Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi
Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering

Basılı/Printed ISSN: 1302-9304. Elektronik/Online ISSN: 2547-958X

Determination of Tympanic Membrane Region in the Middle Ear Otoscope Images with Convolutional Neural Network Based YOLO Method

Konvolüsyonel Sinir Ağı Tabanlı YOLO Yöntemi ile Orta Kulak Otoskop Görüntülerinde Timpanik Membran Bölgesinin Belirlenmesi

Erdal Başaran [1*], Zafer Cömert [2], Yüksel Çelik [3], Subha Velappan [4], Mesut Toğaçar [5]

[1] Ağrı İbrahim Çeçen Üniversitesi Eğitim Fakültesi Bilgisayar ve Öğretim Teknolojileri Eğitimi, İzmir, TÜRKİYE
[2] Samsun Üniversitesi Mühendislik Fakültesi Yazılım Mühendisliği, Samsun, TÜRKİYE
[3] Karabük Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği, Karabük, TÜRKİYE
[4] Manonmaniam Sundaranar University, Department of Computer Science and Engineering, Tamil Nadu/INDIA
[5] Fırat Üniversitesi, Teknik Bilimler Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, Elazığ/TÜRKİYE
Sorumlu Yazar / Corresponding Author *: erdalbasaran085@gmail.com

Geliş Tarihi / Received: 11.09.2019 *Araştırma Makalesi/Research Article*
Kabul Tarihi / Accepted: 28.04.2020 *DOI:10.21205/deufmd.2020226625*
Atıf şekli/ How to cite: BASARAN, E., COMERT, Z., CELIK, Y.,VELAPPAN, S., TOGACAR, M., (2020). Determination of Tympanic Membrane Region in the Middle Ear Otoscope Images with Convolutional Neural Network Based YOLO Method. DEUFMD 22(66), 919-928.

Abstract

Due to inflammation of the middle ear, various deformations occur in the eardrum. In order to diagnose the disease, it is necessary to examine the tympanic membrane in detail with an otoscope. In recent years, deep learning has been applied in many areas including biomedical field and very effective results have been achieved. Deep learning based methods are used successfully in automatic object detection. In this study, a deep learning based object detection method namely You Only Look Once (YOLO), is used for automatic detection of tympanic membrane in eardrum images obtained using otoscope device. To enable automatic detection of tympanic membrane by YOLO, experimental studies were conducted with AlexNet, VGGNet, GoogLeNet and ResNet. According to the performance results, the most efficient results were obtained with ResNet and VGGNet architectures. Tympanic membrane region detection with YOLO, was performed with an accuracy rate of 93%.

Keywords: *Biomedical signal processing, YOLO, tympanic membrane, object detection, convolutional neural network*

Öz

Orta kulak iltihabından dolayı kulak zarında çeşitli deformasyonlar meydana gelmektedir. Hastalığın teşhis edilebilmesi için otoskop cihazı ile kulağa bakıldığı zaman zar bölgesine erişilmesi ve detaylı bir şekilde kulak zarının incelenmesi gerekmektedir. Son yıllarda derin öğrenme birçok alanda uygulanmış ve oldukça etkili sonuçlar elde edilmiştir. Derin öğrenmenin biyomedikal alanda da sık bir şekilde kullanıldığı ve oldukça iyi neticelere varıldığı bilinmektedir. Otomatik nesne tanımlamada da derin öğrenme tabanlı yöntemler başarılı bir şekilde kullanılmaktadır. Bu çalışmada otoskop cihazı ile elde edilen orta kulak imgelerinde zar bölgesinin otomatik tespiti için derin öğrenme tabanlı nesne algılama yöntemi olan YOLO kullanılmıştır. YOLO yöntemi ile ilgili alanın otomatik olarak tespit edilmesini sağlamak üzere, nesne önerileri için evrişimsel sinir ağı mimarilerinden olan AlexNet, VGGNet, GoogLeNet, ve ResNet ile deneysel çalışmalar yapılmıştır. Performans sonuçlarına göre ResNet ve VGGNet mimarileri ile en verimli sonuçlar elde edilmiştir. YOLO ile zar bölgesinin tespiti %93 başarı oranı ile tespit edildi.

Anahtar Kelimeler: *Biyomedikal sinyal işleme, YOLO, kulak zarı, nesne algılama, evrişimsel sinir ağları*

1. Introduction

Otitis Media (OM) is known as inflammation and the presence of fluid in the middle ear [1][2]. In order to diagnose OM, it is necessary to examine the eardrum in the middle ear via an otoscope [3]. The presence of inflammation in the tympanic membrane (TM) or the presence of perforation in the TM indicates OM [4]. In clinical practice, the TM image provided by the video otoscope device is reflected on the monitor and can be seen by both the patient and the physician. OM type is determined according to the change in color and shape of the eardrum. So, it is very important to examine the TM in detail in the diagnosis of OM. OM is divided into three main classes: Acute OM (AOM), OM with Effusion (OME) and Chronic OM (COM) [5].

For the diagnosis of OM, telemedicine, statistical, classification, web-based and simulation studies have been utilized. Natalie Thone et al. have presented the various simulations used in the field of otolaryngology and the advantages of these simulations [6]. Simulations are frequently used during the education of medical students, otolaryngologists and pediatrician doctors. Vincent Wu et al. conducted the experimental studies with 54 medical students to compare diagnostic methods with otoscope. In this study, the effect of otoscope simulation (OtoSim) on the effectiveness of web based learning and classroom learning methods was examined [7]. Modupe Oyewumi et al. have provided separate otoscopy training to three groups of 57 people consisting of family, community and pediatric physicians. Before the training with OtoSim device, the participants were pre-tested and the final test results were compared with the pre-test results after 3 months. At the end of the training, it was observed that the physicians' otoscopic diagnoses were improved considerably [8]. On the other hand Huang Huang et. Al. carried out a classification according to the three types of OM. After taking a picture with the celloscope connected to the smartphone, the images were sent to a remote server via FTP protocol, where the classification procedure based on the Deep First Search Algorithm was performed. Furthermore, the image preceding the classification step was passed through a series of preprocessing procedures. Active contour segmentation took the TM, deep mesh diagram and the swelling of the membrane and k-means mean color values were determined [9].

When literature is examined, it is seen that the desired regions on biomedical images can be detected successfully and also very efficient results can be achieved in classification processes according to the type of diseases [10]–[14]. From this point, in the first step of this study, otoscope images were obtained with video otoscopy device in Otolaryngology Clinic of Özel Van Akdamar Hospital. Ethics committee required to collect data was obtained from the relevant authority. In the next step, the TM regions in the obtained middle ear images were labeled by experts. To detect the TM automatically, You Only Look Once (YOLO) method has been used. YOLO method has been used in various images such as pedestrian detection and mass detection in mammogram images and have yielded successful results [15][16]. YOLO used Convolution Neural Network (CNN) architecture for object suggestions [17][18]. Experimental procedures were performed with pretrained model AlexNet [19], VGG-16, VGG-19 [20], GoogLeNet [21] and ResNet-50 [22] which are widely used in CNN architecture and the performance results were examined.

The rest of the study is organized as follows: In the second part, data set, YOLO algorithm used for object detection and pretrained CNN models used for object suggestion process are given. In the third part, the results of the experimental studies are given. In the fourth chapter, a discussion is presented on the findings.

2. Material and Method

In this study, 282 TM images were obtained from volunteer patients admitted to Özel Van Akdamar Hospital and labeled by experts. Images diagnosed with COM and images of the middle ear with no visible membrane area due to deformation in the membrane region and earwax are not included in this study. After the images are labeled, the region of interest (ROI) was selected and Ground-truth values were obtained in MATLAB (2019a) Image Viewer application. In the CNN architecture, the feature extraction layers and other hyper parameters were tested to obtain the best result while finding the membrane region in the middle ear image. The schematic outline of the proposed model is illustrated in Figure 1.

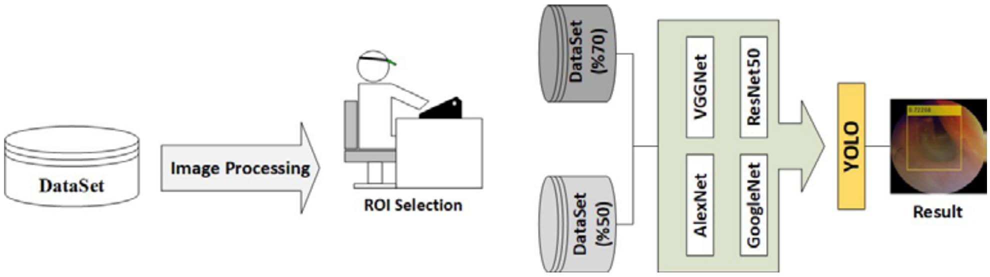


Figure 1. Schematic outline of TM detection with YOLO

2.1. Dataset

For experimental studies, 282 middle ear images were collected during October 2018 - January 2019 in Van Özel Akdamar Hospital Otolaryngology policlinic. These images were obtained from volunteer patients with video otoscope. As mentioned earlier, COM and earwax images were not included in the study. In the data set used in the experimental study, 155 of 224 images belong to Normal TM and 69 AOM classes.

2.2. You Only Look Once (YOLO)

YOLO, an object recognition algorithm based on deep learning, was developed by Joseph Redmon et. al. It uses a single neural network to estimate bounding boxes and class probabilities and re-frame object detection as a single regression problem [23]. Using a single convolutional network, it suggests multiple constraint boxes and class possibilities for these boxes at the same time. YOLO divides the input image into cells with the SxS grid, and the x, y axis and image are evenly divided into the grid. If an object has a central cell, it is responsible for identifying the object. Each grid cell estimates the confidence of the presence of objects. Confidence is shown mathematically in Eq. (1).

$$\text{Conf (Object)} = \text{Pr(Object)}\ xIoU \qquad (1)$$

$Pr\ (Object)$ is the prediction value of the object, and $Pr \in \{0,1\}$ and IoU is the accuracy of the prediction box, and the mathematical model is given in Eq. (2).

$$IoU = \frac{area(box(Truth) \cap box(Pred))}{area(box(Truth) \cup box(Pred))} \qquad (2)$$

Confidence; whether the object exists or the x (Bx) and y (By) coordinates of the midpoint of the object's position exist, and how wide the object width (Bw) and height (Bh) are sure of its coordinates [23][24]. Figure 2 shows the coordinates of TM in the middle ear otoscope image.

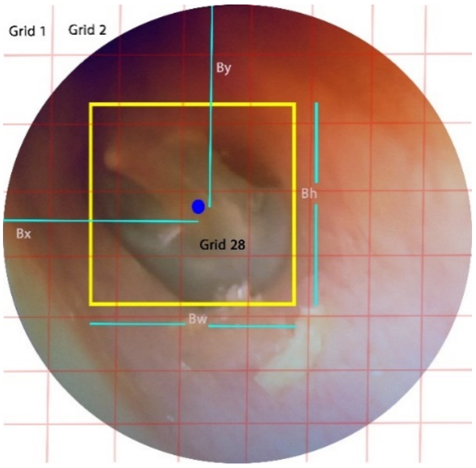


Figure 2. Object location

2.3. Convolutional Neural Network

CNN is a deep learning architecture, which consists of a series of consecutive layers namely convolution, pooling, dropout, and fully connected layers [25]. Convolutional neural networks proposed by By Lecun are one of the most successful methods in pattern recognition [26]. Convolutional layer, the most important layer of CNN, filters the input image and creates a new feature map with f input picture and h filter;

$$f\ x\ h = \sum_{k}\sum_{l} \mathrm{f(k,l)h(i\ -\ k, j\ -\ l)} \qquad (3)$$

The CNN, which consists of several layers, increases the property map with the convolutional layer and therefore increases the cost of computation. For a NxN size I input image, a convolution layer with dxd filter size and stripe value $s\ =\ 1$;

$$O = \frac{(N-d)}{s} + 1 \qquad (4)$$

Eq. (4) determines the size of the output layer. The pooling layer is used after the convolutional layer to reduce the calculation costs. Reducing the size of the output by using average or maximum functions summarizes the subregions. Many convolutional and pooling layers are followed by a fully connected layer. The extracted property vector is processed as input and produces output as many classes [27].

In this study, AlexNet, VGG-16, VGG-19, GoogLeNet and Resnet50 architectures, which are pretrained CNN architectures, were utilized.

2.3.1. AlexNet

AlexNet architecture was proposed by Krizhevsky et al. in the ImageNet competition held in 2012, which consists of 5 convolution layers and 3 fully connected layers [28]. AlexNet, which has an input image size of 227x227x3, has been obtained with 4096 feature vectors and the images consist of 1000 classes.

2.3.2. VGGNets

It is very similar to the AlexNet architecture. In 2014, the CNN architecture proposed by Simonian and Zisserman achieved a remarkable error rate reduction and won in the competition ILSVRC2014 [29]. VGG16 and VGG19 consist of 5 blocks. The first and second blocks consist of 2 convolution layers and 1 pooling layer. Blocks 3 and 4 consist of 3 convolutional layers in VGG-16 model, 4 convolutional layers and 1 pooling layer in VGG-19. 3x3 size stripe filters are used in all convolutional layers [30].

2.3.3. GoogLeNet

GoogLeNet Inception Network is a deep learning model. It consists of 22 layers. 21 of these layers are convolution layers and 1 of them is full connected layer. In GoogleNet, an Inception module has six convolution layers and a pooling layer. The filter of the module consists of 1x1, 3x3 and 5x5 dimensions [31]. Innovation in this design is that the local condensation of sparse matrix operations and parallel processing of properties greatly improve the speed. Inception module deletes the fully connected layers and replaces global average pooling layer. This makes the model faster and reduces the possibility of overfitting [32].

2.3.4. ResNet50

This architecture consists of 49 convolution layers and one fully connected layer and hence it is different from the classic CNN models like AlexNet and VGGNets. The innovation in this model is the depth in its residual blocks and its architecture [33]. The size of input image is 224x224x3. The convolutional layer contains five 1x1 and 3x3 filters. The blocks are followed by the max pooling layer. Other blocks are followed by a few residual layers with a stripe value of 2 [34].

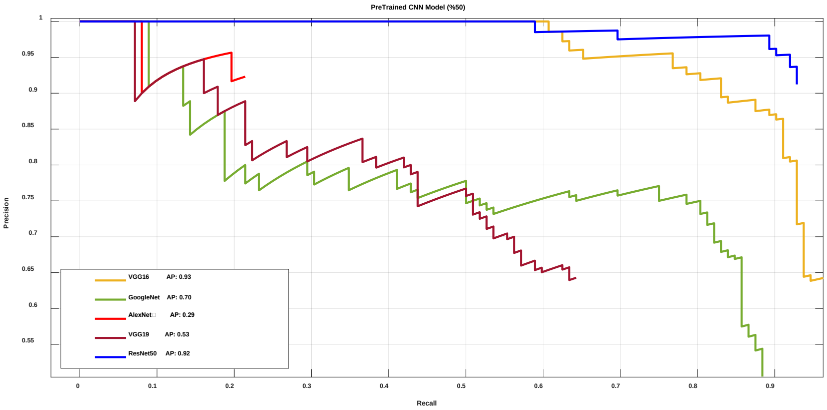
3. Results

Images were resized to 227x277x3 for the CNN architectures AlexNet and 224x224x3 for the VGG-16, VGG-19, GoogLeNet and ResNet50. In CNN architectures, learning rate was determined as 1x(10)^(-3), mini-batch size value 4 and scotastic gradient descent algorithm were used for optimization. Hyperparameters were selected by trial and error method. The data set was first randomly divided into 50% training and 50% test, and then randomly divided into 70% training and 30% test data and the results were analyzed separately. The formulas shown in Table 1 via the confusion matrix are used to measure the performance results. The precision-recall ratios are used to see the relationship between the average accuracy (Accuracy) and the actual positive values (Sensitivity) of the model and the predicted positive values (Precision) to measure the overall accuracy of the model.

Table 1. The performance metrics with short description and their formulas

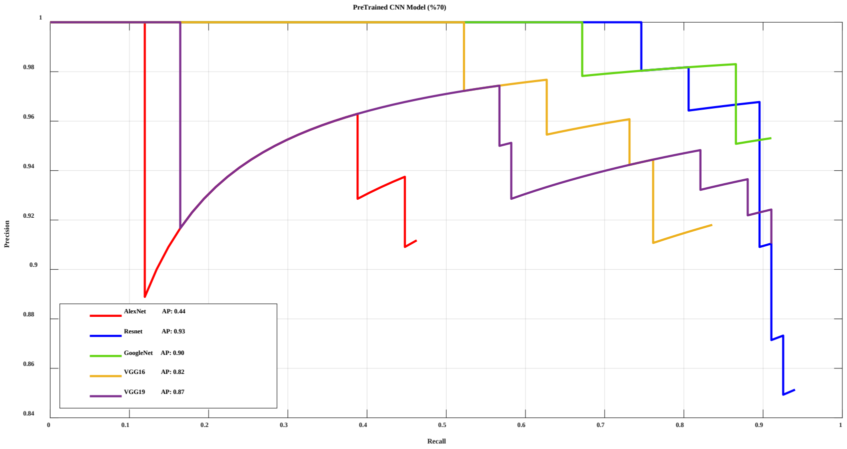
Metrics	Formula	Short Description
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$	Overall effectiveness of the model
Precision	$\frac{TP}{TP + FP}$	The success rate of the positive predicted situation
Recall(Sensitivity)	$\frac{TP}{TP + FN}$	Effectiveness on positive test cases

True Positive (TP) and False Positive (FP) membrane regions are detected correctly and TN and FN are incorrectly detected. Graph 1 shows the Precision-Recall curves obtained by dividing the data set into 50% training and 50% testing. Here, the curves should be as close as possible to the Recall axis which means that the desired region has been successfully detected. Here, the VGG-16 architecture of the CNN models is more successful with 93% accuracy compared to other models. When the data set is divided into 50% training and testing, the highest accuracy rate is the VGG-16 model, whereas the AlexNet, VGG19, GoogLeNet and ResNet50 models have 29%, 53%, 70% and 92% accuracy rates, respectively. The lowest results in the determination of the membrane region with the YOLO method were obtained with the AlexNet model.



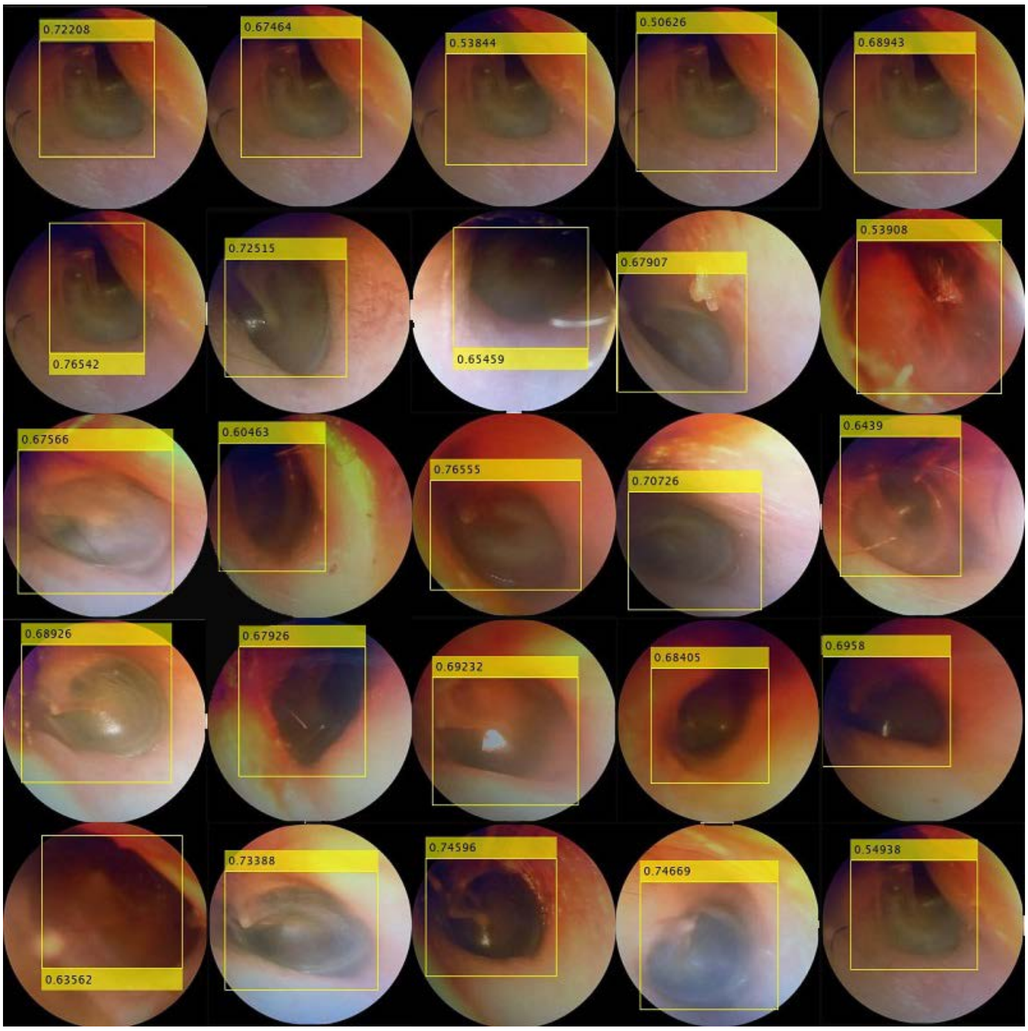
Graph 1. CNN result precision-recall curve (%50)

In order to determine the membrane region, 70% training and 30% test of the dataset were separated and experimental results were examined. Here, the ResNet50 architecture has exhibited an accuracy of 93% compared to other CNN architectures. Precision-Recall curves of CNN architectures are shown in Graph 2. As it can be seen in the graph, Precision and Recall are inversely proportional and Recall increases as the sensitivity of the model decreases. When the data set is divided into 70% training and 30% testing, the highest accuracy rate is obtained with the ResNet50 model, while the AlexNet, VGG16, VGG19 and GoogLeNet models have 44%, 82%, 87% and 90% accuracy rates, respectively. The lowest results in the determination of the membrane zone with the YOLO method were obtained with the AlexNet model.



Graph 2. CNN result precision-recall curve (%70)

Superior results are achieved with VGG16 and ResNet50 architectures when tests are performed by dividing the dataset. In the first experimental study, when the dataset was separated as 50% for training and 50% for test, VGG16 architecture produced an accuracy of 93%, and the same architecture produced an accuracy of 82% when the data set was divided into 70% for training and 30% for test. In the second experimental study by dividing the data set, the ResNet50 architecture resulted the accuracy of 93% and when the dataset was divided into 50% for training and 50% for test, the same architecture resulted the accuracy of 92%. As the number of data used in education increased, the accuracy rate in AlexNet, GoogLeNet, VGG19 and ResNet models increased while it decreased only in VGG 16 model. Picture 1 shows the middle ear images with the membrane region detected.



Picture 1. Detection of TM images

The layers used for feature extraction in CNN architectures for YOLO and the average accuracy rates obtained as a result of these layers are given in Table 2. Table 2 shows the accuracy of 50% of the first value dataset in Accuracy as the test and training, whereas the second value of the dataset is 70% training and 30% test. When the number of data used in education increases, the success rate of the models generally increases, while the VGG16 model decreases. The highest results were obtained in the addition of conv5_3 in the VGG16 model, and in the ResNet50 model in the activation_49_relu layer. The results obtained from other learning layers of CNN models are shown in Table 2.

Table 2. CNN feature extraction layer and accuracy rate

Architecture	Layer	Accuracy (%)	
		50% training & 50% test	70% training & 30% test
AlexNet	conv3	21	44
	conv4	20	39
	conv5	32	37
VGG16	conv5_3	93	82
	conv5_1	90	81
	conv4_2	56	34
VGG19	conv5_4	49	87
	conv5_1	53	86
	conv4_1	52	44
GoogLeNet	inception_5a-output	70	90
	inception_4e-output	57	82
	inception_4d-output	64	89
ResNet50	activation_49_relu	91	93
	activation_31_relu	92	91
	activation_13_relu	63	62

4. Discussion and Conclusion

When investigating in the literature, the conventional methods used for object detection are found to be IFT, LBP, HOG and histogram based methods such as k-Means. Methods such as R-CNN, Fast R-CNN, Faster R-CNN, SSD, YOLO are used for object detection with computer vision along with deep learning. When these studies are examined, though the traditional and deep learning based object-region detection methods are effectively used in studies such as fruit detection and diseased region in pedestrian detection in biomedical images, the deep learning based methods are found to be superior[35][36][37]. In this study, a deep learning based object detection algorithm called YOLO is used which is proven as an efficient method in literature. To determine the membrane region with YOLO, CNN architectures are prepared by trying various layers. The data set was divided in two ways as; 50% for training and 50% for testing and 70% for training and 30% for testing, and then the performance results were examined with layers in CNN architectures. First, the AlexNet architecture conv5 and conv3 layers achieved 32% and 44% accuracy and the Best results were yielded by VGG16. The conv5_3 layer achieved an accuracy of 93% and 82%, respectively. Then VGGNets was built with the other architecture VGG-19. The best results in this architecture were achieved with conv5_1 and conv5_4 layers with an accuracy of 53% and 87% respectively. In the GoogLeNet architecture, 70% and 90% accuracy rates were achieved in the inception_5a-output layer. Finally, it was run with the ResNet50 architecture, which achieved 92% and 93% accuracy with the activation_49_relu layer. In the second experimental study, when the number of data used in education was increased, it was seen that the success rates increased in general, but decreased in the VGG16 model and some layers of ResNet50 models. Learning ability of CNN networks is in line with the increase in the number of data. This situation supports the literature studies [38]. When looking at the data set used in the experimental study, 155 normal TMs and 69 abnormal TMs were used. The images in these classes have different colors and brightness and darkness because of the otoscope. While testing the accuracy of the proposed model, the data in the data set are selected randomly. It affects model success when the data set is unevenly distributed [39].

For the diagnosis of OM, it is necessary to diagnose and treat the patient's ear with an otoscope and after a good examination of the membrane area in the middle ear. 224 images were used in experimental studies for the detection of membrane region in middle ear images. In this context, deep learning based method was used for the detection of membrane region in these images obtained with otoscope device. In this study, AlexNet, VGG-16, VGG-19, GoogLeNet and ResNet50 architectures which are widely used in literature are used. As a result of experimental studies, the best accuracy rate of 93% was achieved with VGG16 conv5_3 layer for the data set containing 50% training data and 50% testing data. When the dataset was divided into 70% for training and 30% for testing, the membrane region was detected with ResNet50 activation_49_relu layer with an accuracy rate of 93%.

In the future studies, it is planned to use the data set containing normal and AOM TM middle ear images where TM are seen and the membrane region is present. Also, it can be attempted to detect the polyzer triangle in other regions of the membrane, such as the handle of malleus.

Acknowledgment

The data set that we used in this study is presented to researchers at http://www.ctganalysis.com/Category/otitis-media link, which we have collected in Van Akdamar Hospital and tagged by specialist doctors with the decision of Bitlis Eren University Ethics Committee.

References

[1] D. K. Marcia Murphy, "A review of techniques for the investigation of otitis externa and otitis media," Clin. Tech. Small Anim. Pract., vol. Volume 16, no. Issue 4, p. Pages 236-241.

[2] T. A. Valdez et al., "Multi-color reflectance imaging of middle ear pathology in vivo," Anal. Bioanal. Chem., vol. 407, no. 12, pp. 3277–3283, 2015.

[3] H. S. a Thorbjörn Lundberg, Leigh Biagio, Claude Laurent and D. W. Swanepoel, "Remote evaluation of video-otoscopy recordings in an unselected pediatric population with an otitis media scale," Int. J. Pediatr. Otorhinolaryngol., vol. 78, pp. 1489–1495, 2014.

[4] T.-I. J. Yong Bin Ji, Hyeon Sang Barg, Dong Woo Park, Sam Kyu Noh, Seung Jae Oh, "Diagnosis Otitis Media Using teahertz Otoscope."

[5] M. Koçyiğit, S. G. Örtekin, and T. Çakabay, "Otitis Media , Sınıflandırma ve Tedaviye Yaklaşım Prensipleri Otitis Media , Classification and Principles of Treatment Approach," vol. 8, no. 2, pp. 65–70, 2016.

[6] N. Thone, M. Winter, R. J. García-Matte, and C. González, "Simulation in Otolaryngology: A Teaching and Training Tool," Acta Otorrinolaringol. (English Ed., vol. 68, no. 2, pp. 115–120, 2017.

[7] V. Wu and J. A. Beyea, "Evaluation of a Web-Based Module and an Otoscopy Simulator in Teaching Ear Disease," Otolaryngol. - Head Neck Surg. (United States), vol. 156, no. 2, pp. 272–277, 2017.

[8] M. Oyewumi et al., "Objective Evaluation of Otoscopy Skills among Family and Community Medicine, Pediatric, and Otolaryngology Residents," J. Surg. Educ., vol. 73, no. 1, pp. 129–135, 2016.

[9] Y. K. Huang and C. P. Huang, "A depth-first search algorithm based otoscope application for real-time otitis media image interpretation," Parallel Distrib. Comput. Appl. Technol. PDCAT Proc., vol. 2017-Decem, pp. 170–175, 2018.

[10] A. Ş. Ümit Budak, Ömer Faruk Alçin, Muzaffer Aslan, "Optic Disc Detection in Retinal Images via Faster Regional Convolutional Neural Networks," pp. 3–5, 2018.

[11] E. Deniz, A. Sengür, Z. Kadiroğuglu, Y. Guo, V. Bajaj, and Ü. Budak, "Transfer learning based histopathologic image classification for breast cancer detection," Heal. Inf. Sci. Syst., vol. 6, no. 1, p. 18, 2018.

[12] [12] Z. Cömert, A. F. Kocamaz, and V. Subha, "Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment," Comput. Biol. Med., vol. 99, pp. 85–97, Aug. 2018.

[13] Z. Zhao, Y. Zhang, Z. Comert, and Y. Deng, "Computer-Aided Diagnosis System of Fetal Hypoxia Incorporating Recurrence Plot With Convolutional Neural Network," Front. Physiol., vol. 10, p. 255, 2019.

[14] Z. Cömert and A. F. Kocamaz, "Fetal Hypoxia Detection Based on Deep Convolutional Neural Network with Transfer Learning Approach," in Software Engineering and Algorithms in Intelligent Systems, 2019, pp. 239–248.

[15] Ö. Algur, V. Tümen, and Ö. Yildirim, "Dış Ortam Görüntülerindeki İnsan Hareketlerinin Hibrit Derin Öğrenme Yöntemleri Kullanarak Sınıflandırılması," vol. 30, no. 3, pp. 121–129, 2018.

[16] M. A. Al-antari, M. A. Al-masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," Int. J. Med. Inform., vol. 117, pp. 44–54, Sep. 2018.

[17] S. Lu, Z. Lu, and Y.-D. Zhang, "Pathological brain detection based on AlexNet and transfer learning," J. Comput. Sci., vol. 30, pp. 41–47, Jan. 2019.

[18] H. Greenspan, M. Frid-Adar, E. Klang, I. Diamant, M. Amitai, and J. Goldberger, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," Neurocomputing, vol. 321. pp. 321–331, 2018.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, 2012, pp. 1097–1105.

[20] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014.

[21] C. Szegedy et al., "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[23] Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection."

[24] C. L. ; Y. T. ; J. L. ; K. L. ; Y. Chen, "Object Detection Based on YOLO Network," in 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), 2018.

[25] Y. Altuntaş, Z. Cömert, and A. F. Kocamaz, "Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach," Comput. Electron. Agric., vol. 163, p. 104874, 2019.

[26] M. Sarıgül, B. M. Ozyildirim, and M. Avci, "Differential convolutional neural network," Neural Networks, vol. 116, pp. 279–287, 2019.

[27] G. Zhao, G. Liu, L. Fang, B. Tu, and P. Ghamisi, "Multiple convolutional layers fusion framework for hyperspectral image classification," Neurocomputing, vol. 339, pp. 149–160, Apr. 2019.

[28] Y. Fu and C. Aldrich, "Flotation froth image recognition with convolutional neural networks," Miner. Eng., vol. 132, pp. 183–190, Mar. 2019.

[29] S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification," Comput. Electr. Eng., vol. 72, pp. 274–282, Nov. 2018.

[30] Y. Seo and K. Shin, "Hierarchical convolutional neural networks for fashion image classification," Expert Syst. Appl., vol. 116, pp. 328–339, 2019.

[31] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," Comput. Biol. Med., vol. 111, p. 103345, Aug. 2019.

[32] S. Lu, B. Wang, H. Wang, L. Chen, M. Linjian, and X. Zhang, "A real-time object detection algorithm for video," Comput. Electr. Eng., vol. 77, pp. 398–408, Jul. 2019.

[33] Y. Altuntaş, Z. Cömert, and A. F. Kocamaz, "Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach," Comput. Electron. Agric., vol. 163, p. 104874, Aug. 2019.

[34] A. Soudani and W. Barhoumi, "An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction," Expert Syst. Appl., vol. 118, pp. 400–410, Mar. 2019.

[35] H. Y. Hulin Kuang, Cairong Liu, Leanne Lai Hang Chan, "Multi-class fruit detection based on image region selection and improved object proposals," Neurocomputing, vol. 283, no. 241, p. 255, 2018.

[36] R. Y. Yoshimasa Kawazoe, Kiminori Shimamoto and H. U. M. F. and K. O. Yukako Shintani-Domoto, "Faster R-CNN-Based Glomerular Detection in Multistained Human Whole Slide Images," J. Imaging, 2018.

[37] N. B. Atakan Körez, "İnsansız Hava Aracı (İHA) Görüntülerindeki Yayaların Faster R-CNN Algoritması ile Otomatik Tespiti," in 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2018.

[38] Ü. Budak, Z. Cömert, M. Çıbuk, and A. Şengür, "DCCMED-Net: Densely connected and concatenated multi Encoder-Decoder CNNs for retinal vessel extraction from fundus images," Med. Hypotheses, vol. 134, p. 109426, 2020.

[39] S. Li, W. Song, H. Qin, and A. Hao, "Deep variance network: An iterative, improved CNN framework for unbalanced training datasets," Pattern Recognit., vol. 81, pp. 294–308, Sep. 2018.