

Main Article

Dr A-R Habib takes responsibility for the integrity of the content of the paper

Presented at the Australian Society of Otolaryngology – Head and Neck Annual Scientific Meeting, 22–24 March 2019, Brisbane, Australia.

Cite this article: Habib A-R, Wong E, Sacks R, Singh N. Artificial intelligence to detect tympanic membrane perforations. *J Laryngol Otol* 2020;**134**:311–315. <https://doi.org/10.1017/S0022215120000717>

Accepted: 29 January 2020
First published online: 2 April 2020

Key words:

Otосcopy; Ear; Tympanic Membrane; Machine Learning

Author for correspondence:

Dr Al-Rahim Habib, Department of Otolaryngology – Head and Neck Surgery, Westmead Hospital, Sydney, Australia
E-mail: ahab1907@uni.sydney.edu.au

Abstract

Objective. To explore the feasibility of constructing a proof-of-concept artificial intelligence algorithm to detect tympanic membrane perforations, for future application in under-resourced rural settings.

Methods. A retrospective review was conducted of otoscopic images analysed using transfer learning with Google's Inception-V3 convolutional neural network architecture. The 'gold standard' 'ground truth' was defined by otolaryngologists. Perforation size was categorised as less than one-third (small), one-third to two-thirds (medium), or more than two-thirds (large) of the total tympanic membrane diameter.

Results. A total of 233 tympanic membrane images were used (183 for training, 50 for testing). The algorithm correctly identified intact and perforated tympanic membranes (overall accuracy = 76.0 per cent, 95 per cent confidence interval = 62.1–86.0 per cent); the area under the curve was 0.867 (95 per cent confidence interval = 0.771–0.963).

Conclusion. A proof-of-concept image-classification artificial intelligence algorithm can be used to detect tympanic membrane perforations and, with further development, may prove to be a valuable tool for ear disease screening. Future endeavours are warranted to develop a point-of-care tool for healthcare workers in areas distant from otolaryngology.

Introduction

Machine learning with deep learning convolutional neural networks represents a form of artificial intelligence (AI) that allows computers to learn from data by recognising patterns and making predictions.¹ Advances in computational speed, power and accessibility have allowed AI to grow in popularity and be applied to many specialties in medicine and surgery.¹ The expansion of data capture, storage and availability has transformed the capacity to interpret large volumes of examination, radiographic and laboratory results, to produce prediction probabilities. Applications of AI in medicine and surgery are diverse, having been applied to image classification, prediction of diagnosis, clinical decision-making, and identification of patients at high risk for complications.^{2,3}

Computer-vision algorithms, such as image classification and object detection, represent clinically relevant applications of AI in medicine and surgery. Supervised learning models can be constructed using image datasets previously graded by experts. Examples of image classification algorithms include those that: differentiate skin lesions such as benign seborrheic keratosis from keratinocytic carcinomas and malignant melanomas;⁴ categorise chest radiographs for pleural effusion, cardiomegaly or mediastinal enlargement;⁵ distinguish tuberculosis from pulmonary nodules on chest radiographs;⁶ determine breast density on mammograms and classify tumours;⁷ and evaluate retinal fundus photographs to detect diabetic retinopathy and macular degeneration.⁸

In the field of otolaryngology, head and neck surgery, image classification algorithms have been constructed to: determine osteomeatal complex inflammation on computed tomography scans;⁹ distinguish squamous cell carcinoma and thyroid cancer from head and neck tissue samples;¹⁰ and detect endolymphatic hydrops in mice cochleae using optical coherence tomography.¹¹ These experiments reveal the feasibility of constructing convolutional neural networks to address clinical queries and to provide a preliminary step towards implementing computer-vision algorithms into otolaryngology, head and neck surgery practice. Various imaging modalities are frequently used during clinical examination and out-patient follow up, providing an ideal opportunity to investigate the application of AI in otolaryngology, head and neck surgery.

Artificial intelligence has the potential to automate routine clinical tasks to improve efficiency, reduce expenditure and provide interpretations of comparable accuracy to specialists in under-resourced settings. Otосcopy is a routine procedure performed by trained practitioners to evaluate the tympanic membrane.¹² In Australia, Indigenous children living in rural and remote areas have the highest rates of ear disease in the world.¹³ In Northern and Central Australia, 91 per cent of Indigenous children suffer from some form of otitis media, including: suppurative otitis media (chronic squamous otitis

media and chronic mucosal otitis media, 48 per cent); otitis media with effusion (31 per cent); acute otitis media without perforation (26 per cent); and acute otitis media with perforation (24 per cent).¹⁴

The key determinant of an accurate diagnosis is a trained clinician who can synthesise examination data and history. Specialist otolaryngologists are able to provide the most accurate diagnoses, but access to such specialists is limited in rural and remote communities, and this poses potential missed opportunities for diagnosis and treatment.^{15,16} Most ear disease screening is performed by local nurses or Indigenous healthcare workers in rural and remote areas, who have less clinical experience and formal training, and limited access to ear specialists (i.e. audiology, otolaryngology, hearing aid services).^{15,17,18} Undiagnosed tympanic membrane perforations increase the likelihood of middle-ear infection, and contribute to impaired hearing and language development.¹⁹ Otoscopy is the primary modality to identify perforations following careful history-taking.²⁰

Whilst otoscopic images can be readily gathered by healthcare workers with limited training and equipment,²¹ subsequent interpretation of gathered otoscopic images by specialist otolaryngologists may be delayed and impair expedient decision-making. Therefore, a computer-vision algorithm capable of identifying pathology has the potential to improve patient care in areas with limited access to otolaryngologists. As a preliminary step towards establishing a comprehensive tympanic membrane computer-vision tool, this study aimed to explore the feasibility of creating a proof-of-concept AI convolutional neural network computer-vision algorithm, to identify tympanic membrane perforations of various sizes.

Materials and methods

Ethics

All patient data in this study were obtained from de-identified publicly accessible online databases. Ethical approval was obtained from the Menzies School of Health Research, Northern Territory, Australia (health research ethics committee number: 2019-3410), and the study was undertaken in compliance with the Helsinki Declaration of 1975.²²

Image library

Otosopic images of human tympanic membranes were retrospectively collected manually using de-identified online image repositories (Google Images; Google, Mountain View, California, USA). Search terms included 'tympanic membrane', 'ear drum', 'otoscope', 'otoscopic' and 'perforation'. The images included in the otoscopic image dataset were required to have an unobstructed view of the entire tympanic membrane and illumination from a speculum light source. Prior to analysis, images were screened by the primary investigator (ARH). Ear canals obstructed with cerumen, grommets or foreign bodies, as well as cases of significant reconstruction or total perforation, were excluded. Images with over- or under-saturation from the light source that distorted interpretation were excluded. Images were cropped to remove unnecessary borders.

Presence and size of perforation

The 'gold standard' classification was defined by trained otolaryngologists, who independently classified each otoscopic image as with or without a tympanic membrane perforation.

Two investigators estimated the size of each perforation independently. The size of the tympanic membrane perforation was quantified as the proportion of pixels across the largest cross-sectional diameter of the perforation divided by the largest cross-sectional diameter of the entire tympanic membrane. Perforation size was categorised into three distinct groups: (a) small – less than one-third, (b) medium – one-third to two-thirds, or (c) large – more than two-thirds of the entire tympanic membrane (Figure 1). Measurements were performed using the GNU Image Manipulation Program, version 2.8.22 (GIMP Development Team, 2017).²³

Training and testing

The convolutional neural network architecture Inception-V3 (Google) was used. The methodology of convolutional neural networks and transfer learning have been previously described.²⁴ The transfer learning method was conducted in Python programming language using TensorFlow deep learning software (Google).

Training images were used to train the convolutional neural network during the transfer learning process. Test images were used to evaluate the performance of the algorithm. Fifty test images were retrospectively gathered from tele-otoscopy outreach campaigns from children in Wadeye, Northern Territory, Australia, between January 2018 and September 2019. As reported previously, the neural network is designed to be trained for 100 000 steps (number of integrations performed during training to compare model predictions, error and weight of each node) at a learning rate of 0.05 (rate at which the algorithm interprets training data or weight of node for each step).⁹

Performance evaluation

The performance of the model was assessed for accuracy using the receiver operating characteristics curve. Accuracy was defined as the percentage of correct responses determined by the algorithm performed on the test images. Output probabilities of each test image were used to construct a receiver operating characteristics curve. The area under the curve was used to evaluate the predictive capacity of the algorithm with corresponding 95 per cent confidence intervals (CIs). Statistical analysis was performed using SPSS software (version 20, 2011; IBM, Armonk, New York, USA).

Results

A total of 233 images (138 left and 95 right) were included in the image library (183 were used for training and 50 were used for testing), consisting of 105 intact and 128 perforated tympanic membranes. There was no disagreement between the two otolaryngologists who classified the images manually. Images for testing ($n = 50$; 27 left and 23 right) consisted of 25 intact and 25 perforated tympanic membranes (7 small, 7 medium and 11 large perforations).

A total of 38 out of 50 test images were correctly classified, yielding an accuracy of 76.0 per cent (95 per cent CI = 62.1–86.0 per cent) and an area under the curve of 0.867 (95 per cent CI = 0.771–0.963) (Table 1). Nineteen of 25 intact tympanic membranes (76.0 per cent) were correctly classified as 'intact'. Nineteen of 25 perforated tympanic membranes (76.0 per cent) were correctly classified as 'perforated'.

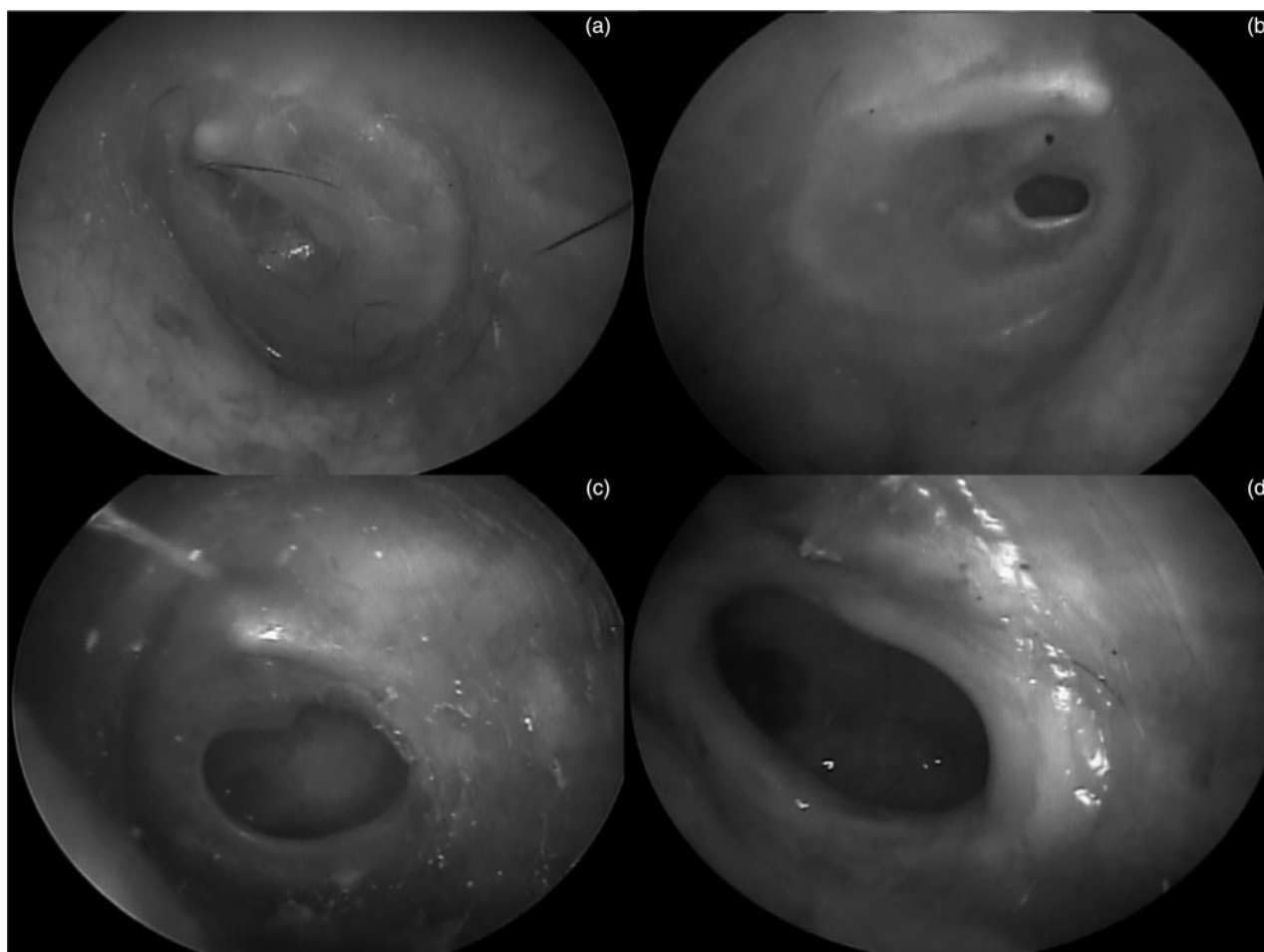


Fig. 1. Examples of tympanic membrane classifications from otoscopic images: (a) no perforation, (b) small perforation, (c) medium perforation and (d) large perforation.

Table 1. Comparison of accuracy and performance for algorithms to classify TM perforations by size

Parameters	TMs correctly classified (<i>n</i>)	Algorithm accuracy (%)	Area under curve (95% CI)
Overall model	38 of 50	76.0	0.867 (0.771–0.963)
Intact TM	19 of 25	76.0	
TM perforation size	19 of 25	76.0	
– Small (<1/3*)	6 of 7	85.7	
– Medium (1/3 to <2/3*)	6 of 7	85.7	
– Large (>2/3*)	7 of 11	63.6	

*Of the total tympanic membrane diameter. TM = tympanic membrane; CI = confidence interval

Stratifying algorithm responses by size of perforation, six of seven small perforations (85.7 per cent) and six of seven medium perforations (85.7 per cent) were correctly classified as ‘perforated’. Seven of 11 large perforations (63.6 per cent) were correctly classified as ‘perforated’.

Twelve of 50 images (24.0 per cent) were classified incorrectly, consisting of 6 false positives (i.e. tympanic membranes were incorrectly classified as perforated when intact) and 6 false negatives (i.e. tympanic membranes were incorrectly classified as intact when perforated). The false negative

classifications (*n* = 6) were for images of one small, one medium and four large tympanic membrane perforations.

Discussion

The key step in diagnosing ear diseases such as perforation, acute otitis media, or otitis media with effusion, following careful history-taking and before audiometry and tympanometry, is visual inspection of the tympanic membrane using an otoscope.¹² The accuracy of otoscopy to diagnose ear disease depends upon the experience of the practitioner.^{20,25,26} Although audiologists can recognise some tympanic membrane conditions with accuracy that is comparable to otolaryngologists, healthcare workers with less training have significantly lower accuracy.¹⁶ The implications of inaccurate diagnosis are significantly more severe in settings with limited access to otolaryngologists and audiologists.²⁷

This study demonstrates the feasibility of using a publicly available convolutional neural network architecture to develop a classification algorithm, using images from online repositories, to identify tympanic membrane perforations, and to test such an algorithm on otoscopic images collected in a rural setting. The algorithm identified both intact and perforated tympanic membranes with similar accuracy. The algorithm was more likely to identify a perforation on otoscopic images of small and medium tympanic membrane perforations. However, false negative classifications (i.e. classified as intact when a perforation was present) were mostly of otoscopic

images of large perforations (greater than two-thirds of the tympanic membrane diameter).

Accuracy differences between different sized perforations may be because of the algorithm's reduced ability to detect focal abnormalities on the tympanic membrane. For example, this image classifier model was constructed to view the entire otoscopic image, rather than focus on key components (e.g. malleus, annulus). It is plausible that as perforation size increased, the algorithm was unable to identify that the deeper structures exposed to view represented an abnormal tympanic membrane rather than an intact tympanic membrane. For smaller perforations, the algorithm likely recognised discrete aberrations in the consistency of the otoscopic image and considered these defects to represent perforations. This may seem counter-intuitive to clinical experience, where it may be easier to detect larger perforations than smaller ones.

In order to circumvent this limitation, object detection and segmentation techniques may be utilised in future iterations of the algorithm. This would allow the algorithm to be trained to focus on searching for specific defects in the tympanic membrane, rather than reviewing the entire otoscopic image. A previous experiment using object detection and segmentation yielded an accuracy of 81.4 per cent to detect suspicious tympanic membrane perforations.²⁸ The authors did not report or measure perforation size. Object detection and segmentation may be promising techniques to improve the capture of tympanic membrane perforations, but consideration of perforation size is warranted, as the threshold to which this affects performance is unclear.

Image classification algorithms may be most useful to detect global pathology on tympanic membranes, such as acute otitis media. For example, previous image classification applications using Inception-V3 and MobileNets yielded accuracy rates of 82.2 per cent and 80 per cent to detect acute otitis media, respectively, based on 108 images.²⁹ Another image classification application for acute otitis media using a decision tree analysis resulted in diagnostic accuracy of 80.0–82.35 per cent, based on 74 images.²⁷

As this was a proof-of-concept study, there are several limitations, which will be addressed in any future large-scale work. Firstly, images gathered from online repositories may not be comparable to images gathered in primary care settings. Despite this, publicly available images have been used for preliminary development and testing, with future endeavours aimed at increasing the image library prospectively, to potentially yield greater accuracy. The images were optimised and did not include features that would often be seen in the clinical setting, such as wax. The images did not represent the ultimate target group of rural and remote Indigenous children. More images, standardised collection methods, and a variety of perforations could improve the algorithm's ability to identify true perforations.

In addition, accuracy varied by size of perforation. This suggests that the algorithm requires more training images prior to testing. Accuracy may be improved in future experiments with the inclusion of segmentation, image augmentation techniques and hyperparameter tuning.

Additionally, this study was performed using still images rather than a series of images or a video. In practice, clinicians can evaluate the tympanic membrane from several angles, for a longer duration, and using pneumatic otoscopy. It is plausible that the diagnostic accuracy of the object detection algorithm could be improved by using a series of images or a video.

Otoscopy was performed by multiple practitioners, and the angle, depth, brightness and clarity could be inconsistent. It is

possible that a set of images performed in a standardised manner could yield greater diagnostic accuracy. However, the clinical value of this algorithm is to improve the diagnosis made by practitioners without specialist training in a variety of settings. Therefore, performance could be improved by collecting images from numerous sites and practitioners, to develop better diagnostic accuracy despite image heterogeneity.

Finally, the gold standard was based on otolaryngologists' assessment of images. Ideally, independent test results, such as tympanometry or findings recorded at the time of examination, would have been used to validate the otolaryngologists' assessment of the images.

- This study applied artificial intelligence (AI) methods, including machine learning and convolutional neural networks, to develop an image classification algorithm to detect tympanic membrane perforations
- A total of 233 (183 training, 50 testing) otoscopic images were used, and the 'gold standard' 'ground truth' was defined by otolaryngologists (small, medium or large perforations)
- The algorithm equally identified intact and perforated tympanic membranes (76.0 per cent, 95 per cent confidence interval (CI) = 62.1–86.0 per cent)
- Model performance was significantly accurate (area under the curve = 0.867, 95 per cent CI = 0.771–0.963)
- This study represents a preliminary step towards applying AI to ear disease screening
- This may be most applicable for healthcare workers in rural and remote communities distant from otolaryngology services

This study utilised technological advancements in AI, by applying a machine learning convolutional neural network algorithm, to identify tympanic membrane perforations using otoscopic images. The use of AI applications in medicine and surgery is expanding rapidly, with emphasis on computer vision, predictive outbreak models, natural language processing algorithms, and automated planning and scheduling applications.³⁰ In the field of otolaryngology, AI has the potential to be directed towards public health problems that persist in under-resourced communities. For example, significant government funding has been applied to ear disease in Aboriginal Australians through various grants, outreach campaigns and public awareness initiatives, without a significant improvement in disease incidence.³¹

Telemedicine has been shown to improve screening rates, expenditure and community engagement.³² However, telemedicine requires an investment of time and experienced otolaryngologists to review otoscopic findings. An alternative option may be to include AI into these strategies by automating the process of otoscopic image review, in real time. An accurate and reliable AI algorithm could serve to review otoscopic images quickly and efficiently, with comparable accuracy to experienced otolaryngologists, as a form of rapid screening, to triage patients requiring further review. An extension of this could be a smartphone application that provides non-specialty-trained healthcare workers in rural settings with instantaneous diagnoses, to potentially initiate early treatment and specialist referral.

To reach this goal, significantly more rigorous and comprehensive investigation is required for computer-vision algorithms. This would include a much larger image library, co-ordination across urban and rural settings, training in multiple otological conditions, in addition to tympanic membrane

perforation, and resolution of discrepancies in the current algorithm. Future endeavours are underway to prospectively collect photographs taken by otolaryngologists using endoscopes and handheld otoscopes, in order to expand the image library, refine model performance and fine-tune the algorithm for application using various instruments.

Conclusion

This study demonstrates the feasibility of constructing a proof-of-concept AI convolutional neural network algorithm to identify tympanic membranes with or without perforation. However, accuracy is affected by perforation size. Future endeavours are warranted to establish a comprehensive library of tympanic membrane pathology and construct a robust algorithm for use in daily practice. The application of an image classification algorithm may be useful to screen for ear disease in under-resourced communities distant from otolaryngology services.

Competing interests. Raymond Sacks is a consultant for Medtronic.

Acknowledgments. We sincerely appreciate and acknowledge Dr Graeme Crossland, Dr Hemi Patel, Dr Robyn Aitken, Christian Aguilar, Royal Darwin Hospital, Top End Health Service and the Department of Health for the Northern Territory, Australia. We thank the community councils and traditional owners for allowing this research in their communities. We acknowledge the assistance of many clinical and administrative staff in the Top End Health Service for their efforts collecting and managing clinical data used for this study.

References

- Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S *et al.* Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017;**2**:230–43
- McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP *et al.* Deep learning in radiology. *Acad Radiol* 2018;**25**:1472–80
- Liu X, Sinha A, Unberath M, Ishii M, Hager G, Taylor RH *et al.* Self-supervised learning for dense depth estimation in monocular endoscopy. *ArXiv* 2018;1–11
- Esteve A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;**542**:115–18
- Swaminathan S, Qirko K, Smith T, Corcoran E, Wysham G, Bazaz G *et al.* A machine learning approach to triaging patients with chronic obstructive pulmonary disease. *PLoS One* 2017;**12**:1–21
- Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017;**284**:574–82
- Wang D, Khosla A. Deep learning for identifying metastatic breast cancer. *ArXiv* 2016;1–6
- Kapoor R, Walters SP, Al-Aswad LA. The current state of artificial intelligence in ophthalmology. *Surv Ophthalmol* 2018;**64**:233–40
- Chowdhury NI, Smith TL, Chandra RK, Turner JH. Automated classification of osteomeatal complex inflammation on computed tomography using convolutional neural networks. *Int Forum Allergy Rhinol* 2018;**9**:46–52
- Halicek M, Little JV, Wang X, Griffith CC, El-Deiry MW, Chen AY *et al.* Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging. *J Biomed Opt* 2017;**22**:60503
- Liu GS, Zhu MH, Kim J, Raphael P, Applegate BE, Oghalai JS. ELHnet: a convolutional neural network for classifying cochlear endolymphatic hydrops imaged with optical coherence tomography. *Biomed Opt Express* 2017;**8**:4579–94
- Rosenfeld RM, Shin JJ, Schwartz SR, Coggins R, Gagnon L, Hackell JM *et al.* Clinical practice guideline: otitis media with effusion executive summary (update). *Otolaryngol Neck Surg* 2016;**154**:201–14
- Dart J. Australia's disturbing health disparities set Aboriginals apart. *Bull World Health Organ* 2008;**86**:245–7
- Morris PS, Leach AJ, Silberberg P, Mellon G, Wilson C, Hamilton E *et al.* Otitis media in young Aboriginal children from remote communities in Northern and Central Australia: a cross-sectional survey. *BMC Pediatr* 2005;**5**:27
- Gibney K, Morris P, Carapetis J, Skull S, Leach A. Missed opportunities for a diagnosis of acute otitis media in Aboriginal children. *J Paediatr Child Health* 2003;**39**:540–2
- Gunasekera H, Miller HM, Burgess L, Chando S, Sheriff SL, Tsembe JD *et al.* Agreement between diagnoses of otitis media by audiologists and otolaryngologists in Aboriginal Australian children. *Med J Aust* 2018;**209**:29–35
- Australian Medical Association. *2017 AMA Report Card on Indigenous Health: A National Strategic Approach to Ending Chronic Otitis Media and its Life Long Impacts in Indigenous Communities*. Barton, ACT: AMA, 2017
- Gunasekera H, Morris PS, Daniels J, Couzos S, Craig JC. Otitis media in Aboriginal children: the discordance between burden of illness and access to services in rural/remote and urban Australia. *J Paediatr Child Health* 2009;**45**:425–30
- Morris PS, Leach AJ, Halpin S, Mellon G, Gadil G, Wigger C *et al.* An overview of acute otitis media in Australian Aboriginal children living in remote communities. *Vaccine* 2007;**25**:2389–93
- Oyewumi M, Brandt MG, Carrillo B, Atkinson A, Iglar K, Forte V *et al.* Objective evaluation of otoscopy skills among family and community medicine, pediatric, and otolaryngology residents. *J Surg Educ* 2016;**73**:129–35
- Elliott G, Smith AC, Bensink ME, Brown C, Stewart C, Perry C *et al.* The feasibility of a community-based mobile telehealth screening service for Aboriginal and Torres Strait Islander children in Australia. *Telemed J E Health* 2010;**16**:950–6
- World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;**310**:2191–4
- GIMP: GNU Image Manipulation Program. In: <https://www.gimp.org/> [9 March 2020]
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 2012;**25**:1097–105
- Asher E, Leibovitz E, Press J, Greenberg D, Bilenko N, Reuveni H. Accuracy of acute otitis media diagnosis in community and hospital settings. *Acta Paediatr* 2005;**94**:423–8
- Buchanan CM, Pothier DD. Recognition of paediatric otopathology by general practitioners. *Int J Pediatr Otorhinolaryngol* 2008;**72**:669–73
- Myburgh HC, van Zijl WH, Swanepoel DW, Hellström S, Laurent C. Otitis media diagnosis for developing countries using tympanic membrane image-analysis. *EBioMedicine* 2016;**5**:156–60
- Seok J, Song J-J, Koo J-W, Kin H-C, Choi BY. The semantic segmentation approach for normal and pathologic tympanic membrane using deep learning. *BioRxiv* 2019;In press
- Kasher MS. *Otitis Media Analysis: An Automated Feature Extraction and Image Classification System*. Helsinki: Helsinki Metropolia University of Applied Sciences, 2018
- Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob Health* 2018;**3**:e000798
- Kenyon G. Social otitis media: ear infection and disparity in Australia. *Lancet Infect Dis* 2017;**17**:375–6
- Nguyen KH, Smith AC, Armfield NR, Bensink M, Scuffham PA. Cost-effectiveness analysis of a mobile ear screening and surveillance service versus an outreach screening, surveillance and surgical service for Indigenous children in Australia. *PLoS One* 2015;**10**:e0138369