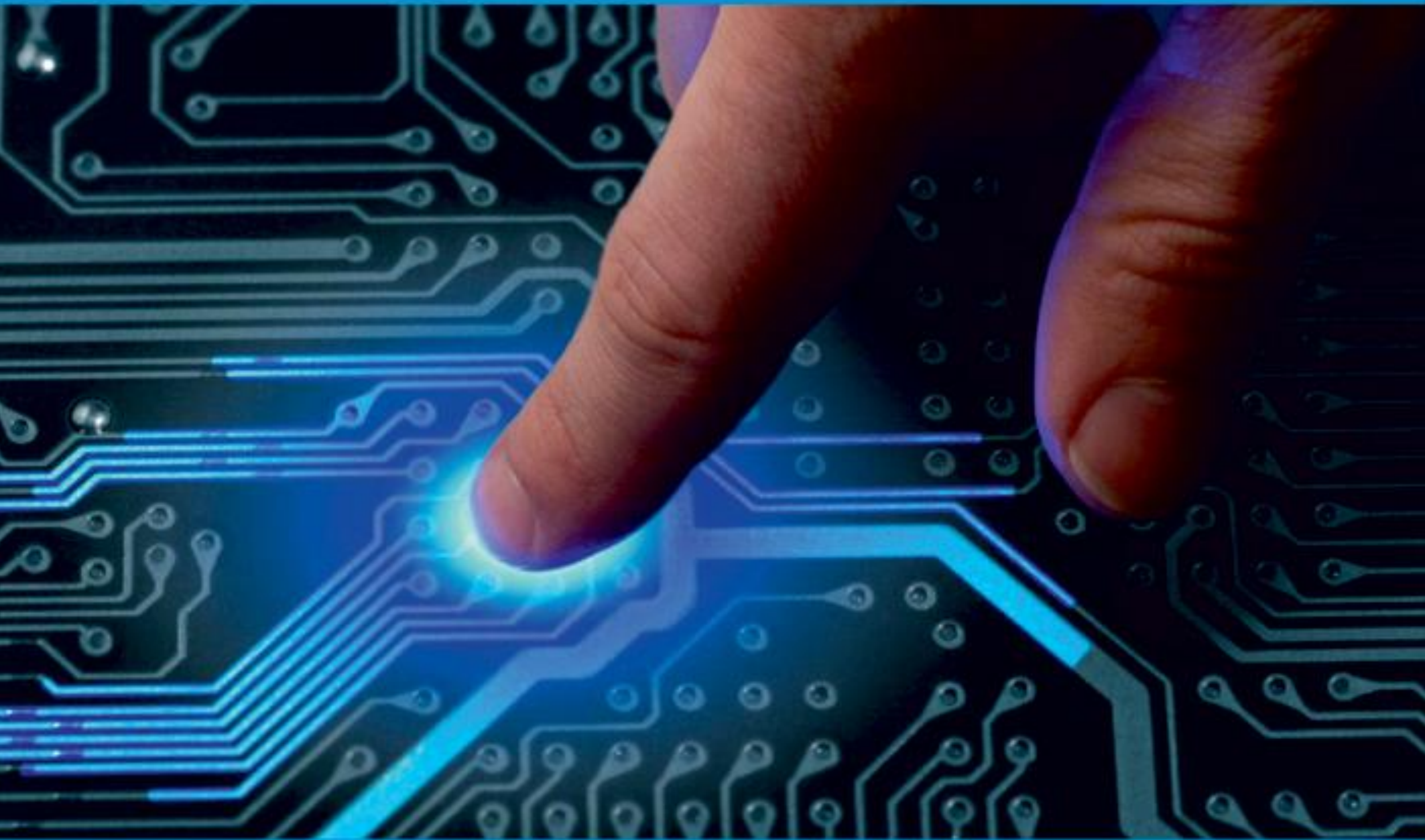
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204180 |

Otoscopy Disease Prediction Using Deep Learning

MR.A.AKILAN, R.ASHIKA, B.PAVITHRA, A.PUNITHA, R.SUMITHRA

Assistant Professor, Department of Computer Science and Engineering, MRK Institute of Technology, Kattumannarkoil, Tamilnadu, India

Department of Computer Science and Engineering, MRK Institute of Technology, Kattumannarkoil, Tamilnadu, India

ABSTRACT: Otoscopy could be a demonstrative strategy utilized to look at the outside sound-related canal and the eardrum (tympanic layer) utilizing an otoscope. Otoscopy is commonly performed to evaluate ear wellbeing, recognize variations from the norm such as diseases, aggravation, wax buildup, remote bodies, auxiliary abandons, tumors, or signs of injury. It is a fundamental portion of the assessment handle of ear infection for patients with ear-related indications like torment, hearing misfortune, or release. Otoscopy plays a pivotal part in diagnosing different ear pathologies, however precise and opportune classification of otoscopy pictures remains a challenging assignment. In this consideration, we proposed a deep-learning-based strategy for computerized otoscopy classification. We curated a different dataset comprising otoscopy pictures enveloping ordinary life structures and a range of pathologies, including diseases, inflammations, tumors, and basic variations from the norm.

KEYWORDS: artificial intelligence; machine learning; convolutional neural network; otitis media; global health; digital imaging

I. INTRODUCTION

Otitis media (OM) is a prevalent ailment in children, presenting symptoms such as fever, sleep disturbances, and acute infections. This illness significantly affects not only children who experience considerable pain but also their caregivers. OM prevalence is high worldwide, with rates of 9.2% in Nigeria, 10% in Egypt, 6.7% in China, 9.2% in India, 9.1% in Iran, and 5.1–7.8% in Russia. Additionally, the incidence of OM in native Australian children is 90%, the highest worldwide. Prior works have discussed OM diagnosis and treatment methods. If OM is inaccurately diagnosed, it can lead to severe consequences, including hearing loss, cognitive development disorders, unnecessary surgeries, antibiotic overuse, and disease exacerbation. Notably, 80% of OM patients receive antibiotics, leading to potential antibiotic resistance and unnecessary expenses. Therefore, accurate diagnosis is essential to mitigate these side effects and provide effective treatment.

Diagnostic techniques for both acute and chronic middle ear infections have long posed challenges. Infants, in particular, present difficulties due to their narrow external ducts, which, coupled with the presence of earwax, can hinder accurate diagnosis using an ear endoscope alone. Furthermore, in primary clinics and pediatrics, the accuracy of diagnosis tends to be low due to a lack of systematic training and unfamiliarity with pneumatic ear endoscopy. To address these challenges, various approaches have been explored in the field. These include specialized training programs for medical students, the development of new otoscopic approaches and techniques, the implementation of absorbance and acoustic admittance measurements, and the integration of impedance-measuring hearing aids. Additionally, clinical trials have been conducted to compare the effects of these various approaches. However, despite these approaches and efforts, the diagnostic success rates among pediatricians and otolaryngologists in primary care settings do not exceed 70%.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204180 |

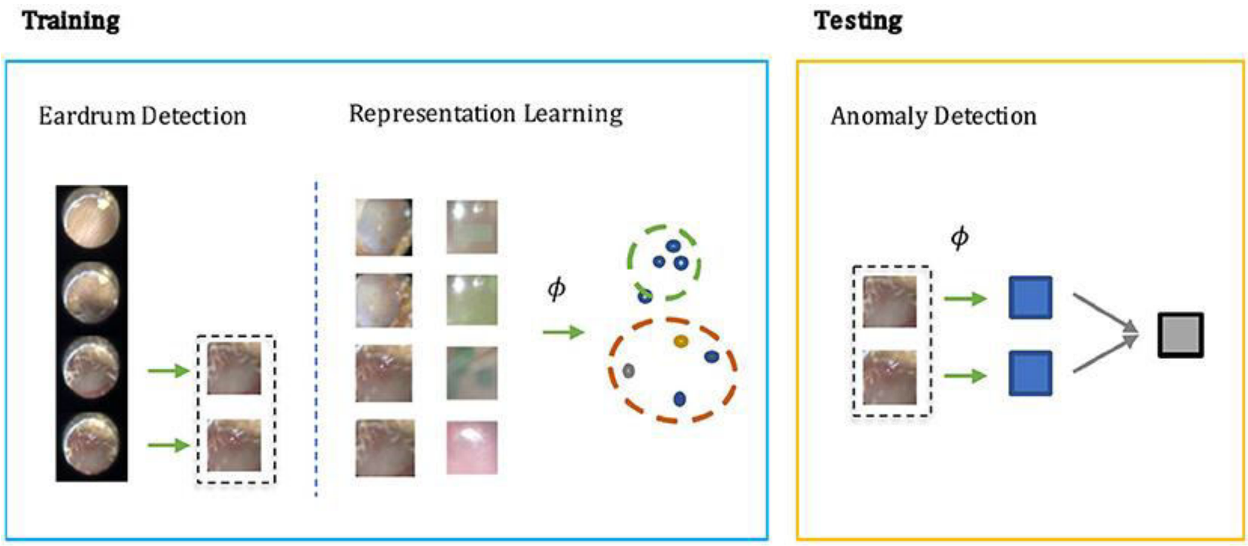


Fig 1: Evaluation of digital otoscopy

Medical image processing is of considerable importance in the analysis and exploration of medical data. However, the complexity inherent in medical images presents challenges to their accurate representation and evaluation using conventional approaches. The use of AI has demonstrated a high level of effectiveness in the analysis of these complex medical images, which has led to its frequent use in medical research. Diagnostic accuracy in otolaryngology can vary based on a physician's training and area of specialization, given the reliance on endoscopic imaging and visual mechanisms. Therefore, the integration of deep learning algorithms in oto-endoscopic imaging is of significant importance. Due to advances in computer science, the utilization of AI in the medical field has seen substantial growth, particularly in studies involving endoscopic images. Despite this progress, the application of an automatic diagnostic system for OM in actual clinical settings remains unimplemented due to uncertainties associated with deep learning, posing a major obstacle as identified in reference.

In this study, we evaluated the diagnostic accuracy according to the AI technology used in automatic OM diagnostic studies based on medical images and the type of OM diagnosed. Based on the findings of the study, we discuss improvement measures in this paper and suggest directions for future research.

II. RELATED WORK

In this prospective study we collected digital images of tympanic membranes and used an expert panel's diagnoses of the images as reference standard in a comparison to the diagnoses of the same images by a CNN. Ethical clearance was provided by the Ethics committee at the University of Pretoria, South Africa, and for data analyses in Sweden by the Ethics committee at Umeå University. Data collection took place between March and May 2016. Written informed consent was obtained from the participants or their parents or legal guardians.

Enrolment of participants took place in low-income communities in the City of Tshwane (Pretoria), Gauteng, South Africa at one district hospital ENT clinic, three primary healthcare clinics and one itinerant clinic for school screening. Anyone in the waiting room could participate in the study, irrespective of ear complaints or not. Exclusion criteria were active draining of ears and persons younger than one year of age. The reason why actively draining ears were excluded was that some participants were also part of an audiological study (reported elsewhere) for which dry ears was an inclusion criterion. All participants gave their written informed consent to participate. All procedures complied with the Declaration of Helsinki.

Images were captured with a handheld digital otoscope (Welch Allyn Digital Macroview Otoscope, Welch Allyn Inc., Skaneateles Falls, New York, NY, USA). A series of images was collected from each ear, the number depending on how well all parts of the tympanic membrane could be visualised. The otoscope was connected to a PC laptop running Microsoft Windows 8 on which images were saved. Different sizes of ear specula were available to fit different widths

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 4, April 2024 ||
DOI: 10.15680/IJIRCCE.2024.1204180 |

of ear canals. A medical intern (JS) with prior training and experience in otoscopy performed all the data collection. A total of 380 ears were included in the study.

The assessment of the 347 ears in the main image set (Figure 1) was performed in identical Google forms as for the calibration. Independently of each other, the panel members made their assessment of each GIF-file. To avoid loss of concentration they assessed a randomised selection of <50 GIF files at a time, in eight subsequent sessions. Not until all five panellists had completed one session was the next made available. A reference standard diagnosis was established when three or more of the five panellists did agree on a diagnosis.

III. METHODS

Instead of normalizing the datasets (or under-sampling by removing images) and applying data augmentation (a form of oversampling by creating new images) as described in Section 2.4, an alternative approach could have been taken to ensure that no data are discarded. Instead of data resampling, the classifier could have been modified by applying thresholding, learning rate adjustment or adjustment of prior class probabilities. It has been shown that oversampling outperforms the other class modification methods and in cases of large class imbalance ratio under-sampling is on par with oversampling. Since both data under-sampling and oversampling were applied, accuracy is expected to be comparable to or slightly better than classifier modification.

Otologists' accuracy approximates to 95% in diagnosing OME. For this study, we set the accuracy of the automated classification software as at least 85%, which resulted in an estimated sample size of 142 ears. To assess the performance of the CNN, a confusion matrix was calculated for each scenario for the different augmentation factors. A confusion matrix is a means of visualising several important quantitative performance measures true positive rate (TPR), false positive rate (FPR), true negative rate and false negative rate—for all diagnostic groups, by comparing actual diagnoses (reference standard) with the CNN diagnoses. The overall accuracy for each situation was also calculated. To evaluate the CNN's screening performance, sensitivity and specificity were calculated using an online calculator.

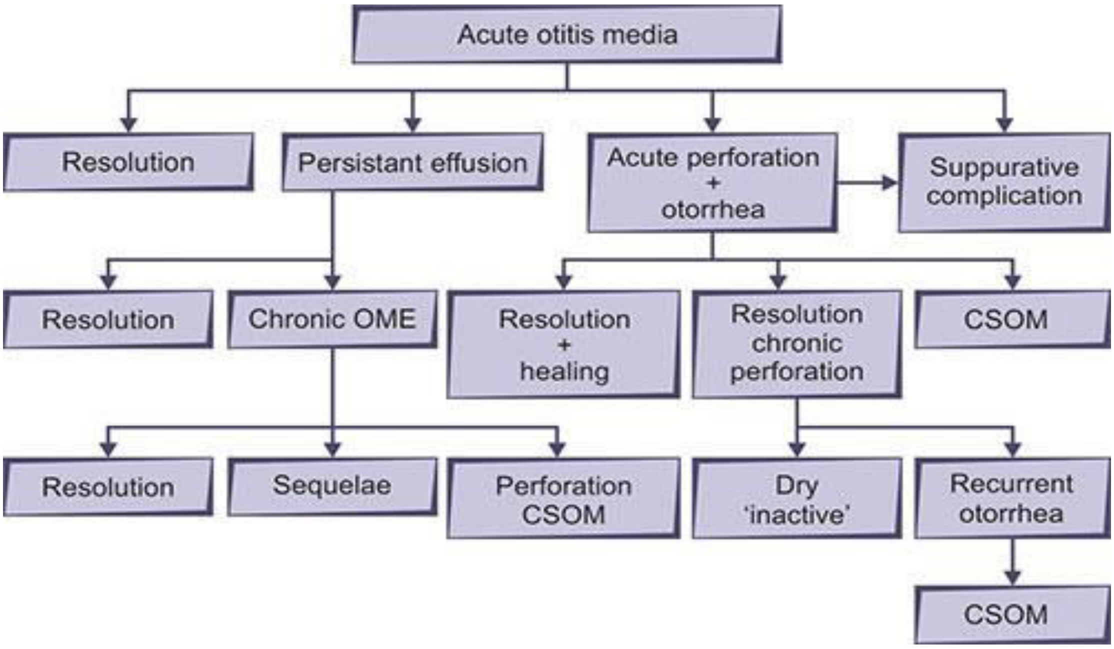


Fig 2: Work Flow

Our results show that augmentation of dataset 1_{norm} enhanced the overall accuracy of the CNN up to an augmentation factor of 10, but with higher augmentation it decreased again. The TPR for Pathological also benefitted from augmentation in the same dataset. Cha et al., instead of normalization, used the approach of merging similar diagnostic classes to reach balance between diagnostic groups, but the effect of this merging was not analysed in their study [14]. To our knowledge, the effect of balancing classes in otological studies, by any method, has never been investigated before, nor has the effect of augmentation of training data been studied, although the modifications are

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204180 |

frequently performed [12,13,14]. Our finding of improvement in the performance for Pathological in dataset 1_{norm} up to a certain level of augmentation, but not thereafter, might be due to more Pathological images being used for training and therefore a better outcome of the augmentation. A subsequent decrease in accuracy when the augmentation factor is higher might be due to usage of too many similar images, or so-called overfitting of the system.

Sensitivity for abnormal (Pathological or Wax) was 100% for augmentation factor 5 in scenario A and B. This indicates that a Normal output from the CNN is most likely true, which is crucial for a screening system. Of clinical interest also is the finding that, in the separate sub-analysis, the CNN classified all nine NPD images as abnormal. This finding indicates that ears difficult to diagnose will most likely be classified as abnormal and as cases for further clinical assessment.

Overall accuracy for the non-normalized dataset was lowered for most augmentation factors when more images were added for training (scenario C and D), and normalization of the new dataset had no major influence. The added images in dataset 2 originate from various clinical materials and were captured with different otoscopes and endoscopes [11]. Furthermore, the reference standard for these images was established with a less solid method as compared to the expert panel's diagnoses of dataset 1. These factors may have contributed to the lower accuracy for scenario C and D. The added images also included more variants of pathological ears. The higher prevalence of Pathological in dataset 2 (53%) compared to dataset 1 (16%) might also have contributed to the lowered accuracy of the CNN in scenario C and D, given the assumption that Normal is easier to diagnose for the CNN. The TPR for Pathological is, however, similar across augmentation factors between scenario A and C, even though scenario C contains more pathological samples. The imbalance in diagnostic groups should benefit scenario C's TPR for Pathological, since CNNs are usually biased towards larger diagnostic groups. The use of only one type of digital otoscope in dataset 1 and the more solid reference standard are probable explanations for the similar performance.

IV. RESULT ANALYSIS

This study demonstrates accurate screening for otitis media using a machine learning approach on digital otoscopic images diagnosed by an expert panel. A CNN was used, and the highest overall accuracy was found for dataset 1 without normalization and augmentation. When adding more images for training of the CNN (scenarios C and D), the overall accuracy slightly decreased, but sensitivity and specificity was still high, also with normalized datasets and across augmentation factors.

To our knowledge, within the field of machine learning in otology this study is unique in terms of its robust reference standard. In a recent meta-analysis on artificial intelligence in diagnosing ear disease by Habib et al., the authors concluded that ground truth assessment is a widespread limitation of existing studies, with limited studies using multiple independent experts' diagnoses [28]. We used five experienced otologists to assess the images and required a minimum of three of them to agree upon the diagnosis to incorporate an image into the reference standard. In comparison, Livingstone et al. used two ENT specialists' consensus diagnoses [12], and in a study by Cha et al. the main author labelled all images, using information from the medical record to double check [14]. A CNN is trained on labelled images and a less solid training material will affect the accuracy of the output.

The expert panel members did not always agree about the diagnosis, which reflects the difficulty in diagnosing tympanic membranes using otoscopy, nor were they confident in their own diagnosis in every case. In 10% (36/347) of cases a majority decision could not be reached. A study by Kuruvilla et al. showed poor agreement between three experts when assigning reference standard diagnoses [7]. In a re-assessment, their panel members were also asked to register their confidence in the chosen diagnosis and only images in which they all agreed and could report a certain confidence level were used in their ground truth set. This process reduced the number of images in their study considerably, from 826 to 181. For our reference standard, we did no discrimination based on confidence in the diagnosis. On the other hand, we employed more panel members and gave the alternative of NPD, most likely eliminating some cases with uncertain diagnoses, and thus maintaining a high quality reference standard.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204180 |

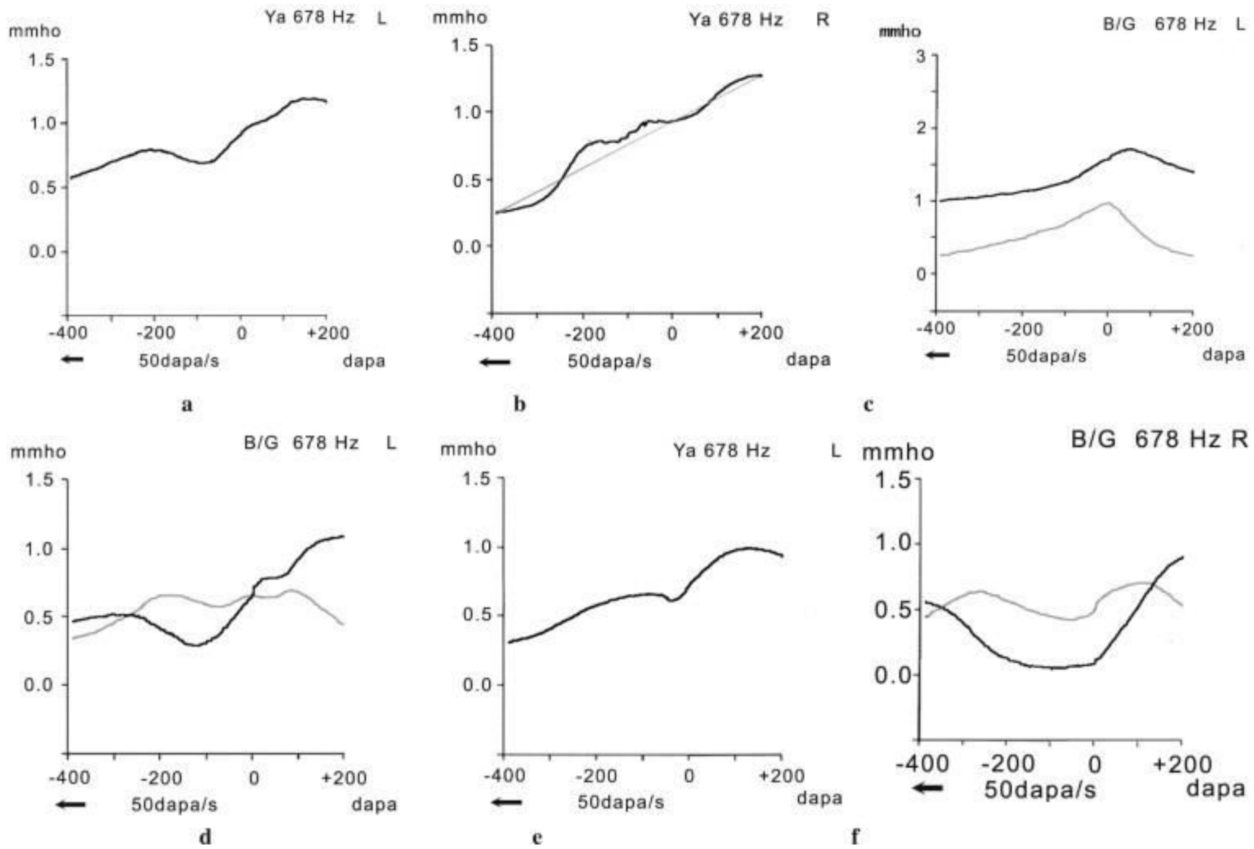


Fig 3: Result analysis otoscopy disease

The overall accuracy dropped when augmentation and normalization were used on dataset 1. Normalization rendered a reduction of images in the larger classification groups, in our dataset 'the Normal'. The change in the proportions of diagnoses may have influenced the performance of the CNN in scenario B. Training and testing images were randomly sampled from the dataset and the difference between scenario A and B could theoretically have been a result of coincidence in the sampling. However, since the pattern was evident for all augmentation factors it seems more probable that the change in proportions of diagnoses was the reason. Diagnostic agreement between expert panel members was higher for Normal than for OME. It may be that Normal are easier to identify for the CNN and reducing the number of Normal through normalization decreases the accuracy of the CNN.

V. CONCLUSIONS

The machine learning approach in this study presented high accuracy in screening for otitis media on images labelled by an expert panel. Sensitivity and specificity in discriminating normal from abnormal ears were satisfactory. Augmentation showed a slight positive effect on overall accuracy for normalized datasets, but we found no apparent advantage of normalization. A CNN like ours could be suitable to assist in basic screening for otitis media in underserved settings.

REFERENCES

1. Homøe, P.; Heidemann, C.H.; Damoiseaux, R.A.; Lailach, S.; Lieu, J.E.; Phillips, J.S.; Venekamp, R.P. Panel 5: Impact of otitis media on quality of life and development. *Int. J. Pediatr. Otorhinolaryngol.* **2020**, *130* (Suppl. 1), 109837. [Google Scholar] [CrossRef]
2. Monasta, L.; Ronfani, L.; Marchetti, F.; Montico, M.; Brumatti, L.V.; Bavcar, A.; Grasso, D.; Barbiero, C.; Tamburlini, G. Burden of Disease Caused by Otitis Media: Systematic Review and Global Estimates. *PLoS ONE* **2012**, *7*, e36226. [Google Scholar] [CrossRef] [PubMed]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

DOI: 10.15680/IJIRCCE.2024.1204180 |

3. Mulwafu, W.; Ensink, R.; Kuper, H.; Fagan, J. Survey of ENT services in sub-Saharan Africa: Little progress between 2009 and 2015. *Glob. Health. Action* **2017**, *10*, 1289736. [**Google Scholar**] [**CrossRef**] [**PubMed**] [**Green Version**]
4. Joshi, R.; Alim, M.; Kengne, A.P.; Jan, S.; Maulik, P.K.; Peiris, D.; Patel, A.A. Task Shifting for Non-Communicable Disease Management in Low and Middle Income Countries—A Systematic Review. *PLoS ONE* **2014**, *9*, e103754. [**Google Scholar**] [**CrossRef**]
5. Suen, J.J.; Bhatnagar, K.; Emmett, S.D.; Marrone, N.; Robler, S.K.; Swanepoel, D.W.; Wong, A.; Nieman, C.L. Hearing care across the life course provided in the community. *Bull. World Health Organ.* **2019**, *97*, 681–690. [**Google Scholar**] [**CrossRef**]
6. Biagio, L.; de Wet, S.; Laurent, C.; Lundberg, T. Video-otoscopy recordings for diagnosis of childhood ear disease using telehealth at primary health care level. *J. Telemed. Telecare* **2014**, *20*, 300–306. [**Google Scholar**] [**CrossRef**] [**PubMed**] [**Green Version**]
7. Kuruvilla, A.; Shaikh, N.; Hoberman, A.; Kovačević, J. Automated Diagnosis of Otitis Media: Vocabulary and Grammar. *Int. J. Biomed. Imaging* **2013**, *2013*, 327515. [**Google Scholar**] [**CrossRef**] [**PubMed**]
8. Myburgh, H.C.; van Zijl, W.H.; Swanepoel, D.W.; Hellström, S.; Laurent, C. Otitis Media Diagnosis for Developing Countries Using Tympanic Membrane Image-Analysis. *EBioMedicine* **2016**, *5*, 156–160. [**Google Scholar**] [**CrossRef**] [**Green Version**]
9. Walczak, S.; Velanovich, V. An Evaluation of Artificial Neural Networks in Predicting Pancreatic Cancer Survival. *J. Gastrointest. Surg.* **2017**, *21*, 1606–1612. [**Google Scholar**] [**CrossRef**]
10. Arcadu, F.; Benmansour, F.; Maunz, A.; Willis, J.; Haskova, Z.; Prunotto, M. Deep learning algorithm predicts diabetic reti-nopathy progression in individual patients. *NPJ Digit Med.* **2019**, *2*, 92. [**Google Scholar**] [**CrossRef**]
11. Myburgh, H.C.; Jose, S.; Swanepoel, D.W.; Laurent, C. Towards low cost automated smartphone- and cloud-based otitis media diagnosis. *Biomed. Signal Process. Control* **2018**, *39*, 34–52. [**Google Scholar**] [**CrossRef**] [**Green Version**]
12. Livingstone, D.; Talai, A.S.; Chau, J.; Forkert, N.D. Building an Otoscopic screening prototype tool using deep learning. *J. Otolaryngol. Head Neck Surg.* **2019**, *48*, 66. [**Google Scholar**] [**CrossRef**] [**PubMed**] [**Green Version**]
13. Lee, J.Y.; Choi, S.H.; Chung, J.W. Automated Classification of the Tympanic Membrane Using a Convolutional Neural Net-work. *Appl. Sci.* **2019**, *9*, 1827. [**Google Scholar**] [**CrossRef**] [**Green Version**]