Koc. J. Sci. Eng., 5(1): (2022) 105-111 https://doi.org/10.34088/kojose.1081402

Kocaeli University

Kocaeli Journal of Science and Engineering

http://dergipark.org.tr/kojose

Classification of Tympanic Membrane Images based on VGG16 Model

Abidin ÇALIŞKAN [1, *]

[1] *Department of Computer Engineering, Batman University, Batman, 72100, Turkey,* ***ORCID:*** *0000-0001-5039-6400*

Article Info

Research paper

Received : *March 01, 2022*
Accepted : *March 31, 2022*

Keywords

Convolution Neural Networks
VGG16
Tympanic Membrane
Otitis Media
Classification
Support Vector Machines

Abstract

Otitis Media (OM) is a type of infectious disease caused by viruses and/or bacteria in the middle ear cavity. In the current study, it is aimed to detect the eardrum region in middle ear images for diagnosing OM disease by using artificial intelligence methods. The Convolution Neural Networks (CNN) model and the deep features of this model and the images obtained with the otoscope device were used. In order to separate these images as Normal and Abnormal, the end-to-end VGG16 model was directly used in the first stage of the experimental work. In the second stage of the experimental study, the activation maps of the fc6 and fc7 layers consisting of 4096 features and the fc8 layer consisting of 1000 features of the VGG16 CNN model were obtained. Then, it was given as input to Support Vector Machines (SVM). Then, the deep features obtained from all activation maps were combined and a new feature set was obtained. In the last stage, this feature set is given as an input to SVM. Thus, the effect of the VGG16 model and the features obtained from the layers of this model on the success of distinguishing images of the eardrum was investigated. Experimental studies show that, the best performance results were obtained for the fc6 layer with an accuracy rate of 82.17%. In addition, 71.43%, 90.62% and 77.92% performance criteria were obtained for sensitivity, specificity and f-score values, respectively. Consequently, it has been shown that OM disease could be accurately detected by using a deep CNN architecture. The proposed deep learning-based classification system promises highly accurate results for disease detection.

1. Introduction

The ear is an anatomical structure which exhibits the function of hearing and contains the balance organ. Sound waves pass through the outer ear and hit the eardrum and make it vibrate. When the sound reaches the middle ear, it is amplified by the ossicles in the middle ear and transmitted to the inner ear. Sound waves coming to the inner ear are also received by all the cells situated here and transmitted to the auditory nerve. Then, it is transferred to the auditory center in the brain via the auditory nerve. The overall structure of the ear is given in Figure 1 [1].

Symptoms of ear diseases and causes of ear diseases can be observed in different ways depending on the condition of the person. Ear diseases can occur due to seasonal reasons, infection or exposure to excessively loud noise [2]. One of the most common diseases among ear diseases is OM whishes an infectious disease caused by viruses or bacteria in the middle ear cavity [3].

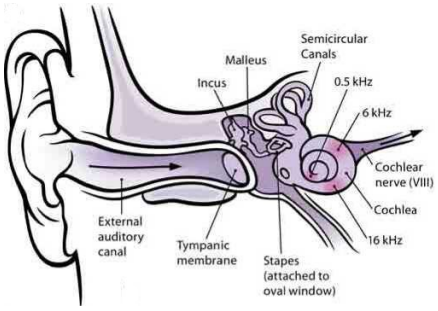

Figure 1. Ear internal structure [1].

The middle ear cavity is the inner part of the eardrum and contains the ear ossicles, which are filled with air. It can be classified as acute, subacute or chronic in accordance with the duration of the infection. Generally, middle ear infections lasting less than three weeks are called acute OM and middle ear infections lasting longer than 3 months are called chronic OM [4].

In the literature, Active Contour Segmentation has

* Corresponding Author: abidin.caliskan@batman.edu.tr

been used to detect the membrane region from middle ear images for the diagnosis of OM. After the membrane region was determined, the feature set was obtained by textural feature extraction algorithms. In addition, the images were classified using the AdaBoost method by adding Gabor features [5]. A portable video otoscopy platform is presented for better visual analysis of the images obtained with the video otoscope device [6]. By targeting and desiring the different regions of the eardrum of the clinical importance, the ossicles and the lesion area in the middle ear were clearly visualized by the linear Unsharp algorithm and low-pass filtering. Hierarchical tree diagram method was used for classification, by using average color value. Canny edge detection algorithm is imployed to in order to monitor the presence of bubbles in the image. An accuracy rate of 84% was obtained by this classification [7]. A new automatic algorithm with a classification success of 89.9% [8] has been proposed to recognize OM under three categories as normal, acute otitis media (AOM) and otitis media effusion (OME). The feature set obtained by subtracting the textural feature extraction algorithm and the average values of the color channels of the common ear images was determined by Başaran [9]. In this study, classification was made with the artificial neural network method with an accuracy rate of 76.14%. A classification model based on the usefulness and performance of a two-stage CNN has been developed for automatic diagnosis of OM from Tympanic Membrane (TM) images [10]. Based on the Class Activation Map, an accuracy of 93.4% was achieved using a two-stage classification model and a three-fold cross-validation network in order to improve accuracy and reliability and also to read TM images. Consequently, literature studies show that CNN has shown great success in the classification of diseases with biomedical images in recent years [11, 12].

In this study, the VGG16 CNN model, which has been proven to be successful in the literature, was used to distinguish OM images [13, 14].

Since different information about images is learned with each layer in CNN architectures, in this study, the success of the learning layers of the VGG16 model in distinguishing OM images was examined. In addition, normal and abnormal TM images were distinguished by combining the OM images and the features obtained with the learning layers.

2. Materials and Methods

Images have been resized in the proposed model to classify normal and abnormal images. After the activation maps of the layers were obtained, they were applied as an input to the SVM classifier model. The feature set obtained is given as an input to the SVM. Thus, the effect of the CNN model and the features obtained from the layers of this model to distinguish images was examined. In order to obtain the most effective results with CNN, tests were carried out with different parameter values. The block diagram of the proposed model is given in Figure 2.

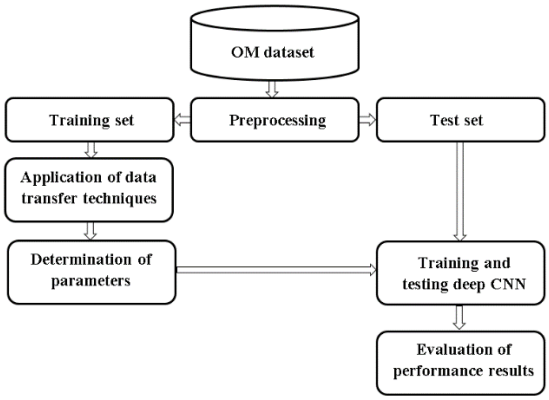

Figure 2. Block diagram of the proposed model.

2.1. Data Set

The data set consists of total 956 middle ear images which were obtained by otoscope device from volunteer patients who were examined in Van Akdamar Hospital in Turkey. The dataset includes 535 images belonging to the normal class and 421 images belonging to the abnormal class. Abnormal images include diseases such as “AOM, CSOM, Earwax, Otitis externa, Tympanoskleros, Ear ventilation tube, Pseudo-membranes, Foreign bodies in the ear” [15]. The number of images in the data set are given in Table 1 and the images of these classes are given in Figure 3.

Table 1. Total number of images in the dataset.

Class	Number of images
Normal	**535**
Abnormal	**421**
AOM	119
CSOM	63
Earwax	140
Otitis externa	41
Tympanoskleros	28
Ear ventilation tube	16
Pseudo-membranes	11
Foreign bodies in the ear	3

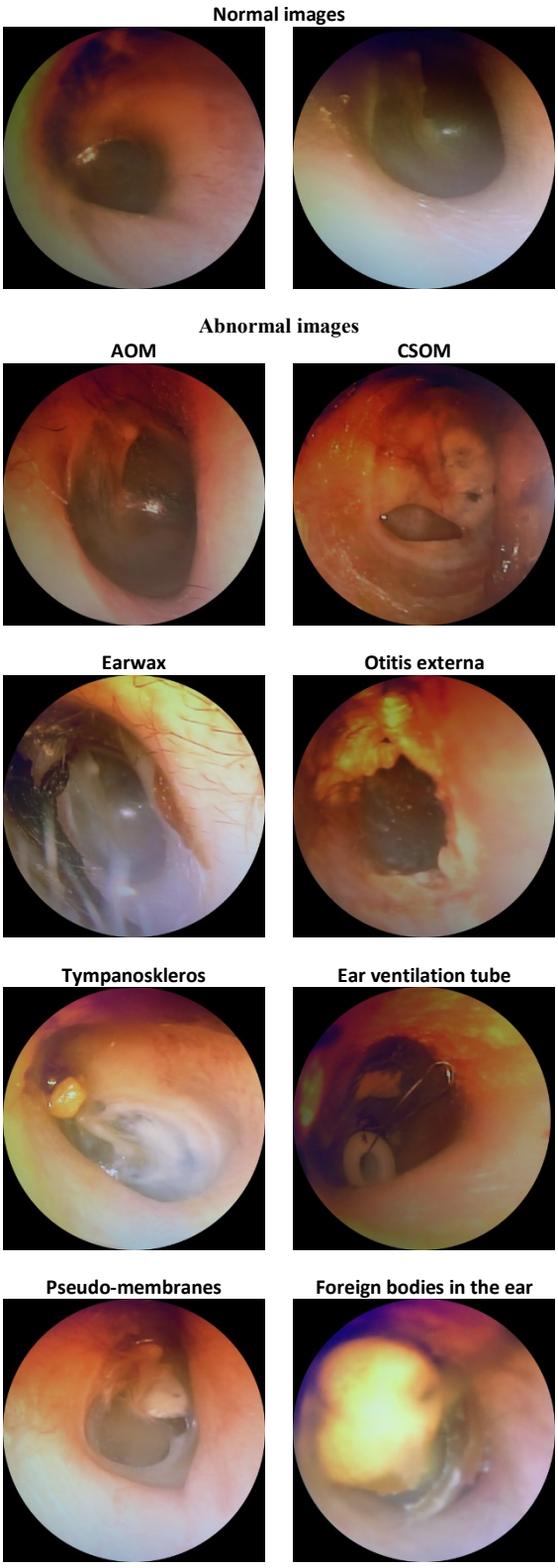

Figure 3. Normal and abnormal Tympanic Membrane images.

2.2. Convolution Neural Networks

CNN is a feed-forward artificial neural network [16], which is a type of multi-layer perceptron. Each layer in the CNN architecture has different tasks. They work in three dimensions in terms of depth, height and width. The CNN architecture performs well in many areas such as disease diagnosis, classification and segmentation in medical images [17, 18]. The architecture of an example VGG16 CNN model consisting of five convolutional layers and the fc6, fc7, fc8 and softmax layers are presented in Figure 4 [19, 20].

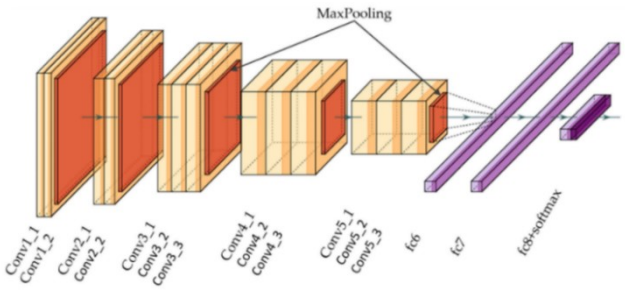

Figure 4. The architecture of the VGG-16 model.

Raw data is taken at the beginning of the CNN model and classification processes are performed after obtaining the layers and feature map. A typical CNN model includes input layer, convolution layer, pooling, fully connected layer and output layer. A feature map is obtained by scanning the image from left to right and top to bottom with the filters that are used to obtain the feature map in the convolution layer.

The first layers learn the edge and color information through the convolution layers of CNN and the kernels within the remaining layers learn the information about the image details [21]. The mathematical expression of the convolution is given in Eq. (1) [13].

$$(f * h)[m,n] = \sum_{j}\sum_{k} h[j,k], f(m-j, n-k) \quad (1)$$

Where f is the input image, h is the filter, m and n are the row and column indexes of the output matrix, respectively. j and k represent the location information of the filter.

The VGG16 network model, a CNN trained on more than one million images, was retrieved from the ImageNet database [22]. This model includes layers with learning weights, followed by layers with ReLu and pooling layers. Learnable layers consist of convolutional layers and fully-connected layers. Activations fc6 and fc7 in layers are used to extract feature vectors [23].

2.3. Support Vector Machines

SVM is designed to classify two-class linear data. Later on, it has been generalized to classify more classes and nonlinear data [24]. SVM tries to minimize the upper

bound of the error with the hyperplane. It is also a classification algorithm that maximizes the boundary separation between training data [25]. Unlike other machine learning methods, the performance of SVM, having a strong learning and generalization ability, is quite better [26]. The purpose of SVM is to find the plane with the largest distance between the two planes expressed in Eq. (2) [27].

$$\begin{cases} w.x_i + b = -1 \\ w.x_i + b = +1 \end{cases} \quad (2)$$

2.4. Performance Metrics

In this study, confusion matrix was used to test the performance of experimental studies. There are four different Performance Metrics in the Confusion matrix where actual and predicted records are kept to calculate various metrics. Here, True positive (TP) and True negative (TN) denote the number of correctly identified positive and negative records. False positive (FP) and False negative (FN) represent the number of records belonging to incorrectly predicted positive and negative classes [2]. Accordingly, the accuracy, sensitivity, specificity and F-Score metrics derived from the error matrix are calculated as shown in Eq. (3)-(6) [13].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

$$\text{F} - \text{Score} = \frac{2 * TP}{2 * TP + FP + FN} \quad (6)$$

In addition to the performance criteria, the Receive Characteristic Curve (ROC) was used to measure the success of the model. The ROC curve is one of the most important evaluation criteria to control the performance of the classification model. A typical ROC curve has the FP rate on the x-axis and the TP rate on the y-axis. Area Under the Curve (AUC) can be considered as a summary of model performance. The size of the AUC area indicates the success of the machine learning model in distinguishing the classes. The ideal value for AUC is 1 [25].

3. Results and Discussion

In the first stage of the experimental study, the images obtained with the otoscope device via the VGG16 CNN model and the deep features of this model were directly used from the end-to-end VGG16 model in order to separate the images as normal and abnormal. In order to optimize the network, the number of data to be trained per iteration (mini-batch size) is set to 32, and the maximum epoch value (max-epoch) to train the entire dataset is 32. The initial learning rate value was $1x10^{-4}$ and the stochastic gradient descent was used as optimization. As a result of the experimental study, an accuracy rate of 78.67% was obtained, while the sensitivity, specificity and f-score values of the model were measured as 79.37%, 78.12% and 76.63%, respectively. In the second stage of the experimental study, the activation maps of the fc6 and fc7 layers consisting of 4096 features and the fc8 layer consisting of 1000 features of the VGG16 CNN model were obtained. Then, the activation maps were given as an input to SVM. After that, deep features from all activation maps were combined and a new feature set was obtained. Finally, this feature set was then given as an input to the SVM and the effect of the features obtained from the VGG16 model and its layers on the success of distinguishing the images of the eardrum was examined. As a result of the experimental study, the best performance results were obtained for the fc6 layer. An accuracy rate of 82.17% was obtained with the fc6 layer. In addition, 71.43%, 90.62% and 77.92% performance criteria were obtained for sensitivity, specificity and f-score values, respectively. The confusion matrix and ROC curve of the fc6 layer, which the best results are obtain illustrated in Figure 5 and Figure 6, respectively.

Normal	100	26
Abnormal	35	125
	Normal	Abnormal

Figure 5. The confusion matrix obtained with the VGG16 model.

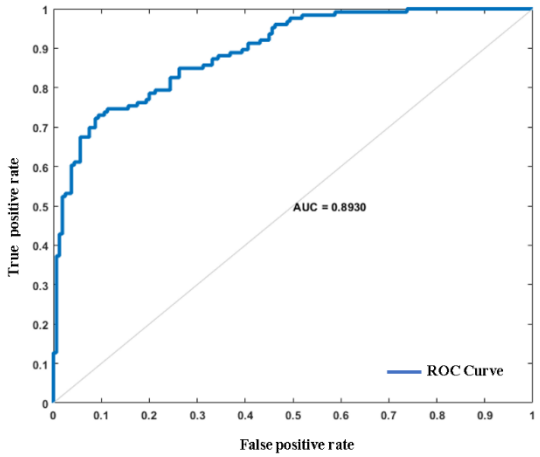

Figure 6. ROC curve obtained with the VGG16 model.

The confusion matrix of the classification of deep features of the fc6 layer by SVM is given in Figure 7. The ROC curve is given in Figure 8.

Normal	90	36
Abnormal	15	145
	Normal	Abnormal

Figure 7. The complexity matrix of the classification of deep features of the fc6 layer by SVM.

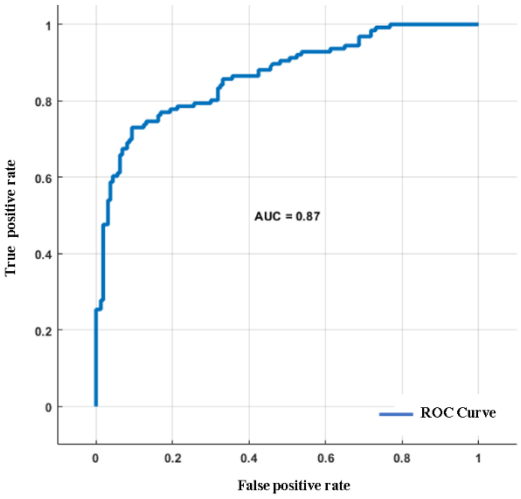

Figure 8. ROC curve of the classification of deep features of the fc6 layer by SVM.

The performance criteria obtained as a result of the classification of fc6, fc7, fc8 and the feature sets created by combining these layers with SVM are given in Table 2. In addition, the accuracy rates directly obtained with the VGG16 model and layers as a result of experimental studies are given in Figure 9.

Table 2. Classification results of the feature layers of the VGG16 model based on SVM.

Layer	Number of Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-Score (%)
fc6	**4096**	**82.17**	**71.43**	**90.62**	**77.92**
fc7	4096	72.73	65.08	78.75	67.77
fc8	1000	77.62	70.63	76.72	73.55
fc6+fc7+fc8	9192	79.37	73.02	84.38	75.72

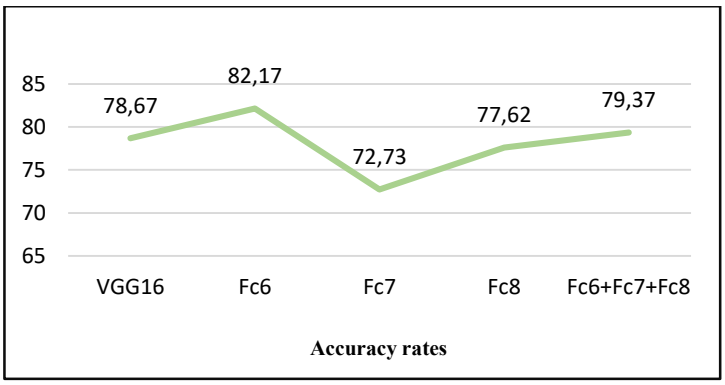

Figure 9. Accuracy rates obtained with VGG16 and layers.

4. Conclusion

Accurate diagnosis of OM, which causes hearing loss, is extremely important for the correct treatment of the disease. In this study, the VGG16 model and its deep features were used to separate the images as normal and abnormal obtained by the otoscope device. Activation maps of fc6, fc7 and fc8 layers, which are the deep features of the VGG16 model, were obtained and given as an input to the SVM. A new feature set was derived by combining the deep features acquired from all activation maps. Then, as an input to the SVM, the success of the VGG16 model and the features obtained from the layers of this model in distinguishing the eardrum images was examined. The experimental study shows that the best performance results were obtained with the fc6 layer as 82.17%. In addition, 71.43%, 90.62% and 77.92% performance criteria were obtained for sensitivity, specificity and f-score values, respectively. In conclusion, a high-accuracy diagnostic model has been developed to help diagnose OM disease from the normal and abnormal images captured by the otoscope device.

Future studies may be conducted on different deep CNN architectures and to develop models that can give higher performance outputs.

Conflict of Interest

No conflict of interest was stated by the author.

Acknowledgements

The author would like to thank the editor and the reviewers for their valuable suggestions and comments.

Declaration of Ethical Standards

The author of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

References

[1] Chittka L., Brockmann, A., 2005. Perception space the final frontier. PLoS biology, **3**(4), pp. 564-568.

[2] Wu Z., Lin Z., Li L., Pan H., Chen G., Fu Y., Qiu Q., 2021. Deep learning for classification of pediatric otitis media. The Laryngoscope, **131**(7), E2344-E2351.

[3] Cetinkaya E. A., Topsakal V., 2022. Acute Otitis Media. In Pediatric ENT Infections, Springer, Cham, pp. 381-392.

[4] Manju K., Paramasivam M. E., Nagarjun S., Mokesh A., Abishek A., Meialagan, K., 2022. Deep Learning Algorithm for Identification of Ear Disease. In Proceedings of International Conference on Data Science and Applications, Springer, Singapore, pp. 491-502.

[5] Shie C. K., Chang H. T., Fan F. C., Chen C. J., Fang T. Y., Wang P. C., 2014. A hybrid feature-based segmentation and classification system for the computer aided self-diagnosis of otitis media. In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, pp. 4655-4658.

[6] Cheng L., Liu J., Roehm C. E., Valdez T. A., 2011. Enhanced video images for tympanic membrane characterization. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, pp. 4002-4005.

[7] Kuruvilla A., Li J., Yeomans P. H., Quelhas P., Shaikh N., Hoberman A., Kovačević J., 2012. Otitis media vocabulary and grammar. In 2012 19th IEEE International Conference on Image Processing, IEEE, pp. 2845-2848.

[8] Kuruvilla A., Shaikh N., Hoberman A., Kovačević J., 2013. Automated diagnosis of otitis media: vocabulary and grammar. International Journal of Biomedical Imaging.

[9] Başaran E., Şengür A., Cömert Z., Budak Ü., Çelik Y., Velappan S., 2019. Normal and acute tympanic membrane diagnosis based on gray level co-occurrence matrix and artificial neural networks. In 2019 international artificial intelligence and data processing symposium (IDAP), IEEE, pp. 1-6.

[10] Cai Y., Yu J. G., Chen Y., Liu C., Xiao L., Grais E. M., Zhao F., Lan L., Zeng S., Zeng J., Wu M., Su Y., Li Y., Zheng Y., 2021. Investigating the use of a two-stage attention-aware convolutional neural network for the automated diagnosis of otitis media from tympanic membrane images: a prediction model development and validation study. BMJ open, **11**(1), e041139.

[11] Albashish, D., Al-Sayyed, R., Abdullah, A., Ryalat, M. H., Almansour, N. A., 2021. Deep CNN model based on VGG16 for breast cancer classification. In 2021 International Conference on Information Technology (ICIT), IEEE, pp. 805-810.

[12] Tripathi, S., Verma, A., Sharma, N., 2021. Automatic segmentation of brain tumour in MR images using an enhanced deep learning approach. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, **9**(2), pp. 121-130.

[13] Başaran E., Cömert Z., Çelik Y., 2020. Convolutional neural network approach for automatic tympanic membrane detection and classification. Biomedical Signal Processing and Control, **56**, 101734.

[14] Rehman, A., Naz, S., Razzak, M. I., Akram, F., Imran, M., 2020. A deep learning-based framework for automatic brain tumors classification using transfer learning. Circuits, Systems, and Signal Processing, **39**(2), pp. 757-775.

[15] Zafer C., 2020. Fusing fine-tuned deep features for recognizing different tympanic membranes. Biocybernetics and Biomedical Engineering, **40**(1), pp. 40-51.

[16] Hiremani V. A., Senapati K. K., 2021. Quantifying apt of RNN and CNN in Image Classification. In Proceeding of Fifth International Conference on Microelectronics, Computing and Communication Systems, Springer, Singapore, pp. 721-733.

[17] Singh S. P., Wang L., Gupta S., Goli H., Padmanabhan P., Gulyás B., 2020. 3D deep learning on medical images: a review. Sensors, **20**(18), 5097.

[18] Li Y., Sixou B., Peyrin F., 2021. A review of the deep learning methods for medical images super resolution problems, IRBM, **42**(2), pp. 120-133.

[19] Kattenborn T., Leitloff J., Schiefer F., Hinz S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, **173**, pp. 24-49.

[20] Lee J. Y., Choi S. H., Chung J. W., 2019. Automated classification of the tympanic membrane using a convolutional neural network, Applied Sciences, **9**(9), 1827.

[21] Tripathi, M., 2021. Analysis of convolutional neural network based image classification techniques. Journal of Innovative Image Processing (JIIP), **3**(02), pp. 100-117.

[22] Dhillon A., Verma G. K., 2020. Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, **9**(2), pp. 85-112.

[23] Yao G., Lei T., Zhong J., 2019. A review of convolutional-neural-network-based action recognition. Pattern Recognition Letters, **118**, pp. 14-22.

[24] Nalepa J., Kawulok M., 2019. Selecting training sets for support vector machines: a review. Artificial Intelligence Review, **52**(2), pp. 857-900.

[25] Uçar M., Akyol K., Atila Ü., Uçar E., 2021. Classification of different tympanic membrane conditions using fused deep hypercolumn features and bidirectional LSTM. IRBM.

[26] Wang Z., Cha Y. J., 2021. Unsupervised deep learning approach using a deep auto-encoder with a one-class support vector machine to detect damage. Structural Health Monitoring, **20**(1), pp. 406-425.

[27] Alhudhaif A., Cömert Z., Polat K., 2021. Otitis media detection using tympanic membrane images with a novel multi-class machine learning algorithm. PeerJ Computer Science, **7**, e405.