M. Shazam Kasher

Otitis Media Analysis: An Automated Feature Extraction and Image Classification System

Helsinki Metropolia University of Applied Sciences

Bachelor of Engineering

Degree Programme in Electronics

Bachelor's Thesis

25.04.2018

Author(s) Title Number of Pages Date	Muhammad Shazam Kasher Otitis Media Analysis: An Automated Feature Extraction and Image Classification System 60 pages 25 April 2018
Degree	Bachelor of Engineering
Degree Programme	Degree Programme in Electronics
Specialisation option	
Instructor(s)	Sakari Lukkarinen, Senior Lecturer (Metropolia UAS) Jani Virtanen, CEO (MDS Finland Oy)

The main goal of the project was to develop novel Artificial Intelligence solutions using images and open source tools based on computer vision and deep learning to analyse and detect otitis media infection. The objective was to develop these solutions to help and support medical professionals in detecting otitis media at an early stage and making the final diagnosis.

Advanced computer vision, deep learning and transfer learning methods were used to develop two software systems. The first software system supports the professional by applying feature engineering to the images and extracting certain features that can help them in classifying the images manually. The second software system is an automated image classification software based on deep learning, which performs classification task for the professional which then can be used to make the final diagnosis.

The computer vision based system performed well on the available data and was successful in extracting all five key features which can be used to manually classify images. The deep learning based image classification software performed fairly good as well and achieved an accuracy of 82.2% with InceptionV3 architecture and 80% accuracy with MobileNets architecture.

It was concluded that the deep learning software system works better overall for classification tasks with high accuracy. In future development, a significantly higher accuracy can be reached using more data and more sophisticated algorithms.

Keywords	Otitis Media, Acute Otitis Media, Feature Extraction, Computer Vision, Deep learning, Machine Learning, CNN

Contents

List of Abbreviations

AOM	Acute Otitis Media
CNN	Convolutional Neural Network
COM	Chronic Otitis Media
CRI	Colour Rendering Index
CV	Computer Vision
DL	Deep Learning
GUI	Graphical User Interface
HSV	Hue, Saturation, Value
MATLAB	Matrix Laboratory
MEE	Middle Ear Effusion
ML	Machine Learning
OM	Otitis Media
OME	Otitis Media with Effusion
PIL	Python Imaging Library
RGB	Red, Green, Blue
ROI	Region of Interest
Tcl	Tool Command Language
TM	Tympanic Membrane

List of Figures

List of Tables

1 Introduction

Otitis Media (OM) is an infection of the middle ear belonging to the group of inflammatory diseases. It is one of the most common childhood illnesses and the second most important reason leading to the loss of hearing. It is most common in developing countries and was ranked fifth on the global burden of disease and affected 1.23 billion people in 2013 [1]. Otitis media (OM) is a leading cause of health care visits and drugs prescription. Despite the advances in medical diagnostic techniques and otoscope technology, OM is often misdiagnosed or not diagnosed at all, especially when it is in the early stages.

The aim of the study is to develop a diagnostic system using tympanic membrane (TM) images and applying computer vision and machine learning to automatically extract certain features and perform image classification which can help diagnose otitis media with greater accuracy. The final goal of the study is to develop two different software systems, one based on feature extraction which will provide visual and numerical data to the professional to manually classify images into the right infection category. The second software will be able to automatically classify images and predict if the TM is normal or infected. The diagnostic system will provide a reliable second opinion for a professional to make his/her analysis and final diagnosis.

Chapter 2 further explores and gives a brief description of different types of otitis media infections. It also provides some statistical data to help understand the importance of dealing with the problem. Chapter 3 of the study provides important features of tympanic membrane. Whereas chapter 4 offers us an overview of the diagnosis methods used by Otologists. It also highlights the difficulties faced by physicians and doctors who are using these diagnosis methods.

Furthermore, chapter 5, provides a better understanding of the whole automated software system and overview of computer vision and machine learning tools used for developing the software. Chapter 6, 7 and 8 explain the idea, methods and algorithms in depth, used for making an automated system as well as the challenges faced during the development. Chapters 9 finalize the study by providing a conclusion of the whole project.

2 Otitis Media

Otitis Media is one of the most common disease to occur during one's lifetime. It is most common amongst infants and young children, although it can also be found in adults. The reason it is so common in children is because a child's eustachian tubes which can be seen below in figure 1 are narrower and shorter in length compared to adults, which easily allows the fluid to get trapped in the middle ear [3].

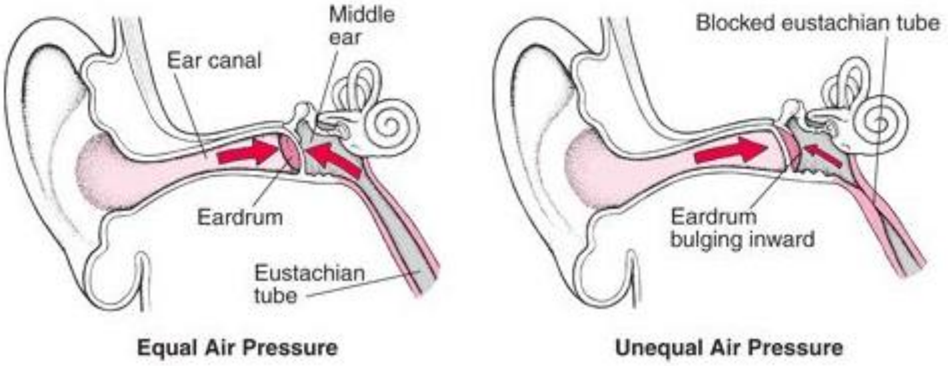


Figure 1 - Eustachian Tube under normal and unequal pressure, Reprinted from healcure.org [4]

Plenty of important factors could lead to otitis media. Age plays an important role as the infection is more common with infant and young children. A child aged between 6 to 36 months is more likely to get ear infection. For example, in infants, it can be caused if they are drinking from a bottle while lying down on the back. The drinking position can cause the eustachian tube to be blocked which can cause the ear infection. Among children, OM can also occur due to attending day-care. One study found out that children who enter daycare in the first year of life, may develop OM at an earlier age [5].

Common illness such as cold or sinus infection can also cause OM infection. Moreover, history of sinusitis or allergies such as rhinitis and nasal can be a factor contributing to the infection. Genetics of a child can also play an important role in increasing chance of developing otitis media. For example, having a family member with ear infection history may imply of the possibility of developing OM. One of the other important factors leading towards an infection between children is exposure to secondhand smoke. The reasons mentioned above, and several others can lead to different kind of Otitis Media, which is further discussed in the chapter.

To better understand the impact of otitis media infection on the population and the significance of a robust diagnostic system, it is important to review the statistical data provided by the research institutes.

OM, the most common specifically treated childhood disease, accounts for approximately 20 million annual physician visits in US. Various epidemiologic studies report the prevalence rate of AOM to be 17-20% within the first 2 years of life, and 90% of children have at least one documented MEE by age of 2 years. OM is a recurrent disease. One third of children experience six or more episodes of AOM by age 7 years. [6.]

Incidence and prevalence in other industrialized nations are like the US rates. In less developed nations, OM is extremely common and remains a major contributor to childhood mortality resulting from late-presenting intracranial complications. International studies show increased prevalence of AOM and chronic OM (COM) among Micronesian and Australian aboriginal children. [6.]

There are many kinds of ear infections, but the most common includes acute otitis media (AOM) and otitis media with effusion (OME). These infections are briefly described below with their individual statistics.

2.1 Acute Otitis Media (AOM)

The most common type of ear infection to occur during childhood is acute otitis media. AOM is a process in which the middle ear shows the signs and symptoms of acute inflammation of the tympanic membrane. AOM is a recurrent disease. More than one third of children experience six or more episodes of AOM by age 7 years [6]. Some of the most common signs and symptoms associated with acute otitis media are fever, irritability, vomiting, otalgia, otorrhea and diarrhea.

A rough categorization of AOM symptoms can be made depending on the age. For example, irritability or feeding problem maybe be the only indication with newborn babies. In older children, fever, and otalgia along with ear tugging can be noticed. In older children and adults, hearing loss becomes a constant feature and otitis media with effusion is noted before the middle ear fluid is detected.

In the United States, 70% of all children experience one or more attacks of AOM before their second birthday. A study from Pittsburgh that prospectively followed urban and rural children for the first 2 years of life determined that the incidence of middle ear effusion episodes is approximately 48% at age 6 months, 79% at age 1 year, and 91% at age 2 years [7].

2.2 Otitis Media with Effusion (OME)

The simplest way of describing OME is when the middle part of the ear is filled with fluid or more specifically presence of fluid behind tympanic membrane. OME can occur after the resolution of other ear infections, especially acute otitis media. Among children who have had an episode of acute otitis media, as many as 45% have persistent effusion after 1 month, but the number decreases to 10% after 3 months [8].

OME is most likely to occur in children since their eustachian tubes are short and narrower which increases the risk of clogging and infection. A blocked tube and excess fluid provides an ideal environment for bacteria to grow which can lead to an ear infection. Otitis media with effusion does not have obvious symptoms. In younger children, hearing problem could be a symptom, whereas in older children and adults muffled hearing or sense of fullness in the ear can be described as a symptom.

Clinical guidelines from a joint commission of specialties document that screening surveys of healthy children between infancy and age 5 years show a 15-40% point prevalence in middle ear effusion (MEE). Furthermore, among children examined at regular intervals for 1 year, 50-60% of child care attendees and 25% of school-aged children were found to have a middle ear effusion at some point during the examination period, with peak incidence during the winter months. [8.]

Approximately 80% of children have had an episode of otitis media with effusion (OME) when younger than 10 years. At any given time, 5% of children aged 2-4 years have hearing loss due to a middle ear effusion that lasts 3 months or longer. [8.]

3 Tympanic Membrane

The tympanic membrane (TM) or commonly known as eardrum is a significant part of the human ear. It is a thin layer of tissue in the human ear which is approximately 0.1mm thick and 8 to 10 mm in diameter [9]. The tympanic membrane is made up of three layers: the outer layer, continuous with the skin; the inner layer, made of mucous membrane; and a layer of fibers between the two, that give the membrane its tension and stiffness. In Figure 2, the structure of human ear can be seen and the placement of tympanic membrane along with other vital components of the ear. In Figure 3, a closeup of tympanic membrane can be seen with it's important parts labelled.

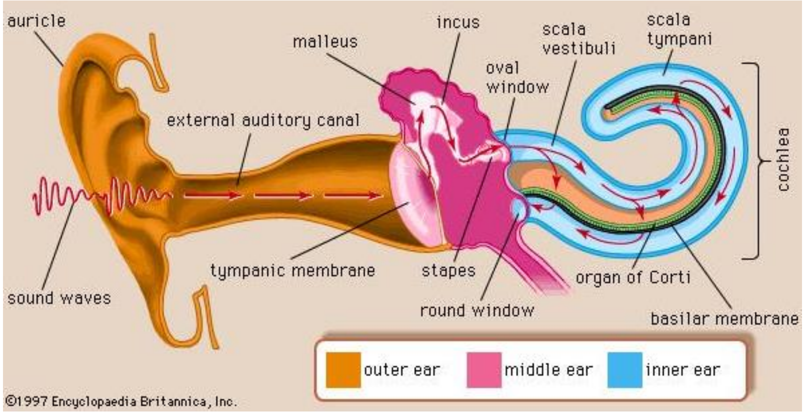


Figure 2 - Structure of Human Ear, Reprinted from Brittanica.com [10]

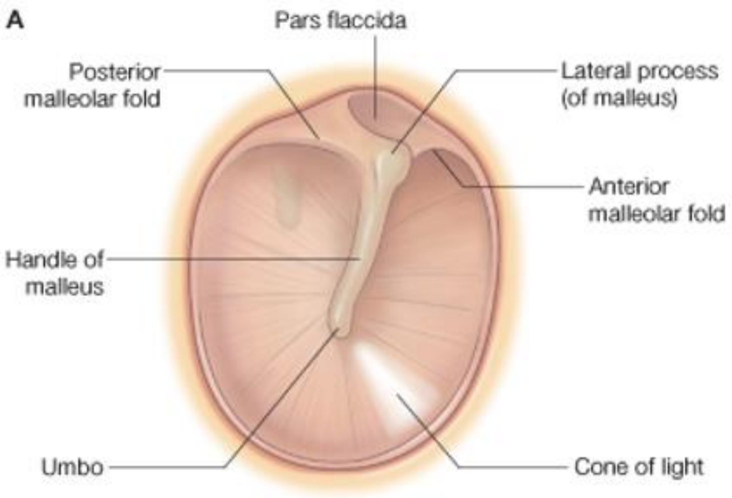


Figure 3 - Tympanic Membrane with all the important parts labelled, Reprinted from drmeu.com [11].

The main function of the tympanic membrane is to assist human in hearing. It receives sound waves from the outer air which then hits the tympanic membrane making it vibrate.

Then the vibrations are transmitted further to the auditory ossicles which are tiny bones in the middle ear cavity. The event can be seen above in Figure 2.

The normal tympanic membrane consists of several features which can help distinguish between middle ear infection and the normal TM. The features are clearly visible for normal TM and tympanic membrane with otitis media. Although, different types of otitis media have different features. For example, acute otitis media has different features compared to otitis media with effusion. The features are further discussed below in detail.

3.1 Features of Normal Tympanic Membrane

Normal tympanic membrane has several properties which makes it distinguishable from the infected TM. Section 3.1 gives an overview of those properties compared to otitis media which can be seen below in Table 1.

One of the key feature that makes normal tympanic membrane unique is the colour of the membrane. Normal eardrum has a unique pearl grey or white colour (see Table 1). Along with the unique colour, the tympanic membrane is also translucent and shaped concave, which is not the case when the ear is infected. Another important feature includes light reflection also known as cone of light, which can be observed easily on a normal tympanic membrane. Moreover, properties like movement can also be noted which is not very common during otitis media. Most of these features such as colour, translucency, shape, and reflection of light can be clearly seen in Figure 4 below.

Figure 4 - Normal Tympanic Membrane, reprinted from EbioMedicine [12]

One of the most significant feature is the Malleus bone visibility. The Malleus bone can be seen in Figure 4 highlighted by a red line. There is no wax, perforation or fluid present in the normal tympanic membrane, which helps us diagnose with better accuracy.

3.2 Features of OM infected Tympanic Membrane

Otitis media has its own key features which can be easily observed on the tympanic membrane. As there are different kinds of otitis media, that means there are different features for each category. Some of the most important types such as AOM and, OME and their features are discussed below in comparison to Normal tympanic membrane.

Acute Otitis Media

Acute otitis media is the most common and has several key features which can be observed and used to distinguish between AOM infected tympanic membrane and normal tympanic membrane. Most of the main features of AOM in comparison to normal TM are listed in Table 1.

The first two most important features to discuss in AOM are shape and colour of tympanic membrane. The shape of tympanic membrane is bulging which compared to normal is very different. Moreover, the colour of AOM tympanic membrane is also most likely to be red in most of the cases in comparison to pearly white in normal ear.

Other features which are unreliable for diagnosis are fluid, light reflection, and wax. The reason being as these features are also seen in normal tympanic membrane and OME. Based on these three features, it is hard to determine as to which category of infection they belong to, so it is better to consider all information or features available to make correct diagnosis. Some of the features discussed above are shown in Figure 5 below which display an AOM infected tympanic membrane.

Figure 5 - AOM infected Tympanic Membrane, reprinted from hawkelibrary.com [13]

Figure 5 clearly shows all the AOM features discussed in the section. Red colour of the tympanic membrane can be noticed and so as the bulged shape. The central concavity in the middle of the tympanic membrane can be clearly seen as well. Moreover, the image also shows the light reflection which is spread, whereas no fluid is spotted.

Otitis Media with Effusion

Otitis Media with Effusion is the type of OM where the fluid is collected behind the tympanic membrane. OME in general does not have many distinct features of its own. Although there are some noticeable properties which can be used to categorize otitis media with Effusion. See Table 1 for the key features of OME compared to normal tympanic membrane.

Some of the most important features of OME are shape of tympanic membrane and fluid. Retracted shape of tympanic membrane during OME is an important feature but the attribute that completely separates OME from normal or AOM is the presence of fluid behind TM. Features such as perforation and wax do not play any significant role in detecting OME from other otitis media.

The fluid is present behind the TM in the form of bubbles which can be noticed. Other features such as malleus bone visibility and light reflection along with presence of fluid can play an important role to support the argument that OME is present.

Figure 6 -OME infected Tympanic Membrane, reprinted from EbioMedicine [12]

Figure 6 shows an image of OME tympanic membrane. It shows small bubbles of fluid behind the tympanic membrane pointed with arrows. Other features, such as light reflection and malleus bone can also be noted.

Table 1: Eardrum Normal vs Infected Properties, adapted from [1]

Properties	Normal TM	AOM	OME
Color	Pearly white	Red	Red or opaque
TM Shape	Concave	Bulging	Retracted
Malleus Bone	Yes	No	Yes
Light Reflection	Yes, well defined	Spread or gone	Spread or gone
Fluid	No	No	Yes
Movement	Sensitive	Impaired or gone	Impaired or gone
Wax	Can be present or not	Can be present or not	Can be present or not
Perforation	No	No	No

Table 1 summarizes the characteristic features of normal, AOM and OME tympanic membrane in order of importance. To detect OM at an early stage the three most important features are the color, shape and malleus bone visibility of the tympanic membrane. Features like fluid, light reflection and movement can also help in OM detection if the decision cannot be based on the first three features.

4 Otoscopes

When a child is suspected of a middle ear infection, usually a paediatrician is consulted. Generally, the symptoms mentioned by parents are ear pain, fever, and discomfort. This leads to an examination by the clinician usually with an otoscope. Chapter 4 discusses different type of otoscopes and help us understand the major concerns related to diagnosis of otitis media.

4.1 Types of Otoscope

An otoscope is a medical device that allows a doctor for checking the process inside an ear. Otoscope enables to visualize and analyse ear related problems more closely. It provides an illuminated magnification of the ear canal and tympanic membrane, which allows the clinician to examine it properly. There are several kinds of otoscopes and some of them are discussed briefly below.

4.1.1 Hand-held Otoscopes

Otoscope design has not changed much over time. It consists of a head part and handle (Figure 7). The head is made up of simple magnifying lens, light source, and a few dioptres. All otoscopes come with disposable or reusable specula in different sizes, which is inserted into the ear and allows to view tympanic membrane closely thru the magnifying glass.

Otoscopes can be wall-mounted or portable. The portable otoscope is powered by a battery and is easy to carry around whereas wall-mounted otoscopes needs an electric outlet where the power cord can be plugged. Figure 7 shows examples of both otoscopes discussed above.

(a) (b)

Figure 7 – (a) Wall Mounted Otoscope, (b) Portable Otoscope, Reprinted from [14] [15]

4.1.2 Pneumatic Otoscope

Pneumatic otoscope is an examination tool which allow for determining the mobility of the tympanic membrane in response to pressure changes. The tool allows the examiner to puff air into the ear canal and observe the movement of tympanic membrane.

The normal tympanic membrane is noted to move in response to pressure. The reason for immobility maybe due to fluid build-up in the middle ear, sterile or infected, a perforation, or tympanosclerosis, among other reasons. In Figure 8 below, a pneumatic otoscope can be seen.

Figure 8 - Pneumatic Otoscope, Reprinted from welchallyn.com [16]

The pneumatic otoscope helps in detection of tympanic membrane mobility which aids in establishing the diagnosis of OME. It is also cheaper than video otoscopes, ranging between $175 to $230 otoscopes which is an advantage and can be used efficiently with appropriate training.

4.1.3 Video Otoscopes

Video-otoscopes are useful as they can be attached to the computer to record images and videos. The clinicians can then view the video feed on a screen while holding the device in the ear. Video otoscope gives advantage over hand-held otoscopes as the images or videos can be seen in real-time. Some of them even has a screen attached to the otoscope for a better view. An example of the video-otoscope can be seen below in Figure 9.

Figure 9 - Video-Otoscope for imaging and video recording, Reprinted from jedmed.com [17]

Some of these video-otoscopes are expensive ranging between $250 to $800 depending on the quality of imaging system they provide but because of advancement in technology, there are otoscopes which are cheaper but due to a low price they do not provide very high-quality image or video feed.

4.1.4 Smart Phone Video Otoscopes

Healthcare technology is progressing fast which has allowed us to make medical devices smaller for everyday use. Such an example is smart phone video otoscopes, these are small devices which can be attached to smart phone and work as an otoscope. Parents can use these devices at home and send the image to a medical profession via an in-app service provided by the manufacturer instead of rushing to the hospitals. The process results in less unnecessary hospital visitations in case there is no infection. If there are any symptoms of infection present, the medical professional via in app service will advise the user to go visit a doctor.

Smart Otoscope devices allow us to take high quality images and videos. The data can be stored on phone or cloud and used to diagnose if there is infection present or not. An example of smart phone otoscope can be seen in Figure 10.

Figure 10 - Cupris Smartphone Otoscope, Reprinted from cupris.com [18]

The small device attached to the cellphone seen in Figure 10 can be used to have a clear view of tympanic membrane and diagnose otitis media. Smartphone otoscopes cost between $150 to $200. For the software development, some of the images were taken using Cupris device.

4.2 Diagnostic Difficulties and Consequences

One of the major challenge faced during the examination of otitis media is misdiagnosis. It means otitis media can either be under-diagnosed or over-diagnosed depending on the conditions, clinicians, symptoms, and many other factors. The uncertainty in the diagnosis arises from factors like examining the middle ear of young children and especially infants, as they cannot express their feeling verbally. Other factors contributing to the problem are narrowness of auditory canal, inability of the child to remain still which makes it difficult for the clinician to use otoscope efficiently which can lead to unclear view of tympanic membrane.

Moreover, presence of the cerumen in the ear canal may add to the difficulty in viewing, thus making the diagnosis difficult. Another factor that adds to the problem is the interpretation of what you see. Meaning that mostly, for example AOM is diagnosed based on the colour of tympanic membrane, in this case red. So, less experienced clinician will call redness of the eardrum an infection, whereas redness in the ear can also be caused by fever or just crying. All these factors provide us an idea as to how easy and problematic misdiagnosis of otitis media can be, which brings our attention to the issues faced by misdiagnosis of middle ear infection.

Unreliable diagnosis leads to over and under-prescription of antibiotics. AOM is the most common over-diagnosed infection due to the reason mentioned above and sometimes OME can also be diagnosed as AOM which leads to unnecessary antibiotic prescriptions which can have adverse effects, especially on children. Over-diagnosis is more common, as doctors with intention of avoiding the possibility of leaving an untreated patient, result in over-prescriptions. Under-diagnosis is also a problem as it may lead to serious complications such as perforation of TM or even in some cases to unnecessary surgical treatment.

Over-prescription of antibiotics leads to another major problem which is antibiotic resistance, which results from misdiagnosis of OM. It is most common in misdiagnosis of AOM. Antibiotic resistance is a very real and concerning problem. In 2014, a report estimated that by year 2050, up to 10 million people per year could die because of ear infections by antibiotic resistant bacteria [19].

The whole process of misdiagnosis of OM and unnecessary prescriptions of medicine leads to increased financial costs. There are many financial costs associated with misdiagnosis of otitis media. Some of those costs are medication costs, emergency department, co-payments etc.

5 Materials and Methods

Section 5 gives an overview of the necessary steps taken to develop the otitis media feature extraction and image classification software system. It briefly explains why these steps are important to have a successful system and helps us understand the importance of such a software.

It is known that otitis media is a serious disease which needs to be diagnosed and treated on time, but the important factor here is just not time but also accuracy. There is no big benefit of diagnosing some disease on time, if it is not diagnosed accurately. Accuracy is a key factor that leads to many problems discussed in chapter 4, such as misdiagnosis which is very resource consuming.

The main goal behind developing an automated feature extraction and image classification system is to achieve a high accuracy rate which is a problem when diagnosed by paediatricians and otolaryngologists. There have been multiple studies to evaluate the accuracy. In a study [20], where these professionals were evaluated by showing them video images and asked to declare their diagnosis results, the average rate of correct diagnosis carried out by paediatricians was 50% and the average rate for correct diagnosis of otolaryngologists was 73%.

Another study [21], showed the results of testing skill level of paediatricians from different countries. The results came out to be concerning, showing correct diagnosis performed by paediatricians in Italy was 54%, Greece 36%, South Africa 53%, and USA 51%. The results and all the key problems discussed above clearly underline the importance of a software system which can help clinicians by providing them a second more reliable opinion when making diagnosis.

5.1 Dataset

Otitis media image dataset is the key element for the development of both software systems. The dataset used during the development process was collected from two sources. The first dataset contained 20 high quality images which was obtained from a researcher in the field of public health and clinical medicine in Sweden. The dataset containing 12 AOM and 8 normal tympanic membrane images mainly helped in developing the feature extraction software.

The second dataset which contained 88 high quality images was collected using Google Images. The dataset contained 48 AOM and 40 normal tympanic membrane images. For image classification software, both datasets were used resulting in a total of 108 images. All the images in the final dataset were between the dimensions of 300 x 300 and 1440 x 1080. Finally, all the images were manually labelled as either AOM or normal.

5.2 Pre-Processing

Pre-Processing is the part where the first software based on image processing and computer vision starts to take shape. Before any features can be extracted from the images, several pre-processing techniques were performed on the images. It was done to standardize the image, remove unnecessary parts from image and reject any unreliable data which could affect the feature extraction results. Pre-processing stage is important as it helps improve the accuracy of the feature extraction stage by eliminating excessive and unreliable data from the images. The methods and algorithms used are discussed in depth in the next chapter.

5.3 Feature Extraction

After the pre-processing of the images, the next step was to extract features from the images which plays the main role in the diagnosis of otitis media made by the professional. Several methods were studied from researches in past years such as [22] and [12] which presented several methods of feature extraction. Also, many other methods were studied from the field of image processing and computer vision. The studied methods were then accessed, and the best suitable methods were then selected to use as extracting features from images. Different feature extraction methods are discussed in detail in chapter 7.

5.4 Automated Classification System

The idea of having an automated system is to have as little human interaction with software system as possible. The automated system was divided into two stages which are discussed in chapter 9. First being the data augmentation stage, and second being the image classification and prediction stage. The goal was to achieve automated diagnostic system using state of the art machine learning techniques involving deep learning and transfer learning. The results of making the software system automatic are discussed further in chapter 9 after examining the methods used for pre-processing and feature extraction stage.

5.5 Software Development Tools

The software was developed using Python programming language [23] and other computer vision, Image Processing, and machine learning libraries and frameworks. Python is a widely used high-level programming language for general-purpose programming. Python version 3.6.5 was used in development. Software development was carried out using free and open source distribution of python called Anaconda [24]. All the packages and libraries needed were installed using Conda, which is a package management system in Anaconda. Furthermore, the most important library used for computer vision was OpenCV [25]. OpenCV is the most widely used computer vision library with over 47000 developer’s community and more than 2500 optimized algorithms.

For basic images processing libraries such as Scikit-image [26] and Python imaging library (PIL) [27] were used. The libraries helped perform different tasks such as image loading, filtering, morphology, segmentation, color conversions, and transformations. The graphical user interface (GUI) was developed using Tkinter library [28]. It is the most commonly used GUI library although it is not the only one, there are several other libraries which can be used. A simple GUI library like Tkinter was used because a very simple interface was needed for the user to analyze each feature separately and generate numerical data corresponding to that analysis. More details about GUI are discussed in chapter 7. The latest versions of all the libraries were used during development.

To develop image classification system, a broader family of machine learning (ML) methods known as DL (deep learning) was used which is further discussed in detail in chapter 9. Moreover, supervised learning method was used which was further broken down into classification problem which helped us divide the images into their specific class or categories. There were mainly two libraries used to implement deep learning algorithms which are TensorFlow and Keras.

TensorFlow is an open-source software library for dataflow programming across a range of tasks. It is a symbolic math library and used for machine learning applications such as neural networks. TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open source license on November 9, 2015. [29.]

Keras is an open source neural network library written in Python. It can run on top of TensorFlow. Designed to enable fast experimentation with deep neural networks, it focuses on being minimal, modular, and extensible. [30.]. Due to its minimalism and fast implementation, keras was used to train and test deep learning algorithms quickly as to understand the model based on our application.

6 Image Pre-Processing Results

The section discusses in depth as to which methods and algorithms were used to develop the pre-processing stage of the otitis media diagnostic system. There were several stages ranging from basic image processing to advanced computer vision techniques. The main idea was to take the image of middle ear and segment the tympanic membrane part from the image. After segmentation an illumination correction stage was applied if needed under certain conditions.

If the image passed all the stages then the pre-processing stage was completed, and the image was ready for further computation, in this case, feature extraction. The stages are further discussed along with the methods used to compute them and results obtained are also shown.

6.1 Automated Segmentation of TM

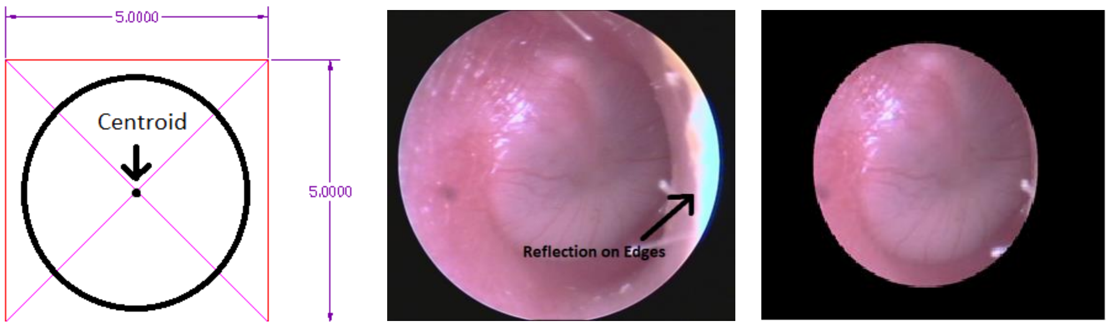
The first and most important stage of the analysis system is to segment out the tympanic membrane area of the middle ear from the rest of the image which may include some part of ear canal or reflected area from specula attached to the otoscope. The two main algorithms used for segmentation stage are Image moments and Snake: Active Contours based segmentation. The idea behind these algorithms and the result obtained from them can be seen below. Although, at the end, for the final system, centre of mass segmentation method was chosen, reasons behind the decision are discussed in section 6.4.

6.1.1 Image Moments Based Segmentation

To explain image moment-based segmentation, it is important to understand the idea of **centroid** which can be calculated using a method in OpenCV called **image moments**. In computer vision and image processing, image moments are often used to characterize the shape of an object in an image. The moments capture basic statistical properties of the shape, including the area of the object, the centroid (i.e., the center *(x, y)*-coordinates of the object)

Centroid or geometric center of a plane figure is the arithmetic mean position of all the points in the shape. As all the images in the dataset are square and the tympanic membrane is always positioned in the middle of the image, we can use this information to calculate centroid or geometric centre of the square. Figure 11 (a) shows how the centroid of a square looks like with a circle inside it. The black circle can be considered as tympanic membrane. The intersection of diagonal lines in a square gives us the centre point of the image which was used as a reference point to draw a circle of a specific radius to segment the tympanic membrane.

The method allowed us to segment images of any size and type taken from any kind of otoscope. Although the radius differs, meaning it becomes bigger or smaller depending on the size of the input image. The radius was chosen after thorough analysis of many images and depending on how much area do we need for further analysis. The actual input image and the resulted image from the segmentation can be seen below in Figure 11 (b) and (c).



(a) (b) (c)

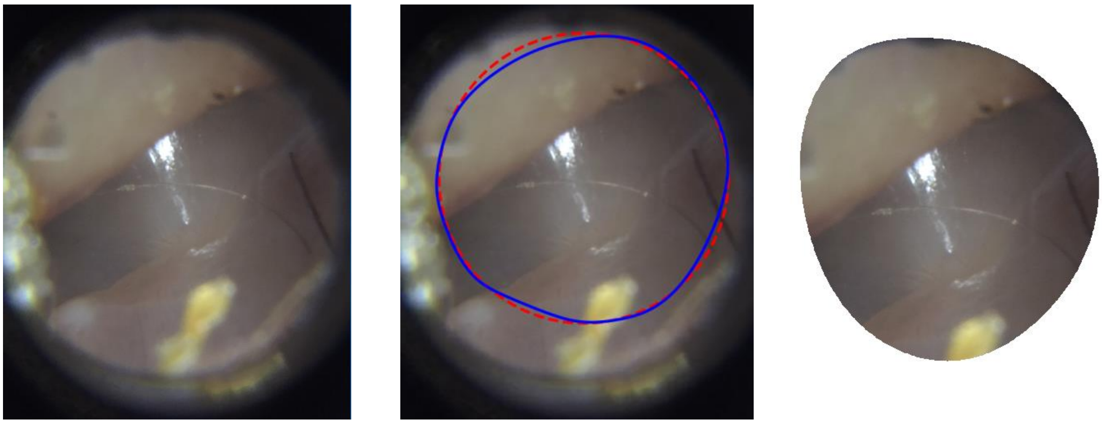
Figure 11 – Image Moment Based Segmentation: (a) How to calculate Centroid of a Square, (b) Original Image, (c) Segmented Image using Image Moment methods

As seen in the Figure 11 above, the tympanic membrane was segmented from the whole image, discarding all the unnecessary information like reflection on the edges from the otoscope specula. Although, output image was left with the extra black pixels on the background, which was dealt with using image cropping in section 6.2.

6.1.2 Snake: Active Contours

Another image segmentation method used during the software development stage was an active-contour based segmentation. The active contour model is a method to fit open or closed splines to lines or edges in an image. It works by minimizing an energy that is in part defined by the image and part by the spline's shape: length and smoothness. The minimization is done implicitly in the shape energy and explicitly in the image energy. The method is based on an algorithm proposed by Michael Kass which can be reviewed here for more details [32].

The algorithm is initialized by a circumference in the center of the image. It then iteratively grows the contour and stops at a predefined convergence standard, which leaves us with an outline which is further processed as a mask which can then be segmented out of an image to obtain tympanic membrane part. Figure 12 below shows an example image on which the algorithm was applied, and the following results were obtained.



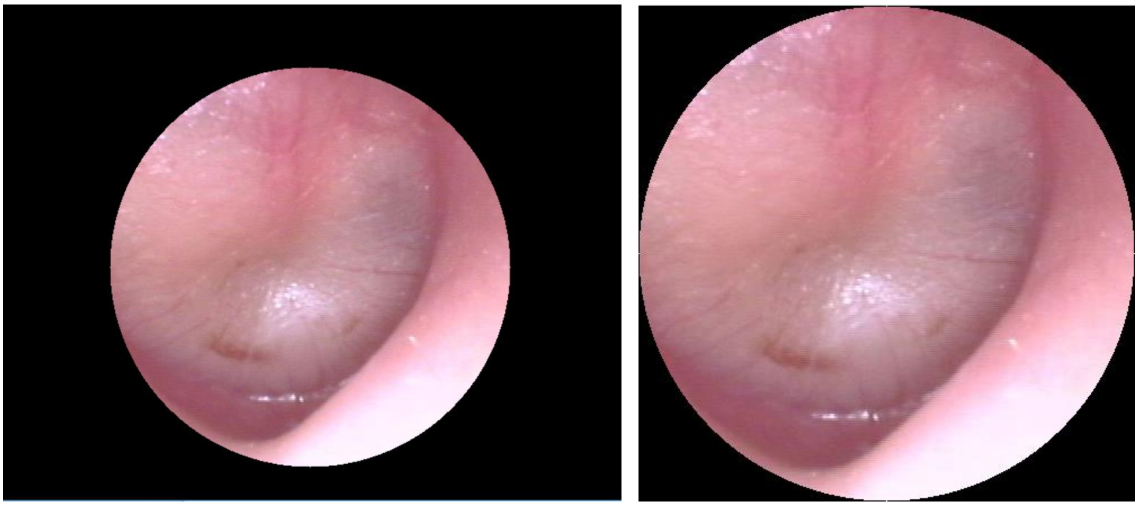
(a) *(b)* *(c)*

Figure 12 - Image Segmentation using Active Contour: (a) Original Image, (b) Image with snake contours drawn, (c) Segmented Image.

Figure 12 (b) shows the initialized circle in red colour, whereas the snake generated can be seen in blue colour outline which was further processed to segment tympanic membrane part shown in Figure 12 (c). There are several parameters that can be changed according to what is needed. For example, number of iterations can be increased or decreased, the initial circle size can be changed and many other factors.

6.2 Image Cropping

After the segmentation of tympanic membrane, the resulting image had unnecessary black background pixels which needed to be removed. The resulting image can be seen below in Figure 13 (a).



(a) *(b)*

Figure 13 – (a) Segmented Image with black borders around TM, (b) Cropped Image

To solve the problem, basic image processing technique called image cropping was used. It was performed using PILLOW library for python. Figure 13 (b) above shows the resulting image from using cropping method.

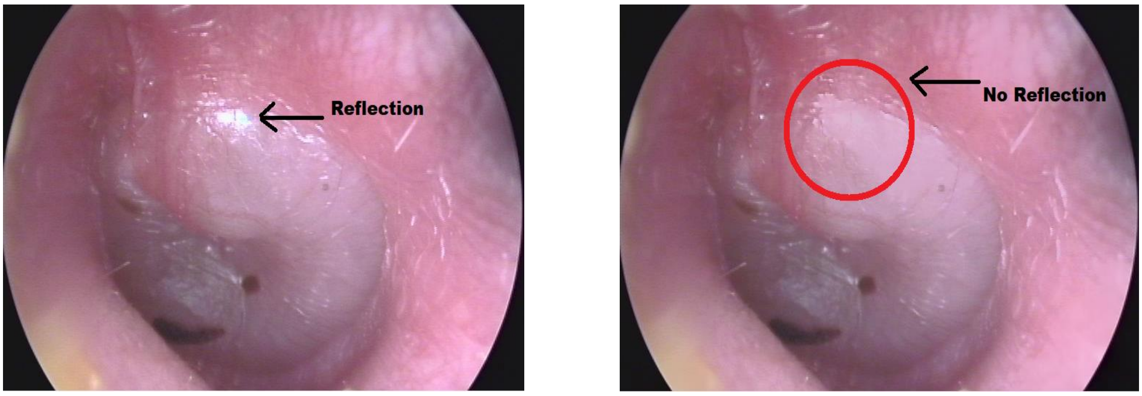
The idea behind the method is that the program searches for the first non-black pixel from each side and marks it as the cropping index. The image is then cropped by keeping only the part of the image that lies within the boundaries of the cropping indices. The black pixels in Figure 13 (b) are less than compared to Figure 13 (a).

6.3 Illumination Correction

A problem very often encountered with tympanic membrane images are specular reflections. Specular reflection can be caused by wax in the ear canal or on the surface of hair. To make sure these reflections do not create a problem while running analysis on image, it needs to be solved. The method adopted to solve illumination problem is Poisson image editing technique, which is further discussed below.

6.3.1 Poisson Image Editing

Poisson image editing technique [33] or more specifically a method called local illumination changes was used to correct the specular highlights in the images. Before applying the method, the white pixel areas were calculated using a simple thresholding technique on image intensities. Furthermore, the illumination change method was applied on the threshold image to gain the results shown below in Figure 14. Since there was no image with specular highlights caused due to wax, instead an image with light of cone was used with high white pixel intensity area which shows the same result as correcting specular reflection caused by wax. It is believed that illumination correction phase will work successfully on images with specular highlights.



(a) *(b)*

Figure 14 - Illumination Correction using Poisson Image Editing: (a) Original Image, (b) Corrected Image after using Poisson Image method

It can be seen in Figure 14 that illumination correction method works fine in correcting the light reflections. The corrected area in Figure 14 (b) can be seen inside the red colour circle which is not as illuminated as original image in Figure 14 (a).

6.3.2 Rejection Due to High Illumination

Due to amount of high illumination in some images which affects the analysis of the diagnostic system, it was decided to reject images which has high illumination due to specular highlights or light reflection from otoscope specula.

To solve this problem, same thresholding method as for illumination correction was used to detect white pixels. Then the percentage of white pixels compared to the whole image was calculated. Based on these calculations and after analysis of images it was decided to set a range of threshold percentage values. Three categories were defined which are as follows:

- Do not apply Illumination correction if percentage of white pixels is less than 10%.
- Apply illumination correction if percentage of white pixels is greater than 10% and less than 25%.
- Reject image and prompt user to retake an image if percentage of white pixels is greater than 25%.

All the pictures which showed an abnormal percentage of white pixels, in this case more than 25%, were all rejected. There were several percentage values tested with the data between 5% to 30%, but these specific percentages were chosen because if the percentage is lower than 10%, the program might consider light of cone which is a feature in diagnostic system as a bad reflection and correct it, which we do not want. If the percentage is too high, the image is rejected as due to high illumination it might affect the other feature.

6.4 Rejected Pre-Processing Method

After testing all the methods and algorithms mentioned above for pre-processing another important factor was considered which is performance of the whole system. The idea of good performance for the analysis system is to minimize the amount of time taken to run pre-processing stage and maximize the accuracy of the results. Based on these two factors, snake: active contour algorithm was rejected.

Snake: Active Contour is a great algorithm for drawing contours and segmentation of desired part of images. It gives you a lot of control over the parameters which can be changed, and desired contour can be drawn. Although the drawback of the contour method in the application is the processing speed.

The method was tested on many TM images and the average processing time just for contour method was around 7 seconds on the computer used for developing analysis system. It was due to the number of iterations the algorithm must run to draw the contour. For the system, the minimum number of iterations that must be run were n = 450. Below 450 iterations, the results were not acceptable. However, the image moment-based segmentation method chosen instead of active contour method showed a remarkable increase in performance and only took 1.5 seconds on average to run the tympanic membrane segmentation process.

Moreover, an increase in performance not just in terms of time but also in accuracy was seen. As active contour method produced irregular shaped circles of different sizes at different points on the images, it was always expected of the algorithm to leave some important information outside the boundary. In the chosen segmentation method, the size and point where the circle is drawn was pre-defined. Every circle was the same size which gave us the same size of segmented image of tympanic membrane.

Illumination correction was rejected not because it has processing time issues but the algorithm in general does not perform well with correcting image as it was first expected. Working of Poisson image editing method varies from image to image, for example, the main idea is to fill the target region with pixel values obtained by interpolation of pixel values along the boundary of the target region. But in this application, where the image is opaque or very dull coloured, when the boundary pixels are interpolated with the target region, not much difference is seen due to not varying intensities between pixels.

However, if the image was brighter, like an image of tympanic membrane which was more red than opaque, a clear difference was seen in the illumination correction algorithm as it interpolates the red pixel intensity values on target region and clear difference can be seen.

The method has several parameters which can be tuned to get the desired result, in this case the two most important alpha and beta were set in a way that good illumination correction was achieved. The conclusion being, the illumination method works fine with more brighter images than compared to opaque coloured, typical for TM images.

7 Feature Extraction Results

The section discusses in detail the feature extraction methods and algorithms used to extract features. Altogether, five different features were extracted. The main idea behind solving the feature extraction problem was to use pre-processed images and apply certain basic to advanced algorithms to them and extract the visual and numerical data needed for the professional to manually classify the images.

To make sense of the visual and numerical data, a range of values or threshold for the normal and infected ear features were analysed and established. By using the values, a professional can compare the values generated by the software with the table 2 provided with values. After the comparison of values, the medical professional can manually classify the images into their respective class.

The table 2 shows the established threshold values for different features. The table includes the values for both normal and AOM infected ears. The threshold values were carefully determined after testing many images. How these values for individual features were analysed and established is further discussed in the chapter.

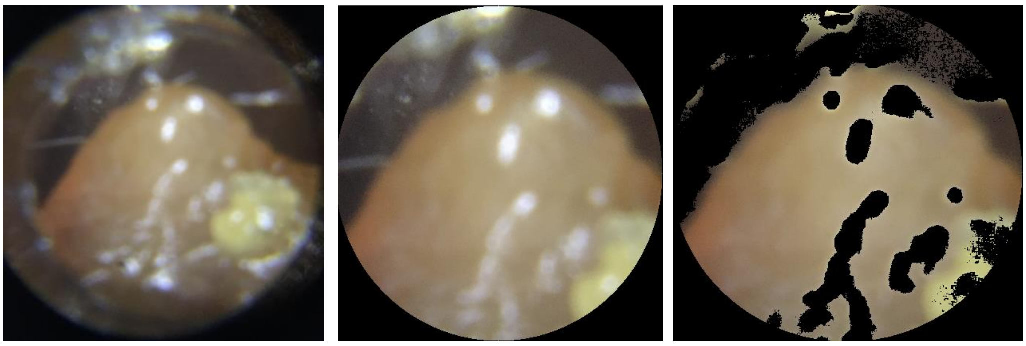
Table 2 - Threshold values for extracted features

FEATURES	NORMAL THRESHOLD	AOM THRESHOLD
Wax/Cerumen	Wax ≤ 10%	Wax > 10%
TM Color	Color ≤ 15%	Color > 15%
TM Shape	Mean Grad ≤ 80	Mean Grad > 80
Light Reflection	Circle Detected ≤ 2	Circle Detected ≥ 3

7.1 Wax Analysis

Detection of wax or cerumen is one of the features in otitis media diagnostic system. Build-up of wax in the ear canal can obstruct the view of tympanic membrane thus leading to inadequate visualization. Diagnostic system run on such tympanic membrane images might lead to unreliable results. The method used to detect wax in the ear canal is discussed below.

A simple and fast method called colour thresholding also known as colour filter was applied to determine the amount of wax in the ear canal. The process was carried out after the tympanic membrane pre-processing stage. The idea behind the method was to apply yellow threshold to HSV (Hue, Saturation, Value) image obtained from the segmented images. The threshold allowed only the yellow or wax part of the image to be visible. The method can be seen applied on the image below in Figure 15.



(a) *(b)* *(c)*

Figure 15 - Wax Detection using Color Detection: (a) Original Image, (b) Segmented Image, (c) Wax detection

Figure 15 (c) demonstrates the colour thresholding method used to detect wax obstructing tympanic membrane. It also shows that only the yellow colour in this case wax could pass thru the filter.

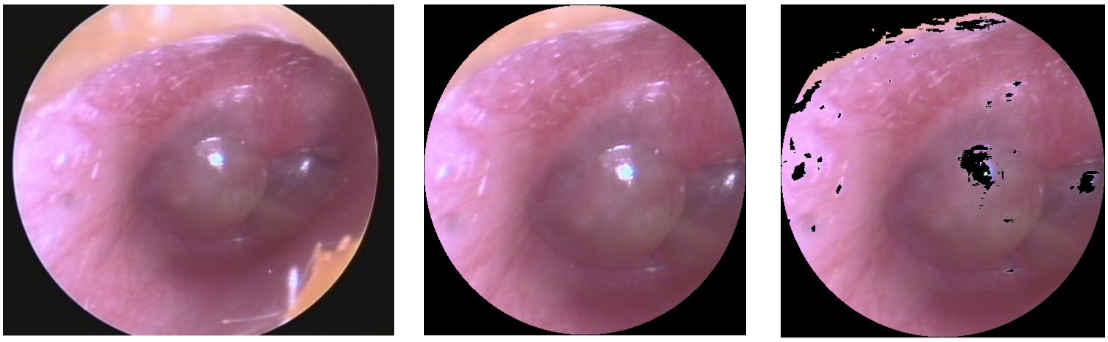
After the wax is detected in the ear canal, another stage is introduced which is responsible for calculating the percentage of yellow pixels in that image and categorizing it as a wax obstructed image, labelling it as the wax image, if it passed a certain threshold level.

After analysis of more than 40 wax obstructed tympanic membrane images, it was decided that the threshold of 10% for allowable amount of yellowness in an image to be passed forward for further processing. Any image that showed an abnormal percentage or above the 10% threshold must be considered as a wax obstructed TM image. The image in Figure 15 was declared unsuitable for making a diagnosis about infection due to an abnormal percentage of wax blocking the view of tympanic membrane and thus labelled accordingly.

7.2 Tympanic Membrane Colour Analysis

One of the most important features in diagnosing an AOM infection from a normal ear is the colour of the tympanic membrane. Among other features mentioned in Table 1, colour feature is easy to perceive as the colour difference between normal and AOM infected ear can be immediately observed by human eyes. But for the computer, it’s not that obvious and for that reason a computer vision approach was used to detect the redness in an image and label it as infection or normal.

The same colour threshold method as in wax analysis was used here. The image was first segmented and then cropped to remove any excess data outside of tympanic membrane. Then an HSV (Hue, Saturation, Value) image was generated from the cropped image on which the colour filter was applied which allowed the software to only let the red colour in the image to pass through the filter, thus giving us an accurate amount of redness in the image. The results of the method applied can be seen in Figure 16 (a - c).



(a) (b) (c)

Figure 16 - Redness Detection using Color Threshold Methods: (a) Original Image, (b) Segmented Image, (c) Redness detected

Figure 16 (c) shows the redness detected in the image with high accuracy, leaving behind other colours such as yellow on the border and white reflection in the middle. To identify an image as normal or infected using colour analysis feature, it was important to set some threshold value.

To set the threshold, first the percentage of red pixels was calculated in the whole image after segmentation and then it was decided after testing several images to set the threshold percentage to 15% for allowable amount of redness in an image. Therefore, if the image has 15% or less red colour, it was labelled as a normal TM. If the threshold is crossed, the image is then labelled as AOM infected. An output snippet generated by the software can be seen in Figure 17 showing the total number of pixels in a segmented image and the total number of red pixels along with the percentage of redness found in the image shown in Figure 16 (c).

```
--------------------------------------------------
Number of Total Pixels =  194481
Number of red Pixels =  142823
Percentage of red Pixels =  73.4380222232506
--------------------------------------------------
AOM Detected
--------------------------------------------------
```

Figure 17 - Output Snippet from TM Colour Analysis Output

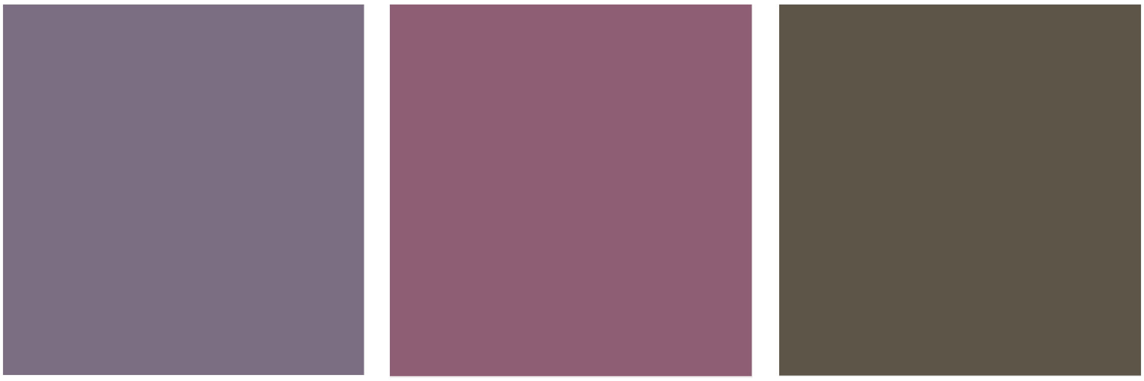
7.3 Average Colour Analysis

Another important feature that can be related to the colour domain of the image is the average colour of an image. The colour of the whole image is analysed rather than just the tympanic membrane which allows us to take in account the surrounding areas of it. It is important as different infections and symptoms demonstrate different colours.

The concept behind extracting average colour feature was simple and fast. It was achieved by first segmenting the image to a specific radius. The radius for the feature was calculated by analysing several images, it was made sure that the radius should be enough to cover as much area as possible only leaving out unnecessary high reflections behind at the border (see Figure 11(a)).

After the image was segmented, the Red, Green and Blue (RGB) channels of the image were extracted, and the average of all pixel was calculated. Moreover, the average mean of the RGB channels was also calculated. At the end, an image was produced with the average colour of the ear which can be seen in Figure 18 (a) to (c).

The Figure 18 (a) shows the average colour of a normal tympanic membrane, Figure 18 (b) shows the average colour of AOM and finally Figure 18 (c) shows the wax obstructed tympanic membrane.



(a) *(b)* *(c)*

Figure 18 – An example of Average Colour of Image: (a) Normal TM, (b) AOM Infected, (c) Wax Obstructed

The difference between the average colour of normal, infected and wax obstructed images can be seen in Figure 18. The information was further used to label the images accordingly.

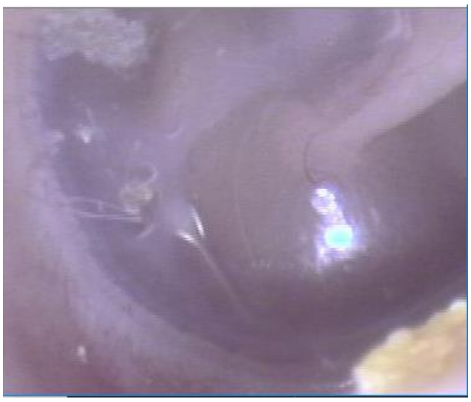
7.4 Light Reflection Analysis

Light reflection also known as cone of light is another feature which can be used to make an analysis on tympanic membrane. Light reflection is normally caused by reflection of the light from otoscope on the TM. Most of the time, light reflection is only found on normal TM compared to the AOM infected as the light reflection is spread or gone on the AOM infected TM which makes the reflection feature valuable during the analysis.

Although sometimes, the light reflection can also be seen on AOM infected TM but most of the time, it is not the case, which is the reason for using several important features to make a final analysis and not depend on just one feature.

Extracting the feature involved several steps, starting by cropping a specific sized rectangular area of the image called ROI (Region of Interest) close to the centre of tympanic membrane. The idea behind cropping ROI was determined by the fact that the reflection always appear around the centre of TM. More specifically, the light reflection appears near the 4 o'clock to 5 o'clock position in the right eardrum and the 7 o'clock to 8 o'clock position in the left eardrum [34]. The original image and the cropped area of a normal tympanic membrane can be seen in Figure 19 (a) and (b).

(a)

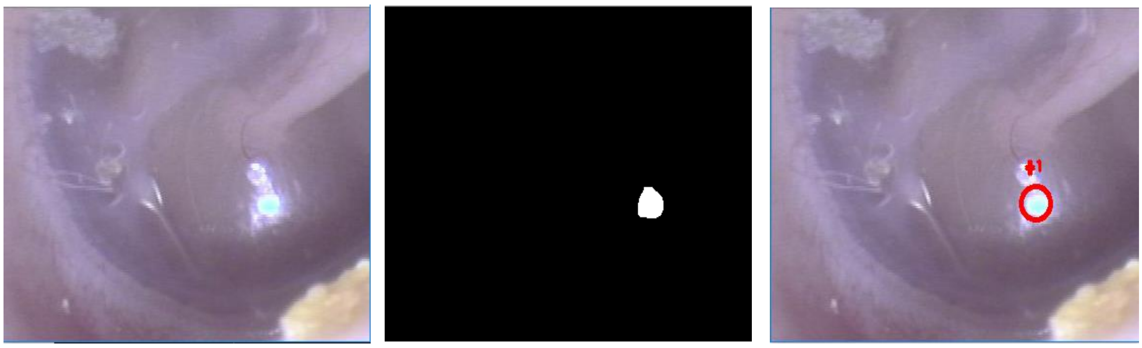


(b)

Figure 19 – ROI for Light Reflection: (a) Normal TM, (b) Cropped ROI

After cropping the ROI of an image, the next step was to use binary threshold method on the ROI to reveal light regions in the image. Moreover, erosion and dilation techniques were used to remove any noise in the form of light present in the image. The number of white pixels were then calculated in the image and by testing of several images it was found that the range for number pixels between which the reflection usually lies is 200 to 600 white pixels.

Using the information, any pixels below or more than the range were discarded and the pixels between these range were used to create a mask. The binary image or mask was then used to find contours and once the contour was found, a circle was drawn around it to show the reflection area. The ROI, mask and the final image with reflection detection in red circle is shown in Figure 20.



(a) (b) (c)

Figure 20 – Light Reflection Analysis: (a) ROI TM, (b) ROI Mask, (c) Reflection Detected

To label the images it was analysed and established that inside the pixel range determined if the image has one or two detected circle, it was labelled as normal, as it is common to see one to two reflection circles in a normal tympanic membrane. If the image has none or more than 2 detected circles, the image was labelled as infected, as sometimes the reflection spread out causing numerous reflection points.

7.5 Shape Analysis

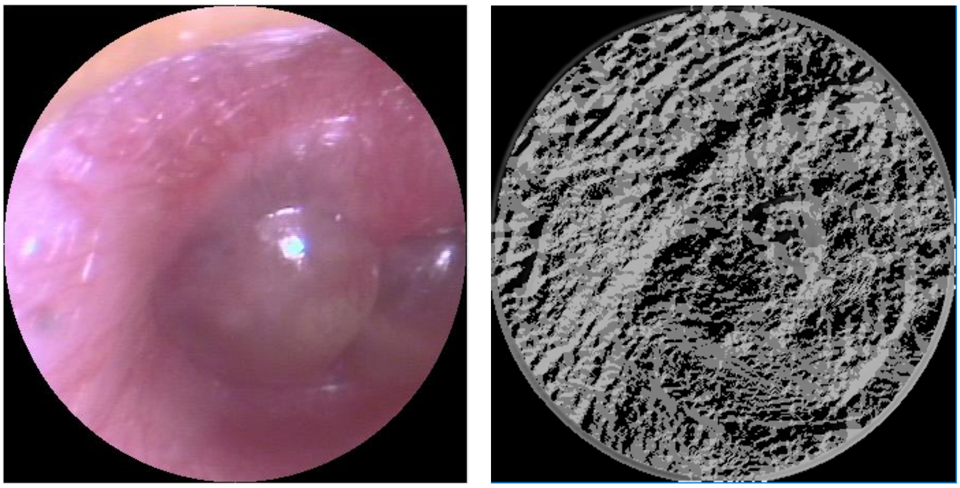
The most significant feature in differentiating between a normal and AOM infected ear is the shape of tympanic membrane. In an AOM infected ear the shape of tympanic membrane is always bulged compared to normal where it's concave. The feature is considered the most important as the shape property of the TM is mostly observed during AOM infection.

To extract the feature, both the normal and AOM infected tympanic membrane were analysed. The method used was named as 'Gradient Analysis'. To perform gradient analysis, first the tympanic membrane area was segmented from the whole image as described in automated segmentation stage. The next step was to use histogram equalization technique to adjust the image contrast or distribute the intensities of the image more evenly.

After the histogram equalization stage, the gradient or the difference in greyness between neighboring pixels of the image in horizontal (X) and vertical (Y) direction was calculated. The gradient was computed using second order accurate central differences in the interior points/pixels and either first or second order accurate differences at the boundaries [35]. Moreover, the magnitude of the gadient was calculated using the following equation:

$$M = \left(\sqrt{gradX^2 + gradY^2}\right) \tag{1}$$

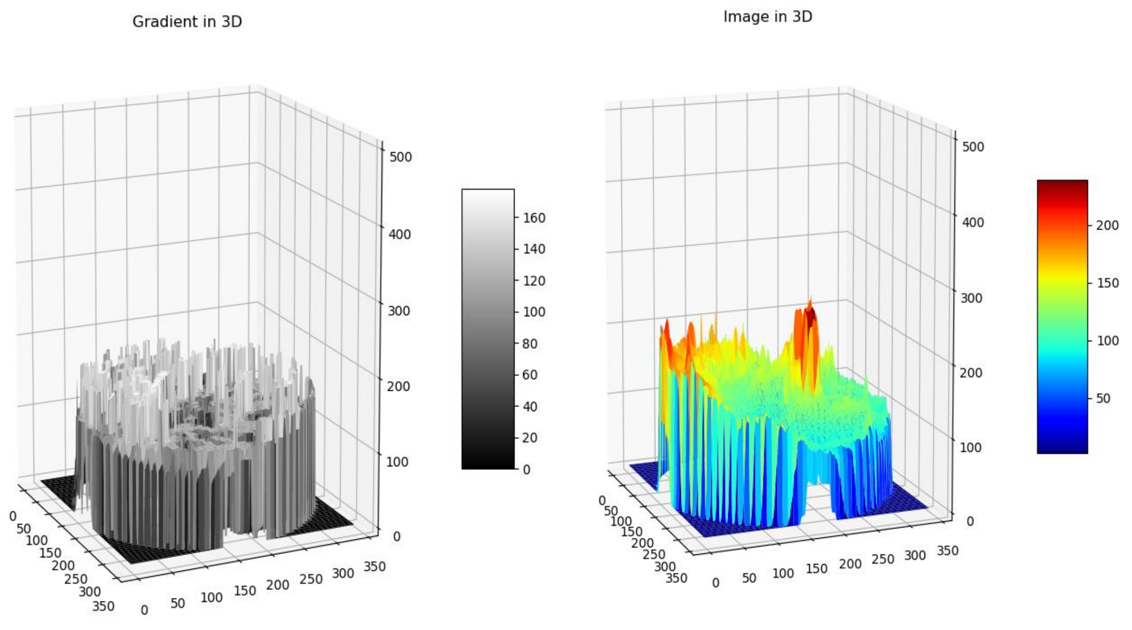
In equation (1), magnitude is denoted by M, whereas gradX and gradY are the gradients in horizontal (X) and vertical (Y) direction. The results of the gradient magnitude calculated on an image can be seen in Figure 21 (b).



(a) *(b)*

Figure 21 – Gradient Magnitude: (a) Original Image, (b) Magnitude of Gradient

Furthermore, the magnitude was used to plot a 3D figure of the gradients and a 3D figure of the original image was also plotted. The figures can be seen in Figure 22 (a) and (b).



(a) (b)

Figure 22 – Shape Analysis 3D plots: (a) Gradient in 3D, (b) Original image in 3D

All the gradient values in X and Y direction can be seen in Figure 22 (a) as a 3D plot and the image coversion from 2D to 3D can be seen in Figure 22 (b). The original image used to draw these 3D plot is the same as in Figure 21 (a). The spikes in both figures are caused by the higher intensity or brighter regions. In terms of gradient it can be explained by the change in level of grayness from dark to bright. The spikes represent the brightest region of the image whereas the darkest region can be seen at value 0 on z-axis as the base of the image.

The final step in the shape feature analysis was to have some numerical data on the bases of which normal ear can be differentiated from the infected ones. In order to do that, the mean value of the gradients was calculated. By analyzing images and using these mean values a range was created for the mean values of normal and AOM infected ears.

As seen in Table 2, all the mean values below and equal to 80 were categorized as normal ear and all the value above 80 were categorized as AOM infected tympanic membrane.

7.6 GUI for Feature Extraction

For users to perform pre-processing stages, analyse features easily and make a final diagnosis based on the visual and numerical data, a graphical user interface (GUI) was developed using python programming language and tkinter library. The main target users were researchers, ENT specialists, nurses and otologists who would like to study or analyse the features of tympanic membrane and classify them between normal or infected.

The GUI developed was very basic with all pre-processing and feature extraction options included. All the stages or options can be used by a click of a button. In total there are eight buttons in the GUI, two of which can be used to run pre-processing stage and see the visual difference between the original and the processed image, the other five buttons can be used for feature extraction and analysis. The feature analysis buttons give out a visual output as well, showing the original image compared to the feature extracted, so clear difference can be analysed, and the image can be classified manually. The last button on the GUI is used to analyse the important numerical data such as percentages of yellow colour for wax and red colour for tympanic membrane. It also shows the numerical value of mean gradient for shape analysis and the average colour for an image. The GUI can be seen in Figure 23, showing the main menu where the features can be accessed.

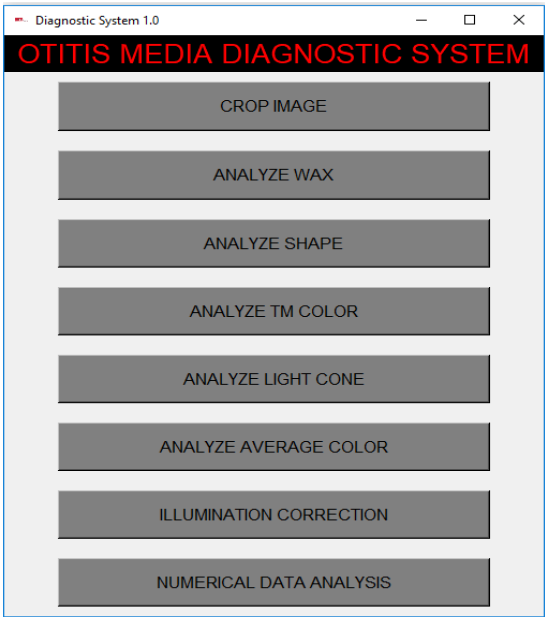


Figure 23 – Graphical User Interface for Analyzing Image Features

Moreover, in Figure 24 and 25, the function of the two of the features (shape analysis and numerical data analysis) running through the GUI can be seen. When both features are run, they open a separate window showing the 3D plots for shape analysis and image analysed, data and a reference table for numerical data analysis. For both these features, the main window shows the original image and segmented image. Furthermore, for the other features, the results are shown in the same main menu window.

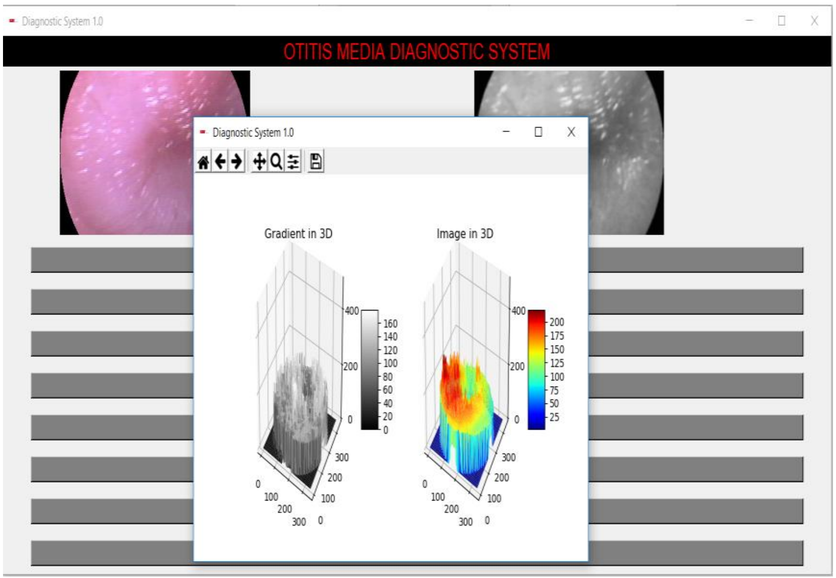
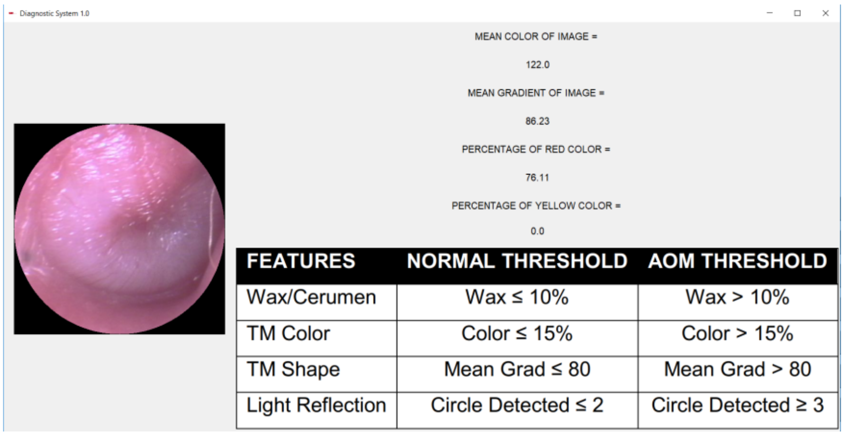


Figure 24 – Shape Analysis Feature running on GUI: showing the 3D plots



FEATURES	NORMAL THRESHOLD	AOM THRESHOLD
Wax/Cerumen	Wax ≤ 10%	Wax > 10%
TM Color	Color ≤ 15%	Color > 15%
TM Shape	Mean Grad ≤ 80	Mean Grad > 80
Light Reflection	Circle Detected ≤ 2	Circle Detected ≥ 3

Figure 25 – Numerical Data Analysis Feature on GUI: showing the numerical data from features extracted

As seen in Figure 25, a reference table is provided to the user so the feature values extracted by the software can be compared with the threshold values set in order to differentiate between normal and AOM infected ears. Based on these values, the user can manually classify the image. The GUI can run multiple features at the same time allowing the user to look at them all at once. Finally, the GUI is supported on all three major operating systems such as Microsoft Windows, MacOs and Linux.

7.7 Challenges of Feature Extraction

From the start to the end of the project there were many challenges faced at different stages during the development of both software systems. Depending on the problem, many were solved and some still exist and will be solved as the development progresses.

Some of the most concerning challenges faced varied from collection of dataset to accuracy of system depending on image quality. The challenges are further discussed in detail in the sub-chapter and the possible way to overcome them are also proposed.

Dataset

One of the biggest challenge faced during the project was the collection of image data for normal and infected tympanic membrane. At the time when development of the first system started, there were only 20 good quality images available from one source of which 12 were AOM infected images and only 8 images were of normal tympanic membrane. But as the project progressed, 88 more good quality images were acquired of normal and AOM infected middle ear from another source.

Regardless of having 108 images, the dataset was not big enough to apply any kind of machine learning algorithm and get a high accuracy. There are two main reasons for such less amount of data for otitis media. The first reason is that otitis media image data is not publicly available on internet to use for developing such automated system as compared to some other medical image data which is available in huge amount online. The second reason for having less data is that the data available in medical institutes is protected by the privacy laws which made it very hard to access and therefore could not help in building the software system.

There are two possible ways to combat this challenge, the fast and most effective of the two is to have a cooperation with a medical institute and provide them with high quality otoscopes to take middle ear images. The cooperation will not just result in more data, but it will also help build a better system as the medical professionals and their expertise will be involved in the project. The second method, which is image augmentation was already used during the development. The method is slow as it takes time to gather data from different resources and after it's acquired, the data is augmented and added to the system for more training.

Image Quality

One of the most important attributes in computer vision and specifically in feature extraction is quality of an image. A bad quality image can lead to wrong results produced by feature extraction algorithms. For example, features like colours, reflection and gradient can be heavily affected by the quality of an image.

Bad quality images are a result of type of cameras and light source used in the otoscopes. Otoscopes that produce high quality images have good quality camera sensor and especially good light source which illuminates the tympanic membrane evenly. High quality otoscopes use fibre optic light source or High Colour Rendering Index (CRI) LEDs which produces close to true colour images.

The challenge was faced during the development of the first feature extraction software system. As an example, three different quality images for normal tympanic membrane can be seen in Figure 26.

(a)

(b)

(c)

Figure 26 – Three different quality Normal Tympanic Membrane Images collected from different sources

As seen in Figure 26, all the three images belong to the normal ear drum category, but they are different in quality which produced different output data when feature extraction was applied. They also required different threshold values or range for extracting features such as colour, reflection and gradient. This raised a problem that all the otoscopes where the images are coming from had to be programmed individually with their own specific threshold and every otoscope features will have a different classification criterion in terms of data generated by feature extraction. Another problem faced due to bad image quality was that sometimes features like reflection were not detected at all such as in Figure 26 (a).

Two possible solutions were concluded for the problem. The first was to take several widely used digital otoscopes and add the features extraction stage for all individual of them in the software system. The method is expensive as multiple otoscopes must be bought to acquire different quality images from them and then extract features for every otoscope. The second solution was easier and more cost effective. The idea was to develop the software system based on images taken from one high quality otoscope. In such a way the feature extraction system will be more robust and accurate, then in future when it has proven to work well after clinical testing, the software can be updated to include more otoscope options.

8 Automated Classification System

Chapter 8 discusses the methods and algorithms used to develop an autonomous system for image classification and prediction of normal ear and otitis media. The main idea behind the system revolves around three main methods, which are data augmentation, deep learning and transfer learning.

In general, the system was trained, validated and tested using two different classes of data, one being normal and the other OM infected ears. About 80% percent of the data was used to train and validate and around 20% was used for testing. The system was able to learn from the training data and later classify and predict which class the test images belonged to. Two different type of training and testing was done, one with augmented data and the other without data augmentation. Further details on how and what kind of pre-processing and training algorithms were used are discussed further in the chapter.

8.1 Data Augmentation

When working with neural networks or deep learning models’ huge amounts of data is required to train the model and achieve high accuracy. Whereas collecting data for training and testing can be time consuming and costly. Especially, in applications like otitis media it becomes even harder when a very small dataset is available online, which is not sufficient for training and testing.

To solve the problem of time and cost related to collection of data for otitis media until it's available in huge amount, data augmentation technique can be used. Data/image augmentation is a process which generates more training data from an existing dataset by manipulating them to create many different versions of the same image. It helps make the classifier more robust as it exposes the classifier to a variety of lighting and color situations and not just generate more images to train.
Image augmentation was one of the pre-processing methods used to prepare the data for training with deep learning model. Two different libraries Keras and OpenCV were used to apply augmentation to the training dataset. Both libraries worked similarly but Keras was used in the actual development process as the deep learning model was based on Keras and TensorFlow.

Keras provides many ways for image augmentation. **ImageDataGenerator** class from Keras library was used to prepare the data and perform augmentation. This one class packs several capabilities ranging from loading the data, performing standardization and augmentation to saving the newly generated images on disk. Different kinds of image augmentation techniques that were performed on the data were rotation, width shift, height shift, shearing, zooming, horizontal flip and vertical flip. The following augmentation techniques applied, and the original image can be seen in Figure 27 (a) – (f).

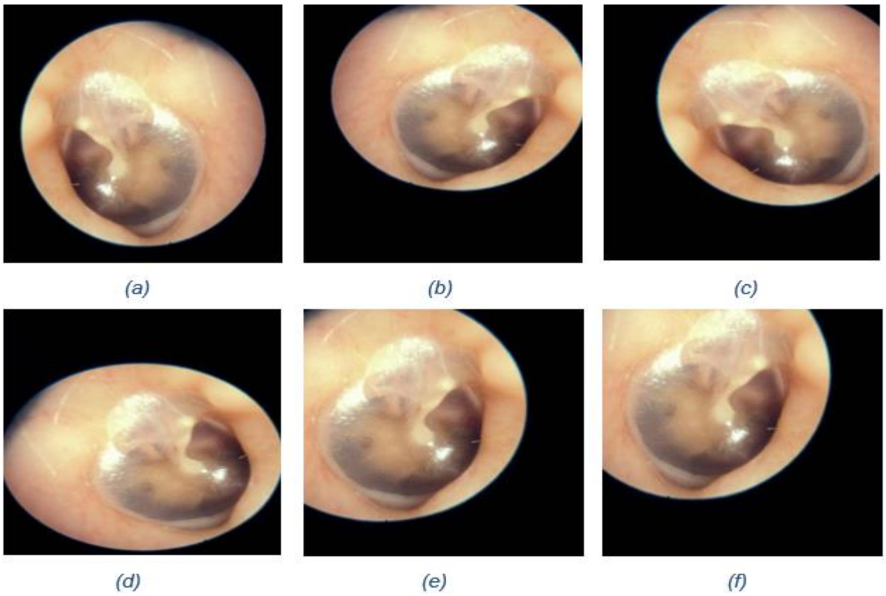


Figure 27 – Date Augmentation applied to Normal TM image: (a) Original Image, (b - f) Random augmentations applied.

In Figure 27 (a) the original image can be seen and in Figure 27 (b – f) the random augmentations applied to the image can be noted as well. Using augmentation technique on 78 original images produced around 3900 images in total, generating 50 variations of each image. Furthermore, image augmentation to the whole dataset can be applied during the training phase which reduces the memory overhead but add additional time during model training.

8.2 Deep Learning and Transfer Learning

Convolutional Neural Networks (CNN) is a type of Deep Neural Networks which are great at image classification tasks. CNNs have proven to exceed human performance in image-recognition on ImageNet dataset. ImageNet is a large visual database designed for use in visual object recognition software research. It has over 14 million images classified into 17 thousand categories. A graph can be seen in Figure 28 comparing error rate of different CNN architectures compared to human performance.

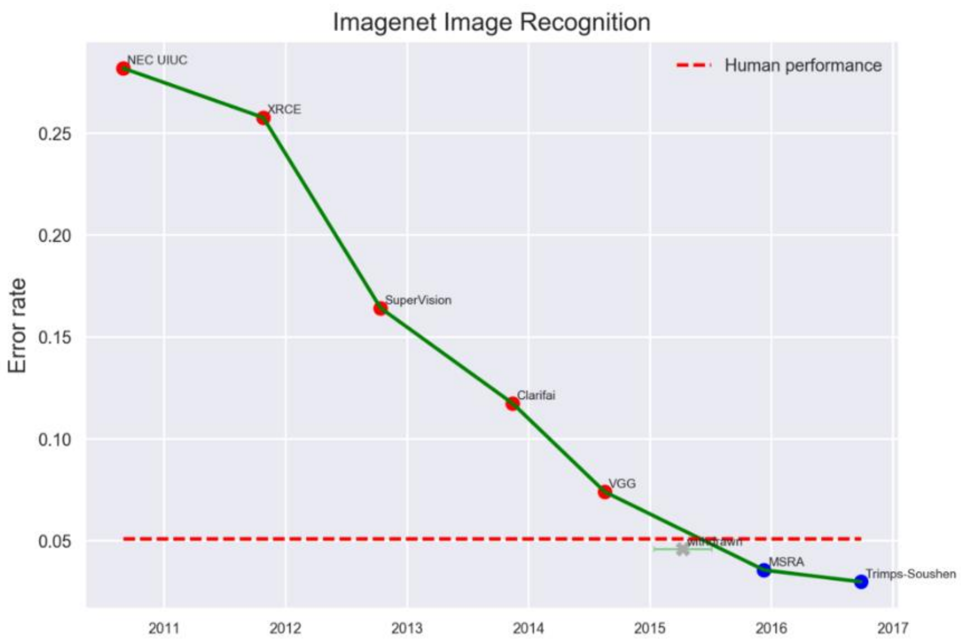


Figure 28 - Imagenet Image Recognition Graph comparing different CNN architecture performance with Human performance, repinted from eff.org [36]

In short, CNNs are multi-layer neural networks that take images as an input. The image is then passed through the hidden layers that can learn increasingly nonconcrete representations of the images provided to the network. For example, given a raw image, the first layer might only learn local edge patterns, whereas then each following layer can learn more complex features. At the end, the last layer can classify the image as normal or infected ear in our application. A very simple and typical CNN architecture can be seen in Figure 29, showing example of layers between input and ouput.

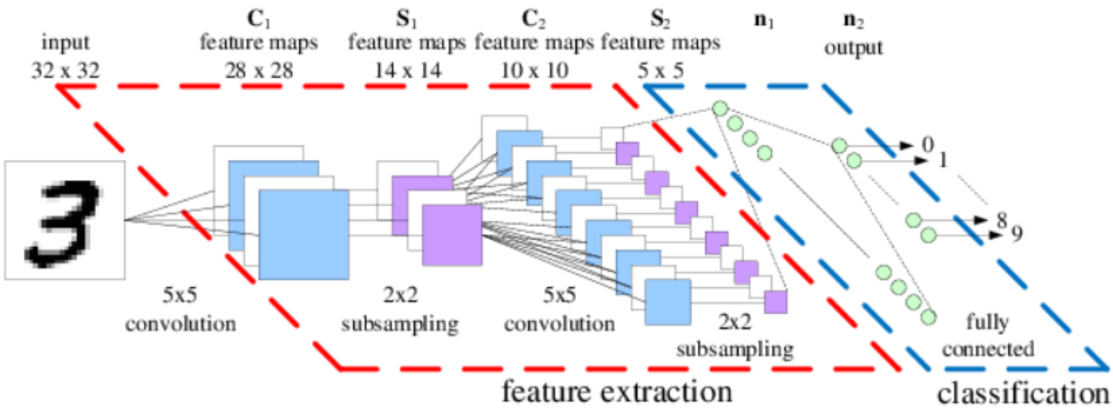


Figure 29 - CNN General Architecture, reprinted from [37]

Figure 29 shows a CNN architecture for classifying hand-written digits. The areas in the red bounding box are the hidden layers and the area in blue bounding box is the last layer where the classification happens. There are many different CNN architecture which can be used according to the application in question. Some of the most common architecture are AlexNet, VGGNet, ResNet, Inception, etc.

Transfer learning is a branch of machine learning with a very simple idea which is to take the knowledge learned in some model in some domain and apply it to another task in another domain. For example the model and weights from CNN trained with ImageNet data can be used and applied to another image dataset, for example in the domain of biomedical image analysis. Transfer learning is a very common appoach when it comes to medical image analysis because to train a CNN from scracth is a very time and hardware resource consuming task. Moreover, it is rare to have a dataset of sufficient size which can produce high accuracy.

Transfer learning can be used in several cases. For example, when the new data set is similar to the initial dataset or new dataset is not similar to the dataset which the CNN model was trained upon or the dateset is different but also very small in size. The last case applies to the otitis media application.

Two of the most known CNN architectures called InceptionV3 and MobileNets were used as a pre-trained model for developing the automated system for classifying normal and infected ears using transfer learning technique. The working and results of the training and tests are further discussed in section 8.2.1.

8.2.1 Deep and Transfer Learning Applied to Otitis Media

An image classifier for otitis media was built upon the ideas of deep learning and transfer learning using the two most widely used CNN architectures named InceptionV3 and MobileNet. InceptionV3 is 42 layers deep compared to MobileNet which is 28 layers deep. Both architectures take images as input which has a fixed size of 299 x 299 pixels for InceptionV3 and 224 x 224 pixels for MobileNet. For more details on how these architectures work, the research paper for InceptionV3 [38] and MobileNet [39] can be further studied. Moreover, the section discusses results of the training and testing of otitis media data using deep learning and transfer learning very briefly due to the confidentiality agreement with the company. The concept behind all the training applied to otitis media dataset was to apply transfer learning to the already trained model which was trained on ImageNet data using deep learning models. More specifically we used the weights obtained from training of these models on ImageNet data and then use those weights to train otitis media data and the last classifier layer. After the new model is trained with otitis media dataset, we have a separate test dataset of 45 normal and infected ear images that the network has never seen before. That test dataset was then used to measure the performance and accuracy of the newly trained model.

The InceptionV3 and MobileNets both models were trained with the same amount of training and validation images. The total amount of augmented images was 3700 out of which 20% were used for validation and 80% for training. The validation also took place during the training of model. The separate test dataset of 45 images was used to measure the accuracy of the model. The training was done on both CPU and GPU to test the time difference it takes to train. For CPU training an AMD 4-core processor was used and for GPU training, an NVIDIA GTX1070 graphics card was selected. The time comparison between both is discussed further where both models are compared.

The training of both models produced some very interesting training and validation accuracy and cross-entropy graphical data using Tensorboard. The data for accuracy graphs for both models can be seen in Figure 30 and 31. Moreover, the cross-entropy data for both models can be seen in Figure 32 and 33. The testing of model with test dataset also produced confusion matrices which were used to calculate the accuracy and the error rate of the model. The confusion matrices can be seen in Table 3.

Training and validation accuracy are two measures that can be seen in Figure 30 and 31. What these graphs shows us are basically how well the networking is doing in terms of accuracy on data it is being trained on; training accuracy usually keeps increasing throughout the time. The validation accuracy numbers tell us how good our model is at predicting outputs for inputs it has never seen before. Validation accuracy is based on the validation set which is 20% of the actual dataset. After each epoch, the model is tested against a validation set which produces validation accuracy. It can be seen in both graphs that the model reached training and validation accuracy of 100% for training and validation both.

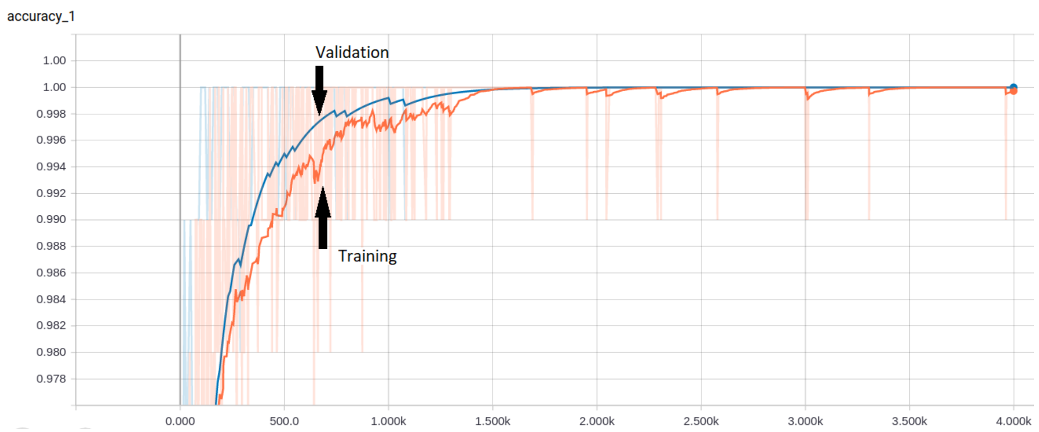


Figure 30 – Inception V3 Accuracy Graph: Orange line for Training and Blue line for validation x-axis = Accuracy and y-axis = number of epoch

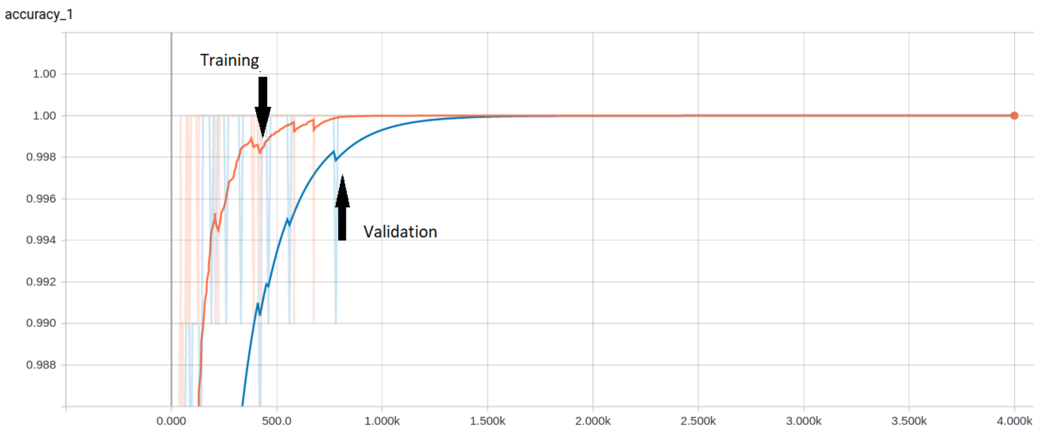


Figure 31 – MobileNets Accuracy Graph: Orange line for Training and Blue line for validation x-axis = Accuracy and y-axis = number of epoch

In Figure 32 and 33 cross-entropy loss graphs are shown for both models which simply explain the idea of measuring the performance of a classification model whose output is a probability between 0 and 1. Cross entropy loss decreases as the predicted values get close to the actual label. The loss graph usually starts at a very high number and over the course of time and epoch it decreases. The lower the cross-entropy loss the better the model. In an ideal case a perfect model would have a log loss of 0. But it is not the case in real life problems, the training and validation loss could get very close to zero as seen in cross-entropy Figures produced by the training.

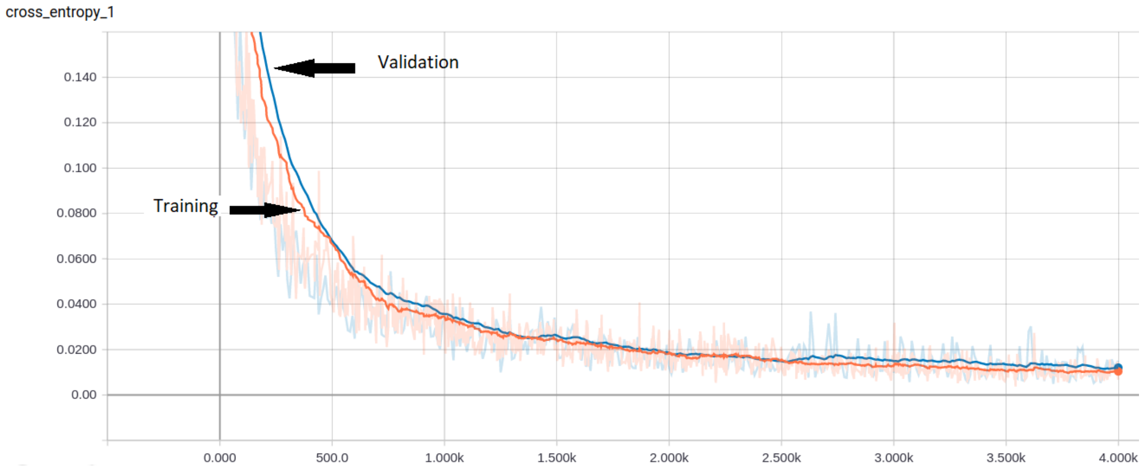


Figure 32 – Inception V3 Cross-Entropy Graph: Orange line for Training and Blue line for validation x-axis = Loss and y-axis = Number of epoch

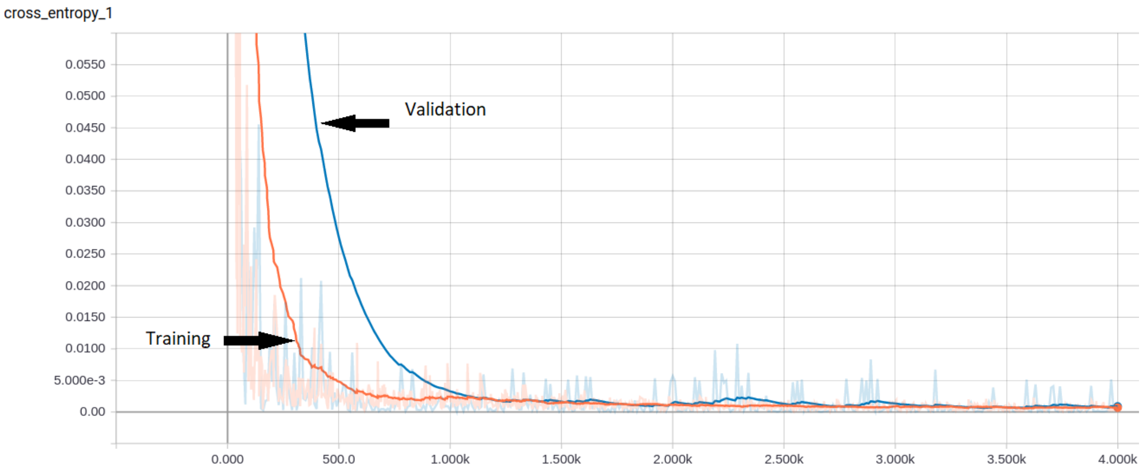


Figure 33 – MobileNets Cross-Entropy Graph: Orange line for Training and Blue line for validation x-axis = Loss and y-axis = Number of epoch

Training of data on both architectures was run using default epoch setting which was 4000 epochs. An epoch means when the whole dataset is passed forward and backward through the neural network only once. So in our case, the entire dataset was run through the neural network 4000 times. It was done in order to observe what would be the most suitable number of epochs to obtain the highest accuracy with the lowest rate in both models. It was decided that 1800 epoch would be an ideal number and hence was chosen to be used in future.

Furthermore, after the training phase was finished, the testing phase took place. The test dataset had 34 infected ear images and 11 normal ear images. The models were then tested on the dataset and the following confusion matrices were produced which can be seen in Figure 3 (a) and (b).

Table 3 - Confusion Matrix for test dataset to calculate accuracy and error rate: (a) Inception V3 Model, (b) MobileNets 1.0 Model

(a)

Architecture: Inception V3			
Confusion Matrix			
		Software Classification	
		Normal	Infected
Ground Truth	Normal	10	1
	Infected	7	27

(b)

Architecture: MobileNet 1.0			
Confusion Matrix			
		Software Classification	
		Normal	Infected
Ground Truth	Normal	10	1
	Infected	8	26

These confusion matrices were used to calculate the accuracy of the model on test dataset and as well the error rate. The accuracy for Inception V3 model was calculated to be 82.2% compared to the accuracy of MobileNets which was 80%. Moreover, the error rate for Inception V3 model was 17.8% and 20% for the MobileNets model. The accuracy of the model was simply calculated by taking the percentage of correctly predicted images for both normal and infected ears. The error rate was calculated by subtracting the accuracy from 100.

The model gave out probability from 0 to 1 as an output when an image was input for inference. It gave a probability of that image as to how close the model thinks it is to the actual label. As a result, the probability and the label were printed onto the image for the user. The label being either normal or infected. The resulting image from the inference can be seen below in Figure 35 (a - c).

(a) (b) (c)

Figure 35 – Inference Output (Label and Classify Images as either Normal or Infected: (a) Normal Predicted Image, (b) Normal Predicted Image, (c) Infection Predicted Image

As seen in Figure 35 (a – c) the images were given as an output with probabilities and their labels printed on them. For example, in Figure 35 (a), the model predicted the image to be normal with a probability of 0.950 or with an accuracy of 95%. It is important to note that all the three images shown above are obtained from different otoscopes. It gives us an idea of how flexible and robust the deep learning software system is, as it is not dependant on one type of otoscope.

8.2.2 Comparison of Inception V3 and MobileNets

The architecture comparison sub-chapter briefly discusses the comparison between two CNN architectures used to develop the model. It discusses some key points which gives a better idea of which architecture is suitable in which condition. Both models used in the project were trained with same settings such as epoch, learning rate, batch size, etc. Both models were also trained and tested with same datasets to make the comparison fairer.

The first major difference between the both architectures is the model size. Inception V3 when trained produces a model of almost 85.5 MB compared to a very small model produced by MobileNets around 16.7 MB. An even smaller model of MobileNets can be produced by using quantized architecture with smaller image size when training the model. More information on different MobileNets model can be found here [40]. The biggest advantage of a small model such as MobileNets is that they can be used on embedded or mobile devices such as smartphones. Which can be further used to do classification on mobile chips in real-time. Although the disadvantage of using smaller models like MobileNets leads to another important issue which is lower overall accuracy. For higher accuracy Inception V3 architectures serves the purpose best if the computations are not to be performed on a mobile device. It can be seen in Figure 34 that Inception V3 was more accurate in classification task.

Apart from the size of the model and accuracy, another key factor plays an important role in the overall process which is the training time related to both architectures. Naturally Inception V3 takes more time to train as it is bigger in size and thus results in more accuracy. As the training was done on both CPU and GPU, the times were recorded for both trainings. The inception V3 model took 3 hours to train on CPU compare to MobileNets which only took 30 minutes. Whereas on GPU Inception V3 took 10 minutes to train compare to 5 minutes of MobileNets training. The conclusion being if the GPU is used to train the models, the difference between training times was not that huge.

9 Conclusions

The main goal of the project was to do research on computer vision and deep learning methods and to develop software which can help support the medical doctor for diagnosing otitis media. The initial idea of using computer vision to extract features was successfully implemented and all the features were extracted in visual and numerical form. The second objective of developing an image classification software using state of the art techniques such as deep learning, was also achieved with fairly good results.

The results of the image classification software can be evaluated by considering the overall accuracy achieved by the software. The accuracy achieved by the developed classification system was 82.2%. The results are very close to a very recent research published [41] on the same problem by renowned researchers who managed to reach an accuracy of 84.4% using very similar techniques. The published research [41] support the idea and algorithms used in the project. It is believed that with more data the accuracy can be pushed much higher and the system can be made more robust.

To sum up, there can be further improvements made by collecting more data and using advanced and newly discovered deep learning methods, which will help make the classification even better by considering the information which current algorithms cannot. Further development on the software will be carried out throughout the year 2018.

References

1. Myburgh H, van Zijl W, Swanepoel D, Hellström S, Laurent C. Otitis Media Diagnosis for Developing Countries Using Tympanic Membrane Image-Analysis [Internet]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4816811. [Accessed 15 Mar. 2018].

2. Daniel, M., Qureishi, A., Lee, Y., Belfield, K. and Birchall, J. (2014). Update on otitis media – prevention and treatment. *Infection and Drug Resistance*, p.15. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3894142/ [Accessed 15 Mar. 2018].

3. Eustachian Tube: Kids-ent.com. (2018). *Kids-ENT: Ear Disease*. [online] Available at: http://www.kids-ent.com/newkids800/earframe.html [Accessed 15 Mar. 2018]

4. Listening Ears. (2018). *Eustachian Tube Obstruction (infant/toddler) | Listening Ears*. [online] Available at: http://listeningears.in/eustachian-tube-obstruction/ [Accessed 15 Mar. 2018]

5. De Hoog M, Venekamp R, van der Ent C, Schilder A, Sanders E, Damoiseaux R Et al. Impact of early daycare on healthcare resource use related to upper respiratory tract infections during childhood: prospective WHISTLER cohort study. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24965189 [Accessed 15 Mar. 2018]

6. Otitis Media: Practice Essentials, Background, Pathophysiology [Internet]. Emedicine.medscape.com. Available from: http://emedicine.medscape.com/article/994656-overview#a2 [Accessed 15 Mar. 2018]

7. Acute Otitis Media: Practice Essentials, Background, Anatomy [Internet]. Emedicine.medscape.com. Available from: http://emedicine.medscape.com/article/859316-overview#a7 [Accessed 15 Mar. 2018]

8. Otitis Media with Effusion: Background, Pathophysiology, Etiology [Internet]. Emedicine.medscape.com. Available from: http://emedicine.medscape.com/article/858990-overview#a3. [Accessed 15 Mar. 2018]

9. Tympanic Membrane Pictures, Function & Anatomy | Body Maps [Internet]. Healthline.com. Available from: http://www.healthline.com/human-body-maps/tympanic-membrane [Accessed 15 Mar. 2018]

10. Tympanic membrane | Definition, Anatomy, Function, & Perforation [Internet]. Encyclopedia Britannica. Available from: https://www.britannica.com/science/tympanic-membrane [Accessed 15 Mar. 2018]

11. Anatomi Histologi: dan Embriologi Organ Pendengaran [Internet]. Medical Science. Available from: http://www.drmeu.com/2016/09/anatomi-histologi-dan-embriologi-organ.html. [Accessed 15 Mar. 2018]

12. Myburgh H, van Zijl W, Swanepoel D, Hellström S, Laurent C. Otitis Media Diagnosis for Developing Countries Using Tympanic Membrane Image-Analysis. Available at: https://www.ncbi.nlm.nih.gov/pubmed/27077122 [Accessed 15 Mar. 2018]

13. Otitis Media Guide: Acute Otitis Media: 1_13 [Internet]. Otitismedia.hawkelibrary.com. Available from: http://otitismedia.hawkelibrary.com/aom/1_13 [Accessed 15 Mar. 2018]

14. Welch Allyn GS 777 Wall System -77794-MSPOT S, HOSPEQ I. Welch Allyn GS 777 Wall System -77794-MSPOT, Ophthalmoscope, Otoscope, Spec Dispenser, Spot Mount [Internet]. A1medicalsales.com. Available from: https://www.a1medicalsales.com/product/77794-MSPOT.html [Accessed 15 Mar. 2018]

15. Welch Allyn: Charger and Desk Sets [Internet]. Welchallyn.com. Available from: https://tinyurl.com/yazuvrud [Accessed 15 Mar. 2018]

16. Welch Allyn: MacroView Otoscope [Internet]. Welchallyn.com. Available from: https://tinyurl.com/zwv2tdg [Accessed 15 Mar. 2018]

17. Video Otoscopy - Audiology [Internet]. JEDMED. Available from: https://www.jedmed.com/collections/video-otoscopy [Accessed 15 Mar. 2018]

18. TYM Otoscope [Internet]. Cupris | Smartphone connected communication platform and medical devices. Available from: http://www.cupris.com/tym-otoscope [Accessed 15 Mar. 2018]

19. Tracking Drug Resistant Infection: [ebook] Available at: http://www.myrolematters.com/cmss_files/attachmentlibrary/Tackling-drug-resistant-infections---An-overview-of-our-work_LR_NOCROPS.pdf [Accessed 15 Mar. 2018]

20. Pichichero, M. and Poole, M. (2018). *Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. - PubMed - NCBI*. [online] Ncbi.nlm.nih.gov. Available at: https://www.ncbi.nlm.nih.gov/pubmed/11576009 [Accessed 15 Mar. 2018]

21. Pichichero, M. Diagnostic accuracy of otitis media and tympanocentesis skills assessment among pediatricians. - PubMed - NCBI [Internet]. Ncbi.nlm.nih.gov. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12938004 [Accessed 15 Mar. 2018]

22. Kuruvilla A, Shaikh N, Hoberman A, Kovačević J. Automated Diagnosis of Otitis Media: Vocabulary and Grammar. Available at: https://dl.acm.org/citation.cfm?id=2537989 [Accessed 15 Mar. 2018]

23. Python.org. (2018). *Welcome to Python.org*. [online] Available at: https://www.python.org/ [Accessed 15 Mar. 2018]

24. Anaconda.org. (2018)*: Anaconda Cloud*. [online] Available at: https://anaconda.org/ [Accessed 15 Mar. 2018]

25. OpenCV library [Internet]. Opencv.org. Available from: http://opencv.org/about.html [Accessed 15 Mar. 2018]

26. Python-Skimage – Python modules for image processing — Debian Neuroscience Package Repository [Internet]. Neuro.debian.net. Available from: http://neuro.debian.net/pkgs/python-skimage.html [Accessed 15 Mar. 2018]

27. Python Imaging Library [Internet]. Available from: https://pillow.readthedocs.io/en/5.1.x/. [Accessed 15 Mar. 2018]

28. Tkinter Documentation. (2018). *24.1. Tkinter — Python interface to Tcl/Tk — Python 2.7.15rc1 documentation*. [online] Available at: https://docs.python.org/2/library/tkinter.html [Accessed 15 Mar. 2018]

29. TensorFlow Documentation. (2018). *TensorFlow*. [online] Available at: https://www.tensorflow.org/ [Accessed 15 Mar. 2018]

30. Keras (2018). *Keras Documentation*. [online] Available at: https://keras.io/ [Accessed 15 Mar. 2018]

31. Sinha, U. (2018). *Image Moments - AI Shack*. [online] Aishack.in. Available at: http://aishack.in/tutorials/image-moments/ [Accessed 15 Mar. 2018]

32. Kass, M., Witkin, A. and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), pp.321-331.Available from: http://web.cs.ucla.edu/~dt/papers/ijcv88/ijcv88.pdf [Accessed 15 Mar. 2018]

33. Patrick P. Poisson Image Editing [Internet]. 2017 [cited 30 July 2017]. Available from: https://www.cs.jhu.edu/~misha/Fall07/Papers/Perez03.pdf [Accessed 15 Mar. 2018]

34. Gelfand, S. (1997). *Essentials of audiology*. New York: Thieme. Available at: https://books.google.fi/books/about/Essentials_of_Audiology.html?id=_tcPcPT-wNQoC&redir_esc=y [Accessed 15 Mar. 2018]

35. Calculating Gradient using Numpy: Docs.scipy.org. (2018). *numpy.gradient — NumPy v1.14 Manual*. [online] Available at: https://docs.scipy.org/doc/numpy/reference/generated/numpy.gradient.html [Accessed 15 Mar. 2018]

36. Electronic Frontier Foundation. (2018). *AI Progress Measurement*. [online] Available at: https://www.eff.org/ai/metrics [Accessed 16 Mar. 2018]

37. Becoming Human: Artificial Intelligence Magazine. (2018). How should I start with CNN? – Becoming Human: Artificial Intelligence Magazine. [online] Available at: https://becominghuman.ai/how-should-i-start-with-cnn-c62a3a89493b [Accessed 21 Mar. 2018]

38. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2018). *Rethinking the Inception Architecture for Computer Vision*. [online] Arxiv.org. Available at: https://arxiv.org/abs/1512.00567 [Accessed 21 Mar. 2018]

39. Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2018). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. [online] Arxiv.org. Available at: https://arxiv.org/abs/1704.04861 [Accessed 21 Mar. 2018]

40. MobileNets: Open-Source Models for Efficient On-Device Vision. [online] Available at: https://research.googleblog.com/2017/06/mobilenets-open-source-models-for.html [Accessed 21 Mar. 2018]

41. Caglar Senaras; Aaron C. Moberly; Theodoros Teknos; Garth Essig; Charles Elmaraghy; Nazhat Taj-Schaal; Lianbo Yua; Metin N. Gurcan. Detection of eardrum abnormalities using ensemble deep learning approaches
Available at: https://tinyurl.com/ybgmeltz [Accessed 21 Mar. 2018