Towards low cost automated smartphone- and cloud-based otitis media diagnosis

Hermanus C Myburgh[a,*], Stacy Jose[a], De Wet Swanepoel[b,c], Claude Laurent[b,d]

[a]*Engineering Building 1, room 15-16, Department of Electronic, Electrical and Computer Engineering, University of Pretoria, Pretoria, South Africa*
[b]*Department of Audiology and Speech-Language Pathology, University of Pretoria, Pretoria, South Africa*
[c]*Ear Sciences Centre, School of Surgery, The University of Western Australia, Nedlands, Australia*
[d]*Department of Clinical Science, University of Umeå, Umeå, Sweden*

Abstract

Otitis media is one of the most common childhood illnesses. Access to ear specialists and specialist equipment is rudimentary in many third world countries, and general practitioners do not always have enough experience in diagnosing the different otitis medias. In this paper a system recently proposed by three of the authors for automated diagnosis of middle ear pathology, or otitis media, is extended to enable the use of the system on a smartphone with an internet connection. In addition, a neural network is also proposed in the current system as a classifier, and compared to a decision tree similar to what was proposed before. The system is able to diagnose with high accuracy 1) a normal tympanic membrane, 2) obstructing wax or foreign bodies in the external ear canal (W/O), 3) acute otitis media (AOM), 4) otitis media with effusion (OME) and 5) chronic suppurative otitis media (CSOM). The average classification accuracy of the proposed system is 81.58% (decision tree) and 86.84% (neural network) for images captured with commercial video-otoscopes, using 80% of the 389 images for training, and 20% for testing and validation.

Keywords: Otitis media, image processing, feature extraction, decision tree, neural network, otoscope, tympanic membrane

Introduction

One of the most common childhood illnesses is otitis media, also known as middle-ear infection. Acute otitis media (AOM) is associated with inflammation of the middle-ear and the tympanic membrane [1]. The two other forms of otitis

―――――――――――――――――

*Corresponding author
Email address:* `herman.myburgh@up.ac.za` (Hermanus C Myburgh)

media are otitis media with effusion (OME) and chronic suppurative otitis media (CSOM) [2]. Acquired hearing impairment in children is usually caused by CSOM, which is usually formed after recurrent AOM [3], but can also follow after a single episode. In most developing countries limited access to medical care increases the prevalence of otitis media [4], which could in some cases lead to permanent hearing loss. The most common method to diagnose otitis media is by an otologist or trained medical professional using an otoscope to examine the tympanic membrane [3]. Because of the potential high cost involved as well as the lack of access to these services, there exists a need for an automated otitis media diagnosis system mainly for use in developing countries. In these countries the children versus the adult population is much higher [5], and this raises the expected benefit of such a system to children.

An automated otitis media classification or diagnosis algorithm for two different forms of otitis media, namely AOM and OME, was proposed by Kuruvilla et al in [6]. This algorithm was validated by three general pediatricians as well as eight classifiers. Using this method otitis media diagnosis is only possible after the removal of obstructing wax or masses in the ear canal, in order to have a clear view of the tympanic membrane [3]. The drawback of this algorithm is that it does not detect wax or foreign body obstructions in order to diagnose the possibility of AOM and OME, and it does not diagnose CSOM.

In a recent paper we have demonstrated that otitis media can be automatically diagnosed using image processing and decision tree classification [7]. This system was able to diagnose AOM, OME, CSOM, wax or foreign body obstruction and a normal tympanic membrane. The image analysis classification system could load tympanic membrane images from a personal computer, running the custom developed Matlab software. The system extracted the visual features using tailor-made feature extraction algorithms, which were then classified using a decision tree. The system also made provision for the use of a low cost custom-made otoscope, also developed by the authors, to be used to take pictures of the tympanic membrane. Such a low cost custom-made otoscope could be provided to underserved communities in to dramatically reduce cost. Reported performance was 81.6% for video-otoscope captured tympanic membrane images, and 78.7% for custom-made video-otoscope images captured in an emergency room.

The smartphone revolution has resulted in unprecedented access to smartphones and smart mobile devices globally. Today one in five people owns a smartphone [8], and it is estimated that in the year 2020, 6.1 billion smartphones with internet access will be in circulation [9]. In fact, currently the fastest global growth of mobile subscribers is experienced in Africa, rising to an estimated 0.5 billion in the next four years [10]. It therefore makes sense to endeavor to develop unique smartphone-based systems to help take health care to underserved populations at a fraction of the cost, especially in developing countries. Mobile health, or mHealth, initiatives and applications improve access to medical supplies, track immunisations, and unlock the potential of remote diagnostics to transform the entire regional healthcare landscape, resulting in higher referrals, faster diagnoses and a dramatically improved patient experience [11].

In this paper the image analysis classification system proposed in [7] is expanded by 1) redesigning the system to be used from an Android smartphone wherever there is an internet connection, 2) redesigning the feature extraction algorithms and the decision tree in [7] to be more compact, and 3) designing a neural network as an improved classification option as a replacement for the decision tree, which results in improved diagnostic performance.

Background and Context

Tympanic Membrane

The tympanic membrane or eardrum is the structure of the ear that mirrors most of middle ear diseases. It is approximately 0.1 mm thick and 8 mm to 10 mm in diameter [12]. This membrane separates the external ear from the middle ear and receives sound vibrations from the air outside, which is transferred to the middle-ear ossicles [13]. Ossicles are three small bones that transmit sound to the cochlea (auditory portion of the inner ear). Figure 1 shows the basic anatomy of the human ear.
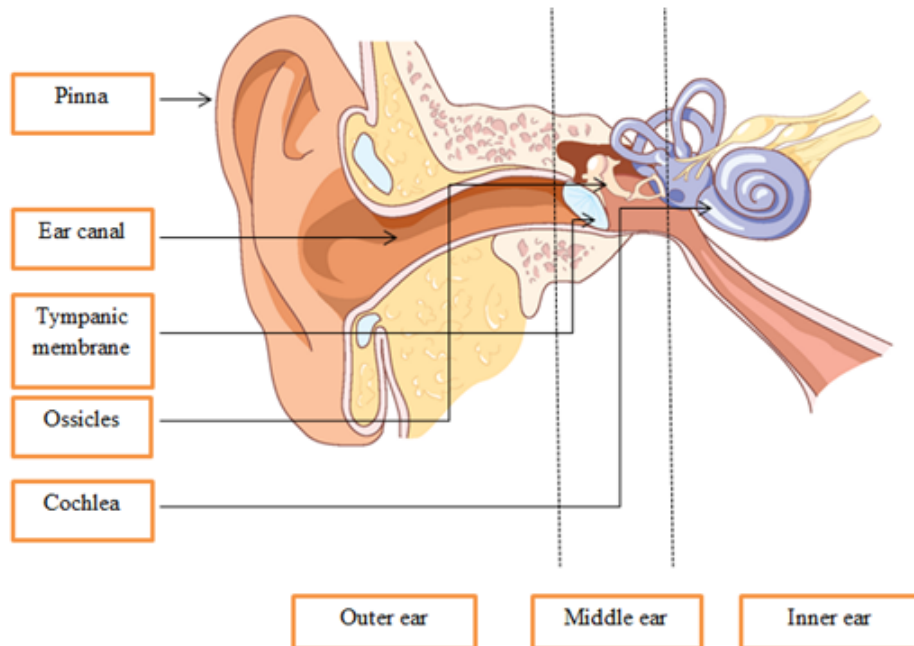

Figure 1: Basic anatomy of the ear. Modified from [14].

Otitis Media Pathologies

A healthy tympanic membrane is pearly grey in colour and slightly conical, with no bulging or retraction. The membrane has a smooth, shiny, translucent

appearance. The malleus handle can be clearly seen as well as the umbo - the central portion of the tympanic membrane where the malleus handle ends. When otoscopy is performed, the light from the otoscope reflects on either the right bottom or left bottom corner below or in front of the umbo (depending on the left or right ear), in the shape of a cone [15], as seen in Figure 2 [16].
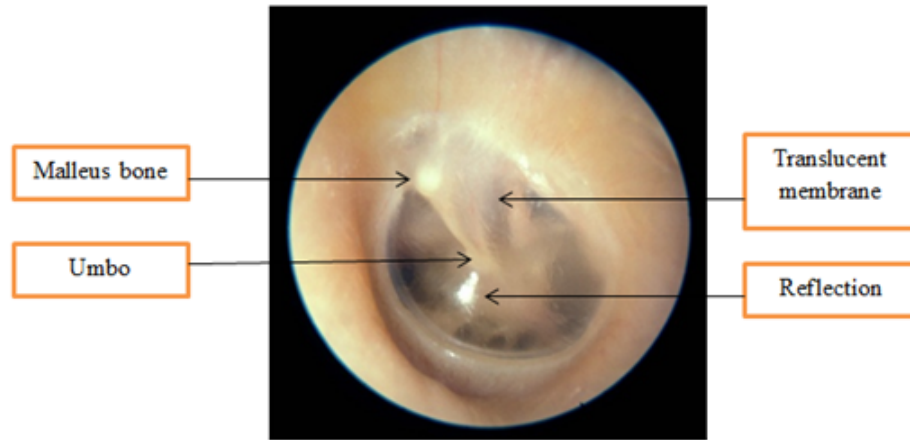

Figure 2: Normal tympanic membrane [16].

AOM is a middle-ear condition that is marked by the presence of infectious material or pus and mainly a red bulging membrane, and the malleus bone is obscured. OME is marked by middle ear fluid of variable character, but the membrane is not bulging and the malleus bone is clearly visible [17]. Figure 3 shows the difference between AOM and OME [18].
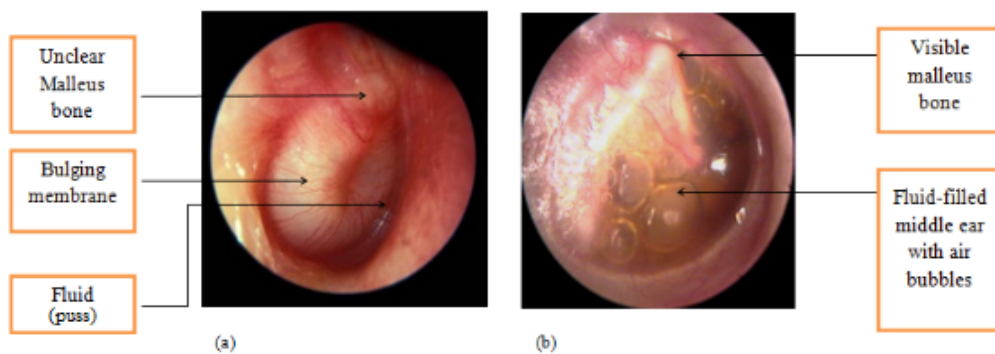

Figure 3: Tympanic membrane with AOM (a) and OME (b) [18].

A common form of CSOM usually occurs after AOM. Its main feature is a

chronic perforation in the tympanic membrane. Figure 4 shows with a huge dry central perforation and visible ossicles in the middle ear [18].
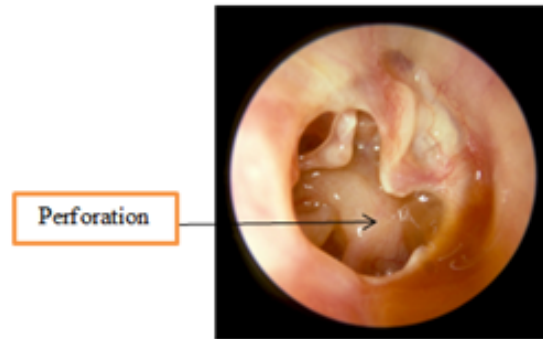

Figure 4: Tympanic membrane with CSOM [18].

Wax obstructions occur when earwax accumulates to a point where it becomes too hard to be cleared naturally. It consists of skin, sweat, hair and debris and its color ranges from dark brown to orange or yellow. Its presence is usually in the outer half of the ear canal [19]. Foreign bodies can also be present amongst the wax; they are usually insects or biological material. These are found most commonly in children [20]. Figure 5 illustrates wax obstruction and foreign body obstructions [21].
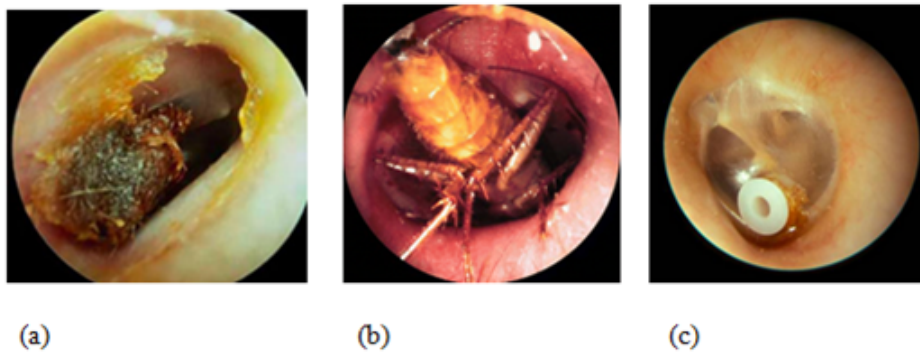

Figure 5: Tympanic membrane with wax obstruction (a) insect obstruction (b) and a grommet surgically placed into the tympanic membrane [20].

Otitis Media Diagnosis using Visual Features

In order to diagnose otitis media from an image taken with a video-otoscope, a number of visual features have to be identified for each diagnosis. Table 1

shows the features associated with each pathology, which were used to develop the feature extraction algorithms of the proposed system.

Table 1: The signs of the different types of diagnostic groups (y - yes, n - no, B - bulging, R - retracted, N - normal, O - other, R - red, W - whitish). W/O designates obstructing wax or foreign bodies in the external ear canal.

	AOM	OME	CSOM	Normal	W/O
Malleus bone visible	n	y	y/n	y	y/n
Tympanic membrane shape	B	R	N/B/O	R/N/B	R/N/B/O
Colour	R	R/W	W	W	W
Perforation	n	n	y	n	n
Wax	y/n	y/n	y/n	y/n	y
Fluid	y/n	y	n	n	n
Reflection	y/n	y/n	n	y	n

Automated Otitis Media Diagnosis

The proposed system uses a smartphone to load tympanic membrane images captured with a video-otoscope, and preprocess the images to prepare them for further processing. After preprocessing, the images are transmitted to a server in the "cloud", where feature extraction and classification are performed by both the newly designed decision tree and the neural network for comparison.

On the macro scale the automated otitis media diagnosis system consists of three main processes, namely *preprocessing*, *feature extraction*, and *classification*. Preprocessing prepares or standarizes the image, feature extraction identifies the visual cues in the image, while classification ultimately diagnosis the image based on the extracted features.

For the proposed system, preprocessing occurs on the Android smartphone, while feature extraction and classification take place on the server, whereafter the diagnosis is sent back to the phone. These processes are discussed in detail in this section.

The video-otoscope images were acquired from a number of sources. In total 562 high-quality images were selected for diagnosis in order to establish ground truth to develop the proposed system from. Thereafter two specialists, including the last author, examined the 562 tympanic membrane images independently, and only the images for which their independent diagnoses were in agreement were retained. This method was also used in [7], where its reliability is explained. This resulted in 489 doubly diagnosed images. These images were further examined, after which another 100 images were eliminated due to image quality, blur, and color deficiencies, which yielded a final set of 389 images.

Preprocessing

Before the features can be extracted from the tympanic membrane image, the image first needs to be prepared, or preprocessed to standarize the image

for use by the feature extraction algorithms. The two processes involved in the preprocessing step are cropping and blur detection.

Cropping

Cropping is an essential preprocessing step; its role is to remove all unwanted image information around the border and thereby isolate the tympanic membrane in the image. The images captured with the video-otoscope have a black border by the interior of the disposable speculum used to examine the ear canal. After cropping the image size is 864×848 pixel^2.

Blur detection

Once the image has been successfully cropped, blurriness of the image is detected in order to establish if the image is well-focused. The accuracy of image processing on the server is directly proportional to the clarity of the image, and as such the system needs to ensure that images are in focus.

Blurriness was detected via thresholding; this is where a single value is calculated to determine if the calculated "blurriness factor" exceeds it or not. A variation of the Laplacian [22] was used. The Laplacian is used to measure the second derivative of an image; it highlights areas where there are rapid intensity changes. To detect if an image is blurry, the image is converted to gray scale. The gray scale image is then convolved with the Laplacian kernel, after which the variance of the response is calculated; a high variance indicates that there is a wide spread of many edges, which is evidence of a focused image. A threshold of 4000 was calculated to determine whether an image is blurry or not. Figure 25 shows a flow diagram of the blur detection process.

Feature extraction

The feature extraction algorithms were designed and tested using a set of video-otoscope images. There were 389 images in total with known diagnoses and suitable image quality, of which 80% were used for training and 20% were used for validation and testing. This ratio was chosen firstly in order to be representative of the problem domain, ie. otitis media diagnosis using image processing of tympanic membrane images, and second to leave enough images for verification and testing of the feature extraction algorithms. Too few training images would result in low diagnostic accuracy, while too many images would leave too few images for validation. Table 2 shows the number of images per category.

The training images were used to design the feature extraction algorithms, in order to ensure that each algorithm can successfully identify its visual cue or feature. These visual cues are described in Table 1. Together the feature extraction algorithms identify the visual features, much like an otologist would, to ensure accurate diagnosis in the final step of the proposed system, namely classification.

7

Table 2: Number of video-otoscope images used to develop the feature extraction algorithms (training)

Categories	Total number of images	Training (80%)	Testing (20%)
AOM	51	41	10
OME	69	55	14
CSOM	86	69	17
Normal	123	98	25
W/O	60	48	12
Total	389	311	78

Color detection

Four out of the seven features required for feature extraction use color detection; namely *wax and foreign body obstructions*, *reflection*, *tympanic membrane color* and *fluid presence*.

Wax and foreign bodies:. To detect wax and foreign body obstructions, the image is cropped further to remove excess black boarder pixels, resulting is a 350×350 pixel^2 image. The image is then converted to a hue, saturation and value (HSV) image. Using the training set of images provided, the color detected in images for wax or foreign body obstructions falls in the color range indicated in Figure 6.

Figure 6: Wax and foreign bodies obstruction color range.

All image pixels that fall in this color range are set to white. The number of white pixels in the image is calculated and expressed as an percentage of the total number of image pixels. Given all the training images, a threshold of 10% was determined, ie. if the percentage of white pixels is above 10% it is deemed as containing excessive wax or a foreign body. Figure 26 shows a flow diagram describing the wax and foreign bodies obstruction color detection process.

Reflection:. To detect if the image contains a reflection from the light of the otoscope (a sign of a normal tympanic membrane) using the training set of images provided, the reflection appears mostly below and in front of the malleus bone (Figure 7 (a)), and therefore the image is first cropped to a 200×200 pixel^2 image (Figure 7 (b)). The image is converted to a gray scale (Figure 7 (c)) where after a brightness filter is applied (Figure 7 (d)). The effect is that

the reflection will be highlighted in the form of a white patch if it is present in the image, while the rest of the image pixels will be black. Each pixel in the image is then evaluated and counted to determine the number of black pixels, and expressed as a percentage of total number of pixel. Given all the relevant training, a threshold of 0.007% was calculated, ie. when the percentage is above 0.007%, a reflection is detected. Figure 27 shows a flow diagram describing the reflection color detection process.
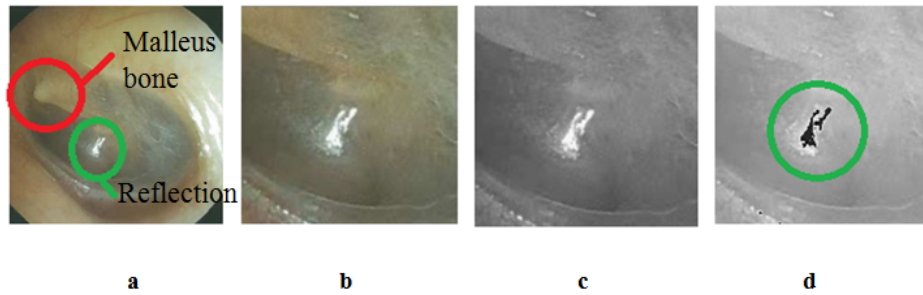
Figure 7: Reflection detection process. a) Original image showing the malleus bone and the reflection. b) The cropped image. c) The gray scale image. d) Gray scale image after brightness filtering.

Tympanic membrane color:. The tympanic membrane color is either pearly white (normal) or a dark shade of red (inflamed). To determine the tympanic membrane color the image was firstly cropped further to remove any excess black boarder pixels thus forming a 350×350 pixel^2 image. The image is analyzed for excessive red in an image, which is done by extracting only the red channel of image. Using the training set of images provided, the color detected in images for a red tympanic membrane ranges in the color range as shown in Figure 8.

Figure 8: Tympanic membrane color range.

All image pixels that fall in this color range will be marked as white. The white pixels in the image are then counted and expressed as a percentage of total image pixels. Given all training data, a threshold of 50% is calculated,

ie. if the white pixel percentage is above 50%, the membrane is noted to have an excessive amount of red. Figure 28 shows a flow diagram describing the tympanic membrane color detection process.

Fluid presence:. Fluid behind the translucent tympanic membrane in the tympanic membrane is usually clear, and white to yellow in color as seen in Figure 9 (a). To detect fluid, therefore, the image is cropped further to remove any access black boarder pixels thus forming a 350×350 pixel^2 image (Figure 9 (b)). The image is then converted to an HSV image (Figure 9 (c)) from which the saturation channel is extracted, where white areas in the RGB image will be black in this channel (Figure 9 (d)). There could also be other white areas such as reflection or transparent membrane, but they are of a different lighter shade than the fluid. By thus applying a brightness filter the less black areas are faded out leaving only the fluid and mass (Figure 9 (e)). The average number of black pixels among all image pixels is then calculated. After evaluating all the training images, a threshold of $60,000$ pixels or 49% was calculated, ie. if the percentage of pixels is above 49% fluid or a white mass is detected. Figure 29 shows a flow diagram describing the fluid color detection process.
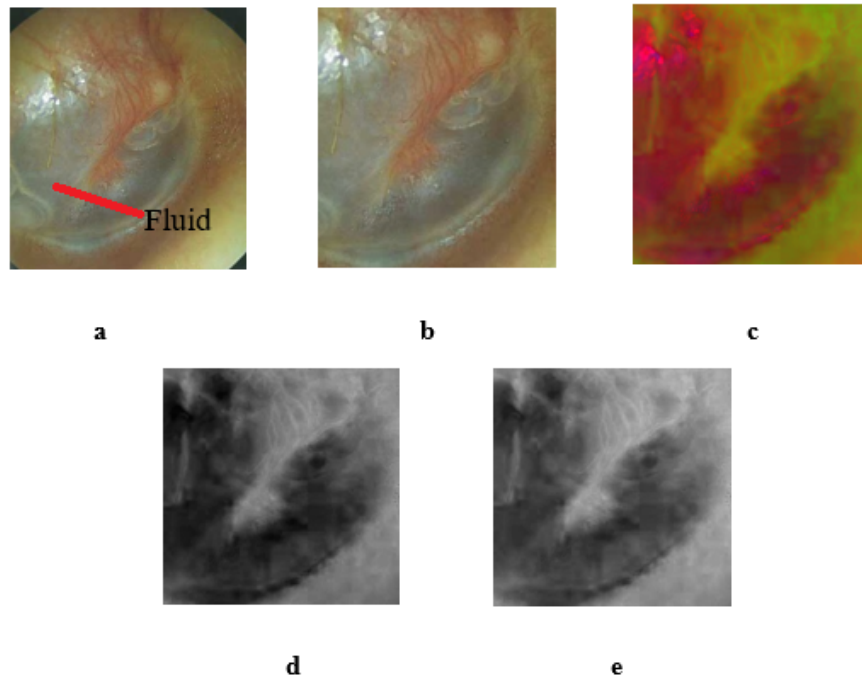

Figure 9: Fluid detection process. a) Original image showing the fluid. b) The cropped image. c) The HSV image. d) Saturation band image. e) Saturation band image after brightness filtering.

Edge detection

Edge detection for feature extraction involves detecting circular sharp edges to identify perforation of the tympanic membrane. Figure 10 (a) shows the original image. The image is firstly cropped further to remove any excess black boarder pixels as before (Figure 10 (b)). The image is then converted to an HSV image (Figure 10 (c)), and the saturation channel is used as edges are more prominent in that channel (Figure 10 (d)). Hough circles are used to mark the identified perforation (Figure 10 (e)).

Canny edge detection was performed to detect the edges caused by perforation [23]. Canny edge detection achieves low error rate, good localization and minimal response. A 3×3 Sobel kernel was set to find the intensity gradients of the image. A lower pixel gradient threshold was set to 20 and upper pixel gradient threshold was set to 60; therefore if the pixel is above the upper threshold, the pixel is accepted as an edge, if the pixel gradient is below the lower threshold then it is rejected. Moreover, if the pixel gradient is in between the two thresholds, then it will be accepted as an edge if it is connected to a pixel that is above the upper threshold [24].

This results in an image that outlines the circular edges of the image. Since there can be other circular images due to smaller particles or reflections, the Hough transform is applied. This enables the algorithm to detect circular edges in the image and marks the circle it detects with a prominent color. The canny edge image is first converted to gray scale as Hough circles detects circles in gray scale images, the minimum radius of the circles detected is set to 20 and the maximum is set to 130 [25]. The image is then converted back to a BGR image to be able to see the colored circle indicators. From the training set of images provided, there is usually only one perforation in the tympanic membrane. Thus the algorithm will count the number of circles present, if it detects one circle it is a true positive that there is a perforation in the image. Figure 30 shows a flow diagram describing the edge detection process.

Blob detection

One of the malleus bone's most prominent features is that the handle at its top has a circular or ball-shaped component with a minimum of 100 pixels wide. It also appears on the first one third of the image from the top. However, there are many other circular features that can be present on the image such as reflection and small perforations. Therefore the image is cropped to a 486x190 image, observing the first third of the image as shown in Figure 11 (a). The image is then converted to a HSV image (Figure 11 (b)), and saturation band is used (Figure 11 (c)). A brightness filter is applied to allow other dark areas to either fade out or be darker, leaving the malleus bone area to be of a certain shade (Figure 11 (d)). Simple blob detection is then applied, which allows circles to be detected using color, area, circularity, inertia and convexity (Figure 11 (e)).

From the results obtained from the training images there is usually a malleus bone with a height ranging between 10-15 pixels and a smaller circle. If one circle is detected, the image is classified as having a visible malleus bone, and if none
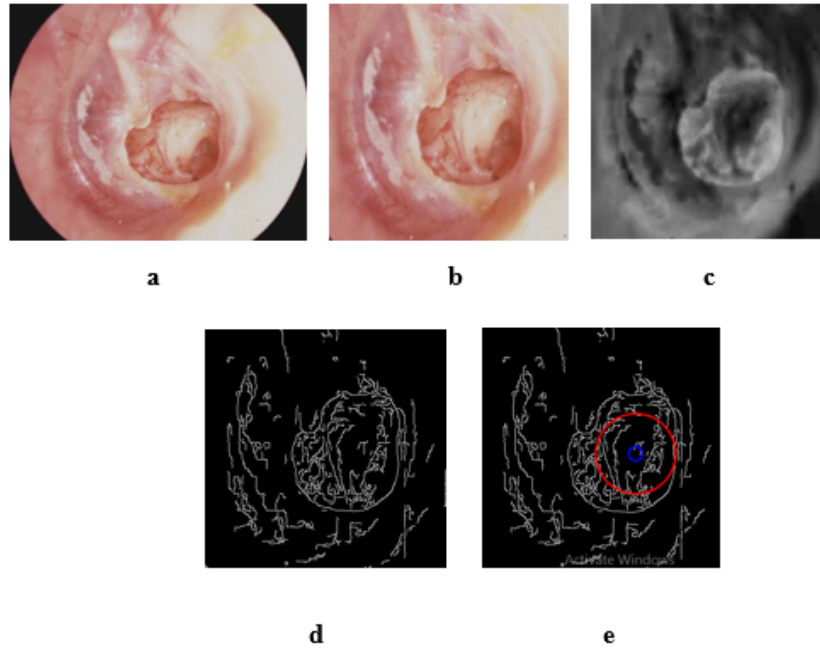
Figure 10: Perforation detection process. a) Original image showing the perforation. b) The cropped image. c) The HSV image. d) Saturation band image with canny edge detection. e) Hough circles indicating the perforation.

is detected the image is classified as not having a visible malleus bone. Figure 31 shows a flow diagram describing the blob detection process.
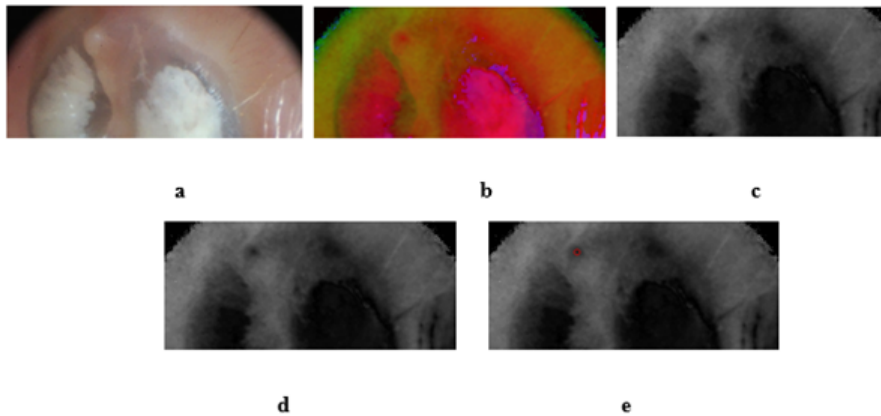

Figure 11: Malleus bone detection process.

Shape detection

In order to detect the shape of the tympanic membrane, the image is cropped to remove any excess black boarder pixels, yielding a 350×350 pixel^2 image. The image is then converted to gray scale so that the gradient of each pixel can be calculated. This is done by subtracting the neighboring pixel values; subtracting each pixel value from the one to the right of it. The average of the all the differences is then calculated to determine the average gradient, which indicates whether the tympanic membrane is bulging, normal or retracted. Figure 32 shows a flow diagram describing the shape detection process.

Classification

The feature detection algorithms are used to extract visual information in Table 1 from the tympanic membrane images. This information is then presented in the form of a feature vector - a vector containing all the visual features identified during feature extraction. This feature vector is the provided as input to the classification algorithms to diagnose the image in question.

In this study two classification algorithms are presented - a *decision tree* and a *neural network* - and their respective classification/diagnostic accuracies are compared.

Decision tree

One of the artificial intelligent algorithms used to classify the feature vectors is a decision tree. It identifies the minimum distance between the input vector and the decision tree branches encoded in the decision tree structure using feature vectors obtained from the training images. It does this by comparing the features contained in the feature vectors iteratively with the features encoded in the decision tree, and with each iteration traverses down the decision tree until it reaches the end of a branch, which constitutes classification. We refer the reader to [7] for a thorough discussion on decision tree construction, training and classification.

There were 389 video-otoscope images in total, 80% (311 images) of those images were used to construct the decision tree and 20% (78 images) were used for testing each category. In order to construct the tree training images were preprocessed and then processed by the feature extraction algorithms, and feature vectors of each image were obtained. A total of 108 vectors were obtained by excluding duplicates. The entropies and information gains were subsequently calculated, and the feature with the highest information gain was selected to be divided first. This sequence was repeated until the feature vector subsets were empty. The resulting decision tree had seven levels. The labels used for features are presented table 3. Figures 12 to 18 show the resulting decision tree.
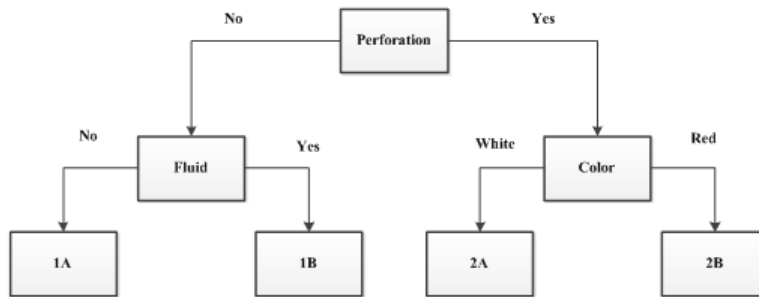
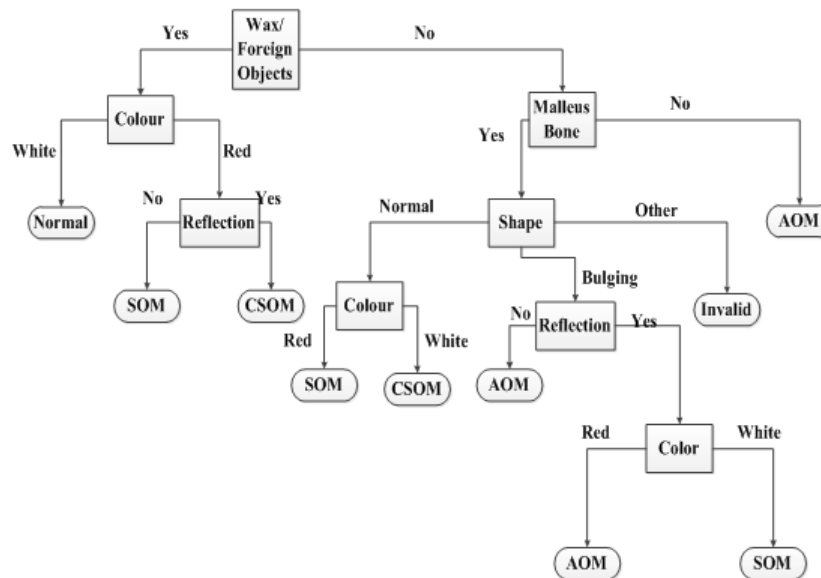Figure 12: First two levels of the decision tree.

Figure 13: Branching out of 1A from figure 12 (acute otitis media (AOM), otitis media with effusion (OME), chronic suppurative otitis media (CSOM), obstructing wax or foreign bodies (W/O)).

Neural network

Apart from the decision tree, an artificial neural network was also used to classify or diagnose the tympanic membrane images. Like the decision tree, the neural network receives the feature vectors as the input, and produces a classification at the output. But before it can be used as a classifier to diagnose the tympanic membrane images' feature vectors, it must first be trained using training images of known diagnoses, and validated and tested with images of unknown diagnoses.

As mentioned, the feature vectors are provided as input to the neural network. However, the various features have to be provided in the ranges $[-1, 1]$
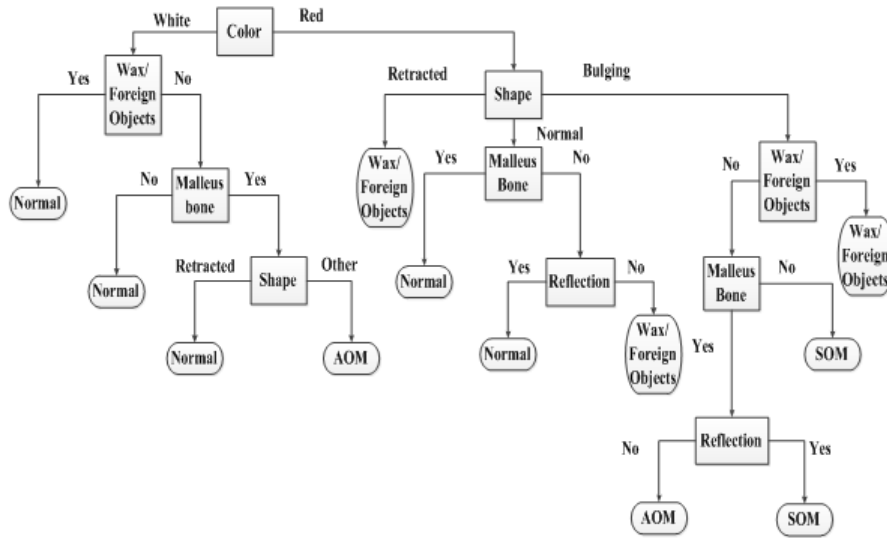
14

Figure 14: Branching out of 1B from figure 12 (acute otitis media (AOM), otitis media with effusion (OME) or secretory otitis media (SOM), chronic suppurative otitis media (CSOM), obstructing wax or foreign bodies (W/O)).
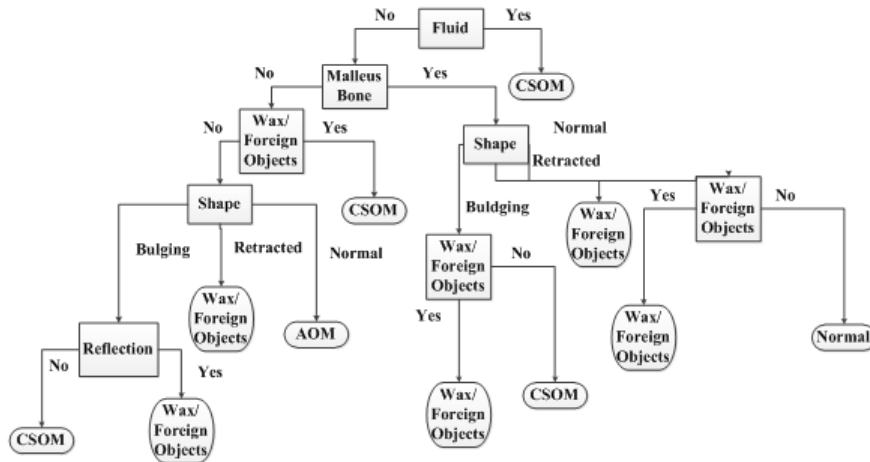

Figure 15: Branching out of 2A from figure 12 (acute otitis media (AOM), otitis media with effusion (OME) or secretory otitis media (SOM), chronic suppurative otitis media (CSOM), obstructing wax or foreign bodies (W/O)).

or $[0, 1]$. Therefore the values representing the various features in the feature vectors calculated from the tympanic membrane images used to train, validate and test the system, have to be normalized.

The design of the neural network architecture used to classify the feature vectors of the tympanic membrane images is as follows (shown in Figure 19):
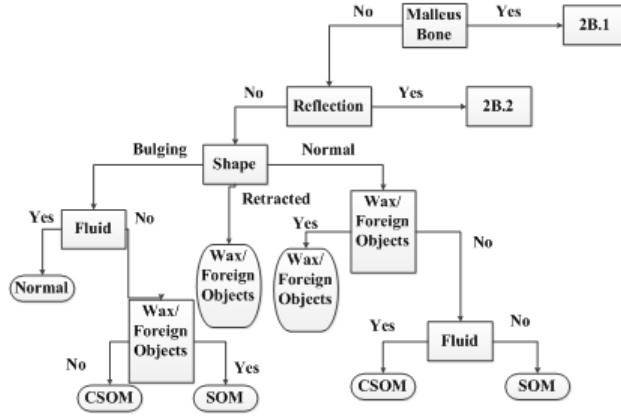
15

Figure 16: Branching out of 2B from figure 12 (acute otitis media (AOM), otitis media with effusion (OME) or secretory otitis media (SOM), chronic suppurative otitis media (CSOM), obstructing wax or foreign bodies (W/O)).
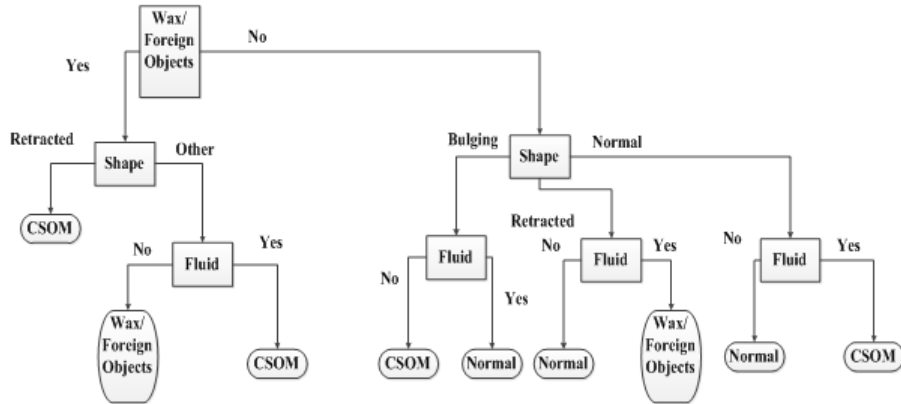

Figure 17: Branching out of 2B.1 from figure 16 (acute otitis media (AOM), otitis media with effusion (OME)or secretory otitis media (SOM), chronic suppurative otitis media (CSOM), obstructing wax or foreign bodies (W/O)) .

- 7 neurons in the input layers; one for each feature.

- 10 neurons in a single hidden layer.

- 5 neurons in the output layer; one for each pathology/diagnosis option.

Table 4 shows the encoding of the outputs of the neurons in the output layer to represent each diagnosis. The hyperbolic tangent function was used as the activation function. Figure 20 shows the hyperbolic tangent function [26] and is expressed as

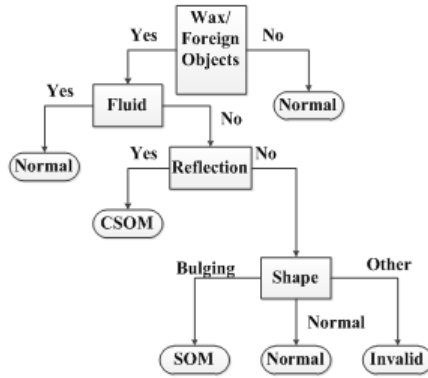$$f(x) = \frac{1 - e^{-2x}}{1 + e^{2x}}. \tag{1}$$

16

Figure 18: Branching out of 2B.2 from figure 16 (acute otitis media (AOM), otitis media with effusion (OME) or secretory otitis media (SOM), chronic suppurative otitis media (CSOM), obstructing wax or foreign bodies (W/O)) .

Table 3: Decision tree labels

Label	Definition
W/O	Wax and foreign body obstructions
R	Reflection
C	Color
S	Shape
MB	Malleus Bone
P	Perforation
F	Fluid
Max	Maximum information gain
PN	Perforation No
PY	Perforation Yes
FY	Fluid Yes
FN	Fluid No
CW	Color White
CR	Color Red
W/OY	Wax/Obstructions Yes
W/ON	Wax/Obstructions No
MBY	Malleus Bone Yes
MBN	Malleus Bone No

Training:. Before the neural network can be used to classify the feature vectors of undiagnosed tympanic membrane images, it first has to be trained using feature vectors of images with known diagnoses. Back-propagation learning is a well-known and efficient neural network training algorithm [27] [50], where example feature vectors are used to adjust the network's weights in order to achieve a realistic mapping between the input features and the output diagnosis. By providing the neural network with input feature vectors with known diagnoses,
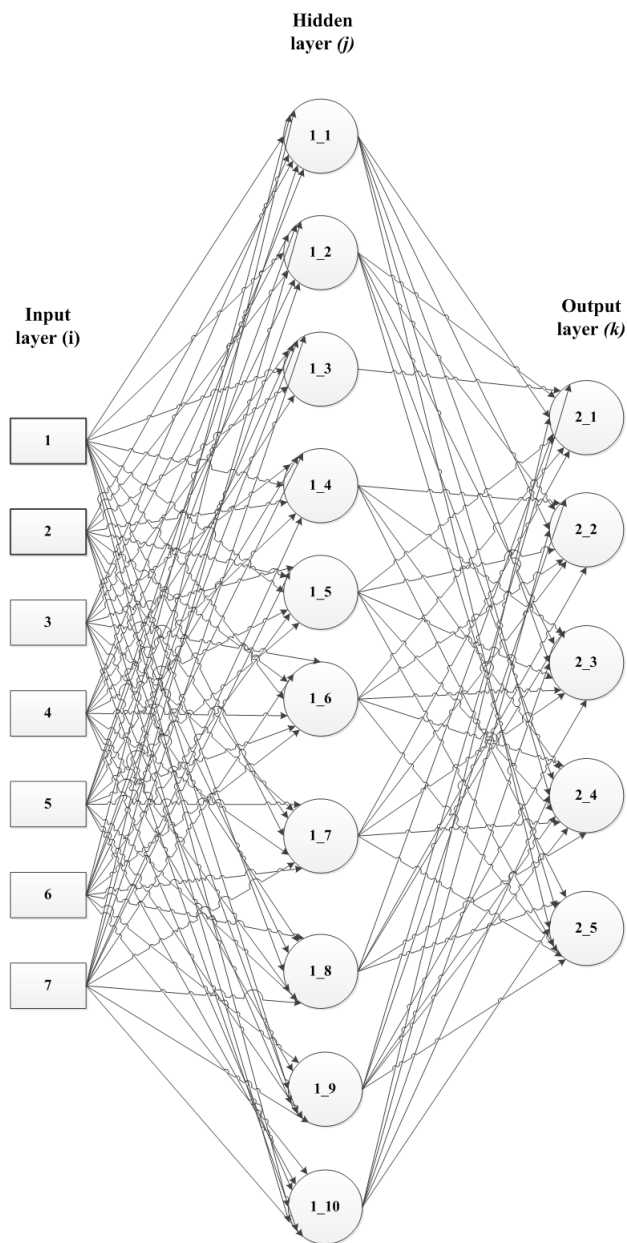
17

Figure 19: The neural network used for classification of the feature vectors, with 7 inputs, 10 neurons in the hidden layer, and 5 output neurons.

the weights are iteratively adjusted by comparing the output of the network to the known output/diagnosis, and feeding the output error back into the net-
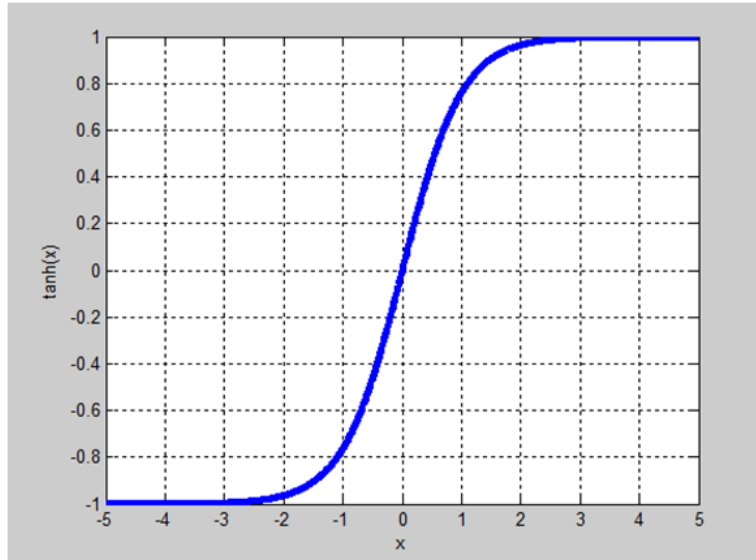
Figure 20: The hyperpolic tangent function used as the decision function of the neurons.

Table 4: Neural network output encoding for five neuron outputs. For each diagnosis only one of the neurons' outputs have a value of 1, while the other neurons' output values are -1.

Diagnosis	Neural network output
AOM	$-1, -1, -1, -1, 1$
OME	$-1, -1, -1, 1, -1$
CSOM	$-1, -1, 1, -1, -1$
W/O	$-1, 1, -1, -1, -1$
Normal	$1, -1, -1, -1, -1$

work. By performing this process on sufficient amounts of training data, the output error decreases to the point where the network's output closely matches the known output/diagnosis.

Back propagation is ideal for pattern recognition and mapping tasks. Training is halted when the error reaches an arbitrary sufficiently low value. This ensures that the classification/diagnoses of unseen tympanic membrane image feature vectors will have a good chance of being accurate. To increase the learning tempo of the neural network, a constant (α) called the learning rate is used. If selected too low, the convergence time (time to reach a low error) will be long. If selected too high, it will result in the weight adjustment increments to be too large, leading to an increase in the output error [28].

As was the case for the decision tree, 311 images were used for training, while 87 images were used for validation and testing respectively. To train the neural network, the training images were processed using the preprocessing and feature extraction algorithms. The resulting feature vectors were normalized to the

range of $[-1, 1]$ and were applied to the input of the neural network, after which the inputs were processed and an output was produced. Because the diagnosis of each image was known, the difference or error between the neural network output and the known output (see table 4) was calculated, and fed back into the system according to the back-propagation learning algorithm. Therefore, for each example or training image, the neural network weights were adjusted to reduce the output error. After the 311 training images were processed once, the validation images were used to calculate the output error separately. If the error threshold had not been reached, the 311 training images would once again be processed. This repetitive training process of all the training images is called an epoch. Training was halted once the error was sufficiently low or if 10 epochs where completed.

Various learning rates and amounts of epochs were experimented with to obtain a minimal error. The final learning rate was selected to be 0.035 and the number of epochs 10. Figure 21 shows the error achieved over 10 epochs using the training, validation and test data. Figures 33 and 34 show the back-propagation learning process as well as the classification process of the neural network respectively.
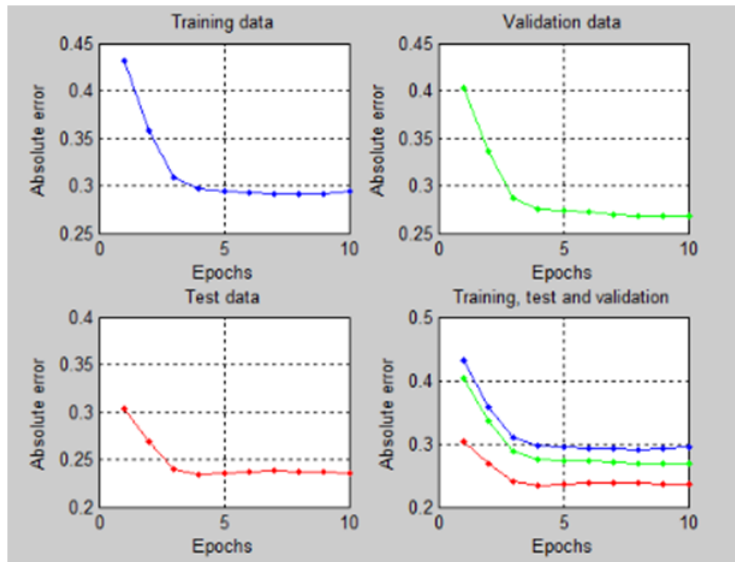

Figure 21: Neural network training errors over 10 epochs using training, validation and test data.

Smartphone application user interface

The systems user interface was designed using Android studio. The application designed is user friendly and simple to use. The application allows a camera view so that the tympanic membrane can be observed. Buttons, status

bars and pop up signs were used to guide and alert the user of any changes or statuses. Users with little technical knowledge can easily use the application.

The user can launch the Android application from the application list. Once the user has launched the application it opens as seen in Figure 22.
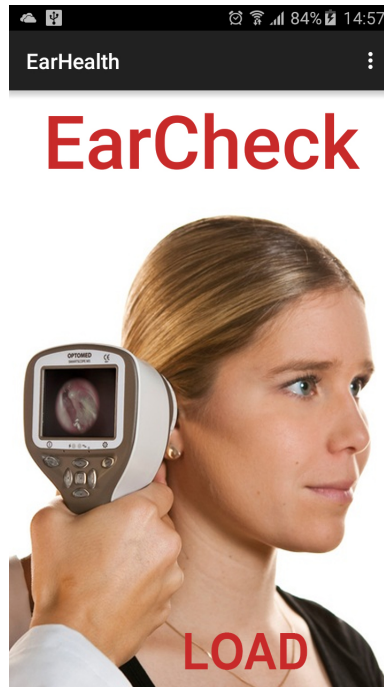

Figure 22: Smartphone application home screen.

Upon tapping on the "LOAD" text, the screen in Figure 23 (a) opens, but only after a tympanic membrane image has been place in the correct folder on the phone. When the user clicks the CROP button the image is cropped to an 866x848 image (Figure 23 (b)), which is ready for upload. If the user clicks the UPLOAD button, the image is uploaded to the server where it will undergo feature extraction and classification or diagnosis. After classification the server sends back the result to the phone as shown in Figure 24.

Results

Decision Tree

The decision tree was designed as explained before using 80% of the video-otoscope image with known diagnoses. The remaining 20% of images, 78 in total, were used to test its diagnostic accuracy.

Table 5 shows the confusion table or classification accuracy table of the decision tree. A confusion table shows the labels of the true diagnoses in the

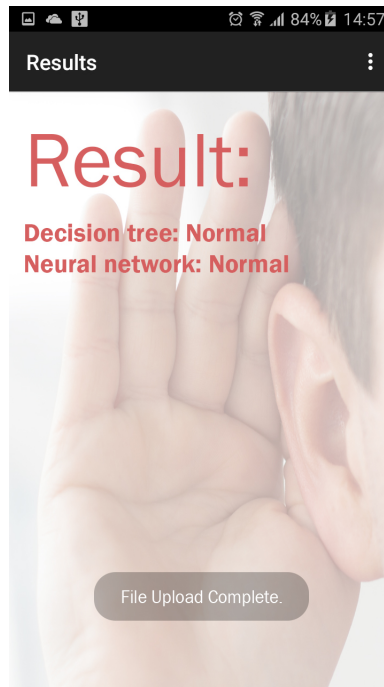Figure 23: Smartphone application a) loaded image and b) cropped image.

Figure 24: Smartphone application result page.

leftmost column, while the first row shows the decision tree's diagnoses labels. The diagonal shows the number of correct diagnoses, while the rest of the matrix

shows the misdiagnoses. The overall weighted diagnosis accuracy of the decision tree was 81.58%.

Table 5: Classification accuracy table using the decision tree for classification/diagnosis. The leftmost column shows the actual diagnoses, with the number of images in each category, while the topmost row shows the predicted diagnoses, with the percentage classifications/diagnoses for each category in its respective column. The diagonal shows the percentage correct diagnose

	Normal	CSOM	AOM	OME	W/O
Normal (25)	83.33	4.17	4.17	8.33	0.00
CSOM (17)	11.76	82.35	0.00	0.00	5.88
AOM (10)	0.00	0.00	80.00	20.00	0.00
OME (14)	7.69	0.00	7.69	84.62	0.00
W/O (12)	0.00	0.00	16.67	8.33	75.00

Table 6 shows the sensitivity, specificity, positive prediction value (PPV) and negative predictive value (NPV) of the system for all the diagnostic groups. The sensitivity ranged between 0.75 and 0.85 for all diagnostic groups. CSOM had the highest specificity (0.99) while AOM had the lowest (0.92). This highest PPV was that of CSOM (0.95), while OME had the lowest (0.70). The NPV ranged between 0.93 and 0.95.

Table 6: Sensitivity, specificity, PPV and NPV using the decision tree for classification/diagnosis.

	Normal	CSOM	AOM	OME	W/O
Sensitivity	0.83	0.82	0.80	0.85	0.75
Specificity	0.94	0.99	0.92	0.90	0.98
PPV	0.81	0.95	0.74	0.70	0.93
NPV	0.95	0.95	0.94	0.95	0.93

Neural Network

The neural network was designed as explained before, and different amounts of training images were used to evaluate its performance, while 80% of training images were still used to develop the feature extraction algorithms. The neural network was trained using 40% (155), 60% (233) and 80% (311) of the 389 prediagnosed tympanic membrane images respectively.

Table 7 to table 9 show the confusion tables of the neural network using 40%, 60% and 80% of the images respectively for training. For 40% training data, the combined weighted diagnostic accuracy is 49.79%. For 60% training data, the diagnostic accuracy is 62.66%. For 80% training data, the diagnostic accuracy is 86.84t%. It is clear that the neural network performs better with an increase in training data, which was expected.

Table 10 shows the sensitivity, specificity, PPV and NPV of the system for 80% training data. AOM had the highest sensitivity (0.90) while W/O had the lowest (0.83). Also, W/O had the highest specificity (1) while AOM had the

Table 7: Classification accuracy table using the neural network for classification/diagnosis with 40% training data. The leftmost column shows the actual diagnoses, with the number of images in each category, while the topmost row shows the predicted diagnoses, with the percentage classifications/diagnoses for each category in its respective column. The diagonal shows the percentage correct diagnoses.

	Normal	CSOM	AOM	OME	W/O
Normal (74)	51.35	27.03	9.46	6.76	5.41
CSOM (52)	17.31	57.69	3.85	7.69	13.46
AOM (31)	6.45	3.23	61.29	22.58	6.45
OME (42)	7.14	30.95	9.52	45.24	7.14
W/O (36)	19.44	16.67	11.11	22.22	30.56

Table 8: Classification accuracy table using the neural network for classification/diagnosis with 60% training data. The leftmost column shows the actual diagnoses, with the number of images in each category, while the topmost row shows the predicted diagnoses, with the percentage classifications/diagnoses for each category in its respective column. The diagonal shows the percentage correct diagnoses.

	Normal	CSOM	AOM	OME	W/O
Normal (50)	60.00	14.00	4.00	10.00	12.00
CSOM (35)	11.43	68.57	5.71	5.71	8.57
AOM (21)	4.76	0.00	71.43	19.05	4.76
OME (28)	3.57	10.71	17.86	60.71	7.14
W/O (24)	12.50	4.17	8.33	20.83	54.17

Table 9: Classification accuracy table using the neural network for classification/diagnosis with 80% training data. The leftmost column shows the actual diagnoses, with the number of images in each category, while the topmost row shows the predicted diagnoses, with the percentage classifications/diagnoses for each category in its respective column. The diagonal shows the percentage correct diagnoses.

	Normal	CSOM	AOM	OME	W/O
Normal (24)	87.50	0.00	4.17	8.33	0.00
CSOM (17)	11.76	88.24	0.00	0.00	0.00
AOM (10)	0.00	0.00	90.00	10.00	0.00
OME (13)	7.69	0.00	7.69	84.62	0.00
W/O (12)	0.00	8.33	8.33	0.00	83.33

lowest (0.94). The PPV and NPV ranged between 0.82 and 1, and 0.95 and 0.97 respectively.

Discussion and Conclusions

In this paper a smartphone- and cloud-based automated otitis media diagnosis system was proposed, using image processing and feature vector classification. This system is an improvement on a similar system recently developed in [7], but with the addition of a neural network used for classification of the

Table 10: Sensitivity, specificity, PPV and NPV using the neural network for classification/diagnosis.

	Normal	CSOM	AOM	OME	W/O
Sensitivity	0.88	0.88	0.90	0.85	0.83
Specificity	0.95	0.98	0.94	0.95	1.00
PPV	0.82	0.91	0.82	0.82	1.00
NPV	0.97	0.97	0.97	0.96	0.95

feature vectors. The proposed system can be operated from an Android smartphone using open source software, whereas the previous system could only be executed on a notebook computer under the Windows operating system using the Matlab software, which increased costs and hindered mobility. The proposed system allows the user to load a video-otoscope captured tympanic membrane image in an Android application, and preprocess the image on the smartphone, after which the preprocessed image is sent to a server, where feature extraction and classification/diagnosis is performed. The diagnosis is then returned to the phone and displays it to the user.

The performance of the automated otitis media diagnosis system is comparable to, if not better than, the diagnostic accuracy of medical specialists or otologists using handheld otoscopes. The average diagnostic accuracy of pediatricians is around 80% [6], that of GP's is $64\% - 75\%$ [29], and that of ENT's 73% [30]. The decision tree classification performance using 80% of the image for training was 81.58, while it was 86.84% when the neural network was used. Although only 311 tympanic membrane images were used for training - 80% of all available images, a fairly low number for these kinds of problems - the performance was still good. It is in principle possible then to use many hundreds or even thousands of images to improve the diagnostic accuracy further.

Although the proposed system is not complete in the sense that the smartphone completely replaces the need for a video-otoscope, it is a major step forward in creating a complete smartphone-based automated otitis media diagnosis system. Future developments will include a low cost extension, which would enable the smartphone to be used to capture high quality tympanic membrane images using either the smartphone camera or an external low cost custom made USB otoscope, after which the proposed system can be used to diagnose the otitis media contained in the images.

Declaration of interest

There is no conflict of interest to declare by any of the authors.

Financial support

No financial support was obtained for the conduct of the research or for the preparation of the manuscript.

References

[1] P. O. Eriksson, C. Mattsson, S. Hellström, First forty-eight hours of developing otitis media: an experimental study, Annals of Otology, Rhinology & Laryngology 112 (6) (2003) 558–566.

[2] C. D. Bluestone, G. A. Gates, J. O. Klein, D. J. Lim, G. Mogi, P. L. Ogra, M. M. Paparella, J. L. Paradise, M. Tos, 1. definitions, terminology, and classification of otitis media, Annals of Otology, Rhinology & Laryngology 111 (3_suppl) (2002) 8–18.

[3] J. Acuin, et al., Chronic suppurative otitis media: Burden of illness and management options.

[4] T. Ibkwe, Otitis media-focusing on the developing world, Irrua Specialist Teaching Hospital, Division of ENT Surgery, Irrua Nigeria.

[5] U. DESA, United nations department of economic and social affairs/population division (2009b): World population prospects: The 2008 revision, Internet: http://esa. un. org/unpp (gelesen am 16.

[6] A. Kuruvilla, N. Shaikh, A. Hoberman, J. Kovačević, Automated diagnosis of otitis media: Vocabulary and Grammar, Journal of Biomedical Imaging 2013 (2013) 27.

[7] H. C. Myburgh, W. H. van Zijl, D. Swanepoel, S. Hellström, C. Laurent, Otitis media diagnosis for developing countries using tympanic membrane image-analysis, EBioMedicine 5 (2016) 156–160.

[8] J. Heggestuen, One in every 5 people in the world own a smartphone, one in every 17 own a tablet (2015).
URL http://www.businessinsider.com/smartphone-and-tablet-penetration-2013-10

[9] I. lunden, 6.1b smartphone users globally by 2020, overtaking basic fixed phone subscriptions (2015).
URL http://techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/

[10] GSMA, New GSMA report forecasts half a billion mobile subscribers in sub-Saharan Africa by 2020 (2014).
URL http://www.gsma.com/newsroom/press-release/gsma-report-forecasts-half-a-billion-mobile-subscribers-ssa-2020/

[11] I. N. Africa, Mobile health solutions enabling efficient healthcare (2015).
URL http://www.itnewsafrica.com/2015/08/mobile-health-solutions-enabling-efficient-healthcare/

[12] H. M. Team, Bodymaps: Tympanic membrane (2015).
URL http://www.healthline.com/human-body-maps/tympanic-membrane

[13] E. Britannica, Eardrum membrane (2015).
URL `http://global.britannica.com/science/tympanic-membrane`

[14] K. S. Audiology, Hearing loss (2016).
URL `http://karenscottaudiology.com/hearing-loss`

[15] D. Davitt, Tympanic membrane (2015).
URL `http://www2.webster.edu/~davittdc/ear/tympanic/tympanic.htm`

[16] L. Hawk, Otitis media guide (2015).
URL `http://otitismedia.hawkelibrary.com/normal/1_G`

[17] C. Limb, L. Lustig, J. Klein, Acute otitis media in adults (suppurative and serous) (2014).
URL `http://www.uptodate.com/contents/acute-otitis-media-in-adults-suppurative-and-serous`

[18] L. AlMaghrabi, Otitis Media(OM), King Saud Medical City.

[19] WebMD, Earwax (2015).
URL `http://www.webmd.com/a-to-z-guides/earwax-topic-overview`

[20] H. insite, Middle ear infection and grommets (2015).
URL `http://www.mydr.com.au/kids-teens-health/middle-ear-infection-and-grommets`

[21] A. Bankaitis, Otoscopes (2011).
URL `http://www.audiologyonline.com/articles/otoscopes-833, urldate={}`

[22] J. L. Pech-Pacheco, G. Cristóbal, J. Chamorro-Martinez, J. Fernández-Valdivia, Diatom autofocusing in brightfield microscopy: a comparative study, in: Pattern Recognition, 2000. Proceedings. 15th International Conference on, Vol. 3, IEEE, 2000, pp. 314–317.

[23] O. development team, Canny Edge Detector (2016).
URL `http://docs.opencv.org/2.4/doc/tutorials/imgproc/imgtrans/canny_detector/canny_detector.html`

[24] O. development team, Miscellaneous Image Transformations (2016).
URL `http://docs.opencv.org/2.4/modules/imgproc/doc/miscellaneous_transformations.html`

[25] O. development team, Hough Circle Transform (2016).
URL `http://docs.opencv.org/2.4/doc/tutorials/imgproc/imgtrans/hough_circle/hough_circle.html`

[26] C. Özkan, F. S. Erbek, The comparison of activation functions for multispectral landsat tm image classification, Photogrammetric Engineering & Remote Sensing 69 (11) (2003) 1225–1234.

[27] S. J. Russell, P. Norvig, Artificial intelligence: a modern approach (international edition).

[28] A. Ng, Sparse autoencoder, Department of Computer Science and Engineering.

[29] K. Blomgren, A. Pitkäranta, Is it possible to diagnose acute otitis media accurately in primary health care?, Family practice 20 (5) (2003) 524–527.

[30] M. E. Pichichero, M. D. Poole, Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media, Archives of pediatrics & adolescent medicine 155 (10) (2001) 1137–1142.
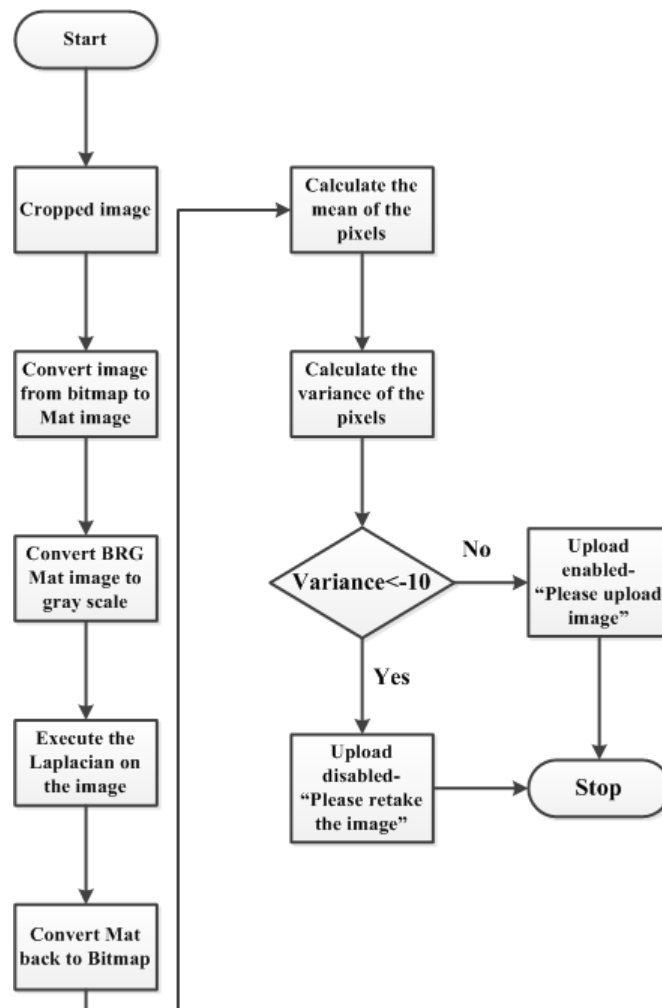
Figure 25: Blur detection process flow diagram.
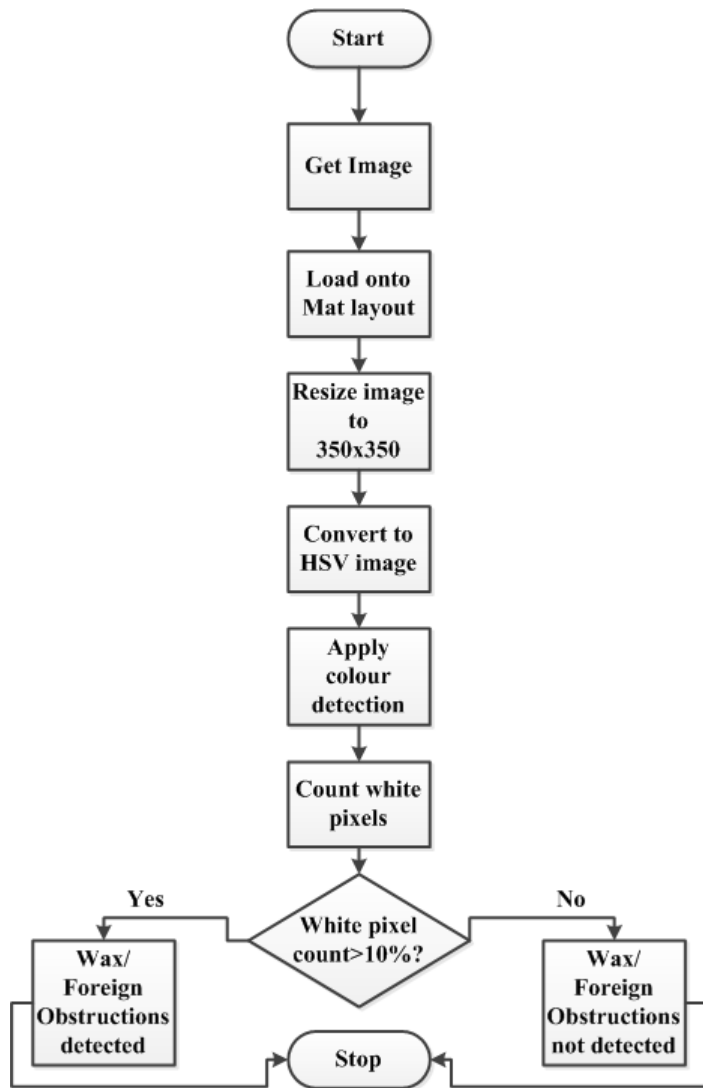
Figure 26: Wax and foreign bodies obstruction color detection process flow diagram.
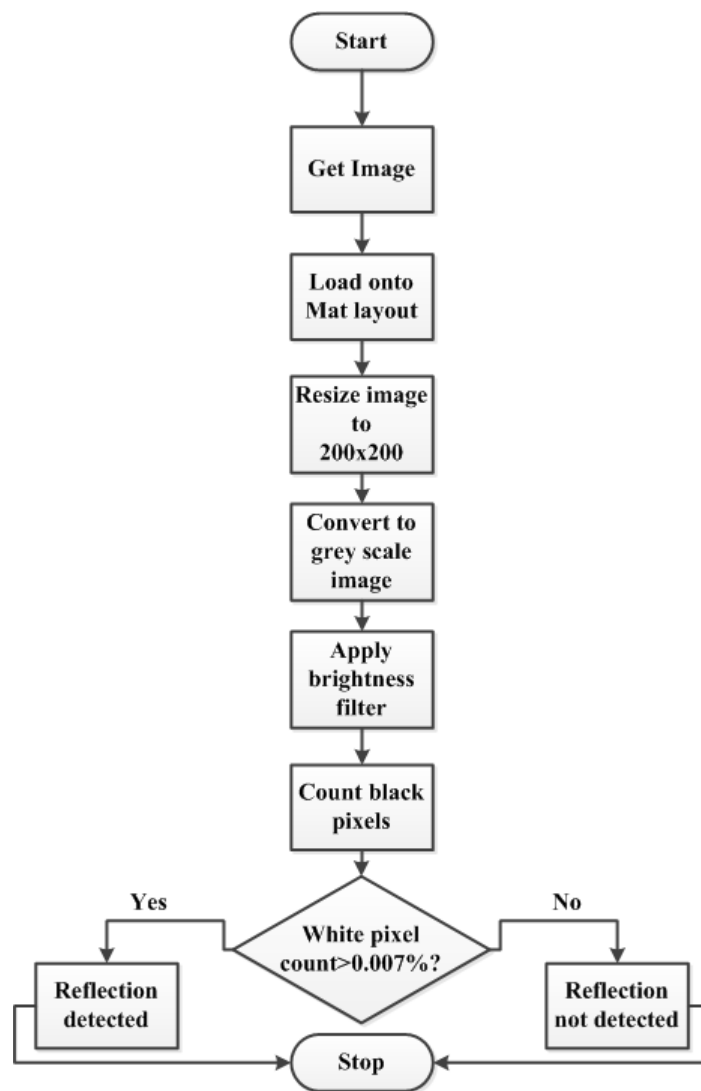
Figure 27: Reflection color detection process flow diagram.
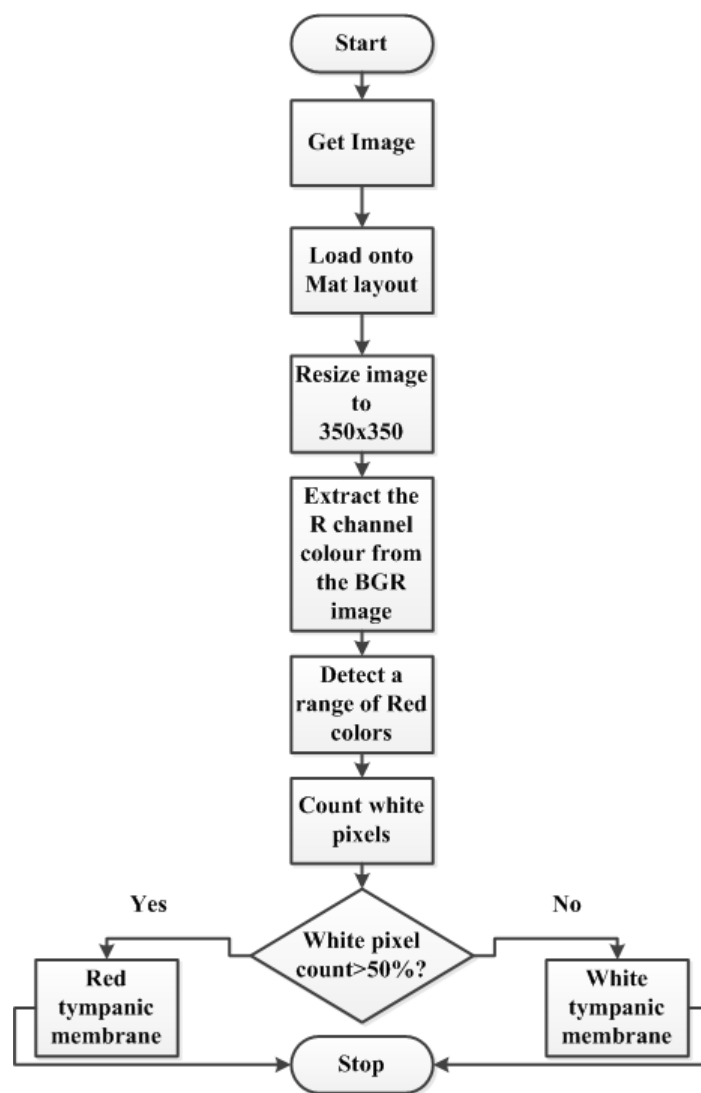
Figure 28: Tympanic membrane color detection process flow diagram.
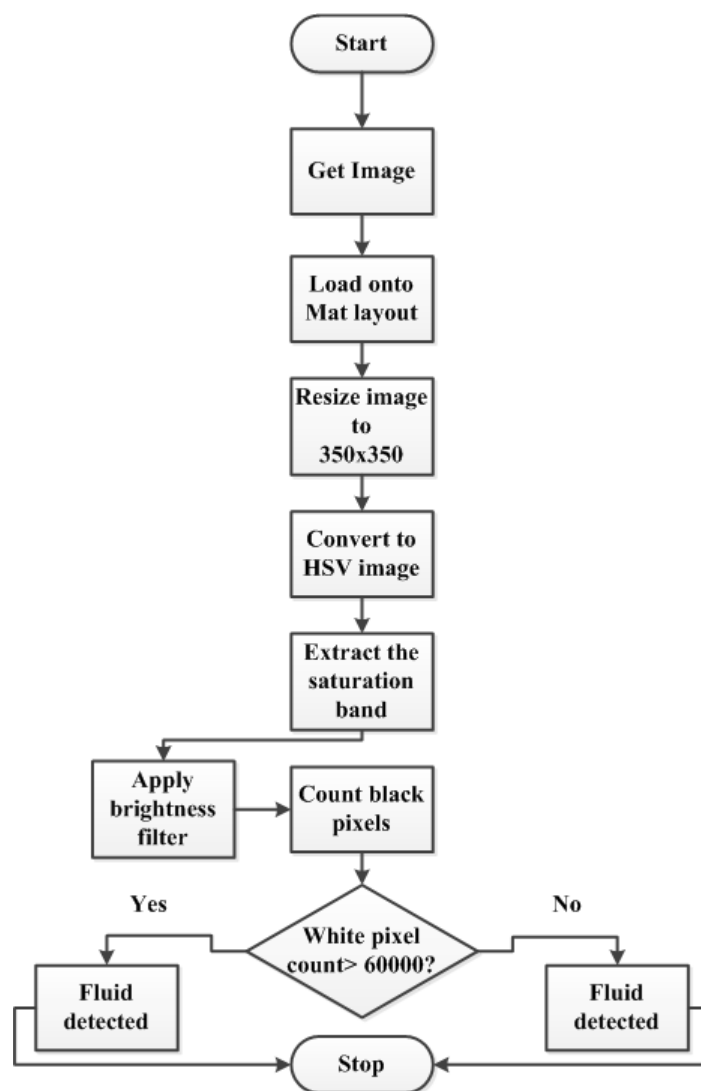
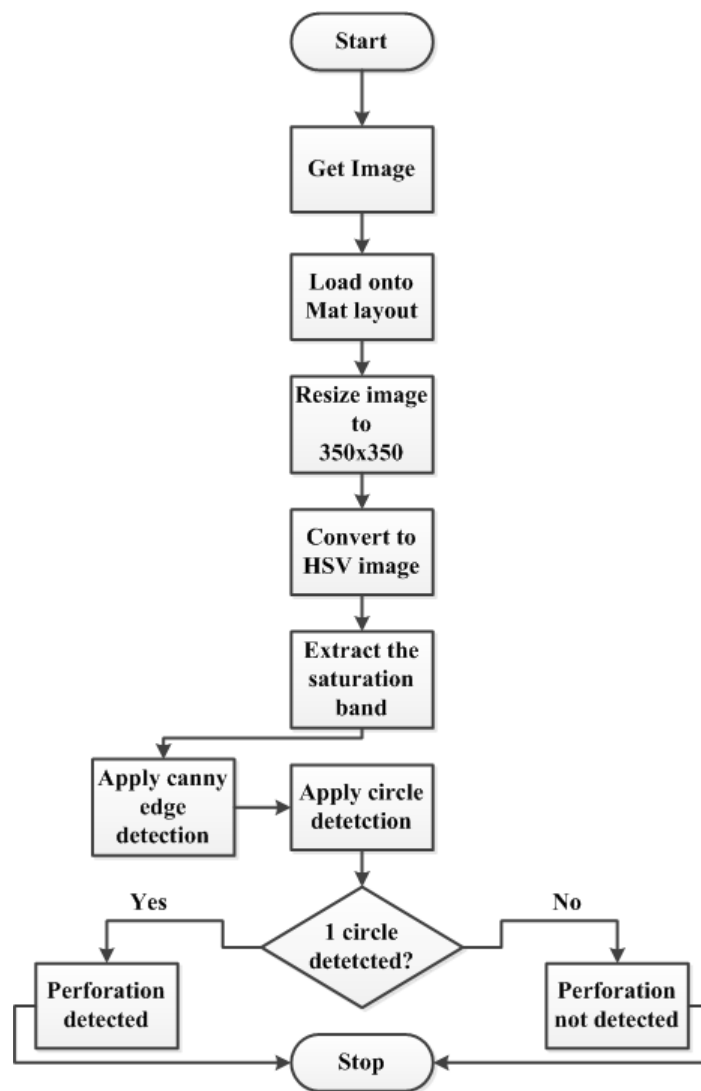Figure 29: Fluid color detection process flow diagram.
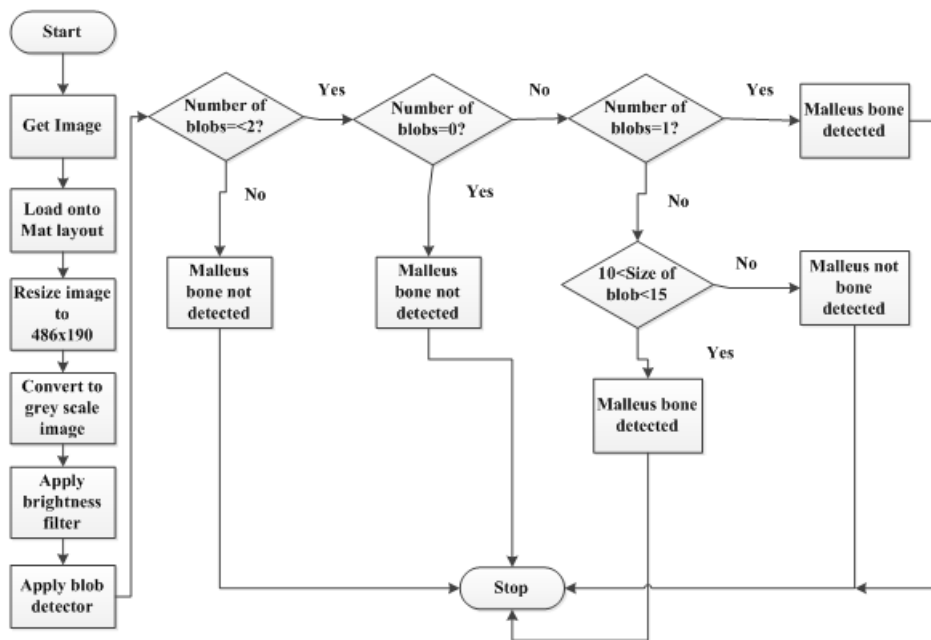
Figure 30: Edge detection process flow diagram.

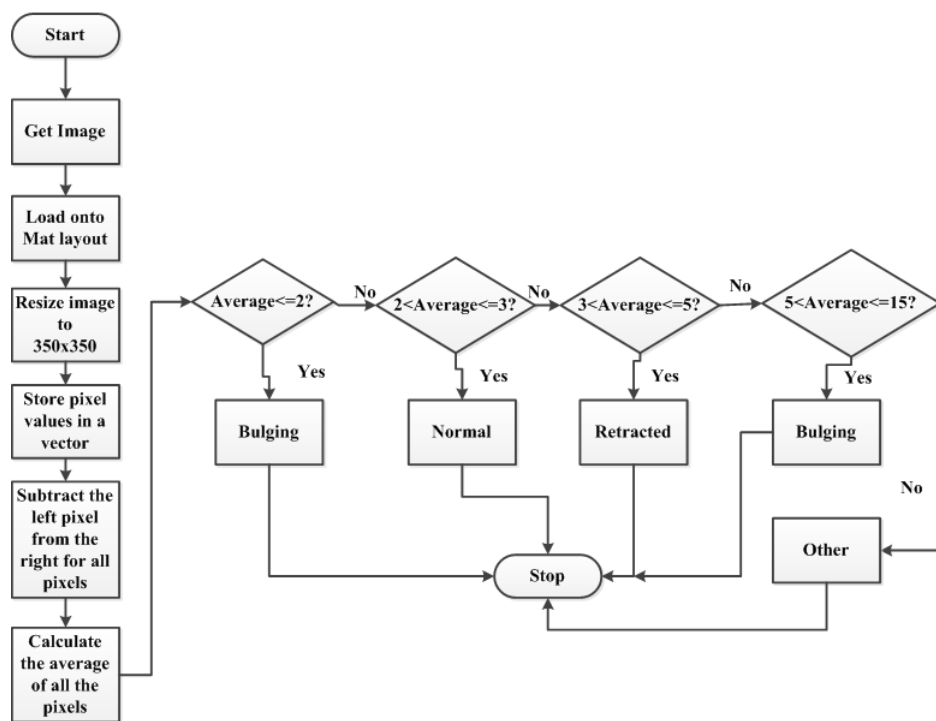Figure 31: Blob detection process flow diagram.

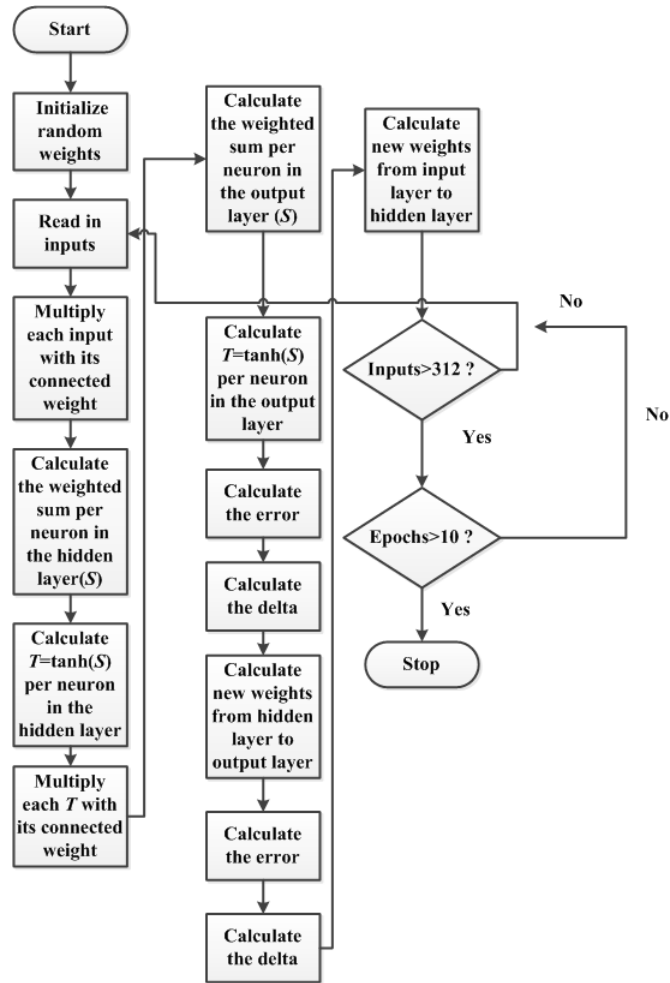Figure 32: Shape detection process flow diagram.
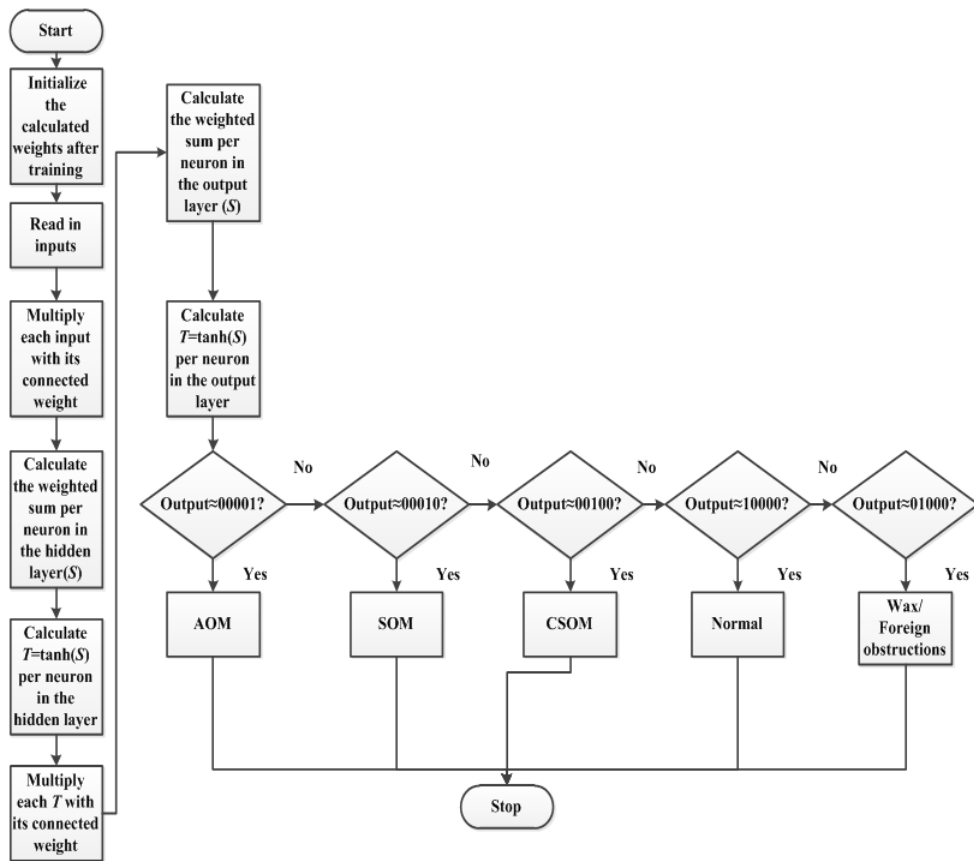
Figure 33: Neural network training process flow chart.

Figure 34: Neural network classification process flow diagram.