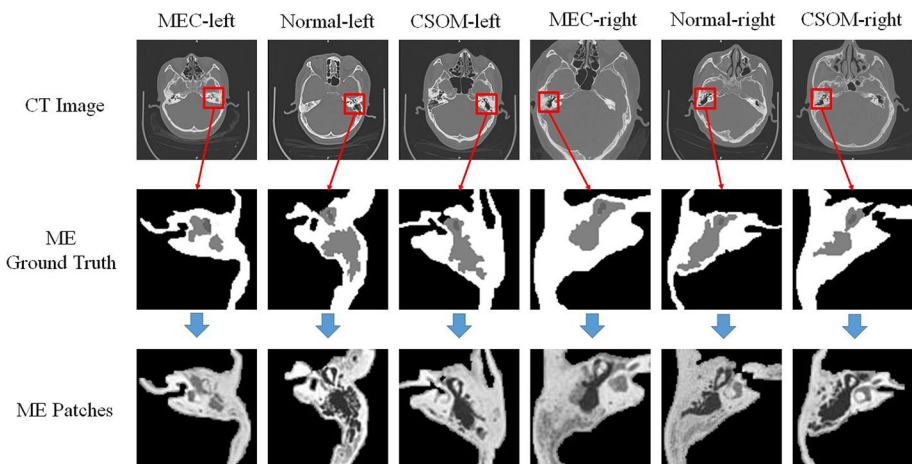


Fig. 2 Preprocessing examples of the OMCN image

credibility, a professional otolaryngologist defined the “window of interest” that we need to select. Therefore, we used U-net [39] to help automatically choose bounding boxes. The U-Net network adopts a U-shaped structure with an encoder-decoder structure, which transmits low-layer feature information to the higher layer through cross-layer connection, avoiding information loss and improving segmentation accuracy. In addition, the U-Net network can maintain feature richness by increasing the number of convolution kernels and reducing the step size of pooling. These features are ideal for the region of interest extraction of medical images such as CT images. At the same time, due to the similarity of CT image density and noise interference of MEC and CSOM, it is very challenging to segment and extract ROI from ME images, and it is necessary to select a verified deep learning network to ensure accuracy and robustness. U-Net is widely used in the field of medical imaging and has achieved excellent performance in a number of medical image segmentation tasks, which can handle the challenges encountered in medical image segmentation and ROI extraction tasks very well [43]. Therefore, by analyzing the specificity of the MED image and verifying the analysis experimentally, the selection of U-Net to extract the region of interest of the ME image can ensure the accuracy and robustness of ME patches. So, we first approximately determine the ME structure through U-net’s search network. Then, we analyzed the advantages and disadvantages of various interpolation methods and the characteristics of the lesion area of our ME patches, and found that bilinear interpolation is more suitable for stabilizing the image clarity and border accuracy of the ME patches, so we use bilinear interpolation of ROI-Align to deal with the misalignment problem in U-net. Finally, we obtained a fixed-size feature map and the correct bounding box. Based on this, we were able to extract the corresponding ME patch for each patient’s CT image.

In Fig. 2, we choose the special ME patches to represent all structures of the ME. In the first stage, we use U-net to obtain the ROI image, and our goal in this phase refers to image segmentation tasks [39]; we give the loss function of our ROI search net in the following formula (Eq. 1):

$$E = \sum_{x \in \Omega} \omega(x) \log(p_{\tau(x)}(x)) \quad (1)$$

$$\omega(x) = w_c(x) + \omega_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right)$$

The energy function is calculated by the following Eq. 1, where $\tau: \Omega \rightarrow \{1, \dots, K\}$ is the true label of each pixel and $\omega: \Omega \rightarrow \mathcal{R}$ is a weight map. The separation boundary is calculated by using morphological operations, where $\omega_c: \Omega \rightarrow \mathcal{R}$ is the weight map to balance the class frequencies, $d_1: \Omega \rightarrow \mathcal{R}$ denotes the distance to the border of the nearest pixel and $d_2: \Omega \rightarrow \mathcal{R}$ the distance to the border of the second nearest pixel. The ROI search net could divided into these steps: firstly, according to the input CT training data and proportionally divide them into training and validation sets, then do the sample mini-batch of data pairs to iterative, and then using Adam as the optimizer and using Eq. 1 as the loss function and via gradient descent to update parameters, finally we obtain the best trained model that we need.

3 Methodology

In the overall framework design of this work, we can divide it into three parts. Firstly, we collect and collate MED datasets from Department of Otorhinolaryngology of Xiangya Hospital, and do a certain preprocessing, reused MEC data and conduct data augmentation

operations, and complete the fine data labeling work with the assistance of Xiangya Otorhinolaryngology experts. Then, we use the U-net algorithm to search ME patches to obtain the ROI, and then we use structure-constrained deep feature fusion to represent the middle ear structure and convert the middle ear structure image into a graph. At last, in order to make better use of the structure-constrained ME patches feature, the framework uses Graph Isomorphism Network (GIN) network to identify ME disease and evaluate the impact of different structures with the middle ear diseases efficiently.

3.1 Structure-constrained deep feature fusion

Image fusion technology has great application value in remote sensing detection, medical image analysis and clear image reconstruction, especially in computer vision [1]. In general, image fusion is divided into pixel-level image fusion, feature-level image fusion and decision-level image fusion [44]. Inspired by the idea of image feature level fusion, we extract the MEC and CSOM feature information from the source ME image, then, the fused image features are obtained by analyzing, processing and integrating the feature information. Therefore, we define an undirected graph $G := (V, E)$ to represent the structure of the ME, where V denotes as a set of vertices and $E = \{(u, v) \mid u, v \in V\}$ denotes as a set of edges. In this approach, we aim to cluster information from patches. For this purpose, we construct an ME graph for each patch image P_i in the dataset. In order to reduce the complexity of construction, we first resize the ME patch into a size of 100×100 and then construct an undirected graph across the smaller patches. Then, we construct different nodes by extracting the homogeneous intensity in the ME patch as the vertices of the undirected graph G and prune vertices outside the mask. And then, we calculate the correlation distance between the two vertices as the weight of the corresponding edge by Eq. 2:

$$\omega = 1 - \frac{(u - \bar{u}) \cdot (v - \bar{v})}{\|(u - \bar{u})\|_2 \|(v - \bar{v})\|_2} \in \mathcal{W} \quad (2)$$

We take the pixel value of the corresponding vertex on the patch as the attribute value of the vertex on the undirected graph G . The coordinates and intensity are formed as the attribute value by using a k-nearest neighbor adjacency matrix with Eq. 3:

$$f_{i,j} \rightarrow a_{i,j} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_x^2}\right) \in \mathcal{A} \quad (3)$$

where vectors x_i and x_j are the spectral signatures associated with the vertices v_i and v_j , and σ_x is the scale parameter defined as the average coordinate distance x_k of the k-nearest neighbors (e.g., $k=8$) for each vertex. Let E_j as a set of edges in v_j , then we let r_j as the feature vector of E_j and we define r_j as Eq. 4:

$$r_j = \phi\left(\sum_{e_{i,j} \in E_j} f_{i,j}\right) \in \mathcal{R} \quad (4)$$

where ϕ is a linear transform function. $f_{i,j}$ is the feature vector of its neighborhoods $e_{i,j}$ with the vertices v_i and v_j . Once $\mathcal{A}$ is given, we can construct the graph Laplacian matrix $\mathcal{L}'$ as following Eq. 5:

$$\begin{aligned}\mathcal{L} &= \mathcal{D} - \mathcal{A} \\ \mathcal{D} &= \sum_j A_{ij} \\ \mathcal{L}' &= \mathcal{D}^{\frac{1}{2}} \mathcal{L} \mathcal{D}^{\frac{1}{2}} \\ &= I - \mathcal{D}^{\frac{1}{2}} \mathcal{A} \mathcal{D}^{\frac{1}{2}}\end{aligned}\quad (5)$$

where I is the identity matrix. The resultant graphs are of sizes 30–100 nodes for an ME patch, we found that our GIN with 75 nodes has achieved the best entire average classification results. The construction detail of the ME graph is given in Alg. 1. Using the proposed structure-constrained deep feature fusion, each ME graph is obtained and is represented in its structure-aware representation.

Figure 3 presents visualization of the structure-constrained deep feature fusion graphs, which indicate the graph representation (i.e., node, coordinate and feature of the node) of MEC. From the ‘node’ column of Fig. 3, the middle ear bone structure of patients with cholesteatoma is incomplete, the sinus opening in the tympanic membrane is enlarged, and the surrounding bone layer is thinner. From the ‘with coord’ column of Fig. 3, we can find that the corresponding location of tympanic node is very sparse, that is, the mastoid process of the middle ear is destroyed. According to the ‘feat and coord’ column of Fig. 3, we can observe that the relation between these nodes is that the more edges there are between two nodes, the closer their relationship.

Algorithm 1 Strategy of Structure-Constrained Deep Feature Fusion

Input: Segmentation based on ME *patch* dataset $P = \{p_1, p_2, \dots, p_L\}$; mask is the ground truth;
Output: structural constrained graph set $G = \{g_1, g_2, \dots, g_L\}$ with adjacent matrix $\mathcal{A}$;

- 1: **Structure constrained map generation, e.g.:**
- 2: **Initialize** $G = (V, E) \leftarrow$ extract homogeneous intensity from P
- 3: **for** each graph $g \in G$ and corresponding *mask* **do**
- 4: **while** node v not in *mask* **do**
- 5: remove node v and corresponding edge e from g
- 6: **end while**
- 7: **Iterative Fusion** with 8-nearest neighbors
- 8: $\omega \leftarrow \text{CorrDist}(u, v)$ via Eq. 2
- 9: $a \leftarrow \text{AdjMatrix}(u, v)$ via Eq. 3
- 10: $r \leftarrow \text{FuseFeat}(a)$ via Eq. 4
- 11: **end for**

3.2 OMCNIC algorithm

The graph Isomorphic Network (GIN) [58] is a classic variant of GNN with great potential, and it has a discriminative power that is equal to that of Weisfeiler-Lehman (WL) graph isomorphism test power [4, 56]. The GIN could iterate the node information by the following Eq. 6:

$$h_i^{k+1} = \text{MLP}^k \left(\left(1 + \xi^k \right) h_i^k + \sum_{j \in N_i} h_j^k \right) \quad (6)$$

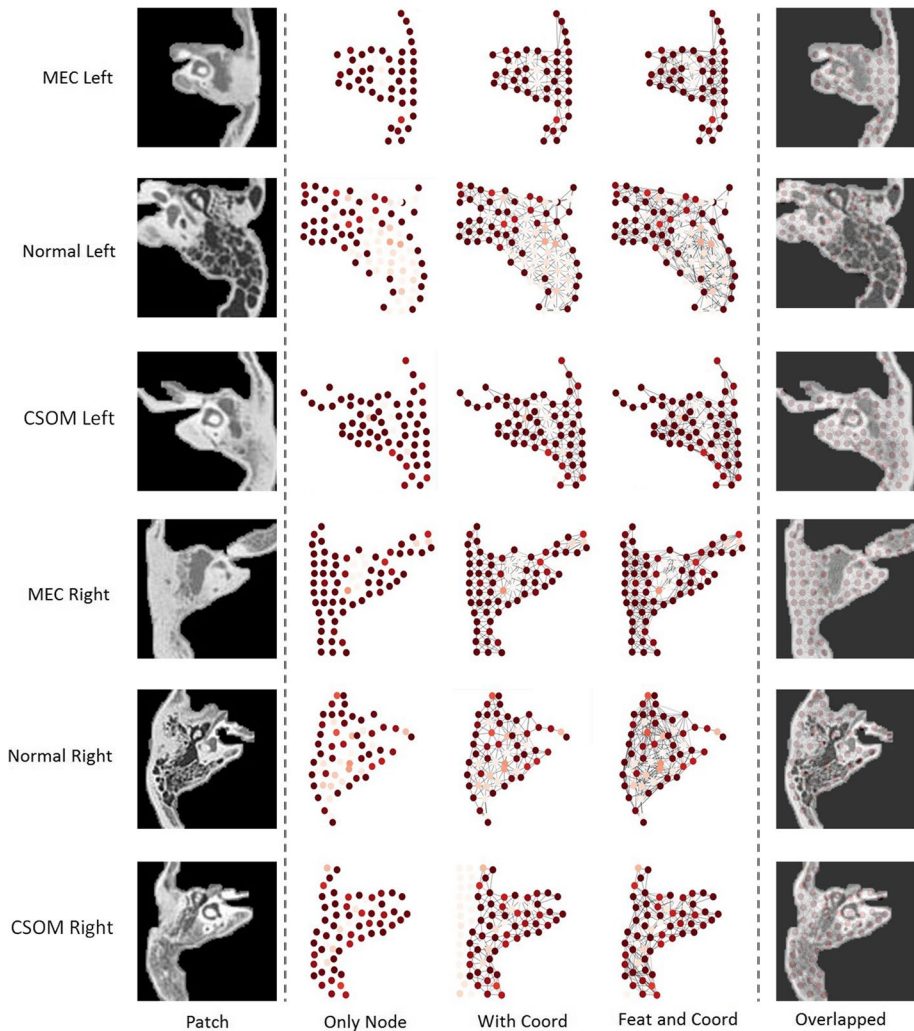


Fig. 3 Illustration of structure-constrained deep feature fusion of the graph

we set $h_i^0 = \mathcal{A}_i$, ξ^k is a parameter that can be improved by learning and MLP is a multilayer perceptron. Moreover, GIN concatenates the information represented by the nodes on all layers of the model according to the Eq. 7 to obtain the final representation:

$$h_G = \text{CONCAT} \left(\sum_{v \in G, k=0}^K h_v^k \right) \quad (7)$$

we set v and G represent as the node and the corresponding graph, respectively. we set $\text{COCATN}(\bullet)$ as the concatenate function. It has been proved from experimentally that the GIN has a more powerful representational power of graph structures than other GNN variant models.

Chronic otitis media and cholesteatoma classification can be regarded as a multi-class identification problem of ME structural graphs in artificial intelligence, which can be denoted as follows: we first given a set of structural graphs $G = \{G_1, G_2, \dots, G_n\}$ and their label set $Y = \{y_1, y_2, \dots, y_m\}$, each structural graph G_i has an attribute vector of vertices ($r \in \mathcal{A}$) and a feature vector of edges ($r \in \mathcal{R}$). Furthermore, we set $\sigma(\bullet)$ to represent the learning function which could learns the corresponding vector $h_G = \sigma(\mathcal{A}, \mathcal{R})$ to aid to predict the labels. At last, we let the labeling function $\zeta(\bullet)$ to allocates the label of the entire structural graph $y = \zeta(h_G)$.

We have improved the ability of GIN to represent the structure graph by two solutions: vertex feature cascading and neighborhood weight changes with using a gate unit. We concatenate r_j (from Eq. 4) to the neighbor's feature vector of the central vertex h_j on every layer of the GIN clustering. Therefore, the Eq. 6 could be replaced by the following formulation Eq. 8.

$$h_i^{k+1} = MLP^k \left((1 + \xi^k) h_i^k + \sum_{j \in N_i} (h_j^k \oplus r_j^k) \right) \quad (8)$$

Where $\oplus$ set as the concatenation. Moreover, Eq. 7 can be replaced by the following Eq. 9:

$$h_G = CONCAT \left(\sum_{i \in V, k=0}^K h_i^k \oplus r_i^k \right) \quad (9)$$

Therefore, the enhance GIN algorithm that we presented can cluster the information of the vertex neighbors and made these patterns change into hidden vectors. In the GIN, all neighbors make the same contribution to vertices updates, which results in ignoring the difference in intensity of impact between the central node and its different neighbors. To solve this problem, we introduced a control gate unit to regulating the role of neighbors in updating the characteristics of the central node [12, 37]. Hence, Eq. 8 can be redefined as the following Eq. 10:

$$h_i^{k+1} = MLP^k \left((1 + \xi^k) h_i^k + \sum_{j \in N_i} \left(\vartheta(h_j^k \mathcal{W}^k + b^k) \otimes h_j^k \oplus r_j^k \right) \right) \quad (10)$$

Where $\otimes$ is elementwise multiplication, we set $\mathcal{W}^k$ (Eq. 2) and b^k represent as the weight matrix and bias of the k-th layer, respectively. By this way, acts as an adjustable, changeable controller for neighborhood weights, which learns the weight matrix to adjust the different intensity of impact between the central node and its neighbors during the training phase. Figure 4 and Alg. 2 outline the main steps of our proposed OMCNIC.

Algorithm 2 Our proposed Graph Isomorphism Network (OMCNIC)

Input: the structure graph $G(V, E)$; the input structural vector $a_i, \forall i \in V$; the input feature vectors $r_i, \forall r \in E$; the learning depth of the cluster K .

Output: the final prediction h_G .

```

1: for each node  $i \in V$  do
2:    $r_i \leftarrow \sum_{j \in N_i} z_{i,j}$ 
3:    $h_i^0 \leftarrow a_i \oplus r_i$ 
4: end for
5: Set  $\vartheta$  as the sigmoid optimizer  $\in [0, 1]$ 
6: for ( $k = 1; k \leq K; k++$ ) do
7:   for each node  $i \in V$  do
8:      $\text{sum}N_i \leftarrow \sum_{j \in N_i} (\vartheta(h_j^{k-1} \mathcal{W}^{k-1} + b^{k-1}) \otimes h_j^{k-1} \oplus r_j^{k-1})$ 
9:      $h_i^k \leftarrow \text{MLP}^k((1 + \xi^{k-1})h_i^{k-1} + \text{sum}N_i)$ 
10:   end for
11: end for
12:  $h_G \leftarrow \text{CONCAT}(\sum_{i \in V, k=0}^K h_i^k \oplus r_i^k)$ 

```

4 Experimental results

The proposed OMCNIC framework is used to perform the training on a ASUS 8460-PLUS server (hexa-core 2.90 GHz processor, 64 GB RAM and one NVIDIA GeForce RTX 2070 SUPER video card). The presented methods were implemented in Python using the Pytorch framework.

4.1 Performance evaluation method

In this paper, we use accuracy (ACC), sensitivity (Sens) and specificity (Spec) as our evaluation metrics. According to the classification of all test samples, the classification results of

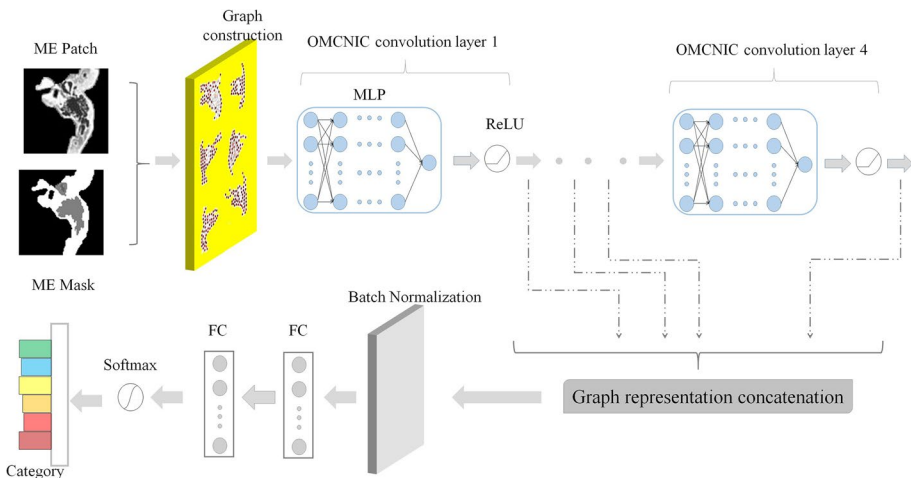


Fig. 4 Illustration of structure-constrained deep feature fusion of the graph

the experiment can be divided into true positive ($\mathcal{M}_{tp}$), false positive ($\mathcal{M}_{fp}$), true negative ($\mathcal{M}_{tn}$) and false negative ($\mathcal{M}_{fn}$). The three evaluation indicators are defined as follows:

$$Acc = \frac{\mathcal{M}_{tp} + \mathcal{M}_{tn}}{\mathcal{M}_{tp} + \mathcal{M}_{fp} + \mathcal{M}_{tn} + \mathcal{M}_{fn}} \quad (11a)$$

$$Sensitivity = \frac{\mathcal{M}_{tp}}{\mathcal{M}_{tp} + \mathcal{M}_{fn}} \quad (11b)$$

$$Specificity = \frac{\mathcal{M}_{tn}}{\mathcal{M}_{fp} + \mathcal{M}_{tn}} \quad (11c)$$

Obviously, the sensitivity refers to the proportion of positive values we predict among all the original label is positive, while the specificity refers to the proportion of negative values we predict among all the original label is negative. The accuracy (ACC) reflects the accuracy of the overall judgment of the classification model.

4.2 Hyperparameter setting

The presented and mentioned networks are trained by the Adam optimizer [21]. The primary idea of Adam's algorithm is to computing the update step size of the parameters. The main idea of the Adam algorithm is used to calculate the updated step of the parameters. The method can automatic adjustment of the parameters of the learning rate, has little memory requirements so greatly improved the speed of training, improved stability, and thus it is suitable as the optimizer of the multi-classification problem. The Adam method is also applicable to non-stationary targets and problems with very noisy or sparse gradients, and the algorithm has stable convergence in theory [21]. Therefore, by analyzing the specificity and sparsity of our ME data and performing experimental validation, we found that the algorithm of stable and efficient Adam is suitable for our MED identification and classification problems. The form of Adam (Eq. 12) is derived as follows:

$$\begin{aligned} \alpha_t &= \alpha \cdot \sqrt{1 - \beta_2^t} / (1 - \beta_1^t) \\ \theta_t &\leftarrow \theta_{t-1} - \alpha_t \cdot m_t / (\sqrt{v_t} + \hat{\epsilon}) \end{aligned} \quad (12)$$

where α is the learning rate with the default set as 0.001, m_t and v_t are the mean and variance of the gradient after deviation correction, respectively, and the parameter $\beta_1, \beta_2 \in [0, 1)$, $\hat{\epsilon}$ default is set as 10^{-8} to prevent a divisor of 0. More details about the setting of each parameter value in Adam hyperparameters, the underlying mathematical mechanism and the proof of convergence can be studied in detail in [21].

Rectified linear unit (ReLU) is constant at partial gradients greater than 0, there is no gradient dispersion, and the derivative of ReLU is calculated faster. The derivative of ReLU in the negative half region is 0, so when the activation value of neurons is negative, the gradient is 0, this neuron does not participate in training and has sparsity. Therefore, by analyzing the specificity of ME data and validating it experimentally, in the intermediate

process of convolution, we choose the rectified linear unit (ReLU) as our activation function [50]. The form of the ReLU (Eq. 13) is derived as follows:

$$\text{ReLU}(t) = \begin{cases} t, & \text{if } t > 0 \\ 0, & \text{if } t \leq 0 \end{cases} \quad (13)$$

The most obvious feature of the Softmax function is that it takes the ratio of the input of each neuron to the sum of all the inputs of the current layer as the output of that neuron; that is, the greater the output of the neuron, the higher the corresponding category of the neuron is more likely to be a true category. At the same time, Softmax has the advantages of monotonicity and non-locality, which can solve the problem of slow learning and so on [17, 24]. Moreover, the training effect of Softmax with a log-likelihood cost function is better than that with a quadratic cost function, and more details of the mathematical form to prove the effectiveness of Softmax can be seen in [33]. Therefore, in the final fine-tuning of the fully connected layer, we use the Softmax function as an activation function. Then, it outputs a probability distribution for the six classes: “MEC left”, “MEC right”, “CSOM Left”, “CSOM right”, “normal left” and “normal right”. The form of the Softmax (Eq. 14) is derived as follows:

$$S_i = \frac{e^{a_i}}{\sum_{j=1}^C e^{a_j}} \quad (14)$$

In the actual test model, when the sample image passes through the Softmax layer, it will take the maximum index in the vector as the prediction label of this sample. Using the cross-entropy loss function, when the error is large, the gradient is also large, the decline is faster; if the error is small, the update is slower; at the same time, the activation function in some cases into the saturation region, the gradient disappeared problem, and more details of the mathematical form to prove the effectiveness of cross-entropy loss function can be seen in [33]. So, by analyzing the characteristics of our classification problems and combining the previous Softmax [17, 24], we choose the categorical form as our loss function, the form of which (Eq. 15) is derived as follows:

$$\text{Loss} = - \sum_{j=1}^{\text{outputsize}} y_j \log \hat{y}_j \quad (15)$$

Where $\hat{y}_j$ is the output value and y_j is the ground truth. For fairness and convenience, in this study, OMCNIC and other compared methods use the same hyperparameter settings as follows. Batch normalization was applied, the Adam optimizer with an initial learning rate of 0.001 and a learning rate delay of 0.5 per every 10 epochs were applied. Additionally, the following settings were used: the number of hidden units $\in \{8, 16, 32, 64\}$, the dropout ratio $\in \{0, 0.3, 0.5\}$ and the batch size $\in \{8, 16, 32\}$.

4.3 Effect of structure-constrained deep feature fusion

Based on previous work, our OMCNIC has achieved good classification results. As listed in Table 1, ‘Category’ indicates the entire six categories composed of chronic suppurative otitis media, middle ear cholesteatoma and normal, ‘*coor*’ indicates the coordinate distance of vertices and ‘*Feat*’ indicates the feature vector of vertices that our GIN algorithm adopted. The experimental results indicate that the OMCNIC algorithm with the coordinate distance and feature vector achieve the best overall classification results, yielding an

Table 1 Our OMCNIC Classification Result

Evaluation Metrics

Category	<i>Coor</i>	<i>Feat</i>	Acc(%)	Val Acc(%)	Sens(%)	Val Sens(%)	Spec(%)	Val Spec(%)
CSOMCN(L&R)	×	×	91.68%	68.49%	90.44%	86.50%	84.48%	78.27%
	✓	×	93.48%	70.92%	99.00%	89.59%	85.44%	80.42%
	×	✓	94.70%	78.05%	99.00%	90.62%	87.67%	90.62%
	✓	✓	96.36%	80.81%	99.00%	91.87%	89.68%	93.57%

The bold entries indicate that the OMCNIC algorithm with the coordinate distance and feature vector achieves the best overall classification results

Coor: Coordinate; Feat: Feature Vector; Val: Validation; Acc: Accuracy; Sens: Sensitivity; Spec: Specificity

average accuracy, sensitivity and specificity on the actual test set of 96.36%, 99.00%, and 89.68%, respectively. In the validation set, the Val accuracy, Val sensitivity and Val specificity are 80.81%, 91.87%, and 93.57%, respectively. When the feature vector concatenation was removed, in terms of accuracy and specificity on the training dataset, OMCNIC performance was reduced by 2.88% and 4.24%, respectively, but still improved by 1.8% and 0.96% compared with GIN, respectively. When the coordinate weight adjustment was removed, in terms of accuracy and specificity on training datasets, OMCNIC performance was reduced by 1.66% and 2.01%, respectively, but still improved by 3.02% and 3.19%, respectively, compared with GIN.

In particular, our OMCNIC obtained dramatically better performance than previous GIN, increasing the performance by 4.68% on accuracy, 8.56% on sensitivity and 5.2% on specificity. This shows the effectiveness of our OMCNIC in taking the structural coordinate distance and feature vector influence between GIN-based vertices and their neighbors. As shown in Table 1, when we remove feature vector concentration or coordinate distance adjustment from the OMCNIC framework, the overall accuracy of OMCNIC decreases, but is still outperformed to GIN. The experimental results clearly support that the structural characteristics, i.e., the feature vector and coordinate distance, which could efficiently make the GNN learning more local structural features of the middle ear, hence boosting the identification of structural ME properties.

We demonstrate the confusion matrix of the testing results in Fig. 5a. The confusion matrix provides a more complete display of the performance of our algorithm. On the one hand, by testing the performance of our algorithm with structure-constrained deep feature fusion, we found that our OMCNIC achieved the best entire average classification results. On the other hand, by comparing the results from Fig. 5a, we can find that the classification accuracy of MEC and CSOM is overall higher than that of normal cases, that is primary owing to the following reasons. First of all, due to the unbalance distribution of the datasets, the pathological details of some middle ear structures of normal and chronic suppurative otitis media cannot be learned well. In contrast, the middle ear structure of middle ear cholesteatoma in the middle is well illustrated by the nodes graph, so it can be learned well by the computer implementation. Therefore, based on the goal of no missed diagnosis, we give a biased mechanism in the classification process that actually made a low accuracy in normal classes. As shown in Fig. 5b, we demonstrate the ROC result of our proposed algorithm.

4.4 Comparison with baseline models

We evaluated the efficiency of our proposed algorithm for constructing the GNN to learn the structural representation for the middle ear, developing comparisons with VGG-16,

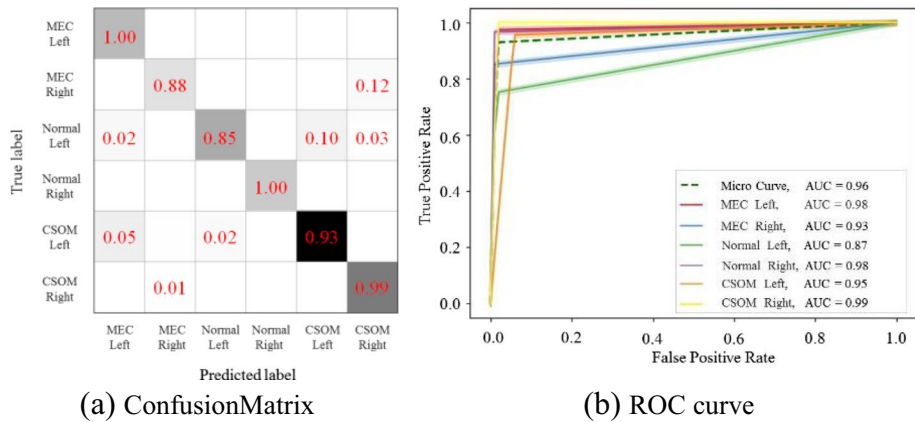


Fig. 5 The testing performance of our OMCNIC. **a** Confusion Matrix. **b** ROC curve

InceptionV3, the graph convolutional network (GCN [22]) and the graph attention network (GAT [48]). We used accuracy for disease identification of the middle ear to evaluate and compare the performance differences between our model and the baseline model. Figure 6 shows the accuracy in the testing dataset of the comparative models on the identification of disorders of the middle ear. All in all, the overall performance of our OMCNIC is significantly better than other SOTA models. Therefore, OMCNIC is a formidable alternative framework for structural middle ear disease identification.

As shown in Fig. 6, VGG-16 and InceptionV3 (the first two rows) perform the worst. It is not surprising because they are the models that have access to spatial representations. Their disappointing performance empirically supports the necessity of spatial dependencies for middle ear identification. Since the other two GNN models are enabled to capture neighboring information, we then compare them to find which method utilizes structure knowledge more effectively, most of the classification results are better than the former two models. Consequently, OMCNIC achieves moderate improvement increments of at least

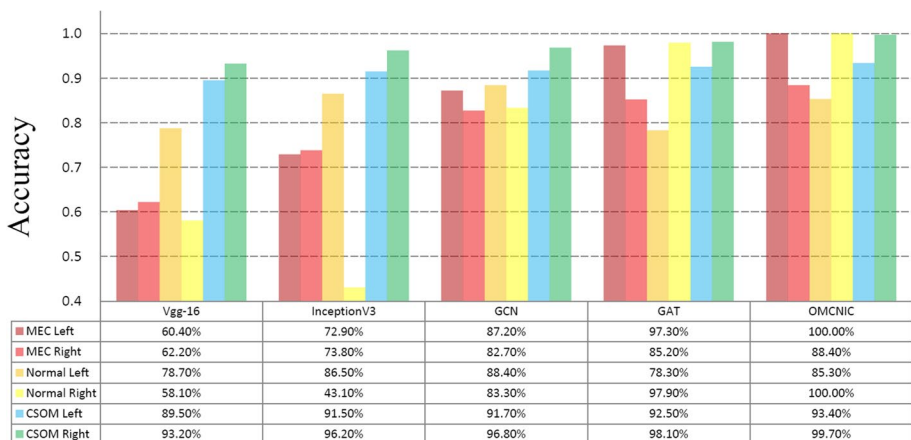


Fig. 6 Comparison of accuracy results in tests of different models on ME patches

Table 2 Comparison of our approach with the state-of-the-art approaches in CSOMCN classification

Methods	Number of patients	Age (mean $\pm$ SD)	Accuracy	Recall	Specificity	Precision	F1-score
Inception-V3 [51]	562	42.70 $\pm$ 13.74	76.70%	83.30%	91.40%	95.55%	89.01%
VGG-16 [55]	573	38.75 $\pm$ 14.38	90.20%	85.40%	93.68%	90.10%	87.20%
Our OMCNIC	499	38.75 $\pm$ 14.38	96.36%	99.00%	89.68%	96.05%	97.50%

The bold entries indicate that the performance of our OMCNIC algorithm is superior to the existing model in most of the evaluation indicators

0.90% with CSOM Left and at most 16.70% with Normal Right over the other two GNN model CSOMCN classification accuracy results, respectively; because the structural relation between ME patches can be excellent modeled as a graph by the GIN. As expected, the OMCNIC outperforms those comparative models, showing its ability to blend the representations of different graph structures.

As shown in Table 2, Inception-V3 [51] performed worst in accuracy and recall measures, while VGG-16 [55] performed worst in Precision and F1-score measures; we found that, compared with other methods, our OMCNIC model is only slightly inferior to the other models in specificity indicators, the other indicators have reached state of the art (SOTA) superior to the other two models. In addition, due to the very similar pathological characteristics of CSOM and MEC and the imbalance between these two types of data, partial symptoms are not evident, and the fact that both CSOM and MEC are prone to misdiagnosis, which leads to a greater number of false positives; low specificity is then obtained. Their disappointing performance in the first two CNN models empirically demonstrates this necessity for MED recognition of pathological graph structure dependence and representation relationships. However, OMCNIC significantly increased by 19.66%, 15.70%, 0.50% and 8.49% compared with the Inception-V3 model in accuracy, recall, precision and F1-score, respectively; and then, compared with VGG-16 model, our OMCNIC model increased by 6.16%, 13.60%, 5.95% and 10.30% in accuracy, recall, precision and F1-score, respectively. Because of the structural constraints between ME patches, the GIN model can be well constructed for unique and easy-to-learn middle ear pathological features, as expected, this shows the importance and high efficiency of the structural information representation features of middle ear pathology chart in the diagnosis of middle ear diseases.

Furthermore, for the class MEC that has very few samples or unbalanced samples, the GCNs could gain higher identification accuracy by take into account the form of a graph structure. In contrast, the CNN models cannot precisely model those classes. Nevertheless, it is notable that the structure-constrained deep feature fusion is capable of better identifying these challenging examples, due to the common use of the coordinate distance and the feature vector of nodes.

5 Discussion

The experimental results indicate that the enhanced GIN algorithm (i.e., OMCNIC) is obviously better than all the baseline models, including GIN, in terms of accuracy, and has obtained the best performance at present. These results show that OMCNIC has powerful graph structure modeling. This is mainly because OMCNIC not only inherits the powerful judgment power of GIN, but also uses structural feature information in graph structural modeling.

In the preprocessing phase, our preprocessing in medicine is faced with the exclusion and inclusion criteria, which need to be based on the clinical diagnostic criteria of CSOM and MEC, it includes reviewing the medical history, preoperative examination results, intraoperative and postoperative pathological findings, screening the patients' data that accord with the standard of CT, and eliminating the patients' data that have too much noise and do not accord with the requirement of CT; to do this, we asked experienced otolaryngology to sift through and remove the ear data that doesn't meet our requirements. And then, when faced data imbalances during the technical pre-processing phase, reasonable data expansion and data enhancement are required, and because MEC patient data are scarce, we reused MEC data, adding inverted left ear MEC case data to the training of the right ear pathology classifier.

Graph isomorphism network (GIN) has the greatest ability to represent from different graph structures and has quantifiable generalization ability, which quickly attracted wide attention from the GNN community and the performance is better than other previous GNN framework. Nevertheless, the GIN has not sufficiently utilizing the features information of the structural graph, cannot to adequately take advantage its structural modeling capability. Based on the GIN, our work primarily utilizing the structural characteristics, i.e., the feature vector concentration and coordinate distance between the vertex and its neighbors, to heighten the graph structural modeling capacity and to promote identification of ME properties. Therefore, we proposed a better method to identify the ME properties, called OMCNIC, which is based on the improvement of GIN. We can obvious find that the OMCNIC performance is dramatically affected by the feature vector concatenation and coordinate weight regulation. Our OMCNIC was realized by cascading the feature vector of the vertex neighbors and the coordinate distance of the vertex, thus adjusting the neighborhood weight in the information cluster by adding a control gate unit.

In the future, we will expand and explore our further work from the following aspects: (1) continue to expand the sample size, cooperate with many medical institutions, carry out multi-center research, and further improve the generalizability and interpretability of the model; (2) single layer was used in this study, and we hope to improve the accuracy of small-scale lesions by analyzing the images from 3D multi-angle; (3) to expand the image of atypical cases, we hope to extend our algorithm to computer-aided diagnostics other middle ear and inner ear diseases. In addition, we look forward to developing a framework for the assistant diagnosis of middle ear diseases, it has great social benefits for the disease management of the ME and CSOM patient database, the construction of a multi-disciplinary team and the multi-center on-line combined diagnosis and treatment.

6 Conclusion

In this paper, our OMCNIC presents a promising method for the identification of large-scale middle ear diseases. The experimental results indicate that our structure-constrained deep feature fusion algorithm can quickly and effectively classify CSOM and MEC. The presented algorithm is also promising for other graph structural modeling problems in medical and biomedical domains. Nevertheless, like to other deep learning methods, our OMCNIC method still has a weakly interpretability. It is difficult to obviously recognize which sub-structures of the ME play significant roles in the identification challenge. In the future, we will continue our research in enhancing the model interpretability.

Acknowledgments This work was supported by the Hunan Province Natural Science Foundation (grant number 2022JJ30673), the Teaching Reform Research Project of Universities in Hunan Province (under grant HNJG-2021-1120), the Hunan Provincial Natural Science Foundation of China (under grant 2022JJ30189), the Hunan Province Natural Science Foundation (Grant No. 2021JJ41017), Foundation of Department of Science and Technology of Hunan Province (grant number 2022GK3003), Foundation of Changsha Intellectual Property Office (grant number 2020GJZ18), Foundation of Department of Transportation of Hunan Province (grant number 2018234-201835).

Declarations

Conflict of interest The authors declare no competing interests.

I confirm I have included a data availability statement in my main manuscript file.

References

1. Aggarwal AK (2020) Fusion and enhancement techniques for processing of multispectral images. In: Avtar R, Watanabe T (eds) *Unmanned aerial vehicle: applications in agriculture and environment*. Springer, Cham, pp 159–175. https://doi.org/10.1007/978-3-030-27157-2_12
2. American Academy of Pediatrics Subcommittee on Management of Acute Otitis Media (2004) Diagnosis and management of acute otitis media. *Pediatr* 113(5):1451–1465. <https://doi.org/10.1542/peds.113.5.1451>
3. Ars B, Ars-Piret N (1994) Middle ear pressure balance under normal conditions. Specific role of the middle ear structure. *Acta Otorhinolaryngol Belg* 48(4):339–342. <https://pubmed.ncbi.nlm.nih.gov/7810303>
4. Babai L, Kucera L (1979) Canonical labelling of graphs in linear average time. In: 20th annual symposium on foundations of computer science (sfcs 1979), pp 39–46. <https://doi.org/10.1109/SFCS.1979.8>
5. Castle JT (2018) Cholesteatoma pearls: practical points and update. *Head Neck Pathol* 12:419–429. <https://doi.org/10.1007/s12105-018-0915-5>
6. Cha D, Pae C, Seong SB, Choi JY, Park HJ (2019) Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. *EBioMed* 45:606–614. <https://doi.org/10.1016/j.ebiom.2019.06.050>
7. Ciano G, Rossi A, Bianchini M, Scarselli F (2021) On inductive-transductive learning with graph neural networks. *IEEE Trans Pattern Anal Mach Intell* 44(2):758–769. <https://doi.org/10.1109/TPAMI.2021.3054304>
8. Cmert Z (2020) Fusing fine-tuned deep features for recognizing different tympanic membranes. *Biocybernet Biomed Eng* 40(1):40–51. <https://doi.org/10.1016/j.bbe.2019.11.001>
9. Colozza P, Anastasio ART (2009) Screening, diagnosing and treating deafness: the knowledge and conduct of doctors serving in neonatology and/or pediatrics in a tertiary teaching hospital. *Sao Paulo Med J* 127(2):61–65. <https://doi.org/10.1590/S1516-31802009000200002>
10. Cunningham LL, Tucci DL (2017) Hearing loss in adults. *N Engl J Med* 377(25):2465–2473. <https://doi.org/10.1056/NEJMr1616601>
11. Garg M, Ubhi JS, Aggarwal AK (2022) Neural style transfer for image steganography and destylization with supervised image to image translation. *Multimed Tools Appl* 2022:1–18. <https://doi.org/10.1007/s11042-022-13596-3>
12. Gers FA, Schraudolph NN, Schmidhuber J (2002) Learning precise timing with LSTM recurrent networks. *J Mach Learn Res* 3(1):115–143. <https://doi.org/10.1162/153244303768966139>
13. Gilberto N, Custidio S, Colao T, Santos R, Escada P (2020) Middle ear congenital cholesteatoma: systematic review, meta-analysis and insights on its pathogenesis. *Eur Arch Otorhinolaryngol* 277(11):987–998. <https://doi.org/10.1007/s00405-020-05792-4>
14. Gioacchini FM, Cassandro E, Alicandri-Ciufelli M, Kaleci S, Cassandro C, Scarpa A, Re M (2018) Surgical outcomes in the treatment of temporal bone cerebrospinal fluid leak: a systematic review. *Auris Nasus Larynx* 45(5):903–910. <https://doi.org/10.1016/j.anl.2018.03.006>
15. Habib AR, Wong EH, Sacks R, Singh N (2020) Artificial intelligence to detect tympanic membrane perforations. *J Laryngol Otol* 134(4):311–315. <https://doi.org/10.1017/S0022215120000717>
16. Jiang X, Zhu R, Li S, Ji P (2020) Co-embedding of nodes and edges with graph neural networks. *IEEE Trans Pattern Anal Mach Intell* 99:1–12. <https://doi.org/10.1109/tpami.2020.3029762>

17. Joulin A, Cissé M, Grangier D, Jégou H et al (2017) Efficient softmax approximation for GPUs. In: proceedings of international conference on machine learning, pp 1302–1310. <https://doi.org/10.48550/arXiv.1609.04309>
18. Kaspar A, Newton O, Kei J, Driscoll C, Swanepoel DW, Goulios H (2018) Prevalence of otitis media and risk-factors for sensorineural hearing loss among infants attending child welfare clinics in the Solomon Islands. *Int J Pediatr Otorhinolaryngol* 111:21–25. <https://doi.org/10.1016/j.ijporl.2018.05.021>
19. Keeler JA, Kaylie DM (2016) Cholesteatoma: is a second stage necessary? *Laryngoscope* 126(7):1499–1500. <https://doi.org/10.1002/lary.25385>
20. Khan MA, Kwon S, Choo J, Hong SM, Kang SH, Park I-H, Kim SK, Hong SJ (2020) Automatic detection of tympanic membrane and middle ear infection from Oto-endoscopic images via convolutional neural networks. *Neural Netw* 126:384–394. <https://doi.org/10.1016/j.neunet.2020.03.023>
21. Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: proceedings of international conference for learning representations, pp 1–15. <https://doi.org/10.48550/arXiv.1412.6980>
22. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*
23. Lingam RK, Bassett P (2017) A meta-analysis on the diagnostic performance of non-echoplanar diffusion-weighted imaging in detecting middle ear cholesteatoma: 10 years on. *Otol Neurotol* 38(4):521–528. <https://doi.org/10.1097/MAO.0000000000001353>
24. Liu W, Wen Y, Yu Z, Yang M (2016) Large-margin softmax loss for convolutional neural networks. In: proceedings of international conference on machine learning, pp 507–516. <https://doi.org/10.48550/arXiv.1612.02295>
25. Luers JC, Hüttenbrink KB (2016) Surgical anatomy and pathology of the middle ear. *J Anat* 228(2):338–353. <https://doi.org/10.1111/joa.12389>
26. Majithia A, Lingam R, Nash R, Khemani S, Kalan A, Singh A (2012) Staging primary middle ear cholesteatoma with non-echoplanar (half-fourier-acquisition single-shot turbo-spin-echo) diffusion-weighted magnetic resonance imaging helps plan surgery in 22 patients: our experience. *Clin Otolaryngol* 37(4):325–330. <https://doi.org/10.1111/j.1749-4486.2012.02502.x>
27. Marom T, Kraus O, Habashi N, Tamir SO (2019) Emerging technologies for the diagnosis of otitis media. *Otol–Head Neck Surg* 160(3):447–456. <https://doi.org/10.1177/0194599818809337>
28. Mason MJ (2016) Structure and function of the mammalian middle ear. II: inferring function from structure. *J Anat* 228(2):300–312. <https://doi.org/10.1111/joa.12316>
29. McCabe BF (1979) Autoimmune sensorineural hearing loss. *Annals Otol Rhinol Laryngol* 88(5):585–589. <https://doi.org/10.1177/000348947908800501>
30. Mittal R, Lisi CV, Gerring R et al (2015) Current concepts in the pathogenesis and treatment of chronic suppurative otitis media. *J Med Microbiol* 64(10):1103–1116. <https://doi.org/10.1099/jmm.0.000155>
31. Morita Y, Tono T, Sakagami M, Yamamoto Y, Matsuda K, Komori M et al (2019) Nationwide survey of congenital cholesteatoma using staging and classification criteria for middle ear cholesteatoma proposed by the Japan otological society. *Auris Nasus Larynx* 46(3):346–352. <https://doi.org/10.1016/j.anl.2018.10.015>
32. Mustafa A, Heta A, Kastrati B, Dreshaj S (2008) Complications of chronic otitis media with cholesteatoma during a 10-year period in Kosovo. *Eur Arch Otorhinolaryngol* 265(12):1477–1482. <https://doi.org/10.1007/s00405-008-0707-8>
33. Nielsen MA (2015) Neural networks and deep learning. Determination press, San Francisco, CA, USA
34. Park M, Lee JS, Lee JH, Oh SH, Park MK (2015) Prevalence and risk factors of chronic otitis media: the korean national health and nutrition examination survey 2010–2012. *PLoS One* 10(5):e0125905. <https://doi.org/10.1371/journal.pone.0125905>
35. Parvaze PS, Bhattacharjee R, Verma YK, Singh RK, Yadav V, Singh A, Khanna G et al (2022) Quantification of Radiomics features of peritumoral vasogenic edema extracted from FLAIR images in glioblastoma and isolated brain metastasis, using T1-DCE perfusion analysis. *NMR Biomed* 2022:e4884. <https://doi.org/10.1002/nbm.4884>
36. Parvaze PS, Bhattacharjee R, Singh A, Ahlawat S, Patir R, Vaishya S, Shah TJ, Gupta RK (2023) Radiomics-based evaluation and possible characterization of dynamic contrast enhanced (DCE) perfusion derived different sub-regions of glioblastoma. *Eur J Radiol* 159:110655. <https://doi.org/10.1016/j.ejrad.2022.110655>
37. Peng Y, Lin Y, Jing XY, Zhang H, Huang Y, Luo G (2020) Enhanced graph isomorphism network for molecular admet properties prediction. *IEEE Access* 8:168344–168360. <https://doi.org/10.1109/ACCESS.2020.3022850>
38. Prasad SC, La Melia C, Medina M, Vincenti V, Bacciu A, Bacciu S, Pasanisi E (2014) Long-term surgical and functional outcomes of the intact canal wall technique for middle ear cholesteatoma in

- the paediatric population. *Acta Otorhinolaryngol Ital* 34(5):354–361 <https://pubmed.ncbi.nlm.nih.gov/25709151>
39. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18. Springer International Publishing, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
 40. Rosito LS, Netto LFS, Teixeira AR, Da Costa SS (2016) Classification of cholesteatoma according to growth patterns. *JAMA Otolaryngol Head Neck Surg* 142(2):168–172. <https://doi.org/10.1001/jamaoto.2015.3148>
 41. Shah-Becker S, Carr MM (2018) Current management and referral patterns of pediatricians for acute otitis media. *Int J Pediatr Otorhinolaryngol* 113:19–21. <https://doi.org/10.1016/j.ijporl.2018.06.036>
 42. Shew M, New J, Wichova H, Koestler DC, Staecker H (2017) Using machine learning to predict sensorineural hearing loss based on perilymph micro RNA expression profile. *Sci Rep* 9(1):1–11. <https://doi.org/10.1038/s41598-019-40192-7>
 43. Siddique N, Paheding S, Elkin CP, Devabhaktuni V (2021) U-net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access* 9:82031–82057. <https://doi.org/10.1109/ACCESS.2021.3086020>
 44. Srivastava A, Aggarwal AK (2018) Medical image fusion in spatial and transform domain: a comparative analysis. *Handbook of research on advanced concepts in real-time image and video processing*. IGI Global 2018:281–300. <https://doi.org/10.4018/978-1-5225-2848-7.ch011>
 45. Tatlipinar A, Tuncel A, Ögredik EA, Gökçeer T, Uslu C (2012) The role of computed tomography scanning in chronic otitis media. *Eur Arch Otorhinolaryngol* 269(1):33–38. <https://doi.org/10.1007/s00405-011-1577-z>
 46. Ubhi JS, Aggarwal AK (2022) Neural style transfer for image within images and conditional GANs for destylization. *J Vis Commun Image Represent* 85:103483. <https://doi.org/10.1016/j.jvcir.2022.103483>
 47. Veldman JE, Hanada T, Meeuwse F (1993) Diagnostic and therapeutic dilemmas in rapidly progressive sensorineural hearing loss and sudden deafness: a reappraisal of immune reactivity in inner ear disorders. *Acta Otolaryngol* 113(3):303–306. <https://doi.org/10.3109/00016489309135813>
 48. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2018) Graph attention networks. In: *Proceedings of the International Conference on Learning Representations*, pp. 1–12. <https://openreview.net/forum?id=rJXMPikCZ>
 49. Wang WQ, Wang ZM, Chi FL (2005) Spontaneous healing of various tympanic membrane perforations in rat. *Acta Otolaryngol* 124(10):1141–1144. <https://doi.org/10.1080/00016480410022921>
 50. Wang P, Ge R, Xiao X et al (2017) Rectified-linear-unit-based deep learning for biomedical multi-label data. *Interdiscipl Sci Comput Life Sci* 9:419–422. <https://doi.org/10.1007/s12539-016-0196-1>
 51. Wang YM, Li Y, Cheng YS, He ZY et al (2019) Deep learning in automated region proposal and diagnosis of chronic otitis media based on computed tomography. *Ear Hear* 41(3):669–677. <https://doi.org/10.1097/AUD.0000000000000794>
 52. Wang Z, Meng Y, Weng F, Chen Y, Lu F, Liu X, Zhang J (2020) An effective CNN method for fully automated segmenting subcutaneous and visceral adipose tissue on CT scans. *Ann Biomed Eng* 48(5):312–328. <https://doi.org/10.1007/s10439-019-02349-3>
 53. Wang Z, Xiao Y, Li Y, Zhang J, Lu F, Hou M, Liu X (2021) Automatically discriminating and localizing Covid-19 from community-acquired pneumonia on chest X-rays. *Pattern Recogn* 110:107613. <https://doi.org/10.1016/j.patcog.2020.107613>
 54. Wang SH, Govindaraj VV, Gorriz JM, Zhang X, Zhang YD (2021) Covid-19 classification by fgcnnet with deep feature fusion from graph convolutional network and convolutional neural network. *Inf Fus* 67:208–229. <https://doi.org/10.1016/j.inffus.2020.10.004>
 55. Wang Z, Song J, Su R et al (2022) Structure-aware deep learning for chronic middle ear disease. *Expert Syst Appl* 194:116519. <https://doi.org/10.1016/j.eswa.2022.116519>
 56. Weisfeiler B, Leman AA (1968) A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Tech Inf* 2(9):12–16 <https://www.bibsonomy.org/bibtex/2d69974889cbfc06859c224edb0c9791d/sxkdz>
 57. Wisotzky EL, Rosenthal JC, Wege U, Hilsmann A, Eisert P, Uecker FC (2020) Surgical guidance for removal of cholesteatoma using a multispectral 3D-endoscope. *Sens* 20(18):1–19. <https://doi.org/10.3390/s20185334>
 58. Xu K, Hu W, Leskovec J, Jegelka S (2019) How powerful are graph neural networks? In: *Proceedings of the International Conference on Learning Representations*, pp. 1–17. <https://openreview.net/forum?id=ryGs6iA5Km>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Cong Cao¹ · Jian Song^{2,3,4} · Ri Su¹ · Xuewen Wu^{2,3,4} · Zheng Wang⁵ · Muzhou Hou¹ 

¹ School of Mathematics and Statistics, Central South University, Changsha 410083, China

² Department of Otorhinolaryngology of Xiangya Hospital, Central South University, Changsha 410008, China

³ Key Laboratory of Otolaryngology Major Disease Research of Hunan Province, Changsha 410008, China

⁴ National Clinical Research Centre for Geriatric Disorders, Department of Geriatrics, Xiangya Hospital, Central South University, Changsha 410008, China

⁵ School of Computer Science, Hunan First Normal University, Changsha 410205, China