Uluslararası Sürdürülebilir Mühendislik ve Teknoloji Dergisi
International Journal of Sustainable Engineering and Technology
ISSN: 2618-6055 / Cilt:9, Sayı:1, (2025), 51 – 57
Araştırma Makalesi – Research Article

EAR PATHOLOGIES USING DEEP LEARNING ON OTOSCOPIC IMAGES

Yasin TATLI[1]

[1]Avrasya University, Vocational School of Health Services, Department of Medical Services and Techniques, Audiometry Programme, Trabzon, Türkiye

(Geliş/Received: 16.05.2025, Kabul/Accepted: 14.06.2025, Yayınlanma/Published: 30.06.2025)

ABSTRACT

In this study, the performance of different deep learning architectures is comparatively analyzed for the classification of ear pathologies based on otoscopic images. The dataset included four basic classes: chronic otitis media, ear wax obstruction, myringosclerosis and normal ear structure. The images were normalized at 224×224-pixel resolution and made suitable for the model, and classification was performed using CNN, CNN-LSTM, DenseNet121, ResNet50 and EfficientNet architectures. During the training and validation phases, performance metrics such as accuracy, F1 score, precision, recall and loss values were calculated, and the class discrimination power of the models was evaluated with ROC curves and complexity matrices. According to the results, CNN+LSTM and DenseNet121 architectures showed the best performance with over 94% accuracy and high F1 score in both training and validation sets. Some transfer learning-based architectures such as EfficientNet and ResNet50 showed low generalization performance. This study demonstrates the effectiveness of deep learning-based models for computerized diagnosis of intra-ear diseases and provides an important basis for decision support systems to be developed in this field.

Keywords: Deep Learning, Ear Diseases, Otoscopic Images, Image Classification.

OTOSKOPİK GÖRÜNTÜLER ÜZERİNDE DERİN ÖĞRENME KULLANARAK YAYGIN KULAK PATOLOJİLERİNİN SINIFLANDIRILMASI

ÖZ

Bu çalışmada, otoskopik görüntülere dayalı kulak patolojilerinin sınıflandırılması için farklı derin öğrenme mimarilerinin performansı karşılaştırmalı olarak analiz edilmiştir. Veri kümesi dört temel sınıf içermektedir: kronik otitis media, kulak kiri tıkanıklığı, miringoskleroz ve normal kulak yapısı. Görüntüler 224×224 piksel çözünürlükte normalize edilerek modele uygun hale getirilmiş ve sınıflandırma CNN, CNN-LSTM, DenseNet121, ResNet50 ve EfficientNet mimarileri kullanılarak gerçekleştirilmiştir. Eğitim ve doğrulama aşamalarında doğruluk, F1 skoru, kesinlik, geri çağırma ve kayıp değerleri gibi performans metrikleri hesaplanmış, modellerin sınıf ayırt etme gücü ROC eğrileri ve karmaşıklık matrisleri ile değerlendirilmiştir. Sonuçlara göre, CNN+LSTM ve DenseNet121 mimarileri hem eğitim hem de doğrulama setlerinde %94'ün üzerinde doğruluk ve yüksek F1 skoru ile en iyi performansı göstermiştir. EfficientNet ve ResNet50 gibi bazı transfer öğrenme tabanlı mimariler düşük genelleme performansı göstermiştir. Bu çalışma, kulak içi hastalıkların bilgisayarlı teşhisi için derin öğrenme tabanlı modellerin etkinliğini göstermekte ve bu alanda geliştirilecek karar destek sistemleri için önemli bir temel sağlamaktadır.

Anahtar kelimeler: Derin Öğrenme, Kulak Hastalıkları, Otoskopik Görüntüler, Görüntü Sınıflandırma.

***Sorumlu Yazar/Corresponding Author:** yasin.tatli@avrasya.edu.tr

ISSN: 2618-6055 / 9, (1), 51 – 57, 2025
DOI: 10.62301/usmtd.1700194

1.Introduction

Ear diseases can cause health problems such as hearing loss, balance disorder and infection, which are especially common during childhood [1]. Early and accurate diagnosis of such pathologies is critical for both the success of the treatment process and the quality of life of the patient [2]. Since traditional diagnostic methods are largely based on physician experience, they are open to subjective evaluations and may vary in diagnostic accuracy [3]. Therefore, the use of image processing and artificial intelligence-based diagnostic support systems is becoming increasingly important [4].

The intra-aural diseases targeted for classification in this study were divided into four main categories with visually distinguishable pathological features based on otoscopic images [5]. Chronic otitis media is characterized by prolonged inflammation of the middle ear and often perforation of the eardrum, which can lead to hearing loss, discharge and balance problems [6]. Earwax plug is a condition in which the auditory canal is completely or partially blocked by cerumen (earwax) and usually causes conductive hearing loss [7]. Myringosclerosis is characterized by scar tissue with lime-like white opacities on the tympanic membrane, mostly due to previous infections or ventilation tube placements [8]. Finally, normal ear images represent the condition in which there are no pathological findings, and a healthy eardrum can be observed with clear anatomical structures [9]. Such otoscopic classifications are the basis for the training of computer-aided diagnostic systems [10].

In recent years, the successful results of deep learning algorithms in the analysis of medical images have paved the way for a new approach in the automatic classification of intra-aural pathologies. In this context, in this study, the performance of different deep learning architectures (CNN, CNN-LSTM, DenseNet121, ResNet50, EfficientNet) in the classification of intra-aural diseases over otoscopic images is systematically evaluated and performance measurement is performed with comparative analyses.

The organization of this study consists of five main sections. The first section, Introduction, describes the purpose, scope and literature of the study, emphasizing the importance of ear diseases and the need for automatic classification. In the second section, the literature on the topics examined in this study and the analyses of previous studies are presented in detail. In the third section, the dataset used, image processing steps, model architectures (CNN, CNN-LSTM, DenseNet121, ResNet50, EfficientNet) and training processes are detailed. The fourth section includes the analysis part where the experimental results are presented with graphs and tables and the findings are compared technically. Finally, in the fifth section, an overview of the study is given, the results are summarized and recommendations for future work are presented.

2. Related Works

In recent years, deep learning-based studies on computerized analysis of otoscopic images have led to significant developments in the field of automatic diagnosis of ear diseases. Image-based artificial intelligence systems aim to increase the accuracy and reliability in the detection of middle ear pathologies, and different neural network architectures and data processing techniques have been tested in this context.

Lee et al. [5], in their in-depth study of artificial intelligence-supported image analysis systems in the diagnosis of middle ear diseases, revealed that models trained with transfer learning techniques can provide high classification success even in limited data sets. Similarly, the CNN-based system developed by Khan and colleagues [9] provided over 90% accuracy in distinguishing pathologies such as chronic otitis media in oto-endoscopic images. Such approaches demonstrate the potential of image-based decision support systems in clinical applications.

In Młyńczak and Kashani's study [7], a machine learning supported approach is proposed for the detection of effusions in images obtained using shortwave infrared otoscopy. The study shows that successful results can be obtained in the integration of alternative imaging technologies with diagnostic systems. In addition, Mahdavi [8], in his literature review on inflammatory middle ear opacities, pointed out the importance of identifying the distinctive visual features of lesions such as myringosclerosis and chronic otitis media in otoscopic diagnosis processes.

The CNN-LSTM architecture proposed by Afify et al [4] was developed with the aim of improving diagnostic accuracy by capturing spatial and sequential relationships in otoscopic images. The

*Sorumlu Yazar/Corresponding Author: yasin.tatli@avrasya.edu.tr

ISSN: 2618-6055 / 9, (1), 51 – 57, 2025
DOI: 10.62301/usmtd.1700194

researchers stated that this model offers a strong classification performance especially in sequential data. In addition, Singh and Raghuvanshi [11] obtained high success rates with a soft computing-based classification approach in otoscopic images with fuzzy boundaries and showed that such algorithms can work effectively even in images with uncertainty.

Viscaino et al [10] tested machine learning algorithms for the classification of various pathologies using an otoscopic dataset presented on the Figshare platform, and significant successes were recorded, especially in classical CNN and SVM-based systems. This dataset is used as a standard source in many studies on otoscopic images.

Chan and Stephenson's clinical guideline [1] discusses the role of otoscopic examination in the diagnosis of diseases such as otitis media, which is common in children, and how this can be enhanced by digitally assisted diagnostic systems. Furthermore, Bone [3], in his comprehensive review of otoscopic examination protocols, argues that AI-assisted analyses can improve reliability, especially in the paediatric population.

On the other hand, studies on the adaptation of current deep learning structures such as the EfficientNet architecture for ear pathologies [12] emphasize that each model should be customized according to the data and the problem. In a systematic review by Rony et al. [13], it was revealed that various architectures offer different levels of performance in pathologies such as cerumen impaction and myringosclerosis and that there is no generalized architecture.

In conclusion, current trends in the literature show that deep learning plays an important role in integrating image-based decision support systems into clinical diagnosis processes and that the choice of architecture should be customized according to the data structure. The present study aims to extend this literature and contribute to the classification of intra-aural pathologies with higher accuracy using otoscopic images.

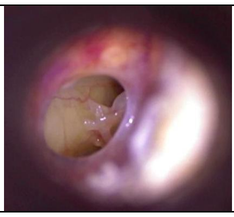
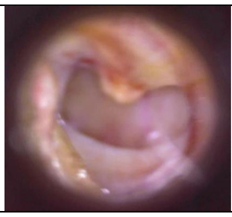
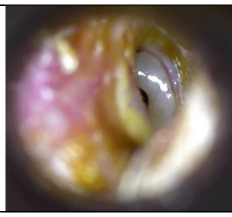
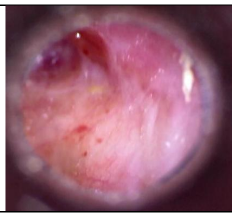
3. Material and Methods

In this section, the dataset used for the classification of intra-ear diseases is introduced. CNN, CNN-LSTM, DenseNet121, ResNet50 and EfficientNet models were applied to train the dataset. The visual data were pre-processed to a fixed size and normalized to be suitable for the model. Each model was trained using the same training and validation set, and the success levels were systematically analyzed with the metrics obtained.

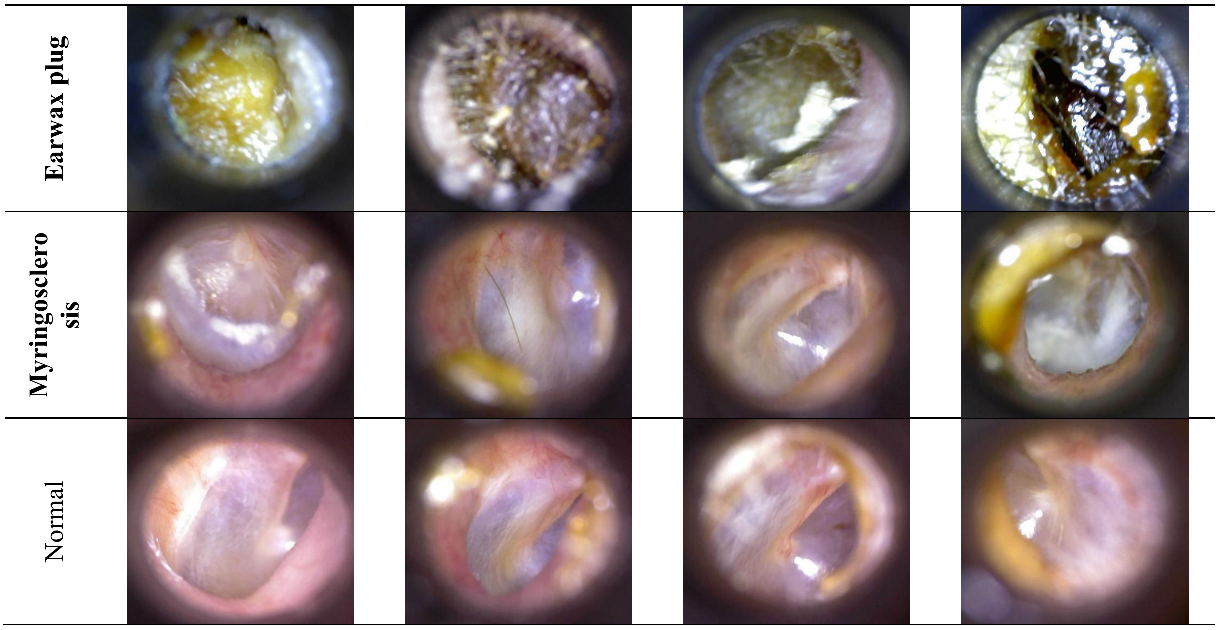
3.1. Dataset

The image data used in this study were obtained from the open access dataset called "Ear Imagery Database", shown in Table 1, shared by Michelle Viscaino and Fernando Auat Cheein on the Figshare platform [14]. The dataset aims to provide a standardized resource for the classification of ear diseases and medical image analysis. The images consist of real patient data acquired with otoscope devices in clinical settings and cover various categories representing different ear disorders. The dataset is organized to include several classes, each representing a specific pathological condition. These classes include Chronic Otitis Media, Chronic Middle Ear Inflammation, Earwax Plug, Myringosclerosis and Normal ear images (800 images). The dataset consisted of 800 otoscopic images, equally distributed among four classes (200 images each for chronic otitis media, earwax plug, myringosclerosis, and normal).

Table 1. Ear Imagery Database samples with classes

Class	Samples			
Chronic otitis media				

ISSN: 2618-6055 / 9, (1), 51 – 57, 2025
DOI: 10.62301/usmtd.1700194



Each model was trained using the Adam optimizer with an initial learning rate of 0.0001, a batch size of 32, and a total of 50 epochs. The categorical cross-entropy loss function was used for multiclass classification. Early stopping was applied with a patience value of 10 epochs based on validation loss to avoid overfitting.

4. Findings and Results

Within the scope of the experimental study, the applicability of different deep learning architectures to the in-ear disease classification problem was tested. The results obtained are shown in Table 2, allowing the performance of each architecture in both training and validation phases to be evaluated on multidimensional metrics. CNN+LSTM and DenseNet121 models achieved remarkably high accuracy, F1 score, precision and recall values in both phases, significantly outperforming the other architectures. This indicates that both the learning and generalisation capabilities of these models are quite strong.

Table 2. Machine learning algorithms training and testing metric values

Metrics & Algorithms	CNN+LSTM	CNN	DenseNet121	ResNet50	EfficientNet
Accuracy	1.0000	0.9842	1.0000	0.5741	0.2698
f1_m	1.0000	0.9832	1.0000	0.5703	0.0293
Loss	0.0029	0.0477	$1.53e^{-06}$	0.8785	1.4864
precision_m	1.0000	0.9842	1.0000	0.6735	0.1152
recall_m	1.0000	0.9823	1.0000	0.5082	0.0172
val_accuracy	0.8562	0.8562	0.9438	0.6187	0.2500
val_f1_m	0.8526	0.8562	0.9395	0.5021	$0.0e^{+00}$
val_loss	0.5708	1.1348	0.5325	0.7633	1.4109
val_precision_m	0.8556	0.8562	0.9417	0.5365	$0.0e^{+00}$
val_recall_m	0.8500	0.8562	0.9375	0.4813	$0.0e^{+00}$

The CNN+LSTM architecture has increased its capacity to represent sequential or complex spatial relationships with the effect of LSTM layers that allow the modelling of the time dimension. The model produced a very low loss value of 0.0029 with 100% training accuracy and F1 score. This value shows that the classification process is almost error-free. Similarly, the DenseNet121 architecture also showed 100% success in the same training metrics. However, the accuracy rate (94.38%) and F1 score (93.95%) of DenseNet121 in the validation phase are slightly superior to the validation performance (85.62%) of CNN+LSTM. This result suggests that CNN+LSTM may have a stronger tendency to overfitting. The classical CNN architecture showed a balanced and stable performance in both training and validation phases. In the training phase, a successful result was obtained with an accuracy of 98.42% and an F1 score of 98.32%, while the performance in the validation phase (accuracy: 85.62%, F1: 85.62%) was lower compared to DenseNet121, but quite adequate.

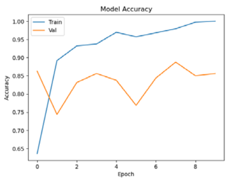
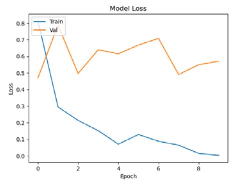
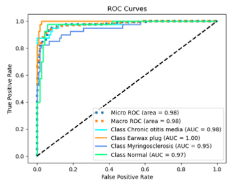
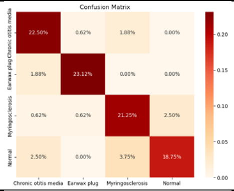
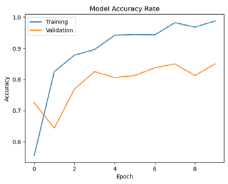
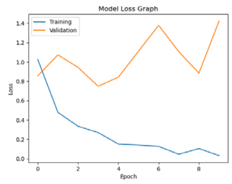
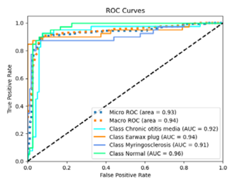
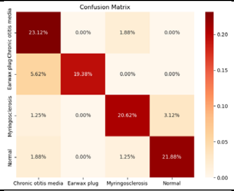
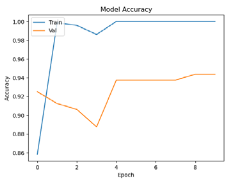
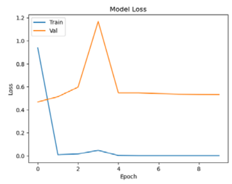
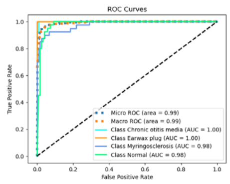
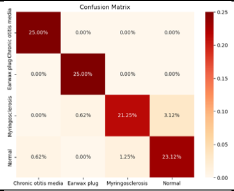
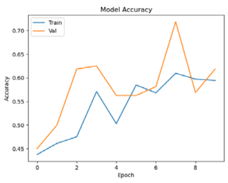
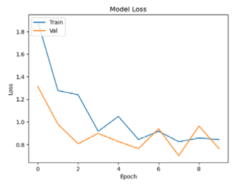
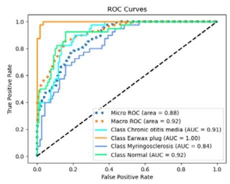
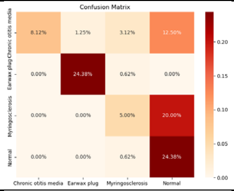
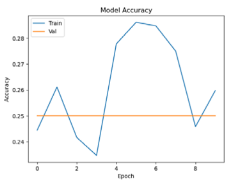
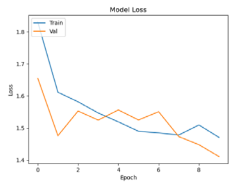
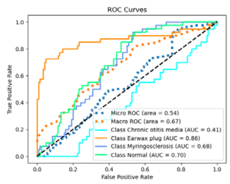
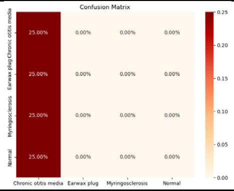
*Sorumlu Yazar/Corresponding Author: yasin.tatli@avrasya.edu.tr

ISSN: 2618-6055 / 9, (1), 51 – 57, 2025
DOI: 10.62301/usmtd.1700194

Furthermore, the validation loss (val_loss: 1.1348) indicates that there is still potential for improvement in the learning curve. The ResNet50 and EfficientNet models did not perform at the expected level within the scope of the study. The training accuracy of ResNet50 was 57.41% and the F1 score was 57.03%, while only 61.87% accuracy and 50.21% F1 score were obtained in the validation phase. This shows that the model does not adapt effectively to either the training data or the validation data. More strikingly, EfficientNet architecture produced very low scores of 26.98% accuracy and only 1.72% sensitivity in training, and almost zero success (F1: 0.00) in the validation phase. These results suggest that EfficientNet is not well suited to this dataset or model configuration and suffers from a serious learning problem.

Machine learning algorithms training and testing graphs shown in Table 3.

Table 3. Machine learning algorithms training and testing graphs

Model	Accuracy Graph	Loss Graph	ROC Curve	Confusion Matrix
CNN+LSTM				
CNN				
DenseNet121				
Resnet50				
EfficientNet				

Accuracy and loss graphs reflect that the CNN+LSTM model shows stable and high performance in both training and validation sets. Accuracy clearly increases throughout the training process, while the loss value remains low and stable. The ROC curve shows that the discrimination power between the classes is quite high (the curves are close to the upper left corner) and the AUC values are around 1.0. The confusion matrix confirms that accurate classification is performed with almost no room for error between classes. The CNN architecture performed an effective learning process with a rapid increase in the accuracy graph and a low loss value in training. In the validation data, although slight fluctuations were observed in the curves, high accuracy and low loss were generally maintained. The ROC curve shows that the discrimination between classes is strong, and the Confusion Matrix shows that the confusion between classes remains at a very low level. The training accuracy graph of the DenseNet121 architecture shows that the model reaches high accuracy very quickly, while the loss graph shows very low values.

The fact that the AUC values are close to 1.0 for almost all classes in the ROC curves proves the discriminative power of the model between classes. Confusion matrix provides almost 100% accuracy

*Sorumlu Yazar/Corresponding Author: yasin.tatli@avrasya.edu.tr

ISSN: 2618-6055 / 9, (1), 51 – 57, 2025
DOI: 10.62301/usmtd.1700194

in almost all classes. In the ResNet50 architecture, accuracy values rise and fall during the training process and fluctuations are observed in the loss graph. This indicates that the model cannot fully adapt to the data and experiences inconsistencies in the learning process. The ROC curves generally have lower AUC values, and the curves are close to the 0.5 line, indicating that the model performs close to randomly predicting some classes. Although the EfficientNet architecture aims for high performance with a low number of parameters in literature, in this study it performed well below expectations. The accuracy graph remains at low levels, and the loss value increases rapidly in the validation phase. The ROC curves clearly show that it cannot distinguish the classes, and the AUC values indicate that the classes are randomly predicted. Since the confusion matrix contains failed predictions in almost all classes, it can be concluded that this model is not compatible with the current data set or problem.

To further investigate the poor performance of the EfficientNet architecture, an ablation study was conducted. We examined different EfficientNet variants (B0, B1) and modified input normalization schemes and batch sizes. None of these adjustments led to significant improvements, suggesting that the model's depth and compound scaling may not align well with the limited dataset size and visual patterns in otoscopic images. Further analysis of misclassified images revealed that EfficientNet tended to confuse wax obstruction and chronic otitis media cases, possibly due to similar background textures.

5. Conclusions

In this study, the performance of various deep learning architectures for the classification of intra-ear diseases on otoscopic images is evaluated in detail. Experimental findings clearly show that CNN+LSTM and DenseNet121 models achieve high accuracy, F1 score, precision and sensitivity values in both training and validation processes. These two models stood out with their capacity to capture and generalise complex spatial patterns in images of intra-aural pathologies. On the other hand, the classical CNN architecture also provided balanced and reliable results, showing that effective classification is also possible with simpler structures. On the other hand, more complex structures based on transfer learning such as ResNet50 and EfficientNet have shown limited success in this context, and EfficientNet could not distinguish between classes in the validation phase. This emphasises that the compatibility of model architectures with specific problem types and datasets is not always directly proportional, and that data set specific configurations are required.

Unlike prior studies that focus predominantly on binary classification (e.g., diseased vs. normal), this work addresses the multiclass classification of four clinically distinct ear pathologies. Moreover, by comparing both custom CNN models and popular transfer learning architectures on a standardized dataset, this study provides a benchmark for future research in otoscopic image analysis.

In conclusion, this study not only demonstrates the significant potential of deep learning in the automatic diagnosis of ear diseases but also contributes to the field by presenting the relative performance of different model approaches in a comparative manner. It is predicted that the classification performance can be further improved in future studies by examining the hyperparameter optimisation of the models in detail.

References

[1] J. Chan, K. Stephenson, Diagnosis and management of middle ear disease in children. *Paediatrics and Child Health 33*(12) (2023) 376-381.

[2] T. Marom, O. Kraus, N. Habashi, S.O. Tamir, Emerging technologies for the diagnosis of otitis media. *Otolaryngology–Head and Neck Surgery 160*(3) (2019) 447-456.

[3] A. Bone, Middle Ear. *Pediatric Physical Examination-E-Book: Pediatric Physical Examination-E-Book* (2023) 192.

[4] H.M. Afify, K.K. Mohammed, A.E. Hassanien, Insight into automatic image diagnosis of ear conditions based on optimized deep learning approach. *Annals of biomedical engineering 52*(4) (2024) 865-876.

[5] D. Song, T. Kim, Y. Lee, J. Kim, Image-based artificial intelligence technology for diagnosing middle ear diseases: a systematic review. *Journal of Clinical Medicine 12*(18) (2023) 5831.

[6] F. Larrosa, L. Pujol, E. Hernández-Montero, Chronic otitis media. *Medicina Clínica (English Edition)*, (2025) 106915.

*Sorumlu Yazar/Corresponding Author: yasin.tatli@avrasya.edu.tr

ISSN: 2618-6055 / 9, (1), 51 – 57, 2025
DOI: 10.62301/usmtd.1700194

[7] R.G. Kashani, M.C. Młyńczak, D. Zarabanda, P. Solis-Pazmino, D.M. Huland, I.N. Ahmad, T.A. Valdez, Shortwave infrared otoscopy for diagnosis of middle ear effusions: A machine-learning-based approach. *Scientific Reports 11*(1) (2021) 12509.

[8] A. Mahdavi, Diagnostic and imaging findings in inflammatory Opacifications of the middle ear: A review of the literature. *The International Tinnitus Journal 27*(2) (2023) 146-153.

[9] M.A. Khan, S. Kwon, J. Choo, S.M. Hong, S.H. Kang, I.H. Park, S.J. Hong, Automatic detection of tympanic membrane and middle ear infection from oto-endoscopic images via convolutional neural networks. *Neural Networks 126* (2020) 384-394.

[10] M. Viscaino, J.C. Maass, P.H. Delano, M. Torrente, C. Stott, F. Auat Cheein, Computer-aided diagnosis of external and middle ear conditions: A machine learning approach. *Plos one 15*(3) (2020) e0229226.

[11] A. K. Singh, A.S. Raghuvanshi, R. Mehta, A Soft Computing Approach for Efficient Diagnosis of Otitis Media Infection by Mucosal Disease Early Detection and Referrals. In *2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) (*2024) (pp. 687-692) IEEE.

[12] Z. Cao, F. Chen, E.M. Grais, F. Yue, Y. Cai, D.W. Swanepoel, F. Zhao, Machine learning in diagnosing middle ear disorders using tympanic membrane images: a meta-analysis, *The Laryngoscope 133*(4) (2023) 732-741.

[13] M. A. H. Rony, K. Fatema, M. A. K. Raiaan, M. M. Hassan, S. Azam, A. Karim, A. Leach Artificial intelligence-driven advancements in otitis media diagnosis: a systematic review (2024) *Ieee access*.

[14] M. Viscaino and F.A. Cheein , “Ear imagery database,” *Figshare*, (2020) [Online] Available: https://figshare.com/articles/dataset/Ear_imagery_database/11886630

***Sorumlu Yazar/Corresponding Author:** yasin.tatli@avrasya.edu.tr