Abstract. - OBJECTIVE: Cholesteatoma (CHO) developing secondary to chronic otitis media (COM) can spread rapidly and cause important health problems such as hearing loss. Therefore, the presence of CHO should be diagnosed promptly with high accuracy and then treated surgically. The aim of this study was to investigate the effectiveness of artificial intelligence applications (AIA) in documenting the presence of CHO based on computed tomography (CT) images.

PATIENTS AND METHODS: The study was performed on CT images of 100 CHO, 100 non-cholesteatoma (N-CHO) COM, and 100 control patients. Two AIA models including ResNet50 and MobileNetV2 were used for the classification of the images.

RESULTS: Overall accuracy rate was 93.33% for the ResNet50 model and 86.67% for the MobileNetV2 model. Moreover, the diagnostic accuracy rates of these two models were 100% and 95% in the CHO group, 90% and 85% in the N-CHO group, and 90% and 80% in the control group, respectively.

CONCLUSIONS: These results indicate that the use of AIA in the diagnosis of CHO will improve the diagnostic accuracy rates and will also help physicians in terms of reducing their workload and facilitating the selection of the correct treatment strategy.

Key Words: Chronic otitis media, Cholesteatoma, Artificial intelligence applications, Computed tomography, Accurate diagnosis.

Introduction

Chronic otitis media (COM) is a chronic inflammatory disease of the middle ear or mastoid cavity. COM manifests itself with recurrent or persistent ear discharge, commonly causing problems such as perforation, atelectasis, retraction, and cholesteatoma (CHO) in the middle ear and eardrum during the process. COM can be divided into subgroups based on the causes of formation as well as the damage to anatomical structures. One of the most serious problems caused by COM is bone erosion, which is at least twice more common and spreads over a wider area in cases of CHO. CHO is considered a benign but destructive epithelial lesion of the temporal bone. It is usually seen unilaterally; if not treated adequately, bone loss due to CHO can easily reach serious dimensions and can lead to hearing loss and intracranial spread. The presence of CHO in COM patients should be detected promptly and treated surgically since it can lead to the aforementioned serious complications.

As in many diseases, the patient’s complaints and physical examination findings are the most important factors guiding the physician. The severity of these findings makes the physician suspicious of the presence of CHO. Although the definitive diagnosis is made by histopathological examination, imaging methods are mostly used for treatment planning and, if necessary, used for distinguishing between CHO and non-cholesteatoma (N-CHO) COM in the preoperative period. Of these imaging methods, computed tomography (CT) is the most preferred method. CT is also highly valuable in the evaluation of bony structures in the middle ear, although it may still be insufficient in diagnosing soft tissue problems. In such cases, diffusion-weighted magnetic resonance imaging (DW-MRI) can be the primary method of choice. Additionally,
early diagnosis of CHO in the preoperative period is of paramount importance. Nevertheless, even when both imaging methods are used, the presence of CHO may not be proven definitively. In that case, the clinician needs to plan surgery upon suspicion. On the other hand, surgical intervention in N-CHO cases may cause unnecessary labor and economic losses. Accordingly, the only way to protect patients from such losses and unnecessary surgical risks is to show the presence of CHO in the preoperative period with more precise methods6-8.

Rapid development of technology has allowed artificial intelligence applications (AIA) to be used in various areas of life, including medical sciences. In the literature, there are studies9-12 reporting the use of AIA in a wide variety of subjects pertaining to the field of otolaryngology. The common purpose of these studies was to increase the rate of accurate diagnosis, reduce the workload of radiologists, and guide the clinicians in the absence of radiologists. Given the difficulties experienced in the diagnosis of CHO, it will be highly beneficial to improve the accuracy of diagnosis with AIA applications.

The aim of this study was to investigate the efficacy of AIA in the differentiation of COM cases with and without CHO based on CT images.

**Patients and Methods**

The study was initiated after obtaining an approval from Dicle University Medical School Ethics Committee (Date: 09.06.2022; No.: 169). The study included patients aged 18-65 years who had been operated on due to COM in 2010 at the departments of Otorhinolaryngology and Head and Neck Surgery in Dicle University Medical School. A total of 200 COM patients were selected by sequential randomization from among the patients who met the inclusion criteria. These patients comprised 100 CHO patients that underwent mastectomy and/or tympanoplasty due to a diagnosis of COM (CHO group) and 100 N-CHO COM cases (N-CHO group). In all the patients in the CHO group, the diagnosis was confirmed histopathologically. Additionally, a control group was formed including 100 patients who had an ear pain ethology and temporomandibular joint disorders in addition to sudden hearing loss and vertigo symptoms and were reported to be completely normal after imaging studies. In total, the study evaluated the CT images of 300 patients. Patients who had previous ear surgery for other reasons, had no temporal bone CT images, and had incomplete postoperative follow-up records were excluded from the study.

**CT Imaging Protocol**

All the CT images were obtained in the axial and coronal planes using the bone window settings [window width: 3,000 Hounsfield Units (HU), window level: 400-500 HU]. A total of 8-10 images were obtained from each patient to create a more homogeneous dataset. The images were prepared blindly by a team of two otolaryngologists and one radiologist who were unaware of primary diseases of the patients. The CT images were evaluated for all three groups. Since the images had varying sizes, they were resized to 224x224 to achieve standardization. The classification method proposed in this study is illustrated in Figure 1. CT images

![Figure 1](image.png)
extracted from the dataset were provided inputs to ResNet50 and MobileNetV2 architectures size of 224x224x3. Subsequently, the images were evaluated by these two models and classified at network output.

**Deep Learning**

Classifications performed using machine learning techniques often consist of three stages. First, the input data is determined. Second, dominant features are extracted from the input data\(^1\). The aim here is to obtain the most distinguishing features to differentiate target variables from each other\(^1,3,4\). In the final stage, classification is performed using these distinguishing features.

Over the recent years, deep neural networks have improved significantly, such that deep learning algorithms can learn from raw data\(^1\).

Deep neural networks perform classification in three stages, as in general machine learning techniques (Figure 2). First, data are fed as input to the deep neural network. In the second step, features are extracted. At this stage, deep neural networks create feature maps from the input data using different layers. These layers generally include a convolutional layer, a pooling layer, and a fully-connected layer\(^1\). In the convolutional layer, image pixels are convoluted with a specified filter. In the next layer, these data are partitioned in the pooling layer. Subsampling is performed by calculating the average or maximum value of the partitioned data. The Rectified Linear Unit (ReLU) function is often employed in the activation process. This function resets negative values to zero. The final layer is the fully-connected layer (FC) where vectorization is performed. The resulting vector length is fed to the classical artificial neural network structure. Classification is performed as the final process\(^1\).

**Transfer Learning**

It is commonly known that the information obtained in a model created by machine learning techniques can be used for other similar tasks\(^1,5,6\). Transfer learning is a machine learning technique allowing the reuse of a previously trained network including its structure and connection parameters\(^5,6\). In this way, the classification process can be performed in a shorter time and in very large networks\(^5,6\). In the present study, ResNet50 and MobileNetV2 architectures were utilized, and by using transfer learning, a fully-connected layer consisting of 1,024 neurons and 3 output layers were added to the last layer. The architectural parameters used in the study were as follows: batch size 16, learning rate 0.001, Adam optimized, and epoch 20.

**Performance Evaluation Metrics**

Performance of the model proposed in the study was measured statistically based on three metrics: Precision, Recall and F1 Score. These metrics were calculated based on the following formulas:

1. Precision = TP/(TP + FP)
2. Recall = TP/(TP + FN)
3. F1_Score = \(2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\)

Of these, TN, FP, TP, and FN denote True Negative, False Positive, True Positive, and False Negative of a confusion matrix, respectively.

Table I shows the confusion matrix used to describe the performance of the classification model. In this matrix, \(\text{T}_{PA}\) denotes the number of True Positive (TP) cases in class A. \(\text{F}_{AB}\) shows the number of cases that are in class A but misclassified as class B. The number of False Negative

---

**Figure 2.** Deep neural network structure.
The number of False Negative (FN) cases in class A is represented by the sum of $F_{AB}$ and $F_{AC}$. The number of False Positive (FP) cases in class A is represented by the sum of $F_{AB}$ and $F_{CA}$. The number of True Negative (TN) cases in class A is represented by the sum of $T_{PB}$, $F_{CB}$, $F_{BC}$ and $T_{PC}$.

### Classification Results

A dataset created based on the CT images obtained from the three groups was used in the study. Each group consisted of 100 images; 80% of the data was used for training and the remaining 20% for testing. In addition, 15% of the training dataset was reserved as validation data. The Shuffle Algorithm was used in dataset parsing. Statistics revealed that the classification performance was 93.33% with ResNet50 and 86.67% with MobilNetV2. Figure 3 and 4 present the training and validation accuracy and loss graphs in the training process performed with ResNet50 and MobileNetV2. Figure 5 presents the confusion matrix obtained as a result of this classification.

In the classification performed with ResNet50, 18 out of 20 images from each group (CHO, N-CHO, and control group) were accurately classified, while the remaining 2 images were misclassified. By contrast, the classification performed with MobileNetV2 accurately classified 17, 19, and 16 images and misclassified 3, 1, and 4 images from the CHO, N-CHO, and control groups, respectively. Table II and III present the performance measurement results obtained for ResNet50 and MobileNetV2. In the classification performed with MobileNetV2, the highest Precision, Recall and F1 score metrics were obtained in the CHO group.

### Gradient-Weighted Class Activation Map (Grad-CAM)

Deep learning architectures extract the features automatically during the classification process. These features are then transferred from
Artificial intelligence modelling for diagnosis of otitis media

the general to the specific layers in the deep network structure. For this reason, it is highly difficult to know what features are extracted from the dataset. However, when performing classification with deep learning architectures, there are tools showing which regions are actively used in predicting the relevant class. One of these tools is named Gradient-weighted Class Activation Map (Grad-CAM), which allows the classification-trained machine learning model to localize class-specific image regions by visualizing the pixels of interest and coloring them in different shades according to their importance relative to the output. Darker shades represent the most distinctive features of the relevant class, while lighter colors represent less distinctive regions. In the present study, Grad-CAM analysis was performed for all three groups. Figure 6 presents examples for the Grad-CAM images in all three groups.

Discussion

The present study investigated the utility of AIA used on CT images in differentiating CHO and N-CHO COM cases. According to the data of the World Health Organization (WHO), COM is a serious health problem affecting 0.3 billion people worldwide. Moreover, this disease is the most important cause of preventable hearing loss, particularly in underdeveloped and developing countries. In addition to hearing loss, CHO may also cause intracranial complications such as...
Facial paralysis, vestibular disorders, meningitis, and abscess formation. Around 20 million people around the world are diagnosed with COM each year, and almost ¼ of them develop CHO. Its annual incidence is 3 per 100,000 children as opposed to 9.2 per 100,000 adults. Due to the different approaches used in the treatment of COM, prompt detection of CHO is of tantamount importance, mainly because COM patients with CHO often require surgical treatment while N-CHO COM cases can be treated with medical therapies.

Imaging methods including CT and DW-MRI are commonly preferred for the diagnosis of CHO in the preoperative period. Although CT has a high sensitivity of 88% in middle ear diseases, this rate decreases in CHO cases. Even so, CT has been the first choice in the diagnosis of middle ear diseases due to its superior properties in documenting bone changes. Nonetheless, it is insufficient in identifying the mass type. By contrast, DW-MRI can distinguish the structure of the mass with high sensitivity and specificity, while it cannot clearly visualize the bone tissue. Therefore, it is not possible to clearly determine the anatomical location of CHO. Due to these reasons, DW-MRI becomes more valuable in the postoperative period.

**Table II. Classification performance of ResNet50 Architecture.**

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (Normal)</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>Non-cholesteatoma</td>
<td>0.90</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td>Cholesteatoma</td>
<td>1.00</td>
<td>0.95</td>
<td>0.98</td>
</tr>
</tbody>
</table>

**Table III. Classification performance of MobilNetV2 Architecture.**

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (Normal)</td>
<td>0.80</td>
<td>0.89</td>
<td>0.84</td>
</tr>
<tr>
<td>Non-cholesteatoma</td>
<td>0.85</td>
<td>0.81</td>
<td>0.83</td>
</tr>
<tr>
<td>Cholesteatoma</td>
<td>0.95</td>
<td>0.90</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Literature indicates that CT is commonly preferred for the diagnosis of CHO in the preoperative period. Studies conducted on this subject have mainly aimed at increasing the diagnostic accuracy of CT. Additionally, some studies reported that the HU measurements or radiomic analyses improve the diagnostic accuracy of CT. In a CT study, Arendt et al reported that CT distinguished CHO and N-CHO COM with 89% accuracy by the aid of radiomic analysis.

There are several studies in the literature performing AIA on CT images in the diagnosis of CHO. In one of these studies, Wang et al compared the diagnostic accuracy rates of AIA diagnosis and physician diagnosis in distinguishing COM cases with and without CHO based on CT images. The overall accuracy rate was 76.7% in AIA diagnosis, as opposed to 73.8% in physician diagnosis. Moreover, the authors also found that the accuracy rate of AIA diagnosis was higher compared to physician diagnosis in all three groups (CHO, N-CHO, and control patients). In a recent study by Eroğlu et al, the overall accuracy rate was 95.4% in all groups, with 92.7% in the CHO group, 93.7% in the N-CHO group, and 98.8% in the control group. The authors investigated the accuracy rate for Alexnet, Googlenet, and Densenet201 models separately and also for the hybrid model created from these models. It was revealed that the highest accuracy rate was obtained in the hybrid model.

In this study, two different models (ResNet50 and MobilNetV2) were used. The overall diagnostic accuracy rate obtained by ResNet50 was 93.33% for the CHO group, while it was 90% for both N-CHO and control groups. In the ResNet50 model, the diagnosis was made correctly in all of the patients in the CHO group. In contrast, the overall accuracy rate obtained by MobilNetV2 was 86.67%, while it was 95% for the CHO group, 85% for the N-CHO group, and 80% for the control group.

Accordingly, it appears that the overall diagnostic accuracy rates obtained in this study are remarkably higher than the rates obtained by Wang et al, whereby our overall rates that were obtained by ResNet50 and MobilNetV2 were 26.20% and 21.20% higher than the accuracy rates obtained for physician diagnosis in that study, respectively. By contrast, our overall accuracy rates are closer to the rates obtained by Eroğlu et al. Of note, an accuracy rate of 100% was achieved in the CHO group, by the ResNet50 model, which indicates that all the CHO patients were diagnosed accurately. In the same model, only two patients in the N-CHO group were misdiagnosed, which accounted for a misdiagnosis rate of 10%. Given the difficulty of diagnosing CHO cases, this rate seems remarkably low. Taken together, the overall accuracy rates obtained in our study implicate that the models used in this study successfully distinguished between CHO and N-CHO cases.

**Study Limitations**

Despite the high accuracy rates obtained in this study, there were several limitations to the study. First, the study had a limited number of cases. Accordingly, it would be beneficial to conduct similar studies using a larger number of images obtained from a larger number of patients. Second, the study had a retrospective nature. Therefore, further prospective studies focusing on hospital conditions may lead to different accuracy rates.

**Conclusions**

In conclusion, the results indicated that CHO can be diagnosed with high diagnostic accuracy rates. Given the difficulties and failures encountered in the diagnosis of this disease, it is clear that AIA will make a remarkable contribution to both patient health and healthcare economy by improving the accuracy of the diagnosis. Our findings also implicated that the AIA used in this study can minimize the error margin in the diagnoses made by radiologists or clinicians or can provide an insight about the actual clinical condition of the patient in healthcare settings where relevant specialists are not available. From all these perspectives, it is clear that the developments in AIA will help physicians in terms of reducing their workload and improving the rate of accurate diagnosis.

**Conflict of Interest**

The Authors declare that they have no conflict of interests.

**Ethics Approval**

The study as approved by the Local Ethics Committee of Dicle University (Date 09.06.2022, Decision number: 169).

**Informed Consent**

All subjects gave written informed consent prior to participation.
Funding
The authors declare that they received no funding for this study.

Availability of Data and Materials
The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Authors’ Contribution
All authors contributed, read, and approved the final version of the manuscript.

ORCID ID
Muhammed Ayral: 0000-0002-2421-4842; Ömer Türk: 0000-0002-0060-1880; Şermin Can: 0000-0003-2688-4927; Davut Esen: 0000-0002-7512-3924; İsmail Topçu: 0000-0001-5479-9311; Ferit Akil: 0000-0002-3402-2625.

References
6) Auino JE, Cruz Filho NA, de Aquino JN. Epidemiology of middle ear and mastoid cholesteato-
Artificial intelligence modelling for diagnosis of otitis media


