EBioMedicine 5 (2016) 156–160

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Otitis Media Diagnosis for Developing Countries Using Tympanic Membrane Image-Analysis

Hermanus C. Myburgh [a], Willemien H. van Zijl [a], DeWet Swanepoel [b,c,g], Sten Hellström [d,e], Claude Laurent [b,f,*]

[a] Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa
[b] Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa
[c] Ear Sciences Centre, School of Surgery, The University of Western Australia, Nedlands, Australia
[d] Department of Audiology and Neurotology, Karolinska University Hospital, Stockholm, Sweden
[e] CLINTEC/Otorhinolaryngology, Karolinska Institutet, Stockholm, Sweden
[f] Department of Clinical Science, University of Umeå, Umeå, Sweden
[g] Ear Science Institute Australia, Subiaco, Western Australia

ARTICLE INFO

Article history:
Received 8 November 2015
Received in revised form 6 February 2016
Accepted 6 February 2016
Available online 8 February 2016

Keywords:
Decision tree
Feature extraction
Automated diagnosis
Otitis media
Video-otoscope
Global medicine

ABSTRACT

Background: Otitis media is one of the most common childhood diseases worldwide, but because of lack of doctors and health personnel in developing countries it is often misdiagnosed or not diagnosed at all. This may lead to serious, and life-threatening complications. There is, thus a need for an automated computer based image-analyzing system that could assist in making accurate otitis media diagnoses anywhere.
Methods: A method for automated diagnosis of otitis media is proposed. The method uses image-processing techniques to classify otitis media. The system is trained using high quality pre-assessed images of tympanic membranes, captured by digital video-otoscopes, and classifies undiagnosed images into five otitis media categories based on predefined signs. Several verification tests analyzed the classification capability of the method.
Findings: An accuracy of 80.6% was achieved for images taken with commercial video-otoscopes, while an accuracy of 78.7% was achieved for images captured on-site with a low cost custom-made video-otoscope.
Interpretation: The high accuracy of the proposed otitis media classification system compares well with the classification accuracy of general practitioners and pediatricians (~64% to 80%) using traditional otoscopes, and therefore holds promise for the future in making automated diagnosis of otitis media in medically underserved populations.

1. Introduction

Otitis media (OM) is the second most important cause of hearing loss, which ranked fifth on the global burden of disease and affected 1.23 billion people in 2013 (Global Burden of Disease Study 2013 Collaborators, 2015). It is one of the most common childhood illnesses and constitutes a major chronic disease in low and middle-income countries (Global Burden of Disease Study 2013 Collaborators, 2015; World Health Organization, 2004). The incidence of OM in sub-Saharan Africa (SSA), South Asia and Oceania is two- to eight-fold higher than in developed world regions with India and SSA accounting for the majority of OM related deaths (Monasta et al., 2012; Acuin, 2004).

Common types of OM include acute otitis media (AOM), otitis media with effusion (OME), and chronic suppurative otitis media (CSOM) (Paparella et al., 1985). OM is often misdiagnosed, or not diagnosed at all, and consequently treated incorrectly which may lead to serious, or even life-threatening complications (Asher et al., 2005; Buchanan and Pothier, 2008; Legros et al., 2008).

Access to ear-, nose- and throat (ENT) specialists and equipment to diagnose OM is severely limited in developing countries (Fagan and Jacobs, 2009; World Health Organization, 2013). In Africa, for example, the majority of countries (64%) report less than one ENT specialist per million people (World Health Organization, 2013). In addition, general practitioners and pediatricians, whom are also scarce in developing countries, are often prone to under- or over-diagnose OM (Asher et al., 2005; Buchanan and Pothier, 2008; Legros et al., 2008). There is therefore a great need to develop systems that can facilitate accurate diagnosis of OM in underserved areas of the world. Kuruvilla et al. (2013) recently reported an automated algorithm based on image-analysis for distinguishing between two types of OM (AOM and OME) and middle ears without effusion with a reported accuracy of 85.61%. While these results are promising, the full range of OM types in addition to other clinical presentations such as ear canal obstructions have not been demonstrated.

* Corresponding author at: Department of Clinical Science, University of Umeå, 901 87 Umeå, Sweden.
E-mail address: claude.laurent@umu.se (C. Laurent).

http://dx.doi.org/10.1016/j.ebiom.2016.02.017

Table 1
Predefined features associated with each diagnosis (O/W — obstructing wax or foreign bodies in the external ear canal; n-TM — normal tympanic membrane; AOM — acute otitis media; OME — otitis media with effusion; CSOM — chronic suppurative otitis media with perforation).

	O/W	n-TM	AOM	OME	CSOM with perforation
Malleus bone visible	Can be either visible or not	Yes	No	Yes	Can be either visible or not
Tympanic membrane shape	Can be categorized as retracted, normal, bulging or irregular	Can be categorized as retracted, normal or bulging or irregular	Bulging	Retracted	Can be categorized as normal, bulging or irregular
Color	Pearly white	Pearly white	Predominantly red	Can be categorized as red or opaque	Pearly white
Perforation	No	No	No	No	Yes
Wax	Can be present or not	Can be present or not	Can be present or not	Can be present or not	Can be present or not
Fluid	No	No	Can be either visible or not	Yes	No
Light reflex	No	Yes	Can be either visible or not	Can be either visible or not	No

The aim of this study was to develop and validate a new image-analysis system to classify images obtained from commercial video-otoscopes into one of the following diagnostic groups: 1) obstructing wax or foreign bodies in the external ear canal (O/W); 2) normal tympanic membrane (n-TM); 3) AOM; 4) OME, and 5) CSOM with perforation. The image-analysis system was also evaluated in a clinical population using a low cost custom-made video-otoscope.

2. Methods

2.1. Development of the Image-Analysis Classification System

A decision tree was employed to classify images into one of the five diagnostic groups. In order for the tree to make an accurate diagnosis, the predefined features associated with each diagnosis (Table 1) had to be accurately identified in the images by feature extraction methods (Supplementary material 1 provides a detailed discussion on the image processing method). It included pre-processing followed by the detection of the malleus bone, tympanic membrane (TM) shape and color, perforation, obstructive wax, middle-ear fluid, and light reflex (Table 1).

Pre-diagnosed color TM photographs (n = 562) from three TM image collections captured with a range of commercially available video-otoscopes were re-evaluated independently by two experienced (>35 years of practice) specialists in otology. From them a set of 489 images (at least 500 × 500 pixels) of the TM and external ear canal were included in the study (examples provided in Fig. 1). Seventy-three images of the original 562 (12.99%) were discarded because of insufficient image quality or disagreement between the two otologists. Only images on which diagnoses were in agreement between the otologists were accepted (approximately 50% of them from children). The final set of images was classified as O/W (n = 120), n-TM (n = 123), AOM (n = 80), OME (n = 80), and CSOM with perforation (n = 86). Eighty percent of all images (n = 391) were randomly selected to develop the feature extraction algorithms and the remaining 20% (n = 98) were used for the validation study. The higher ratio for training images was used to improve classification accuracy while a proportion (20%), not used to train the system, was required for validation and testing. For each image in the training set (n = 391) the algorithms received a diagnosed image as input and analyzed the image using the feature extraction algorithms. It produced an array containing each image's features with the corresponding diagnosis as output. The decision tree was

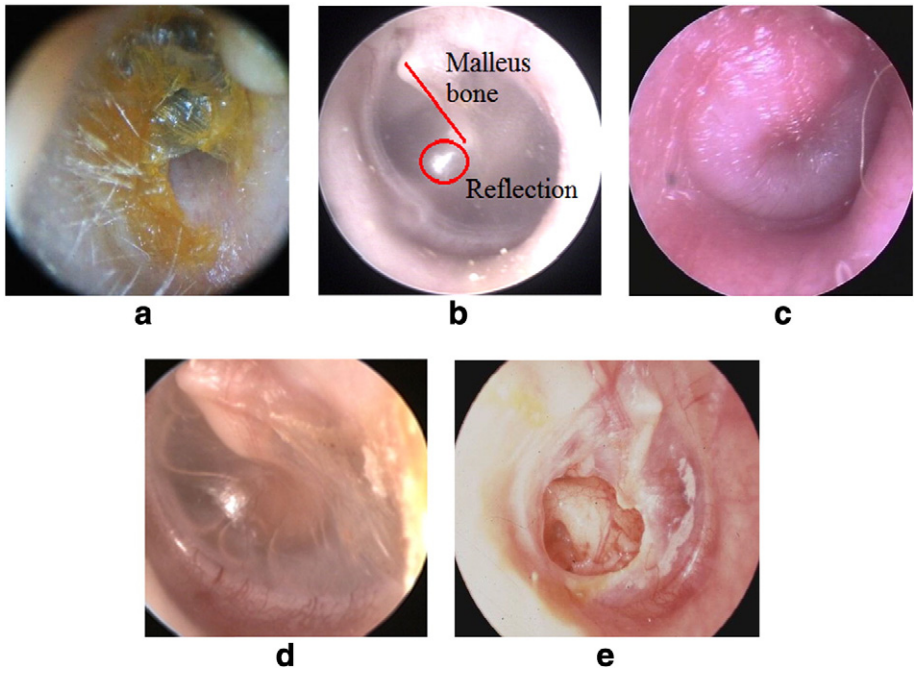


Fig. 1. Examples of the five diagnostic classification categories. a. Obstructing wax or foreign bodies (O/W) in external ear canal precluding visualization of the TM to establish an OM diagnosis; b. a normal TM (n-TM) showing a semi-transparent pearly white TM, triangular shaped light reflex and malleus bone clearly visible (red ring and line, respectively); c. acute otitis media (AOM) showing a bulging TM with red color; d. otitis media with effusion (OME) showing a retracted TM and fluid in the middle ear; e. chronic suppurative otitis media (CSOM) showing a TM perforation.
Images courtesy C. Laurent 2014.

Fig. 2. Side view of the final low cost custom-made video-otoscope.

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 (1986). This division of examples ensures that the examples that are least resembling the feature will eventually be eliminated in order to ensure that the best fit is found. The example feature vectors were subsequently used to manually design the decision tree, which constituted the classification algorithm employed in this study. No additional training was performed. Matlab® (MathWorks, Inc., Natick, Massachusetts, United States) was used as the software due to its processing capabilities, extended set of functions and familiarity. The user interface was designed, using "guide" in Matlab®, to be understandable, reliable and easy to use. (Supplementary material 2 provides a detailed discussion of the creation of the decision tree and the classification methods).

Table 1 shows the five diagnostic groups and their corresponding features, according to which the feature extraction algorithms were developed. To classify an image of unknown diagnosis or pathology, the image is provided as input to the system, where after pre-processing and feature extraction are performed. Once all features are extracted, the decision-tree classifies the feature vector associated with the input 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.

Execution time of the system, i.e. the time it takes to diagnose an image submitted after capturing, is determined by the respective execution times of the feature extraction algorithms and the decision tree classification. Total execution time of the image-analysis classification system was calculated as 3.45 s for a notebook computer with the following specifications: Intel(R) i7-3632QM 2.2 GhZ CPU with 8 GB RAM running on a 64-bit Windows 7 operating system.

2.2. Validation of the Image-Analysis Classification System

The image-analysis classification system was validated on 20% (n = 98) of the 489 diagnosed color images of TM's received from the two experienced otologists. These images had not been used previously to train the system. The image-analysis system classified images via the trained decision tree and a comparison could be made with the original diagnoses. Reliability of the image-analysis classification system was determined by investigating the agreement between diagnosed images and system classifications using Cohen's kappa. Unweighted kappa statistic (κ) was used to quantify "strength of agreement", or diagnostic concordance.

In a separate pilot study, the image-analysis classification system was validated clinically on images captured by a custom-made video-otoscope. This was designed as an example of a low cost replacement for current clinical video-otoscopes. The shell of a commercial ear-cleaning product was used as the body for the otoscope. Internal components were replaced by a battery, a light emitting diode (LED), and resistor. Universal serial bus (USB) communication was implemented to communicate between the otoscope and a notebook computer using an audio/video cable to USB converter. The front part of the custom-made video-otoscope was removed to make space for the camera (2 megapixels) right inside the opening with a disposable ear tip attached, see Fig. 2. Supplementary material 3 provides detailed information about the construction of the custom-made video-otoscope. The final production cost of the custom-made video-otoscope came to around $84.

The low cost custom-made video-otoscope was used to examine ears from patients in an emergency room at a hospital in Pretoria, South Africa. A qualified emergency unit nurse, without previous training in video-otoscopy did the examinations together with an experienced (>20 years of practice) general practitioner (GP). The Institutional Review Board provided ethical clearance before any data collection commenced and each patient signed a consent form explaining the examination procedure. No patient under the age of 18 years was enrolled for the study, which was done in an emergency room for adults.

The GP first made an inspection of the ear with a hand-held otoscope and classified the ear canal, and/or TM status into one of the five diagnostic categories. The nurse then captured images using the custom-made video-otoscope and stored these on a notebook computer. Thereafter the image-analysis system classified the images stored on the computer via the trained decision tree and a comparison was made to the GP's diagnosis.

2.2.1. Funding

No external funding for any of the authors.

Table 2
Diagnostic classification accuracy (n = 98). Shaded cells = % correct diagnosis; unshaded cells = % incorrect diagnosis.

	Number of images (n)	O/W (classified)	n-TM (classified)	AOM (classified)	OME (classified)	CSOM with perforation (classified)
O/W (actual)	24	79.17 (19)	4.17 (1)	0	4.17 (1)	12.5 (3)
n-TM (actual)	25	4 (1)	80.00 (20)	8 (2)	4 (1)	4 (1)
AOM (actual)	16	0	0	81.25 (13)	12.5 (2)	6.25 (1)
OME (actual)	16	0	6.25 (1)	12.5 (2)	81.25 (13)	0
CSOM (actual)	17	5.88 (1)	5.88 (1)	5.88 (1)	0	82.35 (14)

Table 3
Performance parameters for each diagnosed category (PPV = positive predictive value; NPV = negative predictive value).

	O/W	n-TM	AOM	OME	CSOM with perforation
Sensitivity	0.79	0.80	0.81	0.81	0.82
Specificity	0.97	0.95	0.92	0.94	0.94
PPV	0.89	0.83	0.75	0.80	0.78
NPV	0.94	0.94	0.95	0.95	0.95

3. Results

Table 2 shows the number of pre-diagnosed images used for validation in each diagnostic group. Disregarding their diagnoses, all of the validation images were pre-processed, feature extracted, and classified via the decision tree, yielding the results in Table 2, with an average objective classification accuracy of 80.61% (79/98).

The unweighted kappa agreement was 0.76 (95% CI 0.66–0.86) showing a substantial agreement.

Performance characteristics of the image-analysis classification system are presented in Table 3. Sensitivity across all five diagnostic categories was between 0.79 and 0.82. AOM had the lowest specificity (0.92), while O/W had the highest (0.97). AOM had the lowest positive predictive value (PPV) (0.75) and O/W had the highest (0.89). Negative predictive values (NPV) were all between 0.94 and 0.95.

Using the low cost custom-made video-otoscope the nurse was able to capture 108 high quality images in three of the five diagnostic categories (Table 1). The images captured by the nurse were classified using the image-analysis classification system, (Table 4) with an overall diagnostic agreement of 78.70% (85/108). Since the custom-made video-otoscope was tested only in adults, all diagnostic groups were not observed.

4. Discussion

The new image-analysis classification system demonstrated a high degree of agreement (80.6%) to the combined diagnostic correspondence of the two otologists, varying between 79.2 to 82.4% across the five diagnostic categories in the set of test images from adults and children. The accuracy of the image-analysis classification system compares favorably to average diagnostic accuracy of pediatricians (Kuruvilla et al., 2013) (~80%), GP's (Blomgren and Pitkäranta, 2003; Jensen and Louis, 1999) (64–75%) and ENT's (Pichichero and Poole, 2001) (73%) using handheld otoscopes. Based on an individual diagnostic accuracy for ENT's of 73% (Pichichero and Poole, 2001) using a Gaussian statistical analysis, the combined diagnostic accuracy of the two otologists could be calculated to be 88.13%. The ground truth data used to train the image-analysis system may therefore contain an estimate of 11.87% inaccurate diagnoses. This uncertainty can be reduced and the validity of ground truth data improved by increasing the number of experienced specialists performing diagnosis on images used to train the system. For instance if three specialists were used the combined diagnostic accuracy would be 94.40% compared to 98.64% when using 5 specialists.

Compared to an algorithm proposed by Kuruvilla et al. (2013) that only distinguished between two diagnoses, the proposed system is able to diagnose five types of common external ear canal and middle ear conditions, including AOM, OME and CSOM with perforation. Utilization of an automated system for differential diagnosis of OM based on video-otoscopic images can provide a valuable diagnostic crosscheck for health personnel and, in poorly resourced sectors, could serve as the only source for a diagnosis.

The image-analysis system rarely indicated a false negative classification across OM categories with a negative predictive value of 0.95. Normal TM's were only indicated incorrectly in 8% of AOM and 4% of OME and CSOM with perforation. This accuracy is promising since it is similar to and better than reported diagnostic accuracy by pediatricians (Kuruvilla et al., 2013) and GP's (Blomgren and Pitkäranta, 2003; Jensen and Louis, 1999). Minimizing false negatives is important to ensure treatment in cases where it is indicated. Over-diagnosis by clinicians, however, is often a problem in AOM (Pichichero and Poole, 2001) and as a result the American Academy of Pediatrics recently prioritized greater diagnostic specificity to reduce over-diagnosis and unnecessary treatment (Lieberthal et al., 2013). The current system in contrast correctly identified AOM in 81.3% of instances. This could support accurate diagnoses at primary health care levels at which skilled clinical assessment is often unavailable or limited. A computer-assisted diagnostic system, such as this, can also provide a useful second opinion to assist health care workers in making correct diagnoses. Benefits include timely initiation of treatment and reduced health system burden related to over-referrals.

There were, however, some misclassifications of diagnoses noted. For example, O/W was misclassified as CSOM in 12.5% of cases. This is likely attributable to the feature extraction algorithm, which associates yellow features with the O/W category, and identifies similar features in some cases of CSOM. Moreover, the misclassification of AOM as OME and vice versa, in 12.5% of cases, could be attributed to middle-ear fluid not being visible with the TM detected by the system as opaque. These overlaps will be investigated to ensure improved differentiation in future improvements of the system.

The custom-made video-otoscope indicated overall clinical diagnostic accuracy of 78.7% compared to 80.6% on images from commercial video-otoscopes. While being assessed on a smaller sample of ears from adults (108) and only three diagnostic categories represented (Table 4) the results suggest that a custom-made video-otoscope could provide a low-cost method for accurate diagnosis of common ear disease. However, future investigations should evaluate the clinical validity in children compared to adults. The accumulated production cost of the custom-designed video-otoscope was around $84, which is at least 5 times less expensive than commercial entry-level video-otoscopes. Of course, as with regular video-otoscopes it still requires a computer to capture images. The computer together with the Matlab® software adds significantly to the cost. It would therefore come to the region of $1000 to operate the complete systems. However, the use of a netbook and the open source equivalent of Matlab®, (i.e. GNU Octave), will bring the cost down to between $250 and $300. The cost of equipment must be seriously considered if a method for automated diagnosis of OM is to be a viable solution for underserved contexts.

The present study demonstrates the accuracy of an image analysis system for diagnosis of one of the most common childhood illnesses (World Health Organization, 2004; Global Burden of Disease Study 2013 Collaborators, 2015). In light of the shortage of specialists able to accurately diagnose OM, especially in low and middle-income countries, a system like this could be very valuable to ensure that appropriate treatment is provided. For instance, nurses could be trained to take pictures of the TM using a video-otoscope and analyzed by the image analysis system locally or on a cloud-based server. In many world regions this is likely to be the only opportunity for access to a diagnosis of ear disease amongst large populations. In addition to these applications the system could be employed for teaching and training purposes in diagnosing OM to up-skill general health care workers in diagnosing ear disease.

Table 4
Results of system on images obtained from custom-made video-otoscope.

Category	Number of images collected	Number of images correctly diagnosed	Percentage correctly diagnosed
O/W	15	11	73.33
n-TM	72	58	80.56
OME	21	16	76.19
Total	108	85	78.70

In conclusion, we have proposed and validated a new image-analyzing system for the most frequent OM diagnoses and obstructions of the ear canal. The system loaded onto a conventional computer notebook has a diagnostic accuracy that compares favorably to that of pediatricians and GP's. It could be a valuable new tool for health personnel seeing OM patients in developing countries.

5. Limitations

The classification algorithm used in the current system is a decision tree constructed manually by analyzing the feature vectors produced by the feature extraction algorithms used to process the training images. The resulting decision tree is therefore fixed, and will have to be redesigned if more training images are incorporated. A neural network, however can be trained using the same feature vectors used to construct the decision tree and has the ability to be trained further with a single image's feature vector upon assessment by specialists. A comparison of these two approaches could be useful to evaluate for their respective accuracy.

Acquiring anamnestic information from patients, such as occurrence of fever, pain, hearing loss and discharge could increase the diagnostic accuracy of the system and will be considered in future validation studies. The main aim of this study, however was to use only image analysis and feature vector classification to diagnose the various forms of OM and we did not have additional metadata.

Manual examination of misdiagnosed images could help improve the overall accuracy of the system. In its current form the person operating the system has to inspect the image quality before submitting it for diagnosis. If images are of poor quality (e.g. poor focus, motion artifacts, insufficient light, partial TM visibility) the system will still attempt to make a diagnosis which may result in reduced accuracy.

Author Contributions

Hermanus C Myburgh (HCM) and Willemien H van Zijl (WHvZ) have contributed equally to the development of the image-analyzing software but HCM was the main author. DeWet Swanepoel (DWS) conceptualized the project and has contributed to analysis of results and writing of the manuscript. Sten Hellström (SH) was consulted as specialist on otitis media to interpret findings and to correct the manuscript, Claude Laurent (CL) was co-author and corresponding author. Respective contributions: HCM 30%, WHvZ 20%, DWS 20%, SH 10%, CL 20%.

Conflict of Interest

The authors have no conflict of interest to declare.

Acknowledgments

We are indebted to Dr. Björn Åberg, Oslo, Norway and Dr. Thorbjörn Lundberg, Umeå, Sweden for providing us with high quality photographs from their respective TM image collections.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2016.02.017.

References

Acuin, J., 2004. Chronic Suppurative Otitis Media: Burden of Illness and Management Options. WHO, Geneva (http://www.who.int/pbd/deafness/activities/hearing_care/otitis_media.pdf, accessed August 29, 2012).

Asher, E., Leibovitz, E., Press, J., Greenberg, D., Bilenko, N., Reuveni, H., 2005. Accuracy of acute otitis media diagnosis in community and hospital settings. Acta Paediatr. 94 (4), 423–428.

Blomgren, K., Pitkäranta, A., 2003. Is it possible to diagnose acute otitis media accurately in primary health care? Fam. Pract. 20, 524–527.

Buchanan, C.M., Pothier, D.D., 2008. Recognition of paediatric otopathology by general practitioners. Int. J. Pediatr. Otorhinolaryngol. 72 (5), 669–673.

Fagan, J.J., Jacobs, M., 2009. Survey of ENT services in Africa: need for a comprehensive intervention. Global Health Action 2, 1–7.

Global Burden of Disease Study 2013 Collaborators, 2015. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 386, 743–800.

Jensen, P.M., Louis, J., 1999. Criteria, performance and diagnostic problems in diagnosing acute otitis media. Fam. Pract. 16, 262–268.

Kuruvilla, A., Shaikh, N., Hoberman, A., Kovacevic, J., 2013. Automated diagnosis of otitis media: vocabulary and grammar. Int. J. Biomed. Imaging 2013, 1–15.

Legros, J.M., Hitoto, H., Garnier, F., Dagorne, C., Parot-Schinkel, E., Fanello, S., 2008. Clinical qualitative evaluation of the diagnosis of acute otitis media in general practice. Int. J. Pediatr. Otorhinolaryngol. 72 (1), 23–30.

Lieberthal, A.S., Carroll, A.E., Chonmaitree, T., et al., 2013. The diagnosis and management of acute otitis media. Pediatrics 131 (3), e964–e999.

Monasta, L., Ronfani, L., Marchetti, F., et al., 2012. Burden of disease caused by otitis media: systematic review and global estimates. PLoS One 7 (4), e36226. http://dx.doi.org/10.1371/journal.pone.0036226.

Paparella, M., Bluestone, C., Arnold, W., et al., 1985. Definition and classification. Ann. Otol. Rhinol. Laryngol. 94 (Suppl. 116), 8–9.

Pichichero, M.E., Poole, M.D., 2001. Assesing diagnostic accuracy and tympanoscentesis skills in the management of otitis media. Arch. Pediatr. Adolesc. Med. 155 (10), 137–142.

Quinlan, J., 1986. Induction of decision trees. Mach. Learn. 1, 81–106.

World Health Organization, 2004. Chronic Suppurative Otitis Media. Burden of Illness and Management Options. World Health Organization, Geneva, p. 83.

World Health Organization, 2013. Multi-Country Assessment of National Capacity to Provide Hearing Care. Switzerland, Geneva (Retrieved from: http://0www.who.int.innopac.up.ac.za/pbd/publications/WHOReportHearingCare_Englishweb.pdf).

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.

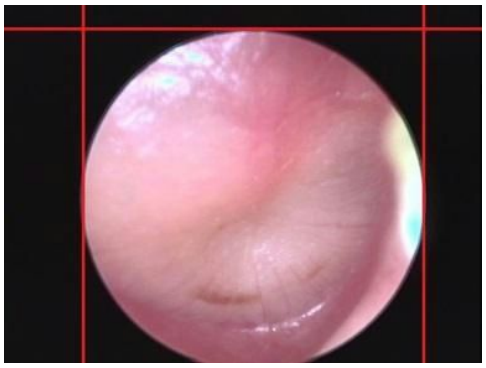


Figure 1. Image showing lines where the figure will be cropped, (image altered from Laurent, 2014)

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

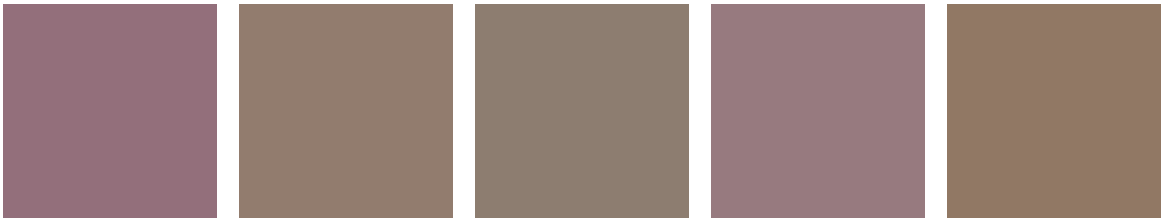
Because different symptoms exhibit different colors, color analysis of the entire image is performed to evaluate the color of each diagnostic category. It is performed by extracting the red, green and blue (RGB) color bands, as seen in figure 3, from the image in figure 2. Each extracted band has the 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



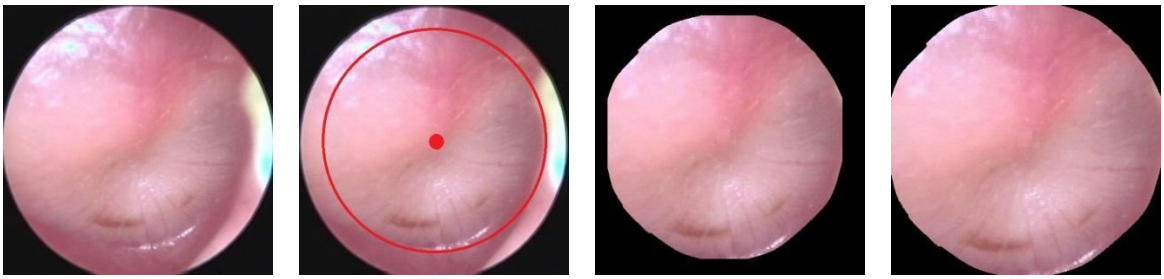
a) **b)** **c)** **d)** **e)**

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)

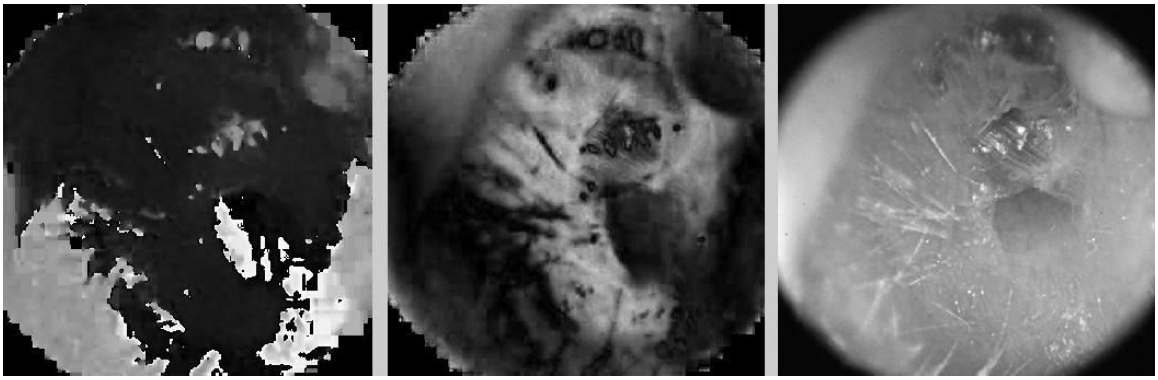
Figure 5 a) Original Image; b) Center and radius of the image shown; c) Image cropped; d) Image with the black sides removed (Laurent, 2014)

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%.

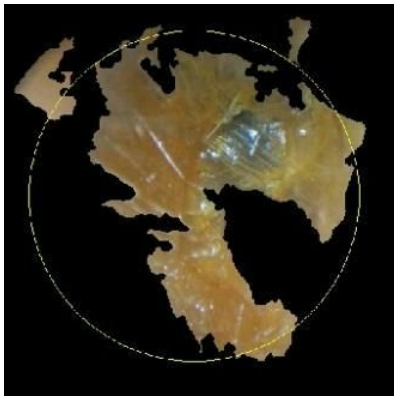


Figure 8. An example of effect of obstructive wax detection, image altered from (Laurent, 2014)

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.

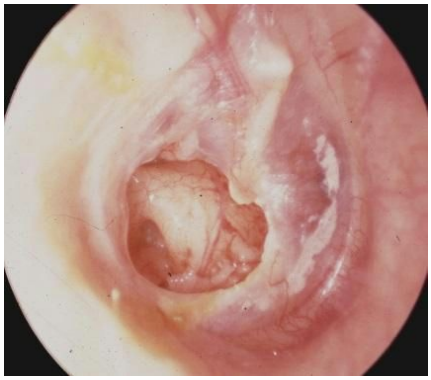
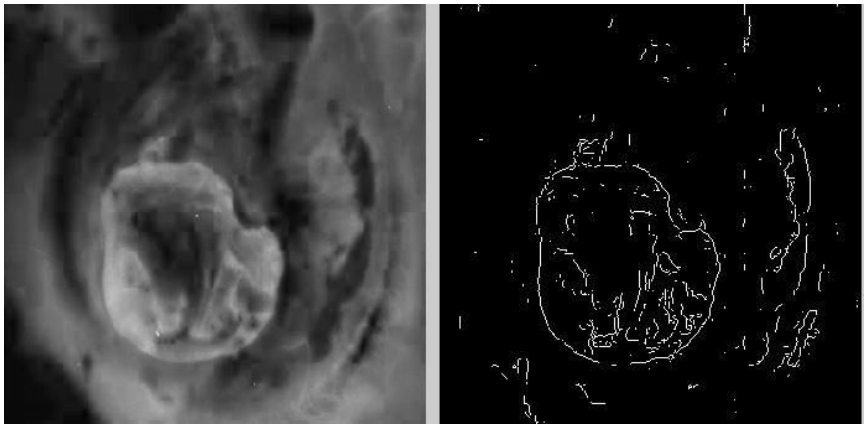


Figure 9. Original RGB CSOM with perforation image (Laurent, 2014)



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

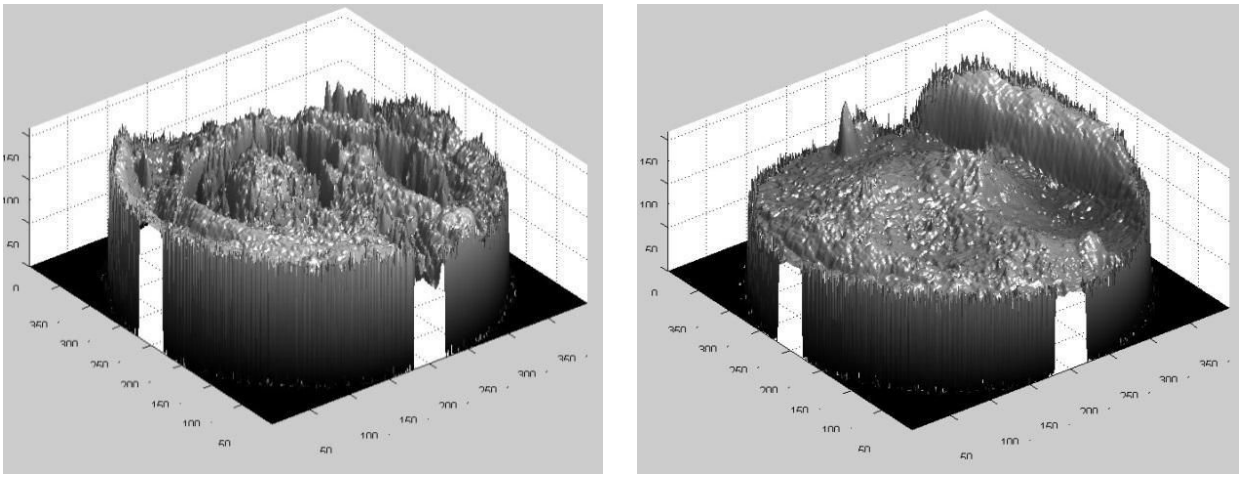
In order to categorize the shape of the TM as retracted, normal, bulging or irregular, a norm first needs to be established. Cropping is done on the image as described for TM color analysis. The image is then converted to a grey scale image, where after the difference in greyness between neighboring 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	$x \geq 15$

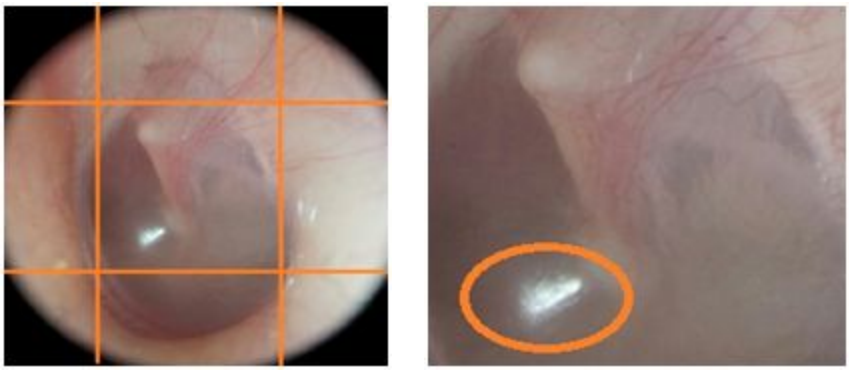


a) b)

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.

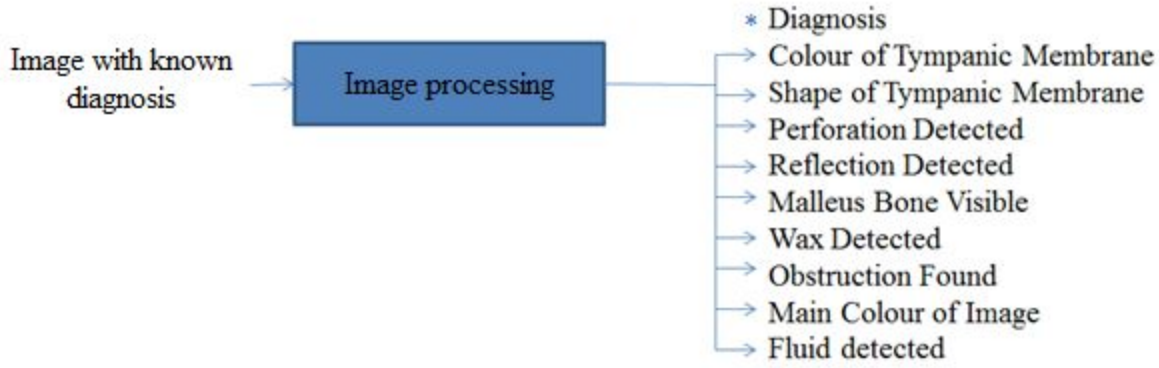


Figure 1. Training feature vector generation

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].

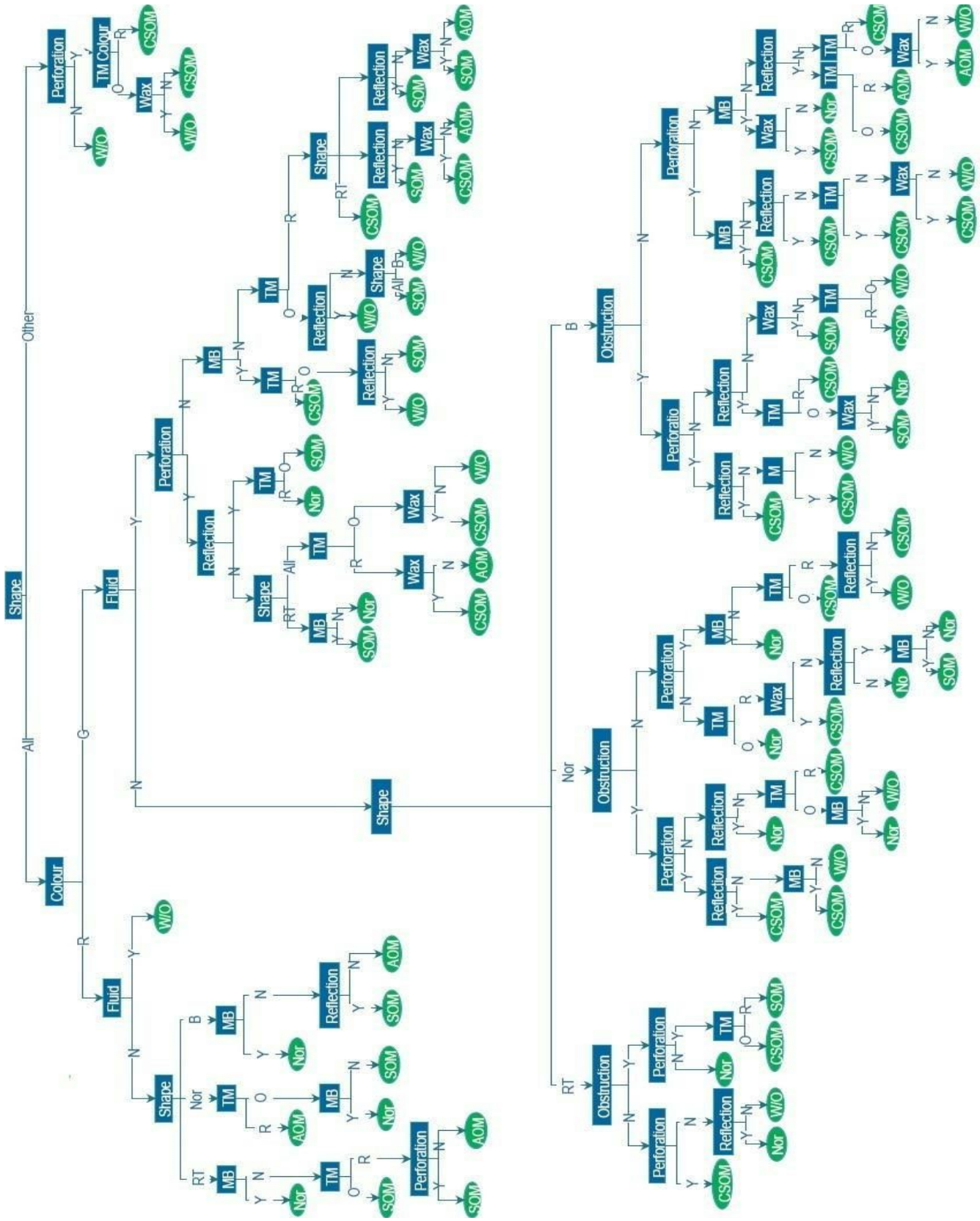


Figure 2. Decision tree

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.

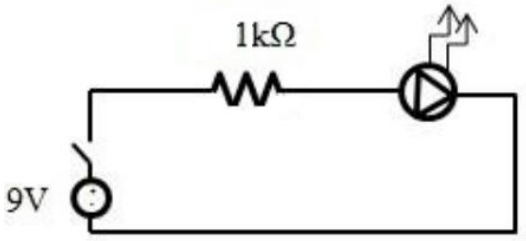


Figure 1. Illumination circuit for the otoscope

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.

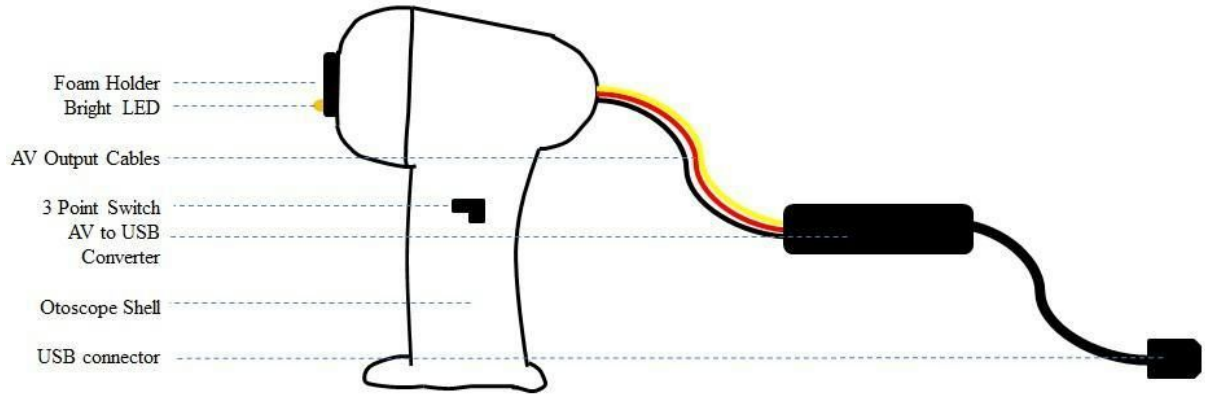


Figure 4. The exterior shell of the video-otoscope

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.

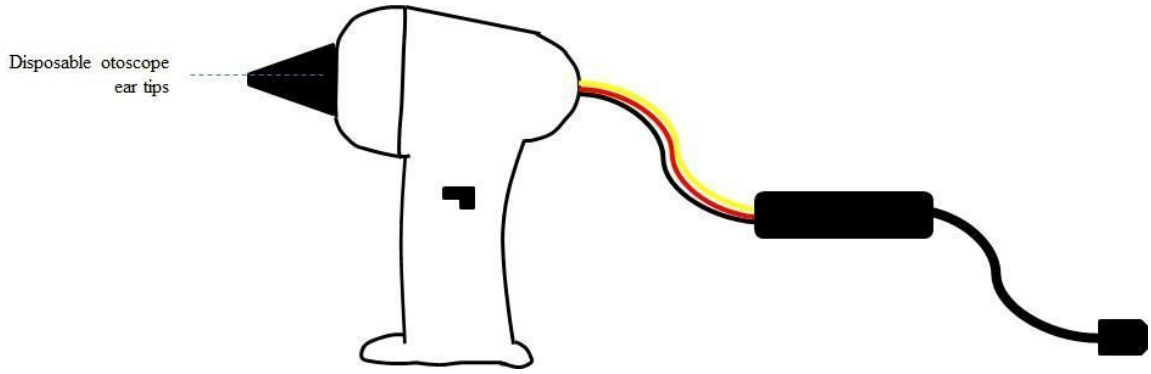


Figure 5. The exterior design of the video-otoscope showing the disposable ear tip.

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.

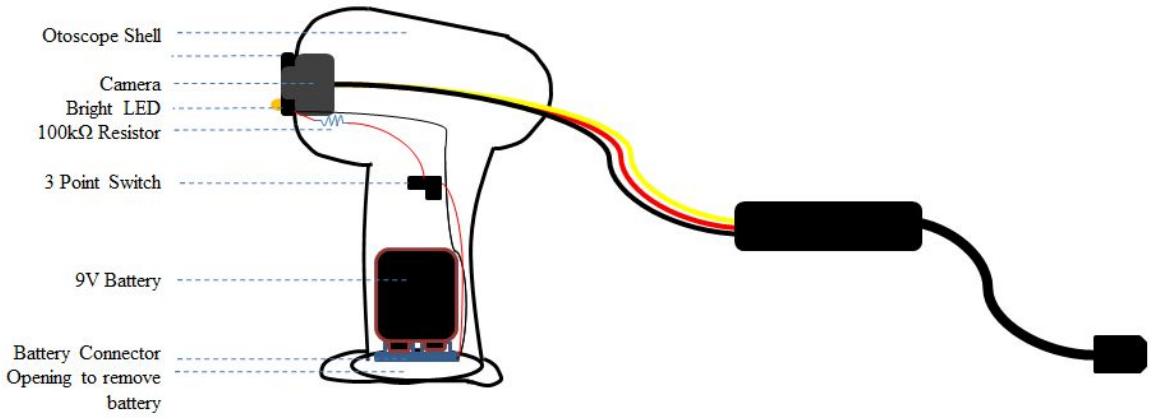


Figure 6. The content of the custom -made video-otoscope

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