Supplementary materials

Supplementary material 1: Image processing (pre-processing and feature extraction)

Before any features can be extracted from the images, a number of pre-processing routines have to be performed on the image. This is done in order to standardize the images for feature extraction.

Pre-Processing

The first pre-processing technique is to remove the black sides in the image. This is done through the cropping function. Cropping involves searching for the first non-black pixel from each side and marks it as the cropping index. The image is then cropped by keeping only the part of the image that lies within the boundaries of the cropping indices. Figure 1 shows the boundary lines.

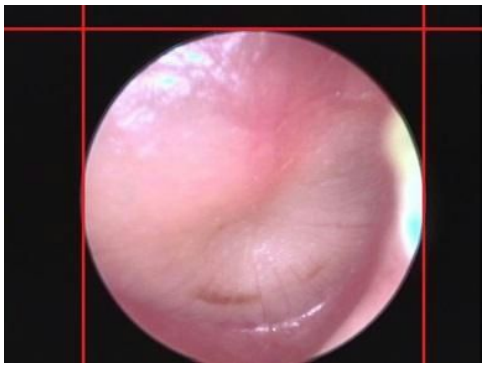


Figure 1. Image showing lines where the figure will be cropped, (image altered from Laurent, 2014)

The next pre-processing technique to be applied is the size standardization of all the images. After analyzing all the images, it was found that no image was smaller than 500 x 500 pixels. Through a random equal number estimation, the normal size that all images should be resized to was chosen as 486 x 486 pixels.

Feature Extraction

A number of methods were developed in order to extract the features in table 1 (main article). These feature extraction methods are discussed below.

a) Main Color Analysis

Because different symptoms exhibit different colors, color analysis of the entire image is performed to evaluate the color of each diagnostic category. It is performed by extracting the red, green and blue (RGB) color bands, as seen in figure 3, from the image in figure 2. Each extracted band has the same dimension as the original image (486 x 486 pixels) with each pixel represented by a single value between 0 and 255, hence the grey appearance. After RGB extraction, the average value of all the pixels in each color band is calculated. The average mean

for each of the RGB channels, as well as difference between each mean is calculated. Figure 4 a) to e) show the colors resulting from the averaged RGB channels for each diagnosis.

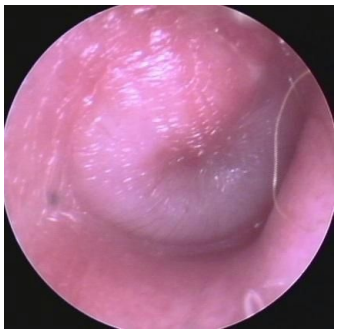
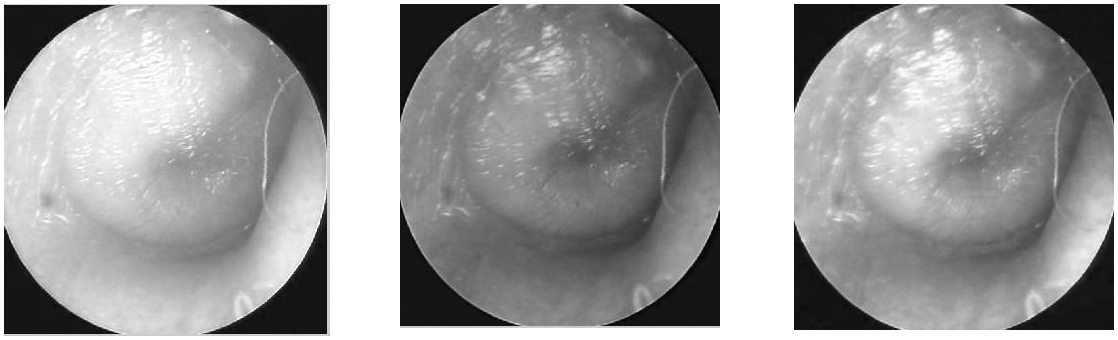
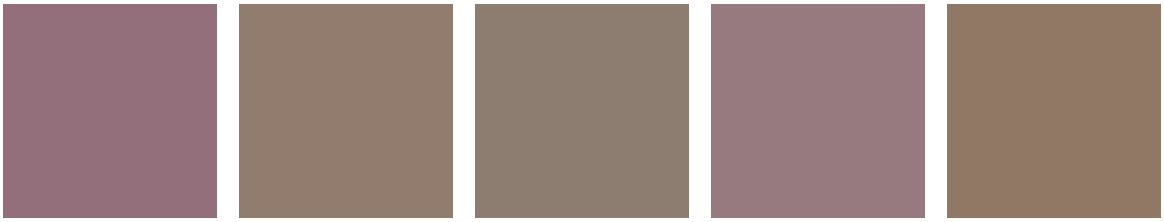


Figure 2. Original RGB AOM image (Laurent, 2014)



a) b) c)

Figure 3 a) Extracted red band; b) Extracted green band; c) Extracted blue band



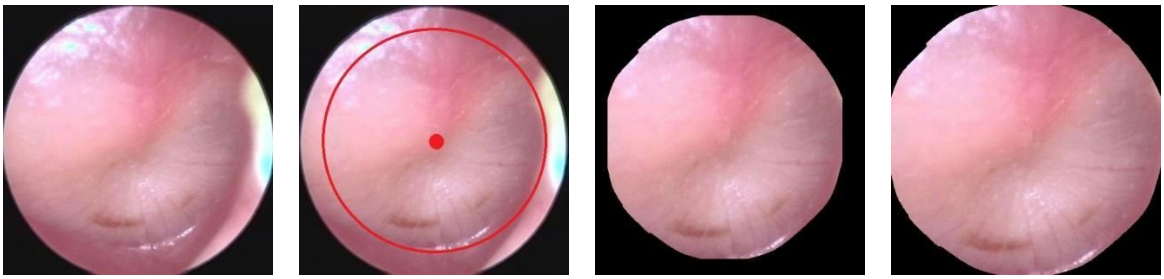
a) **b)** **c)** **d)** **e)**

Figure 4 a) Average color of AOM; b) Average color of CSOM with perforation; c) Average color of Normal tympanic membrane; d) Average color of OME e) Average color of O/W

b) Tympanic Membrane (TM) Color Analysis

In order to determine whether the TM is opaque (normal) or red (inflamed) the image must first be cropped (figure 5 a) so as to use only the TM part of the image. In order to crop this image the center of the image must be calculated (figure 5 b) . The image is then cropped to a circle with a radius of 200 pixels (figure 5 c). This results in a black band of 43 pixels on all four sides of the image, which is subsequently removed using the cropping function explained in the pre-processing section (figure 4 d).

A red threshold is applied to the image in order to filter out all non-red colored pixels. From this image the percentage red can be determined. This procedure is followed for all the images to define an abnormal red percentage threshold in the TM. A threshold of 17% for the allowable amount of red in an image was chosen based on the analysis.



a) b) c) d)

Figure 5 a) Original Image; b) Center and radius of the image shown; c) Image cropped; d) Image with the black sides removed (Laurent, 2014)

d) Wax Detection

In order to determine the amount of wax in the ear canal, a color filter is applied. This is done by extracting the hue, saturation and value (HSV) bands from the image and applying a yellow threshold to each band. This filter will only allow all the yellow parts of the image to be visible, while all the non-yellow parts are filtered out of the image in figure 6, as shown in figure 7. This makes it possible to determine the percentage wax in the image.

Analysis was performed on all the images to define an abnormal threshold of wax in the external ear canal. A threshold of 12% for the allowable amount of wax in an image was identified.

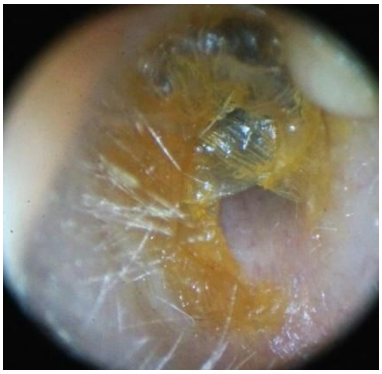
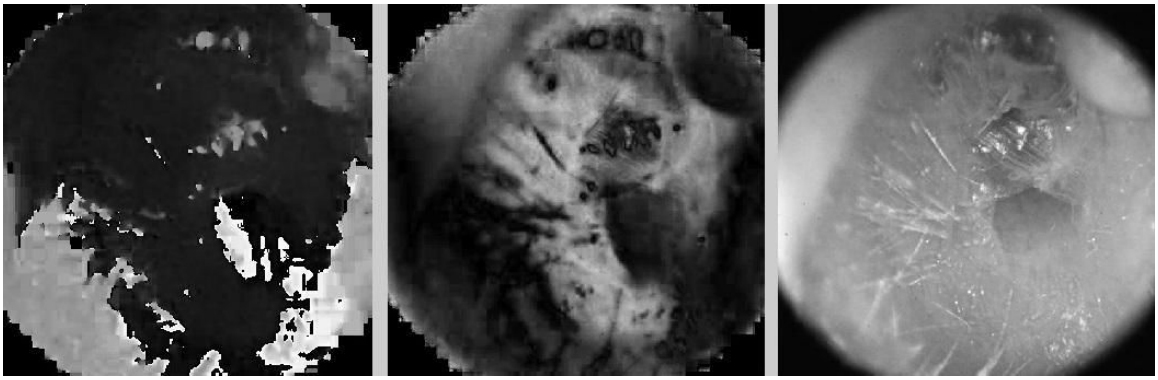


Figure 6. Original RGB wax obscured ear canal (Laurent, 2014)



a) b) c)

Figure 7 a) H band extracted from image; b) S band extracted from image; c) V band extracted from image

e) Obstructive Wax Detection

An image has obstructive wax when visibility of the TM is obscured. The same procedure to crop out the TM, as explained for wax detection, is followed. The cropped image is then used in the same manner as mentioned above to determine if there is an abnormal amount of wax in the image. The resulting filtered image is shown in figure 8. The allowable wax threshold was chosen to be less than 20%.

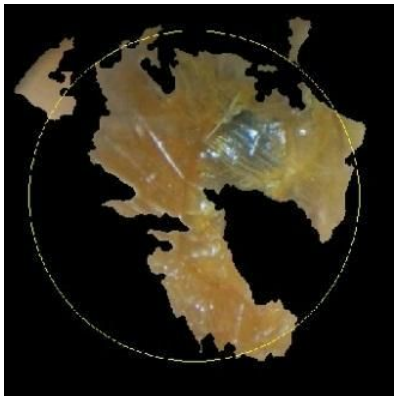


Figure 8. An example of effect of obstructive wax detection, image altered from (Laurent, 2014)

f) Perforation Detection

A perforation in the TM is seen as a hole in the ear drum, with clear edges. Edge detection is therefore used to detect if there is a perforation. It was found that most images, when used in its original form, exhibited too many anomalies and interferences when edge detection was applied. The anomalies and interferences were filtered out by extracting the saturation band (in set of HSV bands) from the image, and performing edge detection on this band only. If a circular shape is detected, the image will be classified as having a perforation, as seen in figure 9 and figure 10.

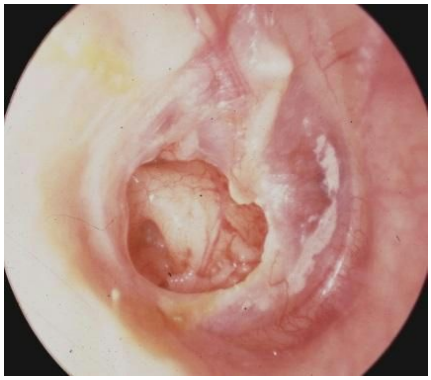
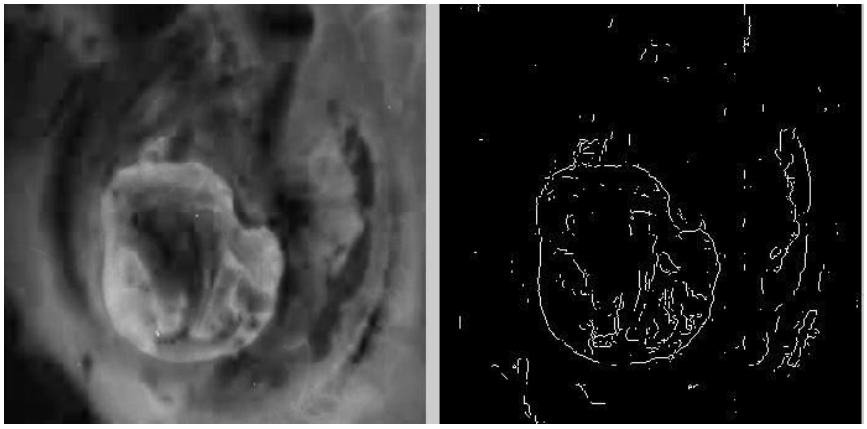


Figure 9. Original RGB CSOM with perforation image (Laurent, 2014)



a) b)

Figure 10 a) Saturation band extracted from the image; b) Edge detection applied to the saturation band

g) Fluid Detection

Fluid presents itself as either a “white mass” or as bubbles behind the TM. In order to determine the presence of fluid behind the TM, the saturation band is again extracted from the image. A brightness filter applied to the saturation band will indicate if there is a “white mass” behind the TM, making it possible to calculate the amount of white in the image. Analysis was performed on the images that had a non-zero white score, to determine the threshold of an abnormal percentage white in the image. The threshold was set at 70000 pixels. Another method for determining fluid behind the tympanic membrane is to use edge detection on the image. If a collection of small circles is detected in the TM, the middle ear most likely contains fluid.

h) Shape Detection

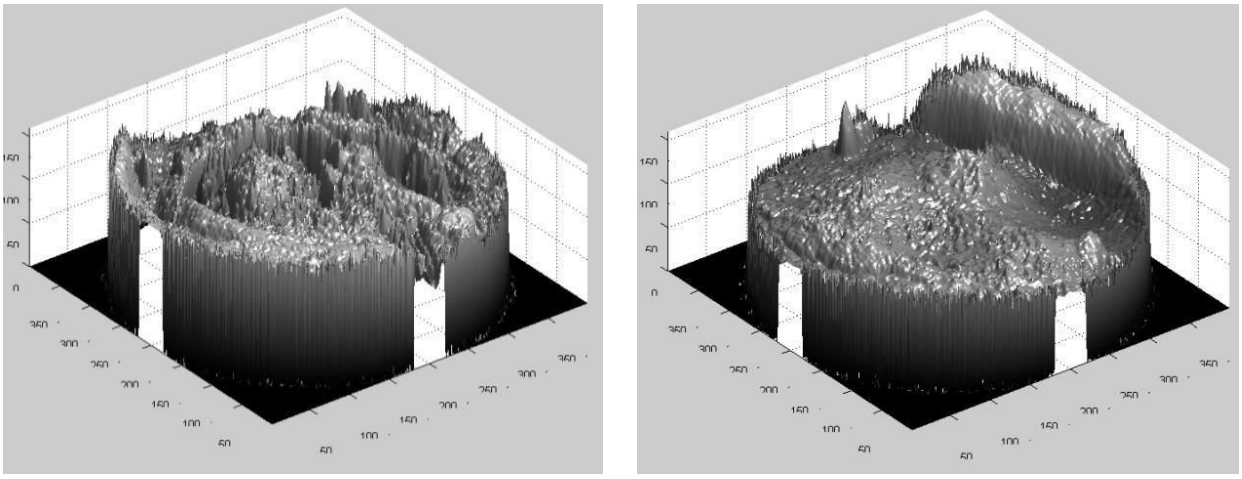
In order to categorize the shape of the TM as retracted, normal, bulging or irregular, a norm first needs to be established. Cropping is done on the image as described for TM color analysis. The image is then converted to a grey scale image, where after the difference in greyness between neighboring pixels is calculated and noted as the gradient of that pixel. This is done by comparing each individual pixel with the one to the right of it. Figure 11 shows the result of

gradient analysis, displaying the image gradients in 3D. After the entire image is completely analyzed, the mean of these gradient values is calculated.

Analysis was performed in order to find the most accurate range for each category (retracted, normal, bulging or irregular). These results can be seen in table S1-1.

Table S1-1. The ranges for each shape category shown

Category	Range
Retracted	$x \leq 9$
Normal	$9 < x \leq 11$
Bulging	$11 < x < 15$
Irregular	$x \geq 15$

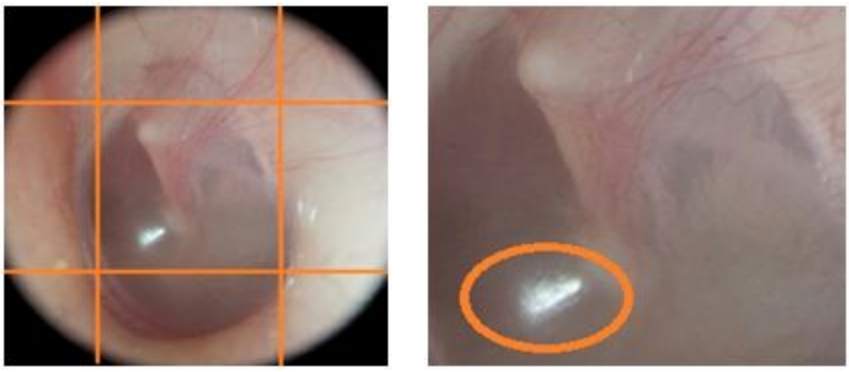


a) b)

Figure 11 a) An image of CSOM with perforation showing the gradient analysis; b) A normal TM image showing the gradient analysis

i) Reflection Detection

The reflection in the image is caused by the reflection of the light from the otoscope on the surface of the TM. Reflections in the normal category images appear close to the center of the image, which necessitates cropping of the image. In fact, the reflection never appears within 30% from the sides of the image as shown in figure 12 a. A brightness filter is applied to the image, allowing only white and near white pixels to be visible in the filtered image. This makes it possible to determine if there is a reflection in the image (figure 12 b). Analyses were performed by calculating the percentage of white in each image, in order to determine the threshold to detect whether the cropped and filtered image contained a reflection. From the analysis it was established that a small range is required to ensure that the diagnosis is accurate. The range was chosen as between 1 and 9 percent of white in the cropped image, due to the fact that most of the normal diagnosed images fell into this range.



a) b)

Figure 12 a) Image showing the boundaries for a reflection on a normal TM; b) Reflection shown in the cropped image, (images altered from Laurent, 2014).

j) Malleus Handle Visibility Detection

A brightness filter, similar to the one described for reflection detection, is applied in order to detect whether the malleus handle is visible or not. After the filter is applied, circle detection is performed to detect whether there are any bright circular areas in the remaining image. Moving circles detect the circular brightness with different radii across the image to determine if it correlates with similar shapes in the image. The circle detection only allows one bright circle to be detected. If multiple circles are detected, however the image will not be classified as having a visible malleus handle. This is done to avoid overlapping features between the malleus bone and fluid.

Supplementary material 2: Classification methods including the decision tree structure

Artificial intelligence is defined by Poole and co-workers[1] as the design and study of intelligent agents. In order to diagnose the various forms of otitis media present in the images, a suitable artificial intelligence technique, namely a decision tree was used to classify the images as belonging to one of these diagnostic groups. In order for the decision tree to make an accurate diagnosis, the predefined diagnostic features associated with each diagnosis had to be accurately identified in the images by the feature extraction methods described in Supplementary material 1.

To classify an image of unknown diagnosis, the image is provided as input to the image-analyzing classification system, where after pre-processing and feature extraction are performed. Once all the features are extracted, the image can be classified using the feature vector of that image. The output, which consists of the extracted features as well as the final diagnosis, is then presented to the user as an output window on the notebook screen.

Training example feature vector generation

Before an undiagnosed image can be classified, the decision tree needs to be constructed from the pre-diagnosed example feature vectors obtained from the images. In order to construct a decision tree, example feature vectors with their corresponding diagnosis have to be created from the diagnosed images. For each image in the example training set, the image processing algorithms receive a diagnosed image as input, and run it through the pre-processing and feature extraction processes. After the image processing algorithms are complete, an array containing each image's symptoms (or features) with the corresponding known diagnosis is produced as output. This process is depicted in figure 1. The feature vectors produced by the image processing algorithms (one feature vector per image) and the corresponding known diagnoses are used as "examples" to construct the decision tree with. The more examples available, the more accurate the decision tree will be able to represent the problem space, implying a more accurate classifier and hence giving a more reliable diagnosis.

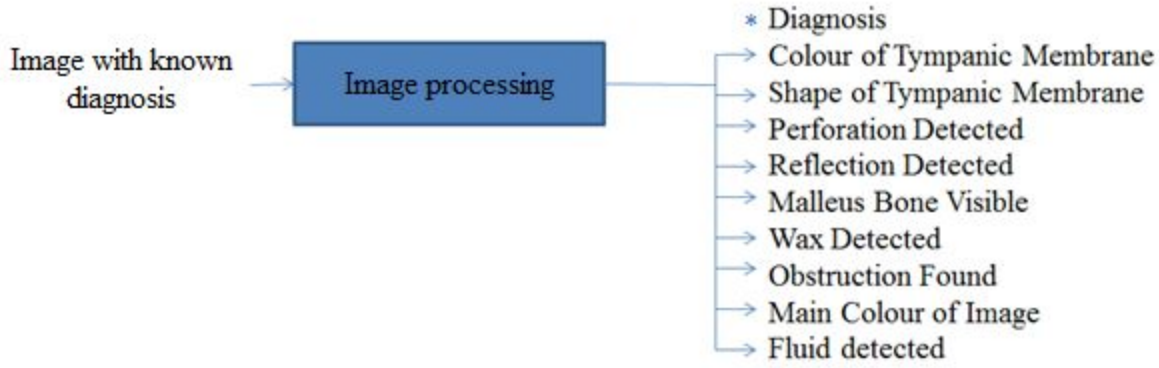


Figure 1. Training feature vector generation

Decision Tree Construction

The decision tree was constructed by continuously dividing the input variables (feature vectors), or examples, produced by the feature extraction algorithms, based on an attribute value test according to Quinlan[12]. This division of examples ensures that the examples that are the least like the feature will eventually be eliminated in order to ensure that the best fit is found. The most common attribute value test is one where the entropy of each attribute is calculated over all examples in the training data set.

The entropy is a measure of uncertainty, while one minus the entropy is a measure of certainty, or potential information gain. Therefore, if the entropy of an attribute is low, the potential information gain from evaluating that attribute is high. It therefore makes good sense that attributes exhibiting lower entropy (less uncertainty, higher information gain) should be evaluated first, followed by the attribute with the second lowest entropy, and so forth (Norvig and Russell[2]). This will ensure that classification is performed in the least possible time, as no unnecessary paths in the decision trees will be explored.

Calculating the entropy of each attribute and ranking the resulting entropies in increasing order is required to determine the order of attribute evaluation and node expansion. The attribute exhibiting the lowest entropy will be placed at the top of the decision tree, followed by the second, and so forth. The decision tree is populated from the top down with the nodes representing attributes (in this case the features produced by the training algorithm for each processed image). Each attribute may have multiple states or discrete values, while continuous variable attributes have multiple predefined ranges. A decision tree creates new branches by evaluating, or "splitting" on the attribute where there is the least uncertainty, or that will lead to the greatest loss of entropy (or the largest information gain), thereby eliminating irrelevant training examples from the training set. This process is repeated until either all the nodes are split or when the splitting of nodes becomes irrelevant to the diagnosis. Following this procedure will ensure a compact decision tree with a low average branching factor, i.e. the average number of branches created by evaluating features using the attribute value test, which in turn will result in low computational complexity, ultimately ensuring classification in the least possible time. Figure 2 shows the resulting decision tree.

The entropy $H(\mathbf{f})$ of a particular feature **f** is calculated by

$$H(f) = \sum_{j=1}^{J} \left(\left(\frac{N^{(j)}}{N} \right) \left(-\sum_{i=1}^{I} \frac{n_i^{(j)}}{N^{(j)}} \log_2 \left(\frac{n_i^{(j)}}{N^{(j)}} \right) \right) \right)$$

where N is the total number of examples in the training set, J is the number of discrete values (or ranges) that the feature can assume (of fall in), $N^{(j)}$ is the total number of examples associated with a given discrete value (or range), I is the total number of classes (or diagnoses), and $n_i^{(j)}$ is the number of examples classified as belonging to class i where the feature is associated with discrete value or range j from a total number of J ranges.

Once the decision tree has been constructed using the training example feature vectors plus their corresponding diagnosis (figure 1), it can be used to classify or diagnose unknown images (images that have not been part of the training set). As was done during the example feature vector generation, the image processing algorithms (pre-processing and feature extraction) are used to analyze the input image, and produces a feature vector as before, but now without a known diagnosis. This feature vector is then used to traverse the decision tree from the top, by testing the feature or attribute values against the corresponding values in the decision tree. Following this process iteratively will result in one unique path through the decision tree, which will terminate at one unique diagnosis, depicted by the green nodes in the decision tree in figure 2. Classification is complete once a diagnosis is determined, i.e. when there is only one obvious diagnosis to choose from according to Norvig and Russell[2].

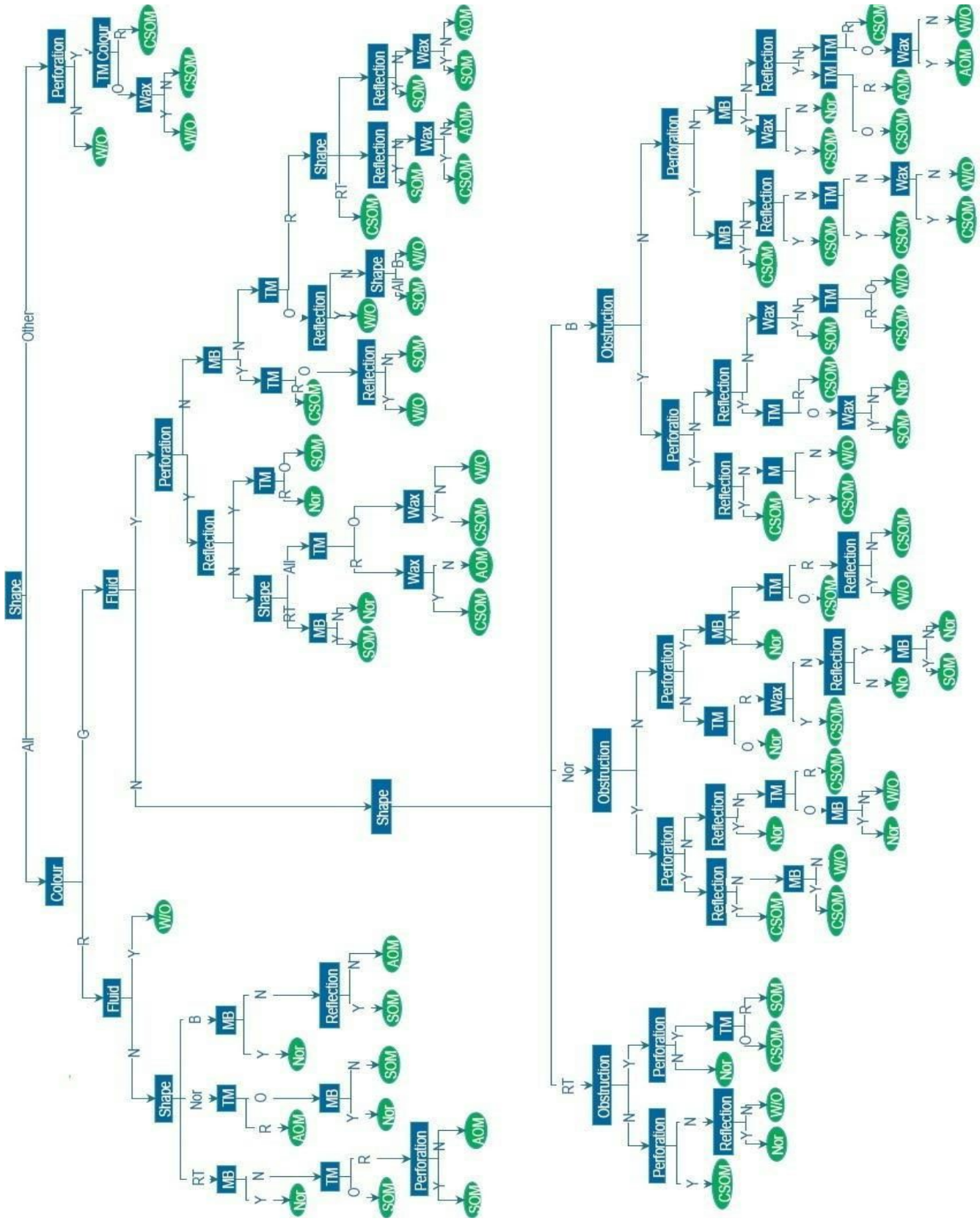


Figure 2. Decision tree

Supplementary material 3: The custom-made video-otoscope

As part of this work, a low cost video-otoscope was designed as a possible replacement for current, rather expensive video-otoscopes available on the market. The illumination circuitry was simplified to only a battery, a 5mm light emitting diode (LED) (maximum current of 100 mA; typical operating current of 20-30 mA, forward voltage of ~3.2 V) and a resistor. A number of resistors were experimented with to find the correct light intensity. With a 100 Ohm resistor the image was very light and the tympanic membrane was difficult to interpret ((9-3.2)/100=58 mA). With a 3 kOhm resistor the image was very dark and the tympanic membrane could not be seen ((9-3.2)/3000=1.93 mA). Using a 1 kOhm resistor the image was correctly displayed ((9-3.2)/1000=5.8 mA). The designed circuit can be seen in figure 1, below.

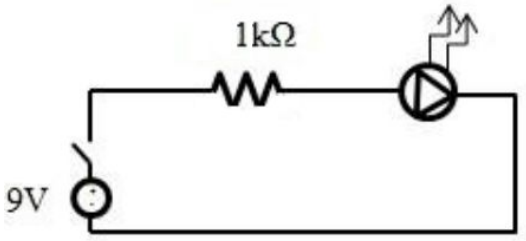


Figure 1. Illumination circuit for the otoscope

The current in the circuit is very small (5.8 mA) and poses no possible harm to a human. Since the video-otoscope was custom-designed, several experiments had to be performed and adjustments had to be made in order to get the best quality image and to ease the process of image pre-processing and feature extraction. No lens was used to save cost. The camera was a SEN-12804 JPEG 2 megapixel color camera with a serial transistor-transistor logic TTL interface, with a resolution of 1600x1200 and a frame rate of 15 frames per second (fps). The price of only the camera was 895 Rand ($54).

USB Communication

The USB communication was implemented to communicate from the custom-made video-otoscope to the notebook computer in the form of an AV-to-USB converter. The camera's output was connected to the audio/video (AV) cable of the converter, while the USB output of the converter was connected to the computer. This converter can be seen in figure 2 below.

Figure 2. AV-to-USB converter cable

The video-otoscope shell

The shell of a commercial ear-cleaning product was used as the body of the otoscope. The package of this ear cleaner can be seen in figure 3.

Figure 3. VAClear ear cleaning product obtained from a local "Dischem store" (South Africa)

The front part of the custom-designed video-otoscope was removed to make space for the camera and disposable ear tips. Figure 4 shows the shell of the video-otoscope together with its components. The three state switch that was obtained from the ear cleaner was reused as the on- and -off switch for the LED.

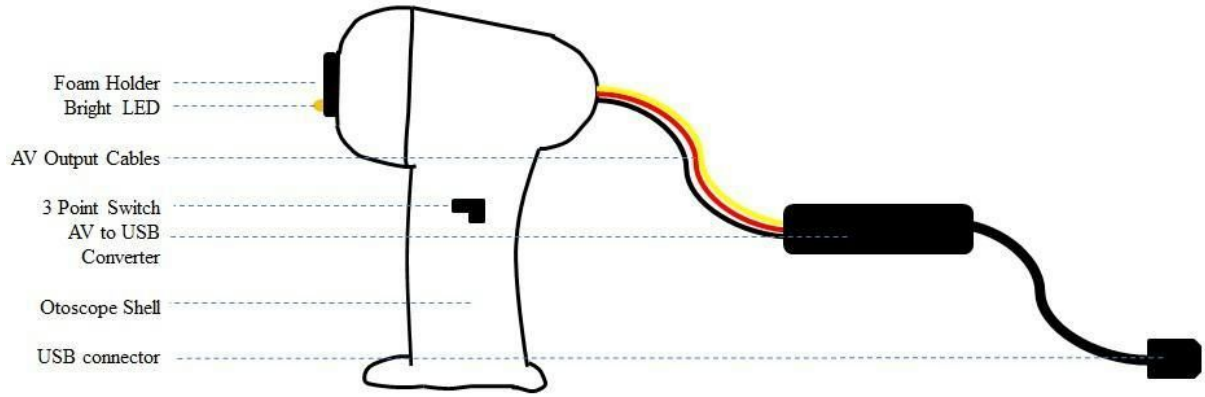


Figure 4. The exterior shell of the video-otoscope

A video-otoscope may carry germs from one person's ear to another. In order to avoid spreading of infections, a disposable ear tip was attached. The disposable ear tip was not only used to prevent contamination, but also to get a better view inside the ear canal. The tip was connected to the video-otoscope with the use of a rubber packing glued onto the front opening. The packing was cut to the exact inside diameter of the disposable ear tip. Another circle was then cut into the washer to make space for the camera. The packing tightly holds the disposable ear tip to prevent it from falling off while capturing images of a patient's ear. The disposable ear tip mounted to the shell is displayed in figure 5.

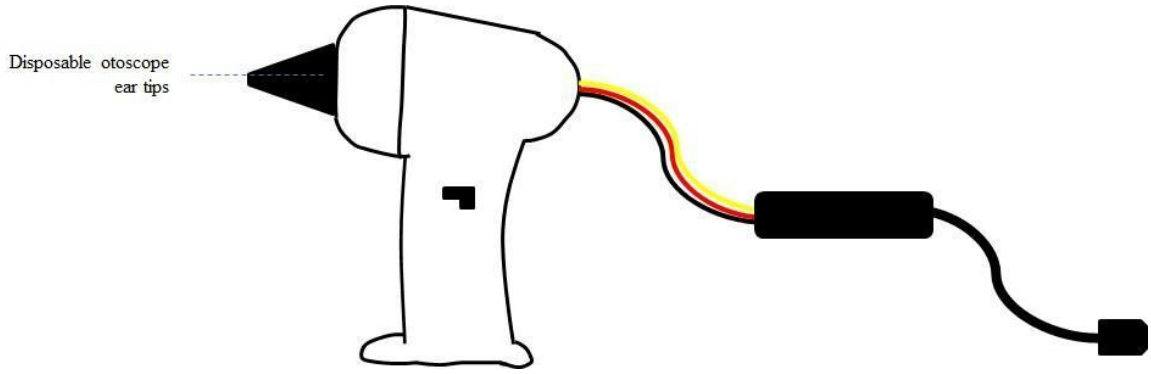


Figure 5. The exterior design of the video-otoscope showing the disposable ear tip.

Video-otoscope content

The video-otoscope was opened in order to determine the space available for all the circuitry. The circuitry was fixed into the otoscope. The 9V battery and bright LED that was fixed to the foam washer as well as the 100 kΩ resistor are shown in figure 6. The final product is shown in figure 7 and 8.

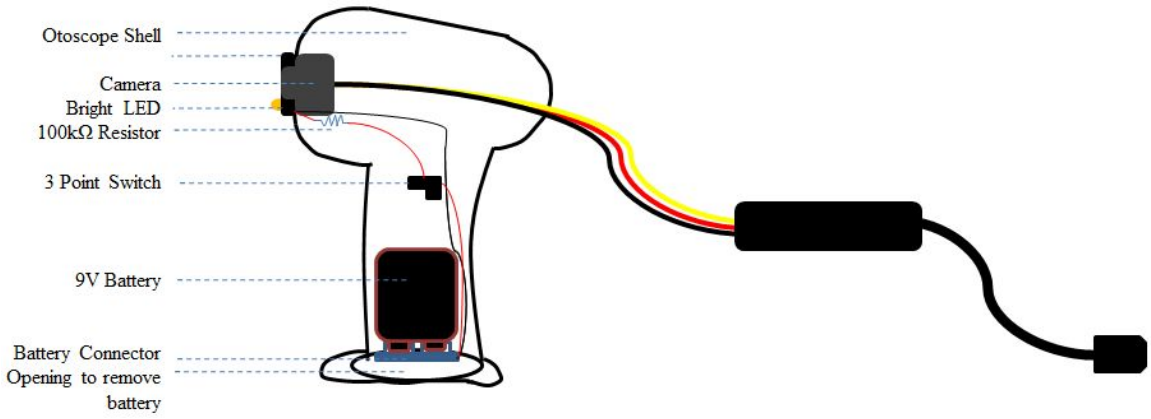


Figure 6. The content of the custom -made video-otoscope

Figure 7. Side view of the final product

Figure 8. Front view of the final product

The custom-made video-otoscope is able to allow the removal and replacement of batteries, as can be seen in figures 9 and 10. This can be done by opening and closing a battery holster made of a piece of Velcro® fabric.

Figure 9. The closed battery holster

Figure 10. The open battery holster

The custom-made video-otoscope captures an image and sends it to the notebook computer via the USB communication. Once the image is loaded into the computer it can be image-processed and classified by the image-analysis classification system as belonging to one of the five diagnostic groups.

References

1. Poole D, Mackworth A, Goebel R. Computational Intelligence - A Logical Approach, New York: Oxford University Press. 1997

2. Norvig P, Russell S. Artificial Intelligence: A Modern Approach. 3rd ed. New Jersey: Pearson. 2002