Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Deep Learning Based Tympanic Membrane Segmentation Using Residual Double Attention UNet

MarySelvi.S1*,Dr.Subha.V2 , Dr.Manivanna Boopathi.A3 , Thanu.S4
1,2,4Department of Computer Science and Engineering, ManonmaniamSundaranar University,Abishekapatti,Tirunelveli,627012,TamilNadu, India.
Email: marymarymsu2023@gmail.com1; subha_velappan@msuniv.ac.in2; tharshinimanian@gmail.com4
3Department of Electrical and Electronics Engineering, Sethu Institute of Technology, Kariapatti, TamilNadu, India. Email: manivannaboopathi@sethu.ac.in3
**Corresponding author: S.MarySelvi, Research Scholar, Department of Computer Science and Engineering, ManonmaniamSundaranarUniversity,Abishekapatti,Tirunelveli,627012,TamilNadu,India.*
Email: marymarymsu2023@gmail.com

ARTICLE INFO

Received: 31 Dec 2024

Revised: 20 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

The airy spaces of the middle ear and temporal bone are home to the Eustachian tube, which is covered by a mucous membrane that is infected and inflamed in Otitis Media (OM). Another of the most prevalent diseases is OM. Otoscope pictures are visually inspected in clinical settings to make the diagnosis of OM. Being subjective and prone to mistakes, this procedure is susceptible. In this study a unique framework Hybrid Colour Residual Double Attention UNet (HCRDAUNet) model is proposed to effectively segment the Tympanic membrane. This model utilizes the strength of three different colour spaces namely RGB, LUV and HSV into a single joined semantic segmentation model with attention mechanism. The proposed attention gate in this approach applies the gating outcome on two different scales of feature map to accurately localize the eardrum. The proposed HCRDAUNet model archives up to 96% of dice co-efficient and 95% of F1-score, which shows that the proposed model attains significant improvements in performance, compared to state art of semantic segmentation models.

Keywords: Deep Learning, Tympanic Membrane segmentation, Otitis Media, Residual Attention U-Net, Attention gate.

INTRODUCTION

Otitis media (OM) is one of the most common diseases in the globe [1,2]. Otoscopic screening, on the other hand, is extremely subspecialized; posing diagnostic challenges for primary care practitioners whose otologic diagnoses are relatively erroneous.Any middle ear inflammation is referred to as otitis media (OM), which can be clinically divided into acute (AOM), chronic (COM), and otitis media with effusion (OME) forms.

Untreated OM may result in a life-threatening intracranial condition or permanent hearing loss. Due to its great incidence, OM is becoming a major global public health issue [3,4].The high risk of recurrence and parental loss of working hours associated with treating OM might result in large healthcare costs [5]. So, in recent years, the healthcare sector has become interested in the telemedicine and home care model for OM.

In actuality, abnormalities in the tympanic membrane (TM) are linked to OM disorders [6,7]. Identification of morphological or colour alterations on the TM is necessary for the diagnosis of OM. The clinical symptoms of OM might include ear discomfort, ear discharge, headache, current or recent upper respiratory tract infection, restlessness, and appetite loss [8].The anomaly of TM causes a variety of effects on patients, including hearing loss and serious infection, if it is not promptly identified and treated. As a result, it is critical to spot the TM anomaly for an early diagnosis of OM, especially in youngsters [9]. Physicians must complete years of training before they can diagnose OM with otoscopic TM results. The state of the ear canal, the diminutive size of the TM, and different anatomical abnormalities can occasionally make a diagnosis difficult. The accuracy of diagnosis may vary across doctors with various educational backgrounds, which further raises questions in regions where access to healthcare is difficult.The regions of the TMs identified from captured picture frames via the TM segmentation approach are significant for subsequent steps in the paediatric otitis media diagnosis process. The segmented region of the TM allowed for the derivation of important diagnostic criteria for OM, includes TM size, texture, geometry, and colour distribution [10].

The detection of OM as well as the precision of OM illness categorization can both be enhanced by effective TM segmentation in otoscopic images. However, because of the unique characteristics of each patient and the diversity of image acquisition procedures, automated and reliable TM segmentation is a difficult task.Additionally, TM pictures are often created from video-otoscopic images [11] that have uneven lighting, making certain image areas brighter or darker than the typical colour of a particular structure [12]. These traits, together with the low contrast of anatomical structure boundaries, make TM automated segmentation tasks challenging, especially when the tympanic membrane is perforated or effused.Additionally, TM pictures typically have intensity inhomogeneity because of influences during the acquisition procedure. Additionally, there are concealed regions present on the TM pictures obtained from video-otoscopic images, which makes TM segmentation even more difficult.

The most effective AI technique for a variety of tasks, including issues with medical imaging, is Deep learning. Recently, deep neural networks have been successfully used for otologic diagnostics. Additionally, several studies using the tympanic membrane (TM) have demonstrated the value of deep learning models for the early diagnosis and treatment of ear disorders [13]. Nevertheless, despite the fact that those models demonstrated highly accurate diagnoses based on the TM, such methodologies have limits when it comes to being effectively used in actual practise. The inherent "black-box *nature"* of deep learning algorithms is to blame for these restrictions.Besides, AI system should be understandable, in order to complement medical professionals and allow them to diagnose and recall their decisions [14].

This study focuses on deep neural network based approach for Tympanic membrane from Otoscope images. This approach combines the properties of three different colour spaces to accurately localize the membrane. The linked Unet with three colour space further improves the proposed double attention model with residual concept.

RELATED WORKS

If the input pictures are noisy, inhomogeneous, or have weak borders, the results might not be sufficient. Thus, Tympanic membrane segmentation is significant to identify the Otitis media. In the literature, a variety of techniques for segmenting time panic membranes have been explored.A large number of models has been developed using different machine learning as well as deep learning techniques. Deep learning approaches performed well in Tympanic membrane segmentation among a large range of techniques. The succeeding section focuses on the previous methods in different techniques.

Tympanic membrane segmentation was hypothesised as a semi-automatic process by Ibekwe et al. [15] and Hsu et al. [16]. When completing segmentation tasks, one must first manually choose a set of points around the subject areas using a computer mouse. However, it might be challenging to create accurate TM limits since it requires clumsy computer mouse manipulation about the appropriate regions. Accordingly, it could result in mistakes. When using a semi-automatic tympanum approach, Comunello et al. [17] construct TM borders by manually adjusting the minor and major axes of an initial ellipse that was set by the user.

To segment TM pictures, Xie et al. [18] use a snake-a parametric active contour model (ACM). Nevertheless, results are inadequate if pictures have weak boundaries since the snakes utilise gradient information, such as image borders, to guide the curves. According to Tran et al.'s research [19], the categorization of acute otitis media and otitis media with effusion uses a segmentation technique based on a level set-based active contour model. The modified double active contour segmentation technique has been presented by Shie et al. [20] to evolve the active contour and eventually terminates on the required boundary condition by minimising an energy function.

Computer-aided systems based on a binary classification technique have been developed for several investigations [21, 22]. They trained several learning models including decision trees, SVM, neural networks, and Bayesian decision approaches, using the colour information of the eardrum picture in order to determine whether the image represents a case of otitis media or a normal ear. Huang et al. [23] used a plug-in otoscope to provide a visual image from the inside and create a system employing the Depth-First Search Algorithm to diagnose otitis media at home.In order to distinguish between several classes, such as the external ear canal's obstructed wax, normal TMs, OME, AOM, and CSOM OM types, Myburgh et al. [24] suggested a decision tree (DT)-based otitis media diagnostic model.Patch-based image classification is accomplished using segmentation methods based on CNNs. In this method, the input picture is segmented into patches, and the CNN model is applied to each patch to get the patch's class label. Due to the heavily overlapping patches of the image, this process requires a lot of processing and is ineffective. Therefore, cutting-edge methods have been created to improve the abilities of deep CNNs for difficult segmentation tasks. Unet was introduced by Ronneberger et al. [25] and is used to segment neural structures in electron microscopic stacks. Fully convolutional networks (FCNs) for semantic segmentation were introduced by Long et al. [26]. Three clinically significant structures were the focus of a segmentation model built by Seok et al. [27] utilising R-CNN and ResNet-50 as the backbone.

For segmenting tympanic membranes (TMs), Pharm et al. [28] added a hybrid loss function to the fully convolutional network that combines the Dice loss and active contour loss. To find anomalies in otoscopic ear pictures, Park et al. [29] used a mask R-CNN method to divide a typical TM into five substructures: the annulus, umbo, malleus, pars flaccida, and cone of light. The effectiveness and dependability of CNNs in identifying the side and perforation of TMs in medical pictures are demonstrated by Lee et al. [30]. A unique application of DenseNet was proposed by Khan et al. [31] for the automated identification of middle ear (ME) and tympanic membrane (TM) infections.

A Deep CNN-based AlexNet model was suggested by Basaran et al. [32] to distinguish between samples with normal TM and chronic otitis media (COM). Basaran et al.[33] utilized an artificial neural network (ANN) and a gray-level co-occurrence matrix (GLCM) to discriminate between normal and acute TMs. A unique SelectStitch model for semantic segmentation technique was put up by Binol et al. [34] to identify the eardrum in each frame of the otoscope footage. The automated detection of eardrum areas in each frame of an otoscope video was trained using a setup of a 4-level depth U-Net architecture. Myburg et al. [35] used decision tree to diagnose otitis media. Viscaino et al. [36] utilizedLaplacian Kernel to segment the regions of an image.The Laplacian operator draws attention to areas of a picture with sudden fluctuations in intensity.

A novel model backed by CBAM and residual blocks was presented by Alhudhaif et al. [37], and the hyper column approach was incorporated for rapid and precise diagnosis. For the purpose of identifying acute otitis media, Sundgaard et al. [38] developed a deep metric learning strategy with five distinct loss functions. The Inception V3 network is the network architecture used in this work. A six-category system of ear illnesses was established by Cha et al. [39] using an ensemble model that included ResNet-101 and Inception V3 to categorise eardrum and external auditory canal data. By combining Faster R-CNN and pretrained CNNs with a Transfer learning technique, Senaras et al. [40] presented an anomaly of the eardrum.

3. PROPOSED WORK- HYBRID COLOUR RESIDUAL DOUBLE ATTENTION UNET

3.1 U-Net

Figure 1 depicts the network architecture [25]. It is made up of a path that shrinks (on the left) and a growing path (on the right). The path of contraction follows the typical topology of a convolutional network. It entails applying a pair of 3x3 convolutions (unpadded convolutions) repeatedly, every time accompanied by a rectified linear unit (ReLU), max pooling operation having filter size2x2, and a stride 2 downsampling operations. We increase the total quantity of feature channels by two at every level of downsampling. The feature map must be upsampled before each step in the expansive path, which also includes a 2x2 convolution (also known as a "up-convolution") that reduces the total quantity of feature channels in one half, concatenation with the appropriately cropped feature map from the path of contraction, and two 3x3 convolutions, every one accompanied by a ReLU. Cropping is necessary because to eliminate the loss of border pixels in each convolution. As the last layer, a 1x1 convolution is employed to divide each 64-component feature vector into the desired number of classes the class will be two either targeted membrane or not. The network as a whole has 23 convolution layers.

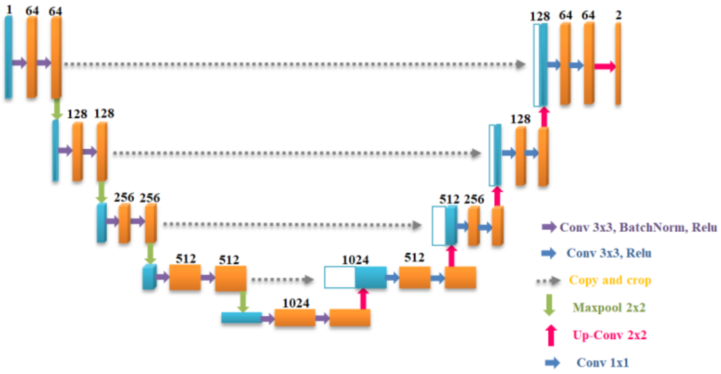


Figure 1. U-Net architecture (example for 32x32 pixels in the lowest resolution)

Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps.

3.2 Attention U-Net

In order to highlight significant features that are sent via the skip connections, the proposed Attention Gates(AG) are shown in Figure 2 incorporated into the standard U-Net design. Data collected on a coarse scale is employed in the gating process to differentiate between unimportant and noisy answers in skip connections. To combine just pertinent activations, this is done just before the concatenation procedure. Furthermore, AGs filter the activations of neurons throughout the forward pass and the backward pass. Gradients that originate from the backdrop are given a lower weight throughout the backward pass. This enables the updating of model parameters in deeper layers depending largely on spatial regions relevant to a specific job.

To determine the result of the skip connection, complimentary information from each sub-AG is retrieved and merged. By performing linear transformations lacking any spatial support (1×1×1 convolutions) and downsampling input feature-maps to the resolution of the gating signal, comparable to non-local blocks, the quantity of trainable parameters and computational cost of AGs are decreased [42]. The associated linear transformations separate the feature-maps for the gating operation and map them to a lower-dimensional space. Low-level feature-maps, or the initial skip connections, as recommended in [41], are not utilised in the gating function due to the fact that they fail to depict the input information in a high dimensional space. We enforce semantic discrimination at each image scale in the intermediate feature-maps using deep-supervision [43]. As a result, attention units are better able to control how people react to a wide variety of visual foreground content at various scales. Thereby, we avoid reconstructing dense predictions from tiny subsets of skip connections.

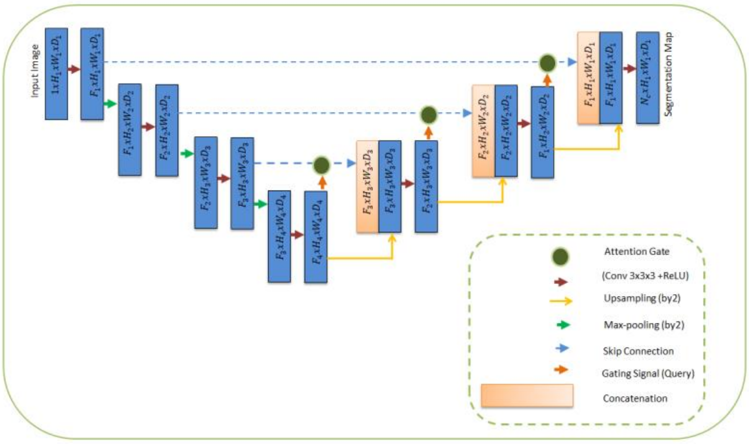


Figure 2. Architecture of Attention U-Net

3.2.1Attention Gates

The attention mechanism, a key technology, has been heavily employed in many domains, including speech recognition, picture detection, statistical learning, and natural language processing (NLP). The idea of Non-local [42] was the initial attempt to apply the attention mechanism for computer vision. We employ the Attention U-Net method, which modifies the network design by including an Attention Gate, to concentrate on the relevant partslinked to the segmentation task. Figure 3 shows the structural representation of Attention Gate, which receives its inputs from the expansion path's upsampling characteristics and the encoder's matching features.Using the former as a gating signal, task-related regions are inhibited and target area learning that is relevant to segmentation is improved [44]. The two inputs are first subjected to the Convolution and BatchNorm processes before being added to produce the output which is further applied by Relu activation layer. Then, this outcome further applied by Convolution (1 x 1), and BatchNorm operation is utilised to obtain second level of outcome. Then it is passed through the second activation function Sigmoid and Resample in order to obtain the attention coefficient (**α**), the encoder feature is finally raised pixel-by-pixel by the attention coefficient.

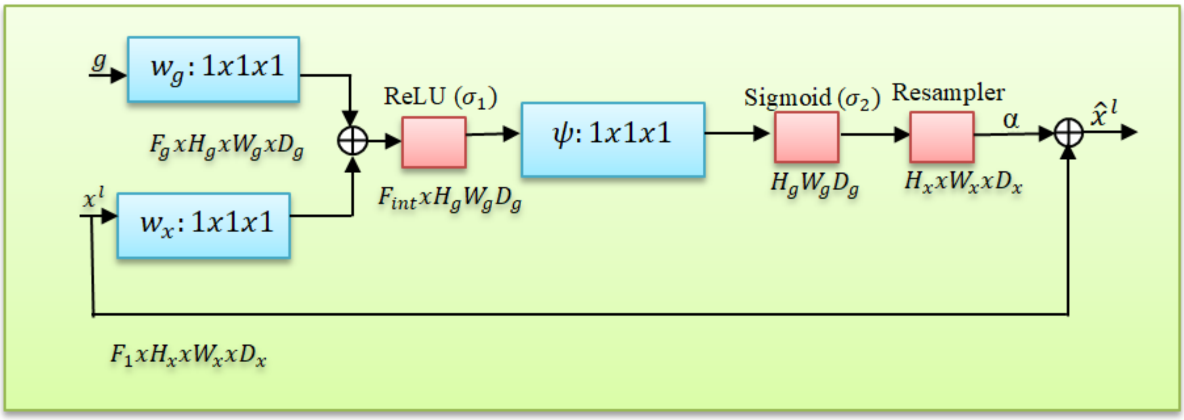


Figure 3. Block Diagram of Attention Gate

3.3 Hybrid Colour Residual Double Attention UNet (HCRDAUNet) WorkFlow

This model takes three colour spaces as input, namely RGB, LUV and HSV. The HSV colour space is more intuitive to how people experience colour than the RGB colour space. As hue (H) varies from 0 to 1.0, the corresponding colours vary from red, through yellow, green, cyan, blue, and magenta, back to red. As saturation(S) varies from 0 to 1.0, the corresponding colours (hues) vary from unsaturated (shades of gray) to fully saturated (no white component).In LUV colour space, L gives luminance and U and V give chromaticity values of colour image. Negative value of U indicates the prominence of red component in colour image and negative value of V indicates the prominence of green component over blue.

The architecture of the proposed Hybrid Colour Residual Double Attention UNet (HCRDAUNet) is depicted in Figure 4.

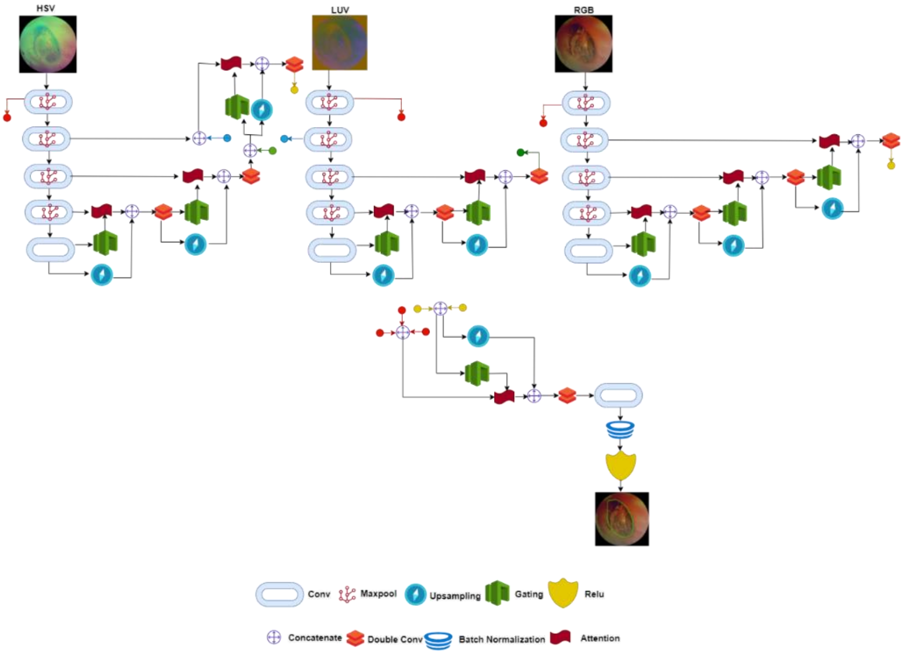


Figure 4. Architecture of the proposed work (HCRDAUNet)

In the proposed model, similar to traditional Unet, the first four level of double convolution with corresponding maximum pool followed by double convolution without maxpooling are applied for the three colour space simultaneously. Let us consider the output produced by the first five levels of HSV colour mode are represented as d_convmp_hsv1, d_conv_mp_hsv2, d_convmp_hsv3, d_convmp_hsv4 and d_conv_hsv5.This process is carried out for other two images of LUV and RGB.

For LUV and HSV colour space model, the last end two level of attention is carried out as similar in traditional ResAttentionUNet model, but the attention is designed by the proposed double attention model which is explained in the section 3.3.2. Whereas for RGB model, all the process are carried out similar to ResAttentionUnet with proposed double attention model. The process of gating and its corresponding doubleattention block process for HSV and LUV for last two levels are presented in the following equation 1 to 4.

att1_hsv = attention(gating(d_conv_hsv5), d_conv_mp_hsv4) (1)

dconv1_hsv = d_conv(concatenate(up(d_conv_hsv5), att1_hsv)) (2)

att2_hsv = attention(gating(dconv1_hsv), d_conv_mp_hsv3) (3)

dconv2_hsv = d_conv(concatenate(up(dconv1_hsv), att2_hsv)) (4)

In the Hybrid Colour Residual Double Attention UNet model the feature maps from HSV and LUV areintegrated as input for the final attention block. Here both the outputs from the encoder section are concatenated and given as input feature map for the Double attention block, whereas both the output from the decoder block after the second level of attention mechanism are given as input for the gating of the third level double attention. It is clearly shown in the Figure 1 and represented by the following equation 5 to 8.

concat_hsv1 = concatenate (dconv2_hsv, dconv2_luv) (5)

concat_hsv2 = concatenate (d_conv_mp_hsv2, d_conv_mp_luv2) (6)

att3_hsv = attention(gating(concat_hsv1), concat_hsv2) (7)

dconv3_hsv = d_conv(concatenate(up(concat_hsv1), att3_hsv)) (8)

Now the HSV and LUV linked features with effective attention co-efficient are further merged with the RGB level of feature map with the help of final attention module which is separatelyshown below the full block diagramofthe

Figure 4. Here, the three red dots represent the first level of the doubleconvolution with max pooling for the three colour spaces which are concatenated to represent input feature map for the double attention block. Similarly, the two yellow dots represent the LUV and HSV linked third level of attention with double convolution outcome and the same level of outcome from the individual RGB mode. Finally, the linked tri colour mode of the attention block is further applied by double convolution followed by the probability prediction of Tympanic Membrane using convolution layer of the filter size as one (Binary Loss) with corresponding batch normalization and Relu activation This is described by the following equations from 9 to 13.

concat_1 = concatenate (dconv3_rgb, dconv3_hsv) (9)

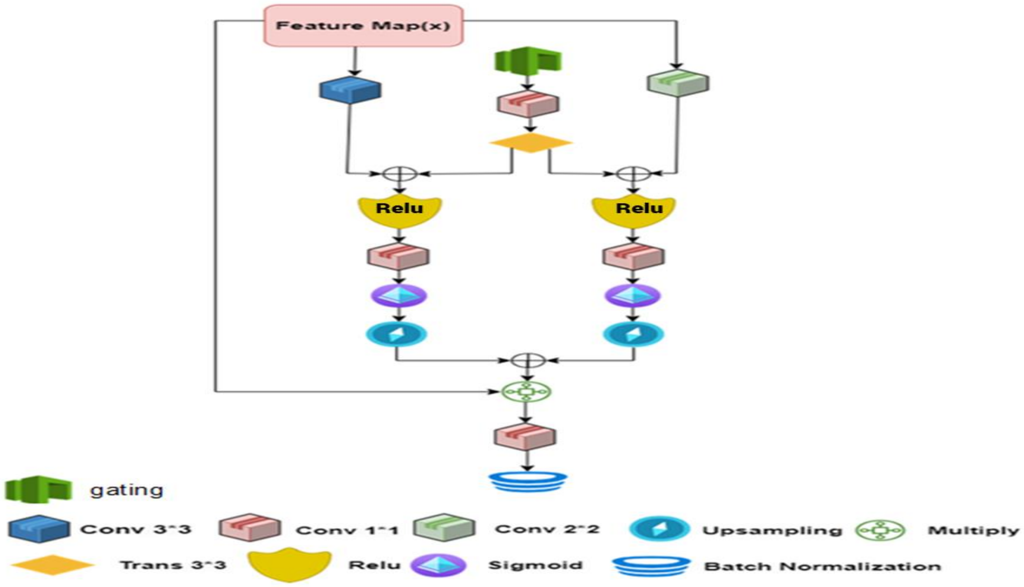
concat_2 = concatenate (d_conv_mp_hsv1, d_conv_mp_luv1, d_conv_mp_rgb1) (10)

att_fin = attention(gating(concat_1), concat_2) (11)

dconv_fin = d_conv(concatenate(up(concat_1), att_fin)) (12)

conv_fin = relu(BN(conv(dconv_fin))) (13)

3.3.1 Proposed Double Attention Gate



`

Figure 5. Structural Overview of the Double Attention Gate

The architecture of Double Attention Gate is illustrated in Figure 5. In order to improve the localization performance compared to the Attention Gate in UNet, an attention mechanism is carried out by double times. The single gating is coupled with two mode of convolution with different size as 3 x 3 and 2 x 2. The process of extracting the attention co-efficient from the two end of process is further added together to get final attention co-efficient.

The process is illustrated as follows, the first concatenation is carried out from the outcome of a convolution layer with a 3×3 filter size applied to the feature map (x), and the result of a transpose layer with a 3×3 filter size that occurs after gating and convolution of filter size 1×1, let's say TransOp1. Contrarily, the TransOp1 combined simultaneously with a 2×2 convolution that is applied on the feature map (x). Both the concatenated layers are supplied separately to Relu, Convolution layer with a filter size of 1×1, Sigmoid, and Upsampling layers, let's say $ConcOp_1$ and $ConcOp_2$. The results of both are then combined once more, multiplied by the feature map (x), placed into a convolution layer with a filter size 1×1, and then subjected to batch normalization. The process of Double Attention Gate is presented in the following equations.

concat1 =concatenate(TransOp1,Conv$_{3\times3}$) (14)

concat2 = concatenate(TransOp1, Conv$_{2\times2}$) (15)

alpha1 = relu(conv(sigmoid(upsample(concat1)))) (16)

alpha2 = relu(conv(sigmoid(upsample(concat2)))) (17)

alpha= alpha1+alpha2 (18)

multiply = featureMap(x) * alpha= (19)

result = BathNorm(Conv$_{1\times1}$(multiply)) (20)

4. RESULTS AND DISCUSSION

4.1 Dataset Description

A database of 1012 OM otoscopic images from children aged 6 months to 12 years old, taken by otologists using a digital otoscope (Karl Storz, Tullingen, Germany), was retrospectively examined using the Institutional Review Board clearance from Cathay General Hospital (No. CGH-P103040). 505 of them have been deemed normal, while 507 have been deemed to have paediatric OM, comprising AOM (100), OME (111), and COM (296). The existence of purulence or effusion in the tympanic cavity, which indicates the suppurative stage or sub-acute stage of AOM or OME, and hyperemic change, bulging, or perforation of the tympanic membrane (TM)[45], that indicate early stage, suppurative stage, and spontaneous performance of AOM, respectively, are additional characteristics of paediatric OM(s).The images were entirely clear and unobstructed by cerumen, allowing for the visualisation of TM.

4.2 Evaluation Metrices

We contrast the findings of the proposed approach with the ground facts (professional manual annotations) in order to evaluate its performance. We employed the Dice similarity coefficient (DC) and Jaccard coefficient (Jac) to assess the quantitative accuracy of segmentation results, as is typical of FCN methods. The following formula is used to determine the Dice coefficient, which assesses how comparable automatic and manual segmentations are:

$$DC = \frac{2S_{am}}{S_a + S_m}$$

The resemblance among two sets is also determined using the Jaccard coefficient, which is defined as:

$$Jac = \frac{S_{am}}{S_a + S_m - S_{am}}$$

Where, the regions that are dynamically delimited, manually segmented, and their intersection are designated as Sa, Sm, and Sam, correspondingly.

Furthermore, we assessed the efficacy of tympanic membrane segmentation algorithms using additional metrics like accuracy, sensitivity, and specificity. The accuracy (Acc) is the percentage of accurate results used to gauge how reliable a diagnostic test is,

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Where, the initials TP, TN, FP, FN stand for the corresponding totals of true positives, true negatives, false positives, and false negatives.

The algorithm's ability to accurately forecast the eardrum regions is indicated by the sensitivity (Sen). The genuine positive rate is specifically indicated by it, and it appears in the following manner:

$$Sensitivity = \frac{TP}{TP + FN}$$

The algorithm's ability to accurately anticipate non-eardrum regions is demonstrated by the specificity (Spe), which is written as:

$$Specificity = \frac{TN}{TN + FP}$$

We employ the Hausdorff distance (HD) and Mean Absolute Distance (MAD) to quantify the inaccuracy among the outlines produced by artificial segmentation and the actual data. The errors associated with the generated boundary, A, and the manually segmented boundary, B, are calculated using the Hausdorff distance, specified [34] as;

$$HD(A,B) = \max\{h(A,B), h(B,A)\}$$

Where $h(A,B) = \underset{a \epsilon A b \epsilon B}{\mathrm{maxmin}}\{dist(a,b)\}$, and dist(a,b) is the Euclidean distance among points a and b.

Table 1.Tympanic membrane segmentation analysis of using single colour model.

	Dice	Accuracy	Precision	Recall	Specificity	HD (Hausdorff distance)	IoU	F1-score
AttentionUnet [50]	0.9010	0.942	-	0.892	0.964	13.293	-	-
AttentionUnet_HSV	0.8748	0.9047	0.8427	0.8624	0.9248	14.476	0.7775	0.8524
AttentionUnet_LUV	0.8621	0.8935	0.8347	0.8561	0.9175	14.871	0.7577	0.8453
ResidualUnet [49]	0.9020	0.932	-	0.896	0.967	13.172	-	-
ResidualUnet_HSV	0.8772	0.9094	0.8449	0.8661	0.9341	14.269	0.7813	0.8554
ResidualUnet_LUV	0.8657	0.8972	0.8416	0.8607	0.9229	14.577	0.7632	0.8510
Att_ResUNet_RGB	0.9137	0.9497	0.8897	0.9076	0.9788	12.454	0.8411	0.8986
Att_ResUNet_HSV	0.8984	0.9293	0.8724	0.8935	0.9583	13.439	0.8155	0.8828
Att_ResUNet_LUV	0.8919	0.9248	0.8643	0.8881	0.9529	13.843	0.8049	0.8760
DAG_ResUNet_RGB	0.9342	0.9698	0.9183	0.9241	0.9871	9.562	0.8765	0.9212
DAG_ResUNet_HSV	0.9182	0.9491	0.8942	0.9093	0.9669	11.886	0.8488	0.9017
DAG_ResUNet_LUV	0.9051	0.9372	0.8917	0.9028	0.9563	11.512	0.8267	0.8972

Table 1 contains the segmentation methods. It demonstrates that the DAG_ResUNet_RGB achieves 0.9342 Dice, 0.9698 Accuracy, 0.9183 precision, 0.9241 Recall, 0.9871 specificity, 9.562 HD, 0.8765 IoU and 0.9212 F1-score which is higher than all the prior methods.

The corresponding graph represented as Dice, HD and F1-score for single colour model

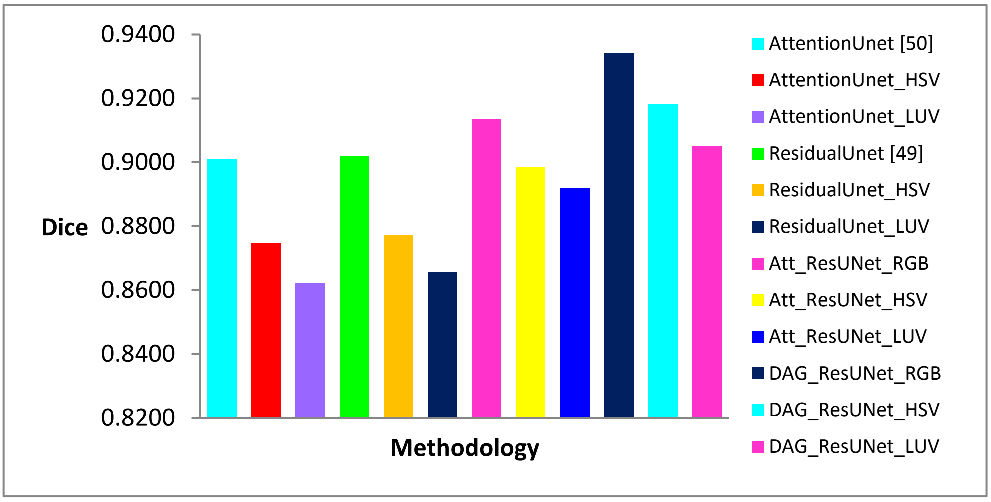


Figure 6. Dice comparison between single colour model

From the Figure 6, it is found that better dice result for single colour model. The DAG_ResUNet_RGBmodelachieves improved result +0.205 thanAtt_ResUNet_RGB, +0.032 than ResidualUnet[49] and +0.033 than AttentionUnet[50].

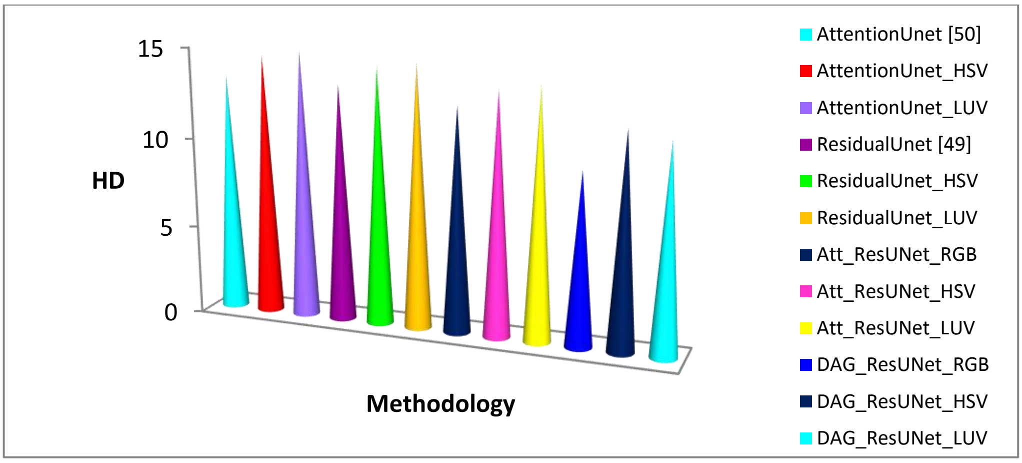


Figure 7. HD comparison of single colour model

From the Figure 7 HD value of the DAG_ResUNet_RGBlower than Att_ResUNet_RGB, ResidualUnet[49] and AttentionUnet[50] by 2.10, 3.61 and 3.73 respectively.

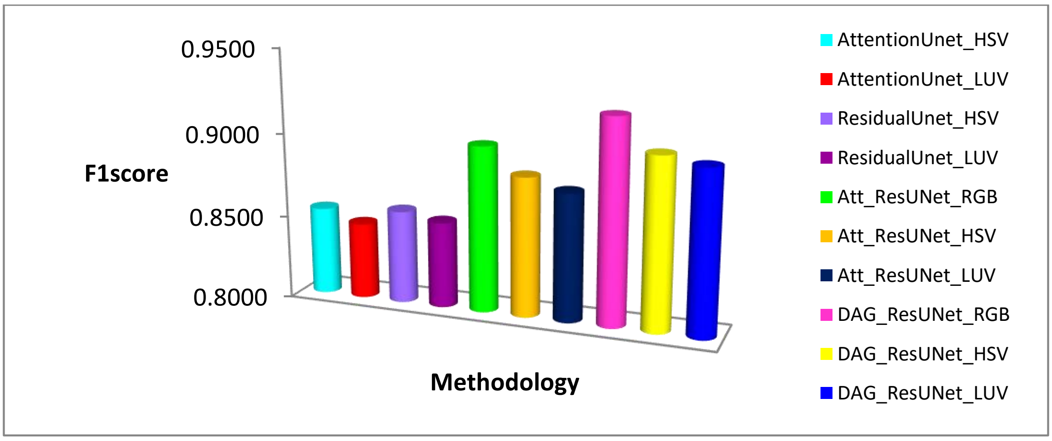


Figure 8. F1-score comparison between single colour model

From the Figure 8 shows the DAG_ResUNet_RGBF1-score value for single colour model. The DAG_ResUNet_RGB better value than +0.022 than Att_ResUNet_RGB.

Table 2. Comparison of Bi and Tri colour models

	Dice	Accuracy	Precision	Recall	Specificity	HD (Hausdorff distance)	IoU	F1-score
Att_ResUNet_RGB_HSV	0.9391	0.9731	0.9178	0.9276	0.9827	9.856	0.8852	0.9227
Att_ResUNet_HSV_LUV	0.9274	0.9578	0.8987	0.9163	0.9618	11.283	0.8647	0.9074
Att_ResUNet_LUV_RGB	0.9205	0.9522	0.8961	0.9132	0.9612	10.728	0.8527	0.9046
Att_ResUNet_RGB_HSV_LUV	0.9387	0.9654	0.9135	0.9224	0.9855	10.121	0.8845	0.9179
DAG_ResUNet_RGB_HSV	0.9572	0.9783	0.9378	0.9483	0.9888	7.244	0.918	0.9430
DAG_ResUNet_HSV_LUV	0.9427	0.9672	0.9278	0.9348	0.9784	8.371	0.8916	0.9313
DAG_ResUNet_LUV_RGB	0.9411	0.9626	0.9241	0.9303	0.9776	8.795	0.8888	0.9272
HCRDAUNet	0.9689	0.9845	0.9475	0.9592	0.9897	6.872	0.9397	0.9533

Table2 contains the segmentation methods. It demonstrates that the DAG_ResUNet_RGB_HSVachieves0.9572 Dice, 0.9783 Accuracy, 0.9378 precision, 0.9483 Recall, 0.9888 specificity, 7.244 HD, 0.918 IoU and 0.9430 F1-score which are higher than all the prior methods.

The corresponding graphs represented as Dice, HD and F1-score comparison of Bi and Tri colour models.

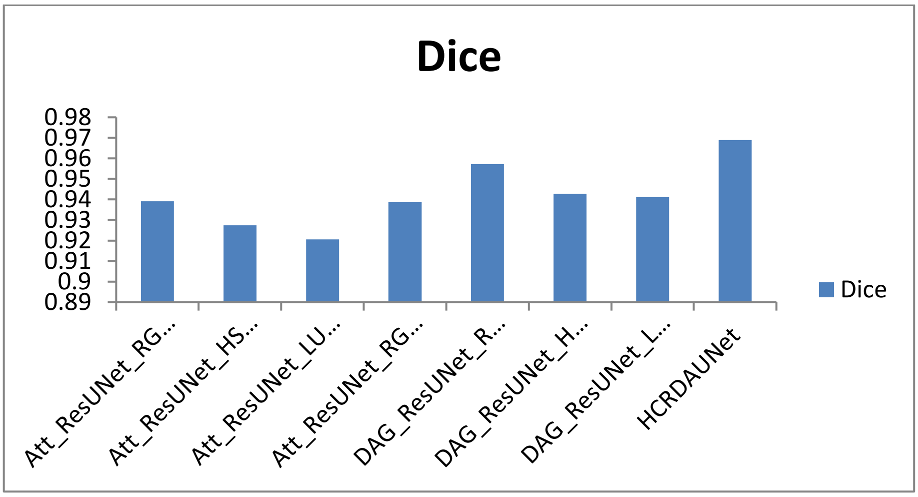


Figure 9. Dice comparison between Bi and Tri colour model

As shown in Figure 9, the HCRDAUNet has a higher dice value for Bi and Ti colour model. The HCRDAUNet better than 0.0298 for Att_ResUNet_RGB_HSV, +0.0415 for Att_ResUNet_HSV_LUV, +0.048 for Att_ResUNet_LUV_RGB, +0.0302 for Att_ResUNet_RGB_HSV_LUV, +0.012 for DAG_ResUNet_RGB_HSV, +0.026 for DAG_ResUNet_HSV_LUV and +0.028 for DAG_ResUNet_LUV_RGB.

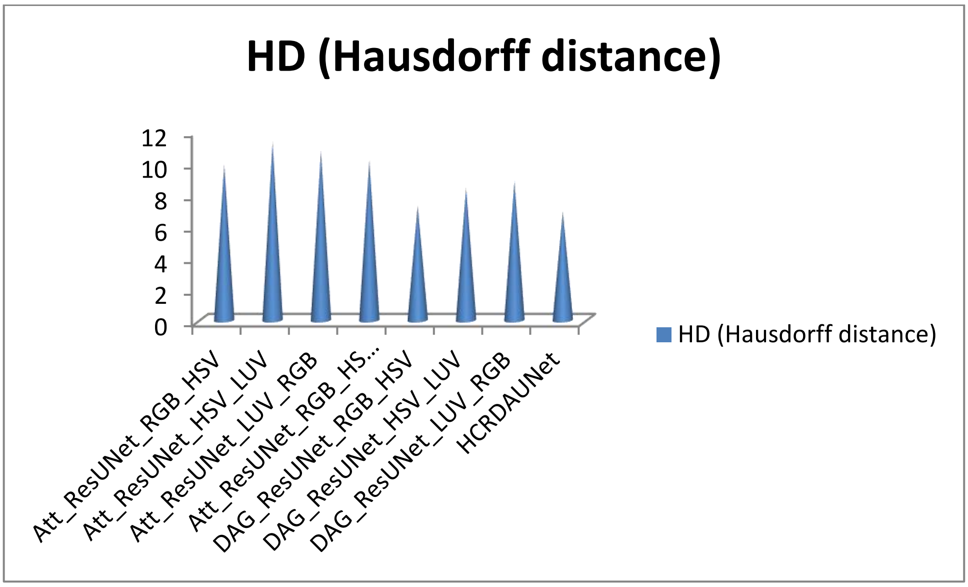


Figure 10. HD comparison between Bi and Tri colour model

From the Figure 10 HD value of the HCRDAUNet lower than comparison model of Bi and Tri colour. The HCRDAUNet lower value than 2.98 for Att_ResUNet_RGB_HSV, 4.411 for Att_ResUNet_HSV_LUV, 3.856 for Att_ResUNet_LUV_RGB, 3.249 for Att_ResUNet_RGB_HSV_LUV, 0.372 for DAG_ResUNet_RGB_HSV, 1.50 for DAG_ResUNet_HSV_LUV and 1.923 for DAG_ResUNet_LUV_RGB.

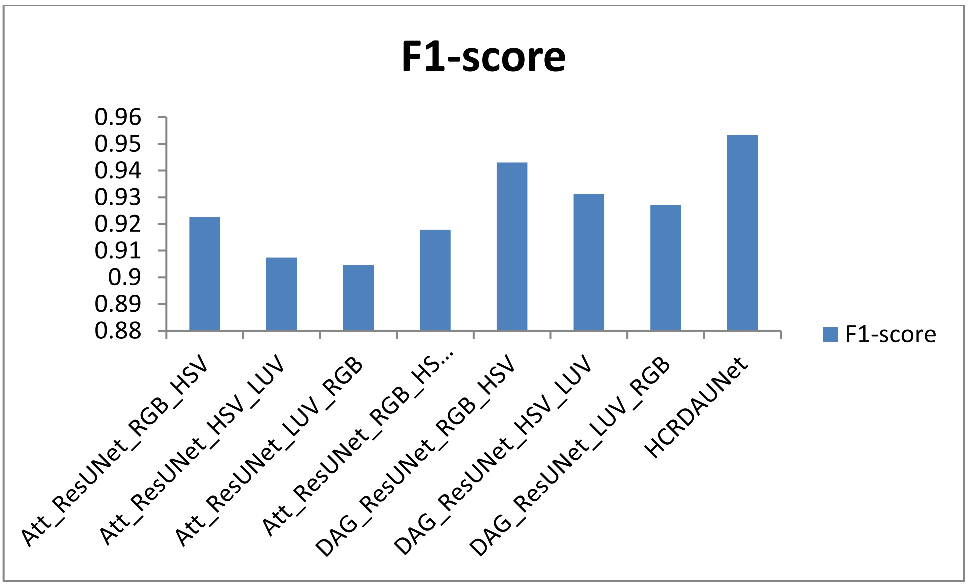


Figure 11. F1-score comparison between Bi and Tri colour model

From the Figure 11 shows the HCRDAUNet F1-score value for Bi and Tri colour model. The HCRDAUNet better value than +0.0306 for Att_ResUNet_RGB_HSV, +0.046 for Att_ResUNet_HSV_LUV, +0.049 for Att_ResUNet_LUV_RGB, +0.0354 for Att_ResUNet_RGB_HSV_LUV, +0.0103 for DAG_ResUNet_RGB_HSV, 0.022 for DAG_ResUNet_HSV_LUV and 0.026 for DAG_ResUNet_LUV_RGB.

Table 3. Comparison of Proposed model with baseline models

	Dice	Accuracy	Precision	Recall	Specficity	HD (Hausdorff distance)	IoU	F1score
FCN [52]	0.8910	0.939	-	0.89	0.961	14.445	-	-
SegNet [51]	0.8920	0.936	-	0.879	0.879	14.304	-	-
Unet [25]	0.89	0.94	-	0.888	0.96	12.789	-	-
Van-Truong Pham[45]	-	0.958	-	0.92	0.976	9.29	-	-
DRLSE [46]	0.8680	-	-	-	-	23.223	0.773	-
CV-vector valued [47]	0.8700	-	-	-	-	22.978	0.776	-
Deep nested LS [48]	0.8944	-	-	-	-	19.192	0.809	-
AttentionUNet [50]	0.9010	0.942	-	0.892	0.964	13.293	-	-
ResidualUNet [49]	0.9020	0.932	-	0.896	0.967	13.172	-	-
HCRDAUNet	0.9689	0.9845	0.9475	0.9592	0.9897	6.872	0.9397	0.9533

Table 3 contains the segmentation methods. It demonstrates that the proposed work achieves 0.9689 Dice, 0.9845 Accuracy, 0.9475 precision, 0.9592 Recall, 0.9897 specificity, 6.872 HD, 0.9397 IoU and 0.9533 F1-score which is higher than all the prior methods

CONCLUSION

Recent success in computer vision and deep learning attains very good success in medical imaging field. The abnormalities in the tympanic membrane (TM) are mainly connected with Otitis Media (OM) disorder. The early diagnosis of OM is very important to avoid hearing loss in all ages. The proposed deep semantic segmentation model helps to segment the mid ear membrane to make further study on it to find the occurrence of inflammation. This method enhances the performance of segmentation with the help of three different colour spaces using a unique linked form with the double end attention block. Initially, the Hybrid Colour Linked model combines the deep features of HSV and LUV model in the decoder or up sampling part and then those information are further joined with the standard RGB mode decoder features with the help of Double Attention Gate to provide the final representation of Tympanic membrane. The proposed HCRDAUNet method achieves96.89% Dice, 98.45% Accuracy, 93.97% IoU and 95.33% F1score for the tympanic membrane dataset. The proposed Hybrid Colour Double Attention Linked form of UNet model significantly rises the performance of membrane segmentation than the state-of-the-art works.

REFERENCES

[1] Joe H, Seo YJ. A newly designed tympanostomy stent with TiO2 coating to reduce Pseudomonas aeruginosa biofilm formation. J Biomater Appl. 2018 Oct;33(4):599-605.

[2] Lee SH, Ha SM, Jeong MJ, Park DJ, Polo CN, Seo YJ, et al. Effects of reactive oxygen species generation induced by Wonju City particulate matter on mitochondrial dysfunction in human middle ear cell. Environ SciPollut Res Int. 2021 Sep;28(35):49244-57

[3] Jabarin B, Pitaro J, Lazarovitch T, Gavriel H, Muallem-Kalmovich L, Eviatar E, et al. Decrease in pneumococcal otitis media cultures with concomitant increased antibiotic susceptibility in the pneumococcal conjugate vaccines era. OtolNeurotol 2017;38:853–9.

[4] Jaisinghani V, Hunter L, Li Y, Margolis R. Quantitative analysis of tympanic membrane disease using video-otoscopy. Laryngoscope 2000;110:1726–30.

[5] Fang T, Rafai E, Wang P, Bai C, Jiang P, Huang SN, et al. Pediatric Otitis Media in Fiji: Survey Findings 2015. Int J PediatrOtorhinolaryngol Extra 2016;(85):50–5.

[6] Lieberthal, A., Carroll, A., Chonmaitree, T., Ganiats, T., Hoberman, A., Jackson, M., Joffe, M., Miller, D., Rosenfeld, R., Sevilla, X., Schwartz, R., Thomas, P., Tunkel, D.: The diagnosis and management of acute otitis media. Pediatrics 131(3), e964–e999 (2013)

[7] Jaisinghani, V., Hunter, L., Li, Y., Margolis, R.: Quantitative analysis of tympanic membrane disease using video-otoscopy. Laryngoscope 110(10 Pt 1), 1726–1730 (2000)

[8] Albu S, Babighian G, Trabalzini F. 1998. Prognostic factors in tympanoplasty. American Journal of Otolaryngology 19(2):136–140 DOI 10.1016/S0196-0709(98)90111-9.

[9] Comunello, E., Wangenheim, A., Junior, V., Dornelles, C., Costa, S.: A computational method for the semi-automated quantitative analysis of tympanic membrane perforations and tympanosclerosis. Comput. Biol. Med. 39(10), 889–895 (2009)

[10] Tran, T., Fang, T., Pham, V., Lin, C.,Wang, P., Lo, M.: Development of an automatic diagnostic algorithm for pediatric otitis media. Otol. Neurotol. 39(8), 1060–1065 (2018)

[11] Jaisinghani V, Hunter L, Li Y, Margolis R. Quantitative analysis of tympanic membrane disease using video-otoscopy. Laryngoscope 2000;110:1726–30.

[12] Cha D, Pae C, Seong SB, Choi JY, Park HJ. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. EBioMedicine. 2019 Jul;45:606-14.

[13] Zeng X, Jiang Z, Luo W, Li H, Li H, Li G, et al. Efficient and accurate identification of ear diseases using an ensemble deep learning model. Sci Rep. 2021 May;11(1):10839.

[14] Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. J Imaging. 2020 Jun;6(6):52

[15] Ibekwe T, Adeosun A, Nwaorgu O. Quantitative analysis of tympanic membrane perforation: a simple and reliable method. J LaryngolOtol 2009;123. e2 Epub 2008 October 22.

[16] Hsu C, Chen Y, Hwang J, Liu T. A computer program to calculate the size of tympanic membrane perforations. ClinOtolaryngol Allied Sci 2004;29:340–2.

[17] Comunello E, Wangenheim A, Junior V, Dornelles C, Costa S. A computational method for the semi-automated quantitative analysis of tympanic membrane perforations and tympanosclerosis. Comput Biol Med 2009;39:889–95

[18] Xie X, Mirmehdi M, Richard Maw R, Amanda Hall A. Detecting abnormalities in tympanic membrane images. Proceedings of the 9th Medical Image Understanding and Analysis 2005:19–22.

[19] Tran T, Fang T, Pham V, Lin C1, Wang P, Lo M. Development of an automatic diagnostic algorithm for pediatric otitis media. OtolNeurotol 2018;39:1060–5.

[20] Shie C, Chang H, Fan F, Chen C, Fang T, Wang P. A hybrid feature-based segmentation and classification system for the computer aided self-diagnosis of otitis media. ConfProc IEEE Eng Med BiolSoc 2014:4655–8.

[21] Fang SH, Tsao Y, Hsiao MJ, Chen JY, Lai YH, Lin FC, et al. Detection of pathological voice using cepstrum vectors: A deep learning approach. Journal of Voice. 2018; 33(5):634–641. https://doi.org/10.1016/j.jvoice.2018.02.003 PMID: 29567049

[22] Mironică I, Vertan C, Gheorghe DC. Automatic pediatric otitis detection by classification of global image features. Proceedings of the E-Health and Bioengineering Conference IEEE. 2011 November 24-26; Iasi Romania; 2011. p.1-4.

[23] Huang Y, Huang CP. A Depth-First Search Algorithm Based Otoscope Application for Real-Time Otitis Media Image Interpretation. Proceedings of the 18th International Conference on Parallel and Distributed Computing, Applications and Technologies. 2017 December 18-20; Taipei Taiwan; 2017. p.170- 175.

[24] Myburgh, H. C., van Zijl, W. H., Swanepoel, D., Hellström, S., & Laurent, C. (2016). Otitis Media Diagnosis for Developing Countries Using Tympanic Membrane Image-Analysis. EBioMedicine, 5, 156–160. doi:10.1016/j.ebiom.2016.02.017

[25] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. ProcIntConf Med Image ComputComput-Assist Intervent 2015:234–41.

[26] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015:3431–40

[27] Seok, Jungirl& Song, Jae-Jin & Koo, Ja-Won & Chan, Kim & Choi, Byung. (2019). The semantic segmentation approach for normal and pathologic tympanic membrane using deep learning. 10.1101/515007.

[28] Pham, VT., Tran, TT., Wang, PC. et al. Tympanic membrane segmentation in otoscopic images based on fully convolutional network with active contour loss. SIViP 15, 519–527 (2021). https://doi.org/10.1007/s11760-020-01772-7

[29] Park YS, Jeon JH, Kong TH, Chung TY, Seo YJ. Deep Learning Techniques for Ear Diseases Based on Segmentation of the Normal Tympanic Membrane. ClinExpOtorhinolaryngol. 2023 Feb;16(1):28-36. doi: 10.21053/ceo.2022.00675.

[30] Lee JY, Choi S-H, Chung JW. 2019. Automated classification of the tympanic membrane using a convolutional neural network. Applied Sciences 9(9):1827 DOI 10.3390/app9091827

[31] Khan MA, Kwon S, Choo J, Hong SM, Kang SH, Park IH, et al. Automatic detection of tympanic membrane and middle ear infection from oto-endoscopic images via convolutional neural networks. Neural Netw 2020;126:384–94.

[32] Basaran E, Cömert Z, Şengür A, Budak Ü, Çelik Y, Togacar M. 2019a. Chronic tympanic membrane diagnosis based on deep convolutional neural network. In: 4th International Conference on Computational Mathematics and Engineering Sciences. 1–4.

[33] Basaran E, Sengur A, Comert Z, Budak U, Celik Y, Velappan S. 2019b. 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. Piscataway: IEEE, 1–6.

[34] Binol H, Moberly AC, Niazi MKK, Essig G, Shah J, Elmaraghy C, Teknos T, Taj-Schaal N, Yu L, Gurcan MN. SelectStitch: Automated Frame Segmentation and Stitching to Create Composite Images from Otoscope Video Clips. *Applied Sciences*. 2020; 10(17):5894. https://doi.org/10.3390/app10175894

[35] Viscaino, M.; Maass, J.C.; Delano, P.H.; Torrente, M.; Stott, C.; Cheein, F.A. Computer-aided diagnosis of external and middle ear conditions: A machine learning approach. PLoS ONE 2020, 15, e0229226

[36] Alhudhaif A, Comert Z, Polat K. Otitis media detection using tympanic membrane images with a novel multi-class machine learning algorithm. PeerJComput Sci. 2021;7:e405.

[37] Myburgh HC, Van Zijl WH, Swanepoel D, Hellström S, Laurent C. 2016. Otitis media diagnosis for developing countries using tympanic membrane image-analysis. EBioMedicine 5(4):156–160 DOI 10.1016/J.EBIOM.2016.02.017.

[38] Sundgaard, J. V., Harte, J., Bray, P., Laugesen, S., Kamide, Y., Tanaka, C., … Christensen, A. N. (2021). Deep metric learning for otitis media classification. Medical Image Analysis, 71, 102034. doi:10.1016/j.media.2021.102034

[39] Cha D, Pae C, Seong S-B, Choi JY, Park H-J. 2019. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. EBioMedicine 45:606–614 DOI 10.1016/j.ebiom.2019.06.050.

[40] Senaras, C., “Detection of eardrum abnormalities using ensemble deep learning approaches”, in <i>Medical Imaging 2018: Computer-Aided Diagnosis</i>, 2018, vol. 10575. doi:10.1117/12.2293297.

[41] Jetley, S., Lord, N.A., Lee, N., Torr, P.: Learn to pay attention. In: International Conference on Learning Representations (2018), https://openreview.net/forum?id=HyzbhfWRW.

[42] 42. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. arXiv preprint arXiv:1711.07971 (2017).

[43] Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets. In: Artificial Intelligence and Statistics. pp. 562–570 (2015).

[44] Li, C.; Tan, Y.; Chen, W.; Luo, X.; Li, F. ANU-Net: Attention-based Nested U-Net to exploit full resolution features for medical image segmentation. *Comput*. Graph. 2020, 90, 11–20. [Google Scholar] [CrossRef].

[45] Van-Truong Pham a,b, Thi-Thao Tran a,b,**, Pa-Chun Wang c,d, Po-Yu Chen c, Men-Tzung Lo b,*EAR-UNet: A deep learning-based approach for segmentation of tympanic membranes from otoscopic images,April2021.

[46] Li, C., Xu, C., Gui, C., Fox, M.D.: Distance regularized level setevolution and its application to image segmentation. IEEE Trans.ImageProcess.19(12), 3243–3254 (2010).

[47] Chan,T.,Sandberg,Y.,Vese,L.:Activecontourswithoutedgesforvectorvaluedimages. J. Vis. Commun. Image Represent. 11(2),130–141 (2000).

[48] Duan, J., Schlemper, J., Bai,W., Dawes, J.W., Bello, G.T., Doumou,G., De Marvao, A., O'Regan, D.P., Rueckert, D.: Deep nestedlevel sets: fully automated segmentation of cardiac MR imagesinpatientswithpulmonaryhypertension. In: International Conference on Medical Image Computing and Computer-AssistedIntervention, pp. 595–603 (2018).

[49] Z. Zhang, Q. Liu and Y. Wang, "Road Extraction by Deep Residual U-Net," in IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp. 749-753, May 2018, doi: 10.1109/LGRS.2018.2802944.

[50] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention UNet: learning where to look for the pancreas. Proc 1st Conf Med Imaging with DeepLearn 2018. p. Available: https://arxiv.org/abs/1804.03999.

[51] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE transactions on pattern analysis and machine intelligence 39.12 (2017): 2481-2495.

[52] Tran PV. A fully convolutional neural network for cardiac segmentation in shortaxis MRI. Available:. 2016. https://arxivorg/abs/160400494.