DrumBeat.ai: Addressing paediatric Indigenous ear disease in rural and remote Australia using artificial intelligence

Al-Rahim Habib

A thesis with publications submitted in fulfilment of the requirements for the degree of Doctor Philosophy

This research reported in this thesis was supported by the award of a Research Training Program scholarship to the PhD Candidate

School of Public Health

Faculty of Medicine and Health

University of Sydney

2023

CULTURAL CARE WARNING NOTICE

Cultural advice
This item may contain culturally sensitive information.

All users are advised that this item may contain images, voices and/or names of people who have died.

Indigenous Cultural and Intellectual Property
This item may contain Indigenous Cultural and Intellectual Property. Please consult with the relevant communities if you wish to use any of the content in this item.

Copyright

TABLE OF CONTENTS

STATEMENT OF ORIGINALITY

This is to certify that, to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Al-Rahim Habib

ACKNOWLEDGEMENTS

I acknowledge the traditional owners of the lands where this research was conducted, their spiritual relationship with the land, and their elders, past, present, and emerging.

I extend my profound appreciation to my primary supervisor, Associate Professor Narinder Singh, and associate supervisor Professor Raymond Sacks. Your guidance, mentorship and leadership have been instrumental in my academic and personal growth. Thank you for the opportunities you provided, the expertise you shared, and the confidence you instilled in me.

Thank you to my co-authors and mentors, including Dr Graeme Crossland, Dr Hemi Patel, Professor Chris Perry, and Professor Kelvin Kong. Your insight and encouragement have enriched both my research and my understanding in immeasurable ways.

I am thankful to the University of Sydney, Westmead Hospital, Royal Darwin Hospital, and Queensland Health for their immense support. Special acknowledgment and sincere gratitude to the Passe and Williams Foundation, Avant Mutual, and the Microsoft AI for Good Lab for supporting this research.

I also acknowledge the clinicians, nurses, audiologists, healthcare workers, and administrative staff for their contributions. The experience you shared with me has been invaluable to understand the real-world implications of this research.

My heartfelt gratitude and deep appreciation to my wife, Faaiza. Your unwavering support, care, and encouragement have been my guiding light through every phase of this journey. Thank you for always radiating positive energy and embracing every new challenge and adventure alongside me. Your constant encouragement and unhesitating support have not only been a source of strength but also an inspiration, fuelling my confidence and dedication to this endeavour.

LIST OF FIGURES

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

LIST OF TABLES

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

ABBREVIATIONS

AHW	Aboriginal healthcare worker
AI	Artificial intelligence
ANN	Artificial neural network
AOM	Acute otitis media
AOMwiP	Acute otitis media with perforation
AOMwoP	Acute otitis media without perforation
AUC	Area under the curve
CHQ	Children's Health Queensland
CI	Confidence interval
CNC	Clinical nurse consultant
CNN	Convolutional neural network
COM	Chronic otitis media
COMwAP	Chronic otitis media with active perforation
COMwDP	Chronic otitis media with dry perforation
DPI	Dots per inch
DT	Decision tree
EAC	External acoustic canal
ENT	Ear, nose, and throat
ETD	Eustachian tube dysfunction
GAN	Generative adversarial network
GPU	Graphics processing unit

Grad-CAM	Gradient-weighted class activation mapping
HL	Hearing loss
k-NN	k-nearest neighbours
MeSH	Medical subject headings
NPV	Negative predictive value
NT	Northern Territory
OE	Otitis externa
OM	Otitis media
OME	Otitis media with effusion
OR	Odds ratio
PPV	Positive predictive value
PROSPERO	Prospective Register of Systematic Reviews
QUADAS - 2	Quality Assessment and Diagnostic Accuracy Studies - 2
RDH	Royal Darwin Hospital
ROC	Receiver operating characteristics curve
SD	Standard deviation
SEN	Sensitivity
SPEC	Specificity
STARD	Standard for Reporting Diagnostic Accuracy Studies
TM	Tympanic membrane
WHO	World Health Organisation

ABSTRACT

Ear disease is an important public health issue for Aboriginal and Torres Strait Islander children. Chronic ear disease can lead to long-term hearing loss. The consequences of long-term hearing loss include impacts on speech and language development, academic performance, behaviour, social skills, and future employment opportunities. Artificial intelligence (AI) is an emerging technology that has the potential to transform ear disease screening and triage.

This thesis is comprised of an introductory chapter (**Chapter 1**), a systematic review and meta-analysis of machine learning techniques for ear disease classification (**Chapter 2**), a comprehensive overview of a telehealth service evaluating ear disease in rural and remote areas (**Chapter 3**), an assessment of inter-rater agreement in ear disease diagnoses using a telehealth approach (**Chapter 4**), an exploration of the performance and generalisability of image classification algorithms for otoscopy (**Chapter 5**), an evaluation of a supervised image classification system to triage otoscopic images from Aboriginal and Torres Strait Islander people in rural and remote areas (**Chapter 6**), and a discussion regarding the significance of findings, clinical and public health implications, strengths and limitations, and future directions (**Chapter 7**).

This thesis provides a framework illustrating an application of AI to address an important clinical need. The implications of this research extend to a diverse audience, including Aboriginal and Torres Strait Islander communities, healthcare professionals involved in ear disease screening and triage, otolaryngologists engaged in telehealth programs and outreach services, data scientists exploring applications of AI in healthcare, and policymakers evaluating innovative solutions to enhance healthcare delivery in resource-constrained settings.

CHAPTER ONE: Overview

1.1 Background

Ear disease is a significant global public health challenge that affects individuals of all ages and socioeconomic backgrounds.[1] The prevalence of ear disease varies across regions, influenced by factors such as access to healthcare, socioeconomic status, and environmental conditions.[2] Developing countries experience high rates of ear infections and hearing loss among children, while developed nations experience increased rates of age-related hearing loss among older adults.[3] This global burden of ear disease contributes to years lived with disability, impacting more than 5% of the world's population.[4] According to the World Health Organization (WHO), over 900 million people are projected to experience hearing loss by 2050.[3] Various factors contribute to hearing loss, including congenital conditions, age-related hearing changes, noise exposure, ototoxicity, and otitis media (OM).[1]

OM is characterised by inflammation of the middle ear, manifesting as either acute otitis media (AOM), otitis media with effusion (OME) or chronic otitis media (COM).[5] The pathophysiology of OM is multifactorial, involving a complex interplay of biological, environmental, and social determinants of health.[6,7] Biological factors, such as genetic predisposition, viral and bacterial infections, and autoimmune disorders, contribute to the risk of developing OM.[6] Social factors, such as poverty, poor living conditions, hygiene, overcrowding, exposure to tobacco smoke, and limited access to healthcare, significantly influence the development and prevalence of OM.[6,8] Limited healthcare access, especially in rural and remote areas, can cause delays in diagnosing and treating ear disease, resulting in recurrent and chronic OM with an increased likelihood of complications.[6,8] Indigenous children experience OM at a younger age with more frequent and persistent infections than non-Indigenous children.[6]

The pathophysiology of OM encompasses microbial factors, host immune responses, and anatomical factors such as Eustachian Tube dysfunction.[9] Complex interactions between otopathogens can colonise the nasopharynx to increase the susceptibility of developing OM.[9,10] Common viruses causing upper respiratory tract infections, such as respiratory syncytial virus, rhinovirus, adenovirus, parainfluenza, and coronavirus, are frequently linked with AOM.[7,11] The predominant bacteria responsible for OM include Streptococcus pneumoniae, Moraxella catarrhalis, and non-typeable Haemophilus influenzae.[10,11]

OM encompasses all forms of inflammation and infection of the middle ear, typically associated with the presence of fluid in the middle ear space, known as middle ear effusion.[12–14] This condition is broadly classified into several types based on the clinical presentation and duration of the symptoms. AOM is characterized by symptoms of ear pain, redness of the tympanic membrane, and at times, purulent discharge.[15] OME involves serous or transudative fluid behind the tympanic membrane which can present with or without symptoms, often following an episode of AOM or appearing independently.[16,17] Chronic forms of OM represent more prolonged conditions where the infection or effusion lasts longer and may present with or with actively discharging perforations.[18]

OME can manifest as episodic or persistent.[19] Episodic OME typically lasts less than three months and can resolve spontaneously, whereas persistent OME lasts longer than three months and may require more intensive intervention.[13,14,20] The diagnosis of OME often involves the use of pneumatic otoscopy and tympanometry, specifically a type B tympanogram, which can be used to detect fluid in the middle ear.[13,19,20] Commonly referred to as "glue ear," "serous otitis media," or "secretory otitis media," OME is significant in paediatric populations where it

can impact hearing and speech development.[21,22] Management strategies may include observation, medical therapy, or surgical interventions such as tympanostomy tube placement.[23]

COM with active perforation is defined by persistent ear discharge through a tympanic membrane perforation lasting more than two weeks, often requiring topical antibiotic treatment due to the size of the perforation which allows penetration into the middle ear space.[24,25] COM with dry perforation refers to a perforation of the tympanic membrane without active infection or discharge.[26] Cholesteatoma, a more severe form of chronic ear disease, involves the abnormal growth of skin cells in the middle ear and can lead to the erosion of surrounding structures, causing significant complications such as hearing loss, facial paralysis, or intracranial issues.[26] Managing these chronic conditions often involves a combination of medical and surgical treatments, tailored to the extent of disease and the specific symptoms presented by the patient.[23,25,27]

Diagnosing OM presents several controversies and challenges, primarily due to the variability in its presentation and the subjective interpretation of its signs and symptoms.[28] One major controversy involves the differentiation between AOM and OME, as both conditions can exhibit similar symptoms such as fluid behind the tympanic membrane but require different management strategies. Accurately distinguishing between these can be challenging in clinical practice within limited user experience.[28] The reliance on pneumatic otoscopy for diagnosing OM is an area of debate; while it is a standard tool, its effectiveness heavily depends on the experience and skill of the clinician, leading to potential variability in diagnoses.[29–31] Tympanometry is diagnostic modality used to assess the presence of fluid in the middle ear by measuring eardrum mobility, but its interpretation can be complicated by factors such as patient age and the presence of ear canal debris.[32] The use of antibiotics for AOM, with differing

guidelines on whether to initiate immediate treatment or adopt a watch-and-wait approach, reflecting uncertainty in diagnostic accuracy and prediction of disease course.[33,34] Diagnosing OM in young children who cannot articulate their symptoms complicates the clinical assessment, often leading to either under or over-treatment.[35]

Misdiagnosing AOM as OME, for example, may lead to unnecessary antibiotic prescriptions, contributing to antibiotic resistance and exposing patients to unnecessary side effects.[36,37] Failing to correctly identify and treat AOM can result in complications such as mastoiditis, hearing loss, or even more severe intracranial infections. For OME, which often does not require antibiotics, misdiagnosis can lead to inappropriate treatment interventions that do not address the actual condition, potentially prolonging the fluid accumulation and impacting a child's hearing and speech development.[22,38] For COM, especially with persistent discharge, inaccurate diagnosis can delay surgical interventions or other targeted treatments that are necessary to prevent ongoing damage to the middle ear structures, which could culminate in permanent hearing loss or, in severe cases, life-threatening complications.[27,39]

Recurrent and chronic OM is a major public health concern for Aboriginal and Torres Strait Islander people in Australia, with this population experiencing some of the highest rates of OM worldwide.[4,40] Aboriginal and Torres Strait Islander children have a three times higher prevalence of OM compared to the national average.[41–43] Indigenous adults are nearly four times more likely to experience hearing loss than non-Indigenous adults.[43] Social and environmental factors, such as overcrowded housing, poverty, exposure to tobacco smoke, and limited access to services, are key contributors to the high prevalence of ear disease among these communities.[6,8,44]

Recurrent and chronic forms of OM can have lasting effects on children, with implications that extend well into adulthood.[45] Long-term hearing loss resulting from recurrent and chronic OM can impede speech and language development in children[45] This impairment can hinder their ability to participate in social activities, contributing to social isolation and reduced quality of life.[46–48] The effects of OM on children's speech and language development can directly influence academic performance.[49] Struggling to hear and understand instructions, engage in classroom discussions, and develop literacy skills can put these children at a disadvantage compared to their peers.[50–52] This disparity in academic achievement can have a downstream impact on future employment opportunities, contributing to a cycle of socio-economic disadvantage..[53]

The implications of ear disease in Aboriginal and Torres Strait Islander people include significant health, economic, and social consequences.[21,44,54,55] Complications of untreated chronic suppurative OM include chronic mastoiditis, facial nerve paralysis, lateral sinus thrombosis, labyrinthitis, meningitis, brain abscess, and potentially fatal outcomes.[40] Economic consequences involve decreased work productivity and increased healthcare costs, while social consequences include stigma, discrimination, and reduced social participation.[56] Early detection and intervention are important to prevent the progression of ear disease and reduce the risk of long-term complications.[57] This can be achieved through community-based screening programs, regular ear health checks, prompt referral to specialist services, and timely provision of appropriate treatment.[57] Aboriginal and Torres Strait Islander healthcare workers, nurses, primary care practitioners, audiologists and Ear, Nose and Throat (ENT) specialists play a vital role in managing and triaging ear diseases, particularly within Indigenous communities.[57] A collaborative approach through a multi-disciplinary team can support culturally appropriate care,

facilitate effective communication between communities and healthcare services, and advocate for improved ear health outcomes.[57] However, practitioners in rural and remote areas encounter a higher burden of children with OM, with limited access to specialist ear health services compared with urban counterparts.[58] The limited availability of local audiology, ENT and hearing aid services can lead to prolonged wait times and a consequently higher risk of complications.[58] One in five Indigenous children wait longer than the recommended three months for audiology testing, and one in eight Indigenous children experience delays longer than the recommended six months for ENT services.[22,58]

Attempts to address the high prevalence of ear disease in rural and remote Indigenous communities include patients and families traveling to access specialist care or outreach clinics, including "fly-in and fly-out" models.[59] Outreach services involve healthcare professionals periodically traveling to rural and remote communities, providing specialised care, consultation, and education through multidisciplinary teams.[59] This approach reduces the need for patients to travel long distances, can reach geographically dispersed communities, minimises the risk for loss to follow up, and promotes patient-centred and culturally appropriate care.[59] However, many remote Indigenous communities are in isolated and hard-to-reach areas, contributing to limited visit frequencies and insufficient coverage. Sustaining a qualified healthcare workforce and ensuring a consistent supply of resources can be difficult due to recruitment challenges and budget constraints.[60] Providing culturally appropriate healthcare services requires a deep understanding of Indigenous cultures and languages.[60] Patient compliance and attendance can be influenced by cultural beliefs, competing priorities, and reluctance to seek medical care.[60] Ensuring continuity of care can be challenging with periodic visits, leading to missed opportunities to manage chronic conditions and regular follow-up.[60]

To address the challenges faced in rural and remote areas, telehealth approaches have been explored as a means of providing ear disease assessments.[61,62] Telehealth refers to the use of telecommunications technology to deliver healthcare services remotely.[63,64] Telehealth encompasses a wide range of healthcare services delivered remotely to connect patients and healthcare providers.[65] Telehealth is classified into two main categories: synchronous and asynchronous approaches.[65] Synchronous telehealth involves real-time interactions between healthcare professionals and patients, enabling immediate communication and monitoring.[65] Asynchronous telehealth, also known as using a store-and-forward approach, utilises electronic recordings and screening of patient-generated health data, allowing for a flexible and time-shifted interaction model where assessment to interpretation may vary at different time points.[65]

Synchronous telehealth, characterised by live interactions, allows healthcare providers to engage with patients in real-time, resembling traditional face-to-face consultations.[66] This approach is particularly valuable for immediate assessments supporting timely and responsive care delivery.[67] Asynchronous telehealth operates on a store-and-forward model, where data and information are captured and transmitted to healthcare providers for later review and analysis.[68] This method is beneficial for scenarios where real-time communication is not essential, allowing for more convenient and efficient healthcare delivery, especially in managing screening, chronic conditions and follow-up care.[68]

The first recorded instance of telehealth occurred in the 1900s, when electrocardiograms were transmitted over telephone lines.[69] In the 1960s, telehealth technology expanded significantly, driven by the space race and the necessity of monitoring astronauts' health remotely.[70] NASA's telemedicine initiatives used advanced telemetry systems to track the physiological parameters of astronauts during space missions.[70] The advent of the internet in the

1990s marked a pivotal shift in telehealth, enabling faster and more secure communication methods.[65,68] The use of email consultations and the development of early forms of electronic health records allowed for better management and accessibility of health data.[71,72] The internet facilitated the exchange of medical information, enhanced collaboration among healthcare providers and had the potential to improve patient care outcomes.[73] The 2000s brought further advancements in telehealth with the introduction of digital health records and improved broadband connectivity.[74] These developments facilitated more robust and interactive telehealth services, enabling high-definition video conferencing and secure data transmission in rural, remote and urban settings.[74]

Technological advancements have enhanced the expansion of telehealth, enabling it to become an important component of modern healthcare, especially during the COVID-19 pandemic where face-to-face examinations were limited.[75–78] High-speed broadband connectivity have been used to enable real-time video consultations, ensuring that patients and healthcare providers can communicate effectively without significant delays or disruptions.[79–81] High-quality video conferencing capability is important for accurate assessments to observe patients' conditions closely and make informed decisions.[82]

Still images are commonly used in store-and-forward telehealth applications, where medical images such as dermatological photographs, wound images, or retinal scans are captured and sent to a specialist for analysis at later time points.[83,84] This method is advantageous for conditions that require expert evaluation but not immediate intervention, enabling specialists to assess cases from remote locations without the need for real-time interaction.[85] Still images require less bandwidth and technological infrastructure, making them accessible and cost-effective, especially in resource-limited settings.[86] However, the static nature of still images

means they may miss transient or progressive symptoms that would be observable in a live examination.[87] For example, the assessment of a chronic skin condition might require observing changes over time or understanding the immediate effects of an applied treatment, nuances that still images alone might not convey.[87,88]

Conversely, video offers dynamic interaction capabilities that can provide valuable insight in clinical scenarios.[89] Through real-time video conferencing or recording videos for review at a later time, healthcare providers can conduct thorough examinations, observe symptoms as they manifest, review pathology at various angles and variations, which may mirroring an in-person clinical examination.[90] This is can be useful in primary care, emergency medicine settings, and post-surgical follow-ups, where visual cues and patient interactions play a significant role in diagnosis and treatment planning.[90–92] Despite these advantages, video consultations come with higher technical demands.[76] They require robust, high-speed internet connections to ensure video quality and continuity during transfer, larger data storage requirements, and instrumentation capable of collecting video assessments, which can be a significant barrier in rural or under-resourced areas.[76]

In the context of ear disease, telehealth enables ENT specialists to remotely assess ear conditions, provide guidance, and facilitate referrals for further care.[62] The asynchronous "store-and-forward" telehealth approach involves collecting clinical history, performing digital video otoscopy, tympanometry and audiometry on-site in the remote location, then transmitting these results to specialists at a different location for review at a later time.[93,94] This technology can bridge the geographical distance between patients and specialists, allowing individuals in underserved communities to access specialised ear health services that may otherwise be difficult

to obtain.[61,94,95] It also reduces the need for patients to travel long distances, which can be burdensome for those living in rural and remote areas.[61,93]

There are limitations with the store-and-forward telehealth approach. Delays between frontline examination, data forwarding, specialist review and final intervention can hinder timely diagnosis and treatment. Limited real-time communication may restrict patients' and healthcare workers' immediate access to guidance or clarifications.[96,97] The reliance on internet connectivity and data storage infrastructure can pose challenges in regions with poor network coverage.[98,99] Data loss or incomplete information during transmission can restrict the capacity for specialists to make accurate assessments.[100] Addressing the limitations of store-and-forward, asynchronous telehealth approaches can enhance service provision to rural and remote communities.

AI has the potential to address some of the limitations of asynchronous telehealth.[101] AI can be defined as training computational algorithms to complete tasks that require human intelligence. Machine learning and deep learning approaches can be used to analyse large datasets and identify patterns in medical images to aid in identification and diagnosis.[102–104] By training algorithms on diverse cases, AI can assist in the automated analysis of visual data, such as clinical images or radiological scans, to detect abnormalities and assist with disease classification.[105–107] Supervised learning involves human experts training algorithms using labelled data to predict outcomes or classify data into predefined categories.[108] Unsupervised learning involves the algorithms themselves identifying patterns and relationships within unlabelled data.[108]

Artificial neural networks (ANNs) are designed based on the structure of biological nervous systems, consisting of interconnected nodes that mimic the functioning of neurons, enabling complex computational tasks.[109] ANNs consist of multiple 'neurons' or interconnected

processes (i.e. nodes) that are comprised of an input, hidden and outer layer. Input layers receive information from previous output layers to produce new outputs that are used to establish a multi-layer, feed-forward system. Hidden layers (between input and output layers) contain individual feature detection algorithms that can recognise increasingly complex patterns through each subsequent node.

Convolutional neural networks (CNNs), a type of ANN, are iterative deep learning networks that automatically extract multiple image features using convolutional, pooling and fully connected sequential layers.[110] Inspired by the organisation of the animal visual cortex, CNNs generate simple-to-complex patterns from spatial hierarchical arrangements on input images.[111] The initial convolutional layers extract features from input images (e.g. kernels) using multi-sized filters or masks to predict the class to which each feature belongs.[111] Convolutional layers can evaluate simple image features, such as the detection of edges, to identify key image features. Multiple convolutional kernels are combined to develop feature maps that represent different characteristics of the input image.[112] Following this, the pooling layer reduces the size of the feature map by combining adjacent pixels into a single representative value.[112] The purpose of the pooling layer is to summarise feature maps to reduce the overall computational cost. In the fully connected layer, convolutional and pooling layers are converted to single vectors that can be used as inputs for the subsequent stages.[113] Pre-trained CNNs can be re-purposed to evaluate various image datasets of interest by way of transfer learning.[114]

In the context of healthcare, image classification projects require a systematic approach including problem identification, data pre-processing, algorithm selection, model training, validation of performance and evaluating generalisability.[114,115] This technology can provide healthcare professionals with additional information and improve the accuracy of diagnoses.[116]

In resource-poor settings, AI applications demonstrate promising potential to improve clinical judgement and decision-making.[117]

In the field of otology, AI has been applied to optimise hearing aid technology, enhance speech processing techniques, and aid in the diagnosis and management of vestibular disorders.[118] For instance, algorithms can analyse audiological data to predict outcomes of sensorineural hearing loss or interpret automatic brainstem responses.[118] Additionally, advanced imaging modalities and image-processing techniques driven by AI have shown promise in improving the visualisation and characterization of ear pathologies using otoscopic images.[119] Image classification AI algorithms may have the potential to address the limitations of the asynchronous store-and-forward telehealth approach to enhance service delivery.

1.2 Purpose and significance of the research

The overarching purpose of this thesis is to explore the use of AI to triage otoscopic images to identify ear disease in Aboriginal and Torres Strait Islander children living in rural and remote areas. Findings from this research can assist community health workers and nurses to triage otoscopic eardrum images to identify OM and recognise cases that require specialist review. The specific aims of this research are to:

1. Conduct a comprehensive literature review on automating otoscopy and the use of AI algorithms for classifying ear disease from otoscopic images (**Chapter 2**).
2. Analyse a tele-otoscopy, telehealth database from the Northern Territory to understand the prevalence, distribution, and progression of ear disease in rural and remote Aboriginal and Torres Strait Islander communities (**Chapter 3**).
3. Evaluate the inter-rater agreement among otolaryngologists to diagnose OM using clinical data collected during telehealth assessments (**Chapter 4**).

4. Investigate the generalisability and performance of deep learning algorithms in detecting middle ear disease using otoscopy (**Chapter 5**).
5. Develop and evaluate a supervised computer-vision algorithm to triage otoscopic images for Aboriginal and Torres Strait Islander people living in rural and remote areas (**Chapter 6**).

1.3 Thesis outline

Chapter 2 provides a systematic review and meta-analysis focusing on the use of various AI methods to classify ear disease from otoscopic images.

Chapter 3 summarises the distribution of ear disease in rural and remote indigenous communities in the Northern Territory, utilising data from a referral-based telehealth program conducted between 2010 and 2019. This chapter examines the prevalence, distribution, and specific challenges faced in rural and remote areas, which will inform the development of the subsequent otoscopic image classifier.

Chapter 4 focuses on evaluating the inter-rater agreement among 13 otolaryngologists to diagnose otitis media in Aboriginal and Torres Strait Islander children using a telehealth approach. This chapter assesses the reliability of the telehealth diagnosis, highlighting the potential of telehealth databases to be used as a training source for an otoscopic image classifier.

Chapter 5 investigates the generalisability and performance of otoscopic image classifiers to detect middle ear disease using otoscopy. This chapter assesses the algorithms' accuracy, reliability, and robustness in diverse populations and real-world settings.

Chapter 6 explores a supervised image classification algorithm to triage otoscopic images of Aboriginal and Torres Strait Islander children living in rural and remote areas of

Australia. This chapter evaluates the algorithm's performance to classify otoscopic images using binary and multi-classification models.

Chapter 7 provides a summary of key findings, discussion of the significance of the studies presented in this thesis, strengths and limitations, implications for clinical practice and policy, and directions for future research.

Additional information that is relevant to the studies presented in this thesis and supplemental material to support the thesis has been included in the **Appendix**.

1.4 References

1. World Health Organisation. World Report on Hearing [Internet]. Geneva; 2021 [cited 2022 Aug 28]. Available from: https://www.who.int/publications/i/item/world-report-on-hearing
2. About ear health | Australian Government Department of Health and Aged Care [Internet]. [cited 2023 Jul 7]. Available from: https://www.health.gov.au/topics/ear-health/about
3. Deafness and hearing loss [Internet]. [cited 2023 Jul 7]. Available from: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
4. Graydon K, Waterworth C, Miller H, Gunasekera H. Global burden of hearing impairment and ear disease. J Laryngol Otol [Internet]. 2019 Jan 1 [cited 2023 Jul 7];133(1):18–25. Available from: https://www.cambridge.org/core/journals/journal-of-laryngology-and-otology/article/global-burden-of-hearing-impairment-and-ear-disease/31921A053C9152DCBD5F6962D8461A0B
5. Australian Institute of Health and Welfare. Ear health and hearing loss among Indigenous children. Canberra; 2018. (16).
6. Jervis-Bardy J, Sanchez L, Carney AS. Otitis media in indigenous australian children: Review of epidemiology and risk factors. Journal of Laryngology and Otology. 2014;128(SUPPL.S1).
7. Jervis-Bardy J, Rogers GB, Morris PS, Smith-Vaughan HC, Nosworthy E, Leong LEX, et al. The microbiome of otitis media with effusion in Indigenous Australian children. Int J Pediatr Otorhinolaryngol [Internet]. 2015;79(9):1548–55. Available from: http://dx.doi.org/10.1016/j.ijporl.2015.07.013
8. Delacy J, Dune T, MacDonald JJ. The social determinants of otitis media in aboriginal children in Australia: Are we addressing the primary causes? A systematic content review. BMC Public Health. 2020;20(1):1–9.
9. Butler CC, Williams RG. The etiology, pathophysiology, and management of otitis media with effusion. Curr Infect Dis Rep [Internet]. 2003 [cited 2023 Jul 7];5(3):205–12. Available from: https://link.springer.com/article/10.1007/s11908-003-0075-x
10. Acharya A, Reid A. Acute Otitis Media. Logan Turner's Diseases of the Nose, Throat and Ear Head and Neck Surgery: Eleventh Edition [Internet]. 2023 Apr 15 [cited 2023 Jul 7];403–9. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470332/
11. Thornton RB, Hakansson A, Hood DW, Nokso-Koivisto J, Preciado D, Riesbeck K, et al. Panel 7 - Pathogenesis of otitis media - a review of the literature between 2015 and 2019. Int J Pediatr Otorhinolaryngol [Internet]. 2020 Mar 1 [cited 2023 Jul 7];130 Suppl 1. Available from: https://pubmed.ncbi.nlm.nih.gov/31879085/
12. Schilder AGM, Chonmaitree T, Cripps AW, Rosenfeld RM, Casselbrant ML, Haggard MP, et al. Otitis media. Nat Rev Dis Primers [Internet]. 2(1). Available from: http://dx.doi.org/10.1038/nrdp.2016.63
13. Rovers MM, Schilder AGM, Zielhuis GA, Rosenfeld RM. Otitis media. The Lancet [Internet]. 363(9417):1325. Available from: http://dx.doi.org/10.1016/S0140-6736(04)16012-1
14. Schilder AGM, Chonmaitree T, Cripps AW, Rosenfeld RM, Casselbrant ML, Haggard MP, et al. Otitis media. Nat Rev Dis Primers [Internet]. 2(1). Available from: http://dx.doi.org/10.1038/nrdp.2016.63

15. Shawabkeh MA Al, Haidar H, Larem A, Aboul-Mahmood Z, Alsaadi A, Alqahtani A. Acute Otitis Media- An Update.
16. Browning G. Otitis media with effusion.
17. Rentschler H, Gibbs JL. Otitis Media with Effusion. The Guthrie Journal [Internet]. 24(1):34–42. Available from: http://dx.doi.org/10.3138/guthrie.24.1.034
18. Morris PS, Leach AJ. Acute and Chronic Otitis Media. Pediatr Clin North Am.
19. Vanneste P, Page C. Otitis media with effusion in children: Pathophysiology, diagnosis, an d treatment. A review. J Otol [Internet]. 14(2):33–9. Available from: http://dx.doi.org/10.1016/j.joto.2019.01.005
20. Li–min L, Min–ming D. Management of Otitis Media with Effusion. J Otol [Internet]. 3(2):68–75. Available from: http://dx.doi.org/10.1016/S1672-2930(08)50017-1
21. Leach AJ. Otitis media in Australian Aboriginal children: An overview. Int J Pediatr Otorhinolaryngol. 1999;49(SUPPL. 1):S173–8.
22. Morris P, Ballinger D, Leach A. Recommendations for Clinical Care Guidelines on the Management of Otitis Media in Aboriginal and Torres Strait Islander Populations. Canberra, ACT; 2010.
23. Wallace IF, Berkman ND, Lohr KN, Harrison MF, Kimple AJ, Steiner MJ. Surgical Treatments for Otitis Media With Effusion: A Systematic Revie w. Pediatrics [Internet]. 133(2):296–311. Available from: http://dx.doi.org/10.1542/peds.2013-3228
24. Deshmukh PT, Khan FQ, Gaurkar SS. Surgical Outcomes of Tympanoplasty with Various Graft Materials in Chr onic Otitis Media: A Retrospective Cohort Study. JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH [Internet]. Available from: http://dx.doi.org/10.7860/jcdr/2024/69598.19284
25. Mahajan R, Abraham N, T. M. N. A clinical study of evaluation of tympanoplasty with mastoidectomy in mucosal chronic otitis media. International Journal of Otorhinolaryngology and Head and Neck Surgery [Internet]. 6(1):163. Available from: http://dx.doi.org/10.18203/issn.2454-5929.ijohns20195709
26. Chang CYJ. Chronic Disorders of the Middle Ear and Mastoid (Tympanic Membrane Per forations and Cholesteatoma). In: Pediatric Otolaryngology for the Clinician [Internet]. Humana Press; 2009. p. 67–73. Available from: http://dx.doi.org/10.1007/978-1-60327-127-1_9
27. Laércio M. Cruz O, Kasse CA, Leonhart FD. Efficacy of Surgical Treatment of Chronic Otitis Media. Otolaryngology–Head and Neck Surgery [Internet]. 128(2):263–6. Available from: http://dx.doi.org/10.1067/mhn.2003.86
28. Takata G, Chan L, Morphew T, Mangione-Smith R, Morton S, Shekelle P. Evidence assessment of the accuracy of methods of diagnosing middle ea r effusion in children with otitis media with effusion. Pediatrics.
29. Pichichero ME, Poole MD. Comparison of performance by otolaryngologists, pediatricians, and general practioners on an otoendoscopic diagnostic video examination. Int J Pediatr Otorhinolaryngol. 2005 Mar 1;69(3):361–6.
30. Shaikh N, Hoberman A, Kaleida PH, Rockette HE, Kurs-Lasky M, Hoover H, et al. Otoscopic signs of otitis media. Pediatric Infectious Disease Journal. 2011;30(10):822–6.
31. Pichichero M. Acute otitis media: Part I. Improving diagnostic accuracy. Am Fam Physician.
32. Helenius K, Laine M, Tähtinen P, Lahti E, Ruohola A. Tympanometry in Discrimination of Otoscopic Diagnoses in Young Ambulat ory Children. Pediatr Infect Dis J.

33. Zon A van, Heijden GJ van der, Dongen T van, Burton M, Schilder A. Antibiotics for otitis media with effusion in children. Cochrane Database of Systematic Reviews.
34. Mittal R, Parrish JM, Soni M, Mittal J, Mathee K. Microbial otitis media: recent advancements in treatment, current chal lenges and opportunities. J Med Microbiol.
35. Grevers G. Challenges in reducing the burden of otitis media disease: An ENT pers pective on improving management and prospects for prevention. Int J Pediatr Otorhinolaryngol [Internet]. 74(6):572–7. Available from: http://dx.doi.org/10.1016/j.ijporl.2010.03.049
36. Jc W, Gr Y, Ld Q. Acute otitis media: making an accurate diagnosis. Am Fam Physician.
37. Pichichero M. Acute otitis media: Part I. Improving diagnostic accuracy. Am Fam Physician.
38. School of Health Research M. 2020 OTITIS MEDIA GUIDELINES FOR ABORIGINAL AND TORRES STRAIT ISLANDER CHILDREN Citation and links to OM app download [Internet]. 2020. Available from: https://play.google.com/store/apps/details?id=com.otitismediaguidelines.guidelines
39. Australian Medical Association. Report Card on Indigenous Health: A National Strategic Approach to Ending Chronic Otitis Media and Its Life Long Impacts in Indigenous Communities. Barton, ACT; 2017.
40. World Health Organization. Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report [Internet]. Geneva; 2017 [cited 2022 Aug 28]. Available from: https://apps.who.int/iris/bitstream/handle/10665/254659/9789241512046-eng.pdf
41. 1.15 Ear health - AIHW Indigenous HPF [Internet]. [cited 2023 Jul 7]. Available from: https://www.indigenoushpf.gov.au/measures/1-15-ear-health
42. Gunasekera H, Knox S, Morris P, Britt H, McIntyre P, Craig JC. The spectrum and management of otitis media in Australian indigenous and nonindigenous children: a national study. Pediatr Infect Dis J [Internet]. 2007 Aug [cited 2023 Jul 7];26(8):689–92. Available from: https://pubmed.ncbi.nlm.nih.gov/17848879/
43. Australian Institute of Health and Welfare. Indigenous hearing health [Internet]. Canberra, Australia; 2020. Available from: https://www.aihw.gov.au/reports/australias-health/indigenous-hearing-health
44. Kong K, Coates HLC. Natural history, definitions, risk factors and burden of otitis media. Med J Aust [Internet]. 2009 Nov 2 [cited 2023 Jul 7];191(S9). Available from: https://pubmed.ncbi.nlm.nih.gov/19883355/
45. Su JY, Guthridge S, He VY, Howard D, Leach AJ. Impact of hearing impairment on early childhood development in Australian Aboriginal children: A data linkage study. J Paediatr Child Health [Internet]. 2020 Oct 1 [cited 2023 May 28];56(10):1597–606. Available from: https://pubmed.ncbi.nlm.nih.gov/32725651/
46. Rosenfeld RM, Goldsmith A, Balzano LT. Quality of life for children with otitis media. Arch Otolaryngol Head Neck Surg [Internet]. 1997 [cited 2023 Jul 7];123(10):1049–54. Available from: https://pubmed.ncbi.nlm.nih.gov/9339979/
47. Homøe P, Heidemann CH, Damoiseaux RA, Lailach S, Lieu JEC, Phillips JS, et al. Panel 5: Impact of otitis media on quality of life and development. Int J Pediatr Otorhinolaryngol [Internet]. 2020 Mar 1 [cited 2023 Jul 7];130 Suppl 1(Suppl 1). Available from: https://pubmed.ncbi.nlm.nih.gov/31883704/

48. Kong KM, Hall ST, Palazzi K, Faulkner J, Hall B, Eisenberg R, et al. Otitis media and quality of life in NSW Aboriginal children. Australian Journal of Otolaryngology. 2022 Mar 1;5.
49. Su JY, He VY, Guthridge S, Howard D, Leach A, Silburn S. The impact of hearing impairment on Aboriginal children's school attendance in remote Northern Territory: a data linkage study. Aust N Z J Public Health. 2019;43(6):544–50.
50. Da Costa C, Eikelboom RH, Jacques A, Swanepoel DW, Whitehouse AJO, Jamieson SE, et al. Does otitis media in early childhood affect later behavioural development? Results from the Western Australian Pregnancy Cohort (Raine) Study. Clin Otolaryngol [Internet]. 2018 Aug 1 [cited 2023 Jul 7];43(4):1036–42. Available from: https://pubmed.ncbi.nlm.nih.gov/29495111/
51. Timms L, Williams C, Stokes SF, Kane R. Literacy skills of Australian Indigenous school children with and without otitis media and hearing loss. Int J Speech Lang Pathol [Internet]. 2014 [cited 2023 Jul 7];16(3):327–34. Available from: https://pubmed.ncbi.nlm.nih.gov/24460058/
52. Williams CJ, Jacobs AM. The impact of otitis media on cognitive and educational outcomes. Med J Aust [Internet]. 2009 Nov 2 [cited 2023 Jul 7];191(9):S69-. Available from: https://appserver-3691d15a-nginx-a3d6607447de4727aff360e4d7177746/journal/2009/191/9/impact-otitis-media-cognitive-and-educational-outcomes
53. He VY, Su JY, Guthridge S, Malvaso C, Howard D, Williams T, et al. Hearing and justice: The link between hearing impairment in early childhood and youth offending in Aboriginal children living in remote communities of the Northern Territory, Australia. Health Justice. 2019;7(1).
54. Leach A, Morris P. Otitis media and hearing loss among Aboriginal and Torres Strait Islander children: A research summaryInquiry into the Hearing Health and Wel. 2017.
55. Coates HL, Morris PS, Leach AJ, Couzos S. Otitis media in Aboriginal children: Tackling a major health problem. Medical Journal of Australia. 2002;177(4):177–8.
56. Closing the Gap Clearinghouse A. Ear disease in Aboriginal and Torres Strait Islander children (full publication; 7 Nov 2014 edition) (AIHW Closing the Gap Clearinghouse) [Internet]. Available from: www.aihw.gov.au/closingthegap
57. Aihw. Ear and hearing health of Aboriginal and Torres Strait Islander people 2021. 2021.
58. 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 [Internet]. 2009 [cited 2019 Feb 20];45(7–8):425–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19722295
59. Gruen RL, Weeramanthri TS, Bailie RS. Outreach and improved access to specialist services for indigenous people in remote Australia: the requirements for sustainability. J Epidemiol Community Health (1978) [Internet]. 2002 Jul 1 [cited 2023 Jul 30];56(7):517–21. Available from: https://jech.bmj.com/content/56/7/517
60. Clair MS, Murtagh DP, Kelly J, Cook J. Telehealth a game changer: closing the gap in remote Aboriginal communities. Med J Aust [Internet]. 2019 Mar 31 [cited 2023 Jul 30];210(6):S36–7. Available from: https://appserver-0c72e448-nginx-8f467156485b482cb12063814dd0b675/journal/2019/210/6/telehealth-game-changer-closing-gap-remote-aboriginal-communities

61. Mathew S, Fitts MS, Liddle Z, Bourke L, Campbell N, Murakami-Gold L, et al. Telehealth in remote Australia: a supplementary tool or an alternative model of care replacing face-to-face consultations? BMC Health Serv Res [Internet]. 2023 Dec 1 [cited 2023 Jul 9];23(1):341. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-023-09265-2
62. Jacups SP, Kinchin I. A rapid review of evidence to inform an ear, nose and throat service delivery model in remote Australia. Rural Remote Health [Internet]. 2021 [cited 2023 May 28];21(1):1–12. Available from: https://pubmed.ncbi.nlm.nih.gov/33601890/
63. D'Onofrio KL, Zeng FG. Tele-Audiology: Current State and Future Directions. Front Digit Health. 2022 Jan 10;3:788103.
64. McElroy JA, Day TM, Becevic M. The Influence of Telehealth for Better Health Across Communities. Prev Chronic Dis. 2020;17.
65. Craig J, Patterson V. Introduction to the practice of telemedicine. J Telemed Telecare [Internet]. 2005 Dec 21 [cited 2024 Jun 16];11(1):3–14. Available from: https://pubmed.ncbi.nlm.nih.gov/15829036/
66. Currell R, Urquhart C, Wainwright P, Lewis R. Telemedicine versus face to face patient care: effects on professional practice and health care outcomes. Cochrane Database Syst Rev [Internet]. 2000 [cited 2024 Jun 16];97(2):35. Available from: https://pubmed.ncbi.nlm.nih.gov/10796678/
67. Thomason C. Effect of telehealth on acute and emergency care. Nurs Times.
68. White LAE, Krousel-Wood M, Mather F. Technology meets healthcare: distance learning and telehealth. Ochsner Journal.
69. Loane M, Wootton R. A Review of Telehealth. Medical Principles and Practice [Internet]. 10(4):163–70. Available from: http://dx.doi.org/10.1159/000050363
70. Doarn CR. Development of Telemedicine and NASA's Contribution. Aerosp Med Hum Perform [Internet]. 2015 May 1 [cited 2024 Jun 16];86(5):504–5. Available from: https://pubmed.ncbi.nlm.nih.gov/25945679/
71. Strode SW. Technical and Clinical Progress in Telemedicine. JAMA [Internet]. 281(12):1066. Available from: http://dx.doi.org/10.1001/JAMA.281.12.1066
72. Demiris G. Electronic home healthcare: concepts and challenges. Int J Electron Healthc [Internet]. 1(1):4. Available from: http://dx.doi.org/10.1504/IJEH.2004.004655
73. Field MJ. Telemedicine and Remote Patient Monitoring. JAMA [Internet]. 288(4):423. Available from: http://dx.doi.org/10.1001/JAMA.288.4.423
74. Wyatt JC. Commentary: Telemedicine trials--clinical pull or technology push? BMJ [Internet]. 313(7069):1380–1. Available from: http://dx.doi.org/10.1136/bmj.313.7069.1380
75. Luo JS. Telemedicine : Is It Time Now?
76. Tanaka MJ, Oh L, Martin S, Berkson E. Telemedicine in the Era of COVID-19. Journal of Bone and Joint Surgery American volume.
77. Ftouni R, AlJardali B, Hamdanieh M, Ftouni L, Salem N. Challenges of Telemedicine during the COVID-19 pandemic: a systematic review. BMC Med Inform Decis Mak. 2022 Dec 1;22(1).
78. Sclafani AP, Shomorony A, Stewart MG, Grunstein E, Overdevest JB. Telemedicine lessons learned during the COVID-19 pandemic: The augmented outpatient otolaryngology teleconsultation. American Journal of Otolaryngology - Head and Neck Medicine and Surgery. 2021 Jul 1;42(4).

79. Chen Z, Yu X, Feng D. A Telemedicine System over Internet. Pan-Sydney Area Workshop on Visual Information Processing.
80. Kailas A, Ingram MA. Wireless communications technology in telehealth systems. International Conference on Wireless Communication, Vehicular Technolo gy, Information Theory and Aerospace & Electronic Systems Technology.
81. Klymash M, Romanchuk V, Krasko O. Telemedicine system model using Internet-technology. International Conference on Modern Problems of Radio Engineering, Tele communications and Computer Science.
82. Janca A. Telepsychiatry: an update on technology and its applications. Curr Opin Psychiatry [Internet]. 13(6):591–7. Available from: http://dx.doi.org/10.1097/00001504-200011000-00021
83. Maglogiannis I. Design and Implementation of a Calibrated Store and Forward Imaging Sy stem for Teledermatology. J Med Syst [Internet]. 28(5):455–67. Available from: http://dx.doi.org/10.1023/B:JOMS.0000041172.70027.a0
84. Jiang SW, Flynn MS, Kwock JT, Nicholas MW. Store-and-Forward Images in Teledermatology: Narrative Literature Revi ew. JMIR Dermatol [Internet]. 5(3):e37517. Available from: http://dx.doi.org/10.2196/37517
85. Zelickson BD. Teledermatology in the Nursing Home. Arch Dermatol [Internet]. 133(2):171. Available from: http://dx.doi.org/10.1001/ARCHDERM.1997.03890380041006
86. Rimner T, Blozik E, Casagrande BF, Overbeck J von. Digital skin images submitted by patients: an evaluation of feasibilit y in store-and-forward teledermatology. EJD European journal of dermatology.
87. Stevenson P, Finnane AR, Soyer HP. Teledermatology and clinical photography: safeguarding patient privacy and mitigating medico-legal risk. Medical Journal of Australia [Internet]. 204(5):198–200. Available from: http://dx.doi.org/10.5694/mja15.00996
88. Witkowski A, Łudzik J, Soyer HP. Telediagnosis with Confocal Microscopy. Dermatol Clin [Internet]. 34(4):505–12. Available from: http://dx.doi.org/10.1016/j.det.2016.05.013
89. Jarvis-Selinger S, Chan E, Payne R, Plohman K, Ho K. Clinical Telehealth Across the Disciplines: Lessons Learned. Telemedicine and e-Health [Internet]. 14(7):720–5. Available from: http://dx.doi.org/10.1089/tmj.2007.0108
90. Stahl JN, Zhang J, Zellner C, Pomerantsev E V, Chou TM, Huang HK. Teleconferencing with dynamic medical images. IEEE Transactions on Information Technology in Biomedicine [Internet]. 4(2):88–96. Available from: http://dx.doi.org/10.1109/4233.845201
91. Vaitheswaran S, Crockett P, Wilson S, Millar H. Telemental health: videoconferencing in mental health services.
92. Zhang J, Stahl JN, Huang HK, Zhou X, Lou S, Song K. Real-time teleconsultation with high-resolution and large-volume medic al images for collaborative healthcare. IEEE Transactions on Information Technology in Biomedicine.
93. Kokesh J, Ferguson AS, Patricoski C. The Alaska experience using store-and-forward telemedicine for ENT care in Alaska. Otolaryngol Clin North Am [Internet]. 2011 Dec [cited 2023 Aug 6];44(6):1359–74. Available from: https://pubmed-ncbi-nlm-nih-gov.ezproxy.library.sydney.edu.au/22032488/

94. Kokesh J, Ferguson AS, Patricoski C. Telehealth in Alaska: delivery of health care services from a specialist's perspective. Int J Circumpolar Health [Internet]. 2004 [cited 2023 Aug 6];63(4):387–400. Available from: https://pubmed-ncbi-nlm-nih-gov.ezproxy.library.sydney.edu.au/15709314/
95. Piette E, Nougairède M, Vuong V, Crickx B, Tran VT. Impact of a store-and-forward teledermatology intervention versus usual care on delay before beginning treatment: A pragmatic cluster-randomized trial in ambulatory care. http://dx.doi.org/101177/1357633X16663328 [Internet]. 2016 Aug 5 [cited 2023 Aug 6];23(8):725–32. Available from: https://journals.sagepub.com/doi/abs/10.1177/1357633X16663328
96. Isautier JMJ, Copp T, Ayre J, Cvejic E, Meyerowitz-Katz G, Batcup C, et al. People's Experiences and Satisfaction With Telehealth During the COVID-19 Pandemic in Australia: Cross-Sectional Survey Study. J Med Internet Res 2020;22(12):e24531 https://www.jmir.org/2020/12/e24531 [Internet]. 2020 Dec 10 [cited 2023 Aug 6];22(12):e24531. Available from: https://www.jmir.org/2020/12/e24531
97. Wiadji E, Mackenzie L, Reeder P, Gani JS, Carroll R, Smith S, et al. Utilization of telehealth by surgeons during the COVID 19 pandemic in Australia: lessons learnt. ANZ J Surg [Internet]. 2021 Apr 1 [cited 2023 Aug 6];91(4):507–14. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/ans.16693
98. Moffatt JJ, Eley DS, Jj M, Ds E. Barriers to the up-take of telemedicine in Australia-a view from providers. 2011 [cited 2023 Aug 6]; Available from: http://www.rrh.org.au
99. Khalil H, Cullen M, Chambers H, Steers N, Walker J. Implementation of a successful electronic wound documentation system in rural Victoria, Australia: A subject of collaboration and community engagement. Int Wound J. 2014;11(3):314–8.
100. Lasierra N, Alesanco A, Gilaberte Y, Magallón R, García J. Lessons learned after a three-year store and forward teledermatology experience using internet: Strengths and limitations. Int J Med Inform. 2012 May 1;81(5):332–43.
101. Pichichero ME. Can machine learning and AI replace otoscopy for diagnosis of otitis media? Pediatrics. 2021;147(4).
102. Montagnon E, Cerny M, Cadrin-Chênevert A, Hamilton V, Derennes T, Ilinca A, et al. Deep learning workflow in radiology: a primer. Insights Imaging. 2020;11(1).
103. Mayo RC, Leung J. Artificial intelligence and deep learning – Radiology's next frontier? Clin Imaging [Internet]. 2018;49(July 2017):87–8. Available from: https://doi.org/10.1016/j.clinimag.2017.11.007
104. Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD. Deep Learning in Radiology: Does One Size Fit All? Journal of the American College of Radiology. 2018;15(3):521–6.
105. Mathenge W, Whitestone N, Nkurikiye J, Patnaik JL, Piyasena P, Uwaliraye P, et al. Impact of Artificial Intelligence Assessment of Diabetic Retinopathy on Referral Service Uptake in a Low-Resource Setting. Ophthalmology Science. 2022 Dec;2(4):100168.
106. Seah JCY, Tang CHM, Buchlak QD, Holt XG, Wardman JB, Aimoldin A, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. Lancet Digit Health [Internet]. 2021;7500(21):1–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/34219054

107. Rahaman MM, Li C, Yao Y, Kulwa F, Rahman MA, Wang Q, et al. Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. J Xray Sci Technol. 2020;28(5):821–39.
108. Bur AM, Shew M, New J. Artificial Intelligence for the Otolaryngologist: A State of the Art Review. Otolaryngology–Head and Neck Surgery [Internet]. 2019;019459981982750. Available from: http://journals.sagepub.com/doi/10.1177/0194599819827507
109. Cömert Z, Kocamaz AF. A study of artificial neural network training algorithms for classification of cardiotocography signals. Journal of Science and Technology [Internet]. 2017 [cited 2023 Jul 9];7(2):93–103. Available from: www.dergipark.ulakbim.gov.tr/beuscitech/
110. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017;2017-Janua:2261–9.
111. Zhang J, Li C, Yin Y, Zhang J, Grzegorzek M. Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. Artif Intell Rev. 2022 Feb 1;1–58.
112. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions [Internet]. [cited 2021 Jun 6]. Available from: https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf
113. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015 Dec 10; Available from: http://arxiv.org/abs/1512.03385
114. Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. BMC Medical Imaging 2022 22:1 [Internet]. 2022 Apr 13 [cited 2023 Jul 9];22(1):1–13. Available from: https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-022-00793-7
115. Yadav SS, Jadhav SM. Deep convolutional neural network based medical image classification for disease diagnosis. J Big Data [Internet]. 2019 Dec 1 [cited 2023 Jul 9];6(1):1–18. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0276-2
116. Bellemo V, Lim G, Rim TH, Tan GSW, Cheung CY, Sadda SV, et al. Artificial Intelligence Screening for Diabetic Retinopathy: the Real-World Emerging Application. Vol. 19, Current Diabetes Reports. Current Medicine Group LLC 1; 2019.
117. 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 [Internet]. 2018;3(4):e000798. Available from: http://gh.bmj.com/lookup/doi/10.1136/bmjgh-2018-000798
118. You E, Lin V, Mijovic T, Eskander A, Crowson MG. Artificial Intelligence Applications in Otology: A State of the Art Review. Otolaryngology - Head and Neck Surgery (United States). 2020;
119. Ngombu S, Binol H, Gurcan MN, Moberly AC. Advances in Artificial Intelligence to Diagnose Otitis Media: State ofthe Art Review. Otolaryngology - Head and Neck Surgery. 2023 Jan 1;168(4):635–42.
120. Habib AR, Kajbafzadeh M, Hasan Z, Wong E, Gunasekera H, Perry C, et al. Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. Clinical Otolaryngology. 2022 May 1;47(3):401–13.

121. 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;5:1–10.
122. World Health Organization - Child and Adolescent Health and Development Prevention of Blindness and Deafness. Chronic suppurative otitis media: Burden of Illness and Management Options [Internet]. Geneva, Switzerland; 2004. Available from: https://www.who.int/pbd/publications/Chronicsuppurativeotitis_media.pdf
123. Gibney KB, Morris PS, Carapetis JR, Skull SA, Smith-Vaughan HC, Stubbs E, et al. The clinical course of acute otitis media in high-risk Australian Aboriginal children: a longitudinal study. BMC Pediatr [Internet]. 2005 Dec 14 [cited 2019 Feb 19];5(1):16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15955251
124. Aihw. Hearing health outreach services to Aboriginal and Torres Strait Islander children and young people in the Northern Territory: 2012–13 to 2014–15. Canberra; 2015.
125. Smith AC, Dowthwaite S, Agnew J, Wootton R. Concordance between real-time telemedicine assessments and face-to-face consultations in paediatric otolaryngology. Medical Journal of Australia. 2008;188(8):457–60.
126. 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. Telemedicine and e-Health [Internet]. 2010;16(9):950–6. Available from: http://www.liebertonline.com/doi/abs/10.1089/tmj.2010.0045
127. Ullah R, Gilliland D, Adams D. Otolaryngology consultations by real-time telemedicine. Ulster Medical Journal. 2002;71(1):26–9.
128. Smith AC. Telemedicine: Challenges and opportunities. Vol. 4, Expert Review of Medical Devices. 2007. p. 5–7.
129. Bennett RJ, Swanepoel D, Manchaiah V. Tele-audiology services in Australia: A shift in clinical practices. Audiology Now [Internet]. 2020;81:11–3. Available from: https://www.researchgate.net/publication/344664862
130. Smith AC, Armfield NR, Wu WI, Brown CA, Perry C. A mobile telemedicine-enabled ear screening service for Indigenous children in Queensland: Activity and outcomes in the first three years. J Telemed Telecare. 2012;18(8):485–9.
131. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. JAMA [Internet]. 2013 Nov 27 [cited 2019 Apr 22];310(20):2191–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24141714
132. Australian Institute of Health and Welfare. Hearing health outreach services for Aboriginal and Torres Strait Islander children in the Northern Territory. Cat no. IH. Canberra: AIHW; 2019.
133. Gaddey HL, Wright M, Nelson TN. Otitis Media: Rapid Evidence Review. Am Fam Physician.
134. Cruz OLM, Kasse C, Leonhart FD. Efficacy of Surgical Treatment of Chronic Otitis Media. Otolaryngology Head & Neck Surgery.
135. Couzos S, Lea T, Culbong M, Mueller R, Murray R. Management of chronic suppurative otitis media.
136. Kenna MA. Treatment of chronic suppurative otitis media. Otolaryngol Clin North Am.
137. Australian Hearing. Australian Hearing Annual Report . Chatswood; 2004.

138. Australian Hearing. Demographic details and aetiology of persons under the age of 17 years with a hearing impairment who have been fitted with a hearing aid. Chatswood; 2005.
139. Wilson D. Disability, Impairment and Quality-of-life. [Adelaide]: University of Adelaide ; 1997.
140. Wormald PJ, Browning GG, Robinson K. Is otoscopy reliable? A structured teaching method to improve otoscopic accuracy in trainees. Clin Otolaryngol Allied Sci. 1995;20(1):63–7.
141. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33(1):159–74.
142. Nash K, Macniven R, Clague L, Coates H, Fitzpatrick M, Gunasekera H, et al. Ear and hearing care programs for First Nations children: a scoping review. BMC Health Serv Res [Internet]. 2023 Dec 1 [cited 2023 Oct 16];23(1):1–12. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-023-09338-2
143. Foreman A. The Aetiology and Prevention of Otitis Media in Indigenous Children in the Northern Territory, Australia. [Darwin, Northern Territory]: University of Sydney; 1987.
144. Roberts D, Gracey M, Spargo RM. Growth and morbidity in children in a remote aboriginal community in north-west Australia. Med J Aust [Internet]. 1988 [cited 2023 Oct 17];148(2):68–71. Available from: https://pubmed.ncbi.nlm.nih.gov/3336337/
145. Australian Institute of Health and Welfare. Northern Territory Emergency Response Child Health Check Initiative: Follow- up Services for Oral and Ear Health. Canberra; 2012.
146. Beigh Z, Lattoo M, Yousuf A, Pampori R. Topical nasal steroids for hearing loss associated with otitis media w ith effusion in children. Indian Journal of Otology [Internet]. 19(3):132. Available from: http://dx.doi.org/10.4103/0971-7749.117474
147. Salmen M, Saleh E, Abd Einaeem M. The Efficacy of Nasal Steroids in Treatment of Otitis Media with Effus ion: A Comparative Study. Egyptian Journal of Neck Surgery and Otorhinolaryngology [Internet]. 0(0):0–0. Available from: http://dx.doi.org/10.21608/EJNSO.2021.158601
148. Chaffee JR, Anna L. Are nasal steroid sprays effective for otitis media with effusion.
149. Tan ACW, Emmerton L, Hattingh L, Jarvis V. Medication supply and management in rural Queensland: views of key informants in health service provision. Res Social Adm Pharm [Internet]. 2012 Sep [cited 2024 Jun 17];8(5):420–32. Available from: https://pubmed.ncbi.nlm.nih.gov/22222342/
150. D'Onofrio KL, Zeng FG. Tele-Audiology: Current State and Future Directions. Vol. 3, Frontiers in Digital Health. Frontiers Media S.A.; 2022.
151. Kuziemsky C, Maeder AJ, John O, Gogia SB, Basu A, Meher S, et al. Role of Artificial Intelligence within the Telehealth Domain. Yearb Med Inform. 2019 Aug 1;28(1):35–40.
152. Schünke LC, Mello B, da Costa CA, Antunes RS, Rigo SJ, Ramos G de O, et al. A rapid review of machine learning approaches for telemedicine in the scope of COVID-19. Vol. 129, Artificial Intelligence in Medicine. Elsevier B.V.; 2022.
153. Habib AR, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, et al. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images From Australian Aboriginal and Torres Strait Islander Children. Otology and Neurotology. 2022 Apr 1;43(4):481–8.
154. Habib AR, Perry C, Crossland G, Patel H, Kong K, Whitfield B, et al. Inter-rater agreement between 13 otolaryngologists to diagnose otitis media in aboriginal and Torres

Strait islander children using a telehealth approach. Int J Pediatr Otorhinolaryngol. 2023 Mar;111494.
155. Habib AR, Xu Y, Bock K, Mohanty S, Sederholm T, Weeks WB, et al. Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. Nature Scientific Reports. 2023;13(5368).
156. Crowson MG, Hartnick CJ, Diercks GR, Gallagher TQ, Fracchia MS, Setlur J, et al. Machine Learning for Accurate Intraoperative Pediatric Middle Ear Effusion Diagnosis. Pediatrics. 2021;147(4):e2020034546.
157. Gates GA, Cooper JC. Effect of anesthetic gases on middle ear pressure in the presence of effusion. Ann Otol Rhinol Laryngol Suppl [Internet]. 1980 [cited 2021 Aug 17];89(3 Pt 2):62–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6778352
158. McCalman J, Bainbridge R, James YC, Bailie R, Tsey K, Matthews V, et al. Systems integration to promote the mental health of Aboriginal and Torres Strait Islander children: protocol for a community-driven continuous quality improvement approach. BMC Public Health [Internet]. 2020 Dec 1 [cited 2023 Oct 5];20(1):1–12. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-09885-x
159. Butler T, Gall A, Garvey G, Ngampromwongse K, Hector D, Turnbull S, et al. A Comprehensive Review of Optimal Approaches to Co-Design in Health with First Nations Australians. International Journal of Environmental Research and Public Health 2022, Vol 19, Page 16166 [Internet]. 2022 Dec 2 [cited 2023 Oct 5];19(23):16166. Available from: https://www.mdpi.com/1660-4601/19/23/16166/htm
160. Tembo D, Hickey G, Montenegro C, Chandler D, Nelson E, Porter K, et al. Effective engagement and involvement with community stakeholders in the co-production of global health research. BMJ [Internet]. 2021 Feb 16 [cited 2023 Oct 5];372. Available from: https://www.bmj.com/content/372/bmj.n178
161. Abdalla SM, Koya SF, Jamieson M, Verma M, Haldane V, Jung AS, et al. Investing in trust and community resilience: lessons from the early months of the first digital pandemic. BMJ [Internet]. 2021 Nov 29 [cited 2023 Oct 5];375. Available from: https://www.bmj.com/content/375/BMJ-2021-067487
162. Abdollahi B, Tomita N, Hassanpour S. Data augmentation in training deep learning models for medical image analysis. Intelligent Systems Reference Library [Internet]. 2020 [cited 2023 Oct 5];186:167–80. Available from: https://link.springer.com/chapter/10.1007/978-3-030-42750-4_6
163. Bhuse P, Singh B, Raut P. Effect of Data Augmentation on the Accuracy of Convolutional Neural Networks. Lecture Notes in Networks and Systems [Internet]. 2022 [cited 2023 Oct 5];191:337–48. Available from: https://link.springer.com/chapter/10.1007/978-981-16-0739-4_33
164. He C, Liu J, Zhu Y, Du W. Data Augmentation for Deep Neural Networks Model in EEG Classification Task: A Review. Front Hum Neurosci. 2021 Dec 17;15:765525.
165. Savage N. Synthetic data could be better than real data. Nature. 2023;
166. Delaney AM, Brophy E, Ward TE. Synthesis of Realistic ECG using Generative Adversarial Networks. 2019 Sep 19 [cited 2023 Oct 5]; Available from: https://arxiv.org/abs/1909.09150v1
167. Odaibo SG, D. M, S. M, S. M. Generative Adversarial Networks Synthesize Realistic OCT Images of the Retina. 2019 Feb 18 [cited 2023 Oct 5]; Available from: https://arxiv.org/abs/1902.06676v1

168. Suresh K, Cohen MS, Hartnick CJ, Bartholomew RA, Lee DJ, Crowson MG. Generation of synthetic tympanic membrane images: Development, human validation, and clinical implications of synthetic data. PLOS Digital Health [Internet]. 2023 Feb 24 [cited 2023 Oct 5];2(2):e0000202. Available from: https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000202
169. Krawczyk B. Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence. 2016;5:221–32.
170. Pot M, Kieusseyan N, Prainsack B. Not all biases are bad: equitable and inequitable biases in machine learning and radiology. Insights Imaging [Internet]. 2021 Dec 1 [cited 2023 Oct 5];12(1):1–10. Available from: https://insightsimaging.springeropen.com/articles/10.1186/s13244-020-00955-7
171. Guo Y, Wang X, Xiao P, Xu X. An ensemble learning framework for convolutional neural network based on multiple classifiers. Soft comput [Internet]. 2020 Mar 1 [cited 2023 Oct 5];24(5):3727–35. Available from: https://link.springer.com/article/10.1007/s00500-019-04141-w
172. Yazdizadeh A, Patterson Z. Ensemble Convolutional Neural Networks for Mode Inference in Smartphone Travel Survey.
173. Zeng X, Jiang Z, Luo W, Li H, Li H, Li G, et al. Efficient and accurate identification of ear diseases using an ensemble deep learning model. Sci Rep [Internet]. 2021;11(1):1–10. Available from: https://doi.org/10.1038/s41598-021-90345-w
174. Cha D, Pae C, Seong SB, Choi JY, Park HJ. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. EBioMedicine [Internet]. 2019;45:606–14. Available from: https://doi.org/10.1016/j.ebiom.2019.06.050
175. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak [Internet]. 2020 Dec 1 [cited 2023 Oct 5];20(1):1–9. Available from: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6
176. Suara S, Jha A, Sinha P, Sekh AA. Is Grad-CAM Explainable in Medical Images? 2023 Jul 20 [cited 2023 Oct 5]; Available from: https://arxiv.org/abs/2307.10506v1

CHAPTER TWO: Literature Review

2.1 Preface to the chapter

This chapter presents a published literature review aimed at exploring the potential of AI-based computer vision algorithms to classify ear diseases from otoscopic images. The primary objectives of this research are to summarise the accuracy of AI algorithms in differentiating between normal and abnormal otoscopy images, as well as their performance in classifying multiple ear disease diagnoses. The purpose of this study is to evaluate the effectiveness of AI-based image classification models and compare their ability to human, non-expert assessors in ear disease diagnosis.

This chapter, consisting of the published paper[120], addresses aim #1 of this thesis as described in Chapter 1. Dissemination of this research and author contributions for this paper are described below.

2.2 Research dissemination

Published peer-reviewed paper:

Habib AR, Kajbafzadeh M, Hasan Z, Wong E, Gunasekera H, Perry C, Sacks R, Kumar A, Singh N. Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. Clin Otolaryngol. 2022 May;47(3):401-413. doi: 10.1111/coa.13925. Epub 2022 Mar 15. PMID: 35253378; PMCID: PMC9310803.

Journal impact factor: 2.7

Conference presentations:

Habib AR*, Kajbafzadeh M, Hasan Z, Wong E, Gunasekera H, Perry C, Sacks R, Kumar A,

Singh N. Artificial intelligence to classify ear disease from otoscopy: a systematic review and meta-analysis. Australian Society for Otolaryngology – Head and Neck Surgery Annual Conference. Adelaide, Australia. 2022. [Oral presentation].

Habib AR*, Kajbafzadeh M, Hasan Z, Wong E, Gunasekera H, Perry C, Sacks R, Kumar A, Singh N. Artificial intelligence to classify ear disease from otoscopy: a systematic review and meta-analysis. New Zealand Society for Otolaryngology – Head and Neck Surgery Annual Conference. Christchurch, New Zealand. 2022. [Oral presentation].

*Presenting author

2.3 Author attribution statement

I, Al-Rahim Habib, was responsible for designing the study, interpreting data, writing drafts of the manuscript, submitting the manuscript, responding to reviewers' comments, and coordinating submission and publication of the manuscript.

My co-authors, Ashnil Kumar and Narinder Singh helped design the study. Al-Rahim Habib and Majid Kajbafzadaeh extracted and analysed the data. Al-Rahim Habib prepared the manuscript; Al-Rahim Habib, Majid Kajbafzadaeh, Zubair Hasan, Eugene Wong, Hasantha Gunasekera, Chris Perry, Raymond Sacks, Ashnil Kumar, and Narinder Singh were involved in synthesis, revision and approval of the manuscript. All authors have read and approved the manuscript.

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statement above is correct.

Signature:

Associate Professor Narinder Singh

2.4 Paper in published format

Received: 16 November 2021 | Accepted: 27 February 2022

DOI: 10.1111/coa.13925

META-ANALYSIS

WILEY

Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis

Al-Rahim Habib[1,2,3] | Majid Kajbafzadeh[1] | Zubair Hasan[3] | Eugene Wong[3] | Hasantha Gunasekera[1,4] | Chris Perry[2,5] | Raymond Sacks[1] | Ashnil Kumar[6] | Narinder Singh[1,3]

[1]Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia

[2]Department of Otolaryngology – Head and Neck Surgery, Princess Alexandra Hospital, Woolloongabba, Queensland, Australia

[3]Department of Otolaryngology - Head and Neck Surgery, Westmead Hospital, Westmead, New South Wales, Australia

[4]The Children's Hospital at Westmead, Westmead, New South Wales, Australia

[5]University of Queensland Medical School, Brisbane, Queensland, Australia

[6]School of Biomedical Engineering, Faculty of Engineering, University of Sydney, Sydney, New South Wales, Australia

Correspondence
Al-Rahim Habib, Faculty of Medicine and Health, University of Sydney, Camperdown, NSW, Australia.
Email: al-rahim.habib@sydney.edu.au

Funding information
Research Scholarship from the Garnett Passe and Rodney Williams Memorial Foundation. Avant Foundation Doctor-in-Training Research Grant

Abstract

Objectives: To summarise the accuracy of artificial intelligence (AI) computer vision algorithms to classify ear disease from otoscopy.

Design: Systematic review and meta-analysis.

Methods: Using the PRISMA guidelines, nine online databases were searched for articles that used AI computer vision algorithms developed from various methods (convolutional neural networks, artificial neural networks, support vector machines, decision trees and k-nearest neighbours) to classify otoscopic images. Diagnostic classes of interest: normal tympanic membrane, acute otitis media (AOM), otitis media with effusion (OME), chronic otitis media (COM) with or without perforation, cholesteatoma and canal obstruction.

Main outcome measures: Accuracy to correctly classify otoscopic images compared to otolaryngologists (ground truth). The Quality Assessment of Diagnostic Accuracy Studies Version 2 tool was used to assess the quality of methodology and risk of bias.

Results: Thirty-nine articles were included. Algorithms achieved 90.7% (95%CI: 90.1–91.3%) accuracy to difference between normal or abnormal otoscopy images in 14 studies. The most common multiclassification algorithm (3 or more diagnostic classes) achieved 97.6% (95%CI: 97.3–97.9%) accuracy to differentiate between normal, AOM and OME in three studies. AI algorithms outperformed human assessors to classify otoscopy images achieving 93.4% (95%CI: 90.5–96.4%) versus 73.2% (95%CI: 67.9–78.5%) accuracy in three studies. Convolutional neural networks achieved the highest accuracy compared to other classification methods.

Conclusion: AI can classify ear disease from otoscopy. A concerted effort is required to establish a comprehensive and reliable otoscopy database for algorithm training. An AI-supported otoscopy system may assist health care workers, trainees and primary care practitioners with less otology experience identify ear disease.

KEYWORDS
artificial intelligence, computer vision, diagnosis, machine learning, otoscopy

Meeting information: Verbal presentation at the 2021 Australian Society of Otolaryngology – Head and Neck Surgery Annual Scientific Meeting on Friday, 17 September 2021, in Melbourne, Victoria, Australia and 2021 New Zealand Society of Otolaryngology - Head & Neck Surgery Annual Scientific Meeting virtual conference on Friday, 25 February 2022.

1 | INTRODUCTION

Otoscopy is a routine component of the ear assessment that visualises the external auditory canal and tympanic membrane (TM). It is used to identify common conditions, including infection (otitis externa, otitis media—acute and chronic), otitis media with effusion (OME), perforation, cholesteatoma, tympanosclerosis, foreign body, tympanostomy tube presence/position and cerumen impaction. Ear examinations are often conducted in primary care settings by local community health workers, nurses, medical students, general practitioners and emergency physicians. Concerning findings typically then lead to intervention or to specialist referral, where an otolaryngologist will examine the ear.

Otoscopy accuracy in primary care settings varies based on user training and experience. However, the reported literature describes diagnostic accuracy estimates between 30% and 67.5% when compared to otolaryngologists (ground truth) for specific diagnoses.[1] Previous efforts to improve performance have focussed on educational techniques, including online teaching modules comparing normal and abnormal otoscopic TM images, practical tutorials conducted by otolaryngologists and simulation with artificial ear models.[2,3] Comparisons made before and after education sessions demonstrate short-term improvements in otoscopy performance, although benefits are not sustained long-term and can decrease from initial assessments.[3] The frequency of otoscopy in primary care settings and the observed performance inconsistencies may provide an opportunity for artificial intelligence (AI) to assist in the accurate identification of ear disease.

AI can replicate the ability of human cognition to recognise patterns, identify anomalies and construct rational solutions to potential obstacles.[4] Popularised applications of AI in health care include the use of computer vision to differentiate benign versus malignant skin lesions, identify diabetic retinopathy from fundoscopic images, assist radiologists to interpret chest x-rays and predict infectious disease outbreaks, as in the case of the novel coronavirus (COVID-19) pandemic.[5–8]

AI-based computer vision algorithms are an emerging technology that can be used to classify ear disease using otoscopic images.[9] The use of AI-based computer vision algorithms as an adjunct to otoscopy performed in primary care settings may be most relevant in rural and remote areas where access to otolaryngologists is limited. In these scenarios, ear examinations are often performed by nurses and community health workers with less clinical experience than otolaryngologists in accurately recognising ear disease. In rural and remote areas, telemedicine initiatives, such as tele-otoscopy, are feasible strategies to capture otoscopic images for subsequent metropolitan specialist review but are often disadvantaged by delays in clinical decision making and implementation of interventions.[10]

The aim of this review was to evaluate the performance of AI-based computer vision algorithms to classify ear disease from otoscopy. Our objectives were to synthesise existing literature related to the use of AI-based computer vision algorithms for otoscopy, assess the performance of existing models and propose a guide for future algorithm development.

Key Points

1. AI-based computer vision algorithms can differentiate between binary (2 diagnosis options) and multiple ear disease diagnoses (3 or more diagnosis options).
2. AI-based computer vision algorithms have been shown to classify otoscopy images more accurately than human, nonexpert assessors.
3. AI-based computer vision algorithms have been developed using various machine learning techniques of which, models using CNNs achieve the greatest classification accuracy.
4. Substantial heterogeneity was found between studies reflecting, in part, diverse sources of images and collection practices, machine learning methods, diagnostic classes and ground-truth definitions.
5. Future efforts are needed to establish a standardised, comprehensive and validated database to develop clinically relevant AI-based computer vision algorithms for otoscopy.

2 | METHODS

The present systematic review was conducted in accordance with the PRISMA guidelines[11] and registered with the International prospective register of systematic reviews (PROSPERO) on 18 February 2021 (ID number: CRD42021202594). Ethics approval and patient consent were not required for this review.

2.1 | Literature search

A systematic search of online databases (including Google Scholar, MEDLINE, Embase, PubMed, Scopus, ProQuest, ACP Journal Club, Health Technology Assessment and the Cochrane Library) for articles, abstracts or conference proceedings published in the past 10 years through 31 October 2021 that used AI-based computer vision approaches to classify otoscopic TM images was conducted. Searches were limited to those involving human subjects and those published in the English Language.

Medical subject headings (MeSH) terms and non-MeSH terms related to AI approaches included: ‘artificial intelligence’, ‘machine learning’, ‘deep learning’, ‘convolutional neural networks’, ‘support vector machines’, ‘image recognition’, ‘image classification’, ‘object detection’ and ‘computer-assisted diagnosis’. MeSH terms and keywords related to otoscopic TM images included: ‘otoscopy’, ‘ear’, ‘eardrum’, ‘tympanic membrane’, ‘ear disease’, ‘acute otitis media’,

'otitis media', 'chronic otitis media' and 'chronic suppurative otitis media'.

2.2 | Selection criteria

Titles and abstracts were reviewed for eligibility by two independent investigators (ARH and MK). Discrepancies between the two investigators were resolved by the senior author (NS), a board-certified otolaryngologist. Reference lists of available full-text articles were also manually screened for further studies eligible for inclusion in this review. Diagnostic observational studies describing the development of an autonomous, supervised algorithm to classify otoscopic TM images using the AI approaches described above were included. Articles that were excluded consisted of those that did not use AI approaches of interest, utilised imaging modalities other than otoscopy or were review articles or editorials. Study inclusion/exclusion is summarised in a PRISMA flow diagram (Figure 1).

2.3 | Data extraction

Two investigators (ARH and MK) independently extracted data from included studies for analysis. The following characteristics were extracted from included studies: primary author, year, study objective, AI technique to achieve study objective, data source, otoscope type and manufacturer, image labelling method, source for ground-truth classification, diagnostic categories of interest, image size and quality, number and distribution of training images, number and distribution of test images, number and distribution of validation images, ratio of training/test/validation images by diagnostic categories and performance characteristics. Additional characteristics extracted included the type of deep learning models, batch size, learning rate, method used to standardise input images and use of image augmentation methods, segmentation, hyperparameter tuning and cross-validation techniques.

2.4 | Critical appraisal and risk of bias assessment

The quality of included studies was assessed using the Quality Assessment and Diagnostic Accuracy Studies-2 (QUADAS-2) tool, as per the Cochrane Collaboration for critical appraisal of diagnostic test accuracy evaluations.[12] The QUADAS-2 assessment tool is composed of patient selection, index test, reference standard and flow and timing. Study quality was assessed by two independent investigators (ARH and MK). Uncertainties or discrepancies were discussed with the senior author (NS) to achieve consensus. As described by Whiting et al.,[12] applicability concerns were determined for patient selection, the index test and the reference standard. For patient selection, reviewers considered whether the subjects recruited for the study and used to train the algorithm were applicable to the target population where algorithms would be implemented. For the index test, reviewers considered whether the classification categories of the algorithm were applicable to its

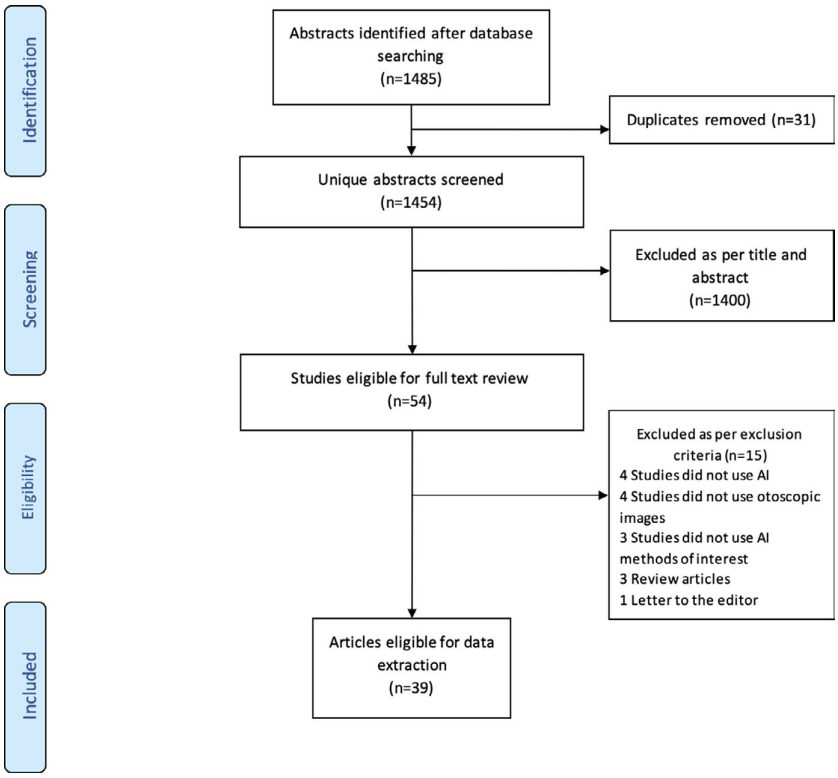


FIGURE 1 PRISMA study flow diagram [Colour figure can be viewed at wileyonlinelibrary.com]

intended use. For the reference standard, reviewers considered whether the method used reflected clinical practice for future applications.

2.5 | Outcomes

A systematic review and meta-analysis were performed with the primary outcome assessed being algorithm diagnostic accuracy in classifying ear disease from otoscopic images. Secondary outcomes included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score and area under the curve (AUC).

2.6 | Statistical analysis

The statistical analysis was performed using ReviewManager (RevMan 5.3, Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014) in accordance with the Cochrane Handbook[13] and Stata 17 (StataCorp LLC, 2021). Heterogeneity of model performance was summarised with the I^2 statistic, H^2 statistic, Cochrane's Q statistic and chi-square test. Heterogeneity values >75% were considered as substantial heterogeneity. Inverse variance weighting was used for pooling model performance, and fixed effects were applied.

3 | RESULTS

3.1 | Study selection

The search strategy yielded a total of 1485 relevant articles and abstracts. Following full-text review, 39 articles met the inclusion criteria (Figure 1).

3.2 | Characteristics of included studies

Table 1 provides a summary of characteristics for included studies. Thirty studies collected images from outpatient and inpatient primary care settings, six studies combined images from primary care settings and online sources (Google Images), one study collected images from Google Images alone and two studies did not report the source of images.

3.3 | Risk of bias of included studies

Figure 2 and Figure S1 illustrate the risk of bias and applicability concerns of included studies using the QUADAS-2 tool. High risk of bias (37 of 39 studies) was identified in patient selection criteria due to failure to utilise consecutive or random sampling. High risk of bias (15 of 39 studies) was also observed in use of the reference standard (ground truth), as these articles did not utilise more than 1 otolaryngologist to review otoscopic images to confirm class labels.

3.4 | Binary classification algorithms

Nine unique binary classification algorithms to classify otoscopy images were reported (Table S1).

3.4.1 | Normal versus abnormal

AI algorithms achieved a pooled accuracy of 90.7% (95%CI: 90.1–91.3%) to difference between normal or abnormal otoscopy images with substantial heterogeneity between studies (n = 14 studies, I^2 = 96.9%, p = .001, Figure 3).

Four studies[14–17] used the Özel Van Akdamar Hospital otoscopic image database to train binary algorithms using various classification techniques. Simon et al.[14] demonstrated that pretrained convolutional neural networks (CNNs) and support vector machines (SVMs) could achieve greater classification accuracy than k-nearest neighbours (k-NNs), artificial neural networks (ANNs), decision trees (DTs) or the Naïve Bayes technique. Basaran et al.[15] utilised multiple pretrained CNNs to evaluate the effect of segmentation, distribution between training and test data and cross-validation on algorithm performance. In this study, pretrained CNNs achieved enhanced accuracy by applying basic image augmentation techniques and using region of interest patches, rather than full otoscopic images.[15] The pretrained CNNs VGGNet-16 and VGGNet-19 achieved the greatest classification accuracy to differentiate normal from abnormal otoscopic images (90.5% and 90.1% respectively).[15] Mironica et al.[18] demonstrated that the greatest classification accuracy was achieved by CNNs and SVMs in 186 images collected from outpatient primary care settings. Enhanced performance was achieved by adding a colour coherence vector (CCV) to the algorithms (models with CCV vs without: CNN – 73.1% vs. 68.8%, SVM – 72.0% vs. 64.5%).[18]

Using intraoperative assessment of children taken to the operating room with the intent of myringotomy to determine the ground truth, Crowson et al.[19] differentiated between normal and OME with 83.8% accuracy using ResNet-34 and Monte Carlo cross-validation resampling with 5 repetitions.

3.5 | Multiclassification algorithms

Seventeen unique multiclassification algorithms to classify otoscopy images were identified (Table S2). Overall, multiclassification algorithms achieved a pooled accuracy of 96.2% (95%CI: 96.1–96.4%) with substantial heterogeneity between studies (n = 18 studies, I^2 = 98.8%, p = .001, Figure 4).

TABLE 1 Summary of included studies

Author/year	Image source	Ground truth	Classes	Total images	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F1 score	AUC
Wang 2021[14]	H	2 otolaryngologists	2	100	81.7	83.3	80.0	80.6	NR	NR	0.88
Sundgaard 2021[15]	H	1 otolaryngologist	3	1336	86.0	NR	NR	NR	NR	NR	NR
Tsutsumi 2021[16]	H[1], O	O: NR, H[1]: 2 otolaryngologists, 1 paediatrician	2	400	77.0	70.0	84.4	81.4	73.7	NR	0.90
Tsutsumi 2021[16]	H[1], O	O: NR, H[1]: 2 otolaryngologists, 1 paediatrician	5	400	66.0	55.4	NR	79.8	NR	NR	0.88
Byun 2021[17]	H	3 otolaryngologists	4	2372	97.2	NR	NR	NR	NR	NR	NR
Zeng 2021[18]	H	6 otologists	8	20542	95.5	NR	NR	NR	NR	NR	0.99
Crowson 2021[19]	H	1 of 5 paediatric otolaryngologists	2	338	83.8	NR	NR	NR	NR	80.0	0.93
Alhudhaif 2021[20]	H[1]	2 otolaryngologists, 1 paediatrician	4	857	98.2	97.7	99.3	NR	NR	96.9	NR
Cai 2021[21]	H	otolaryngologists (number not reported).	4	6066	93.4	NR	NR	NR	NR	96.8	0.98
Camalan 2021[22]	H[1]	1 otologist, 1 paediatric otolaryngologist	3	300	85.8	NR	NR	NR	NR	NR	NR
Ucar 2021[23]	H[1]	1 otolaryngologist	4	880	98.1	98.1	99.4	98.2	NR	NR	0.99
Camalan 2020[24]	H[1]	NR	3	454	88.1	NR	NR	NR	NR	NR	NR
Wu 2020[25]	H	2 otologists	3	12203	97.8	96.8	98.0	96.9	98.4	NR	0.99
Basaran 2020[26]	H[1]	2 otolaryngologists, 1 paediatrician	2	282	90.4	86.8	93.5	NR	NR	87.3	0.95
Goshtasbi 2020[27]	H[1], O	O: NR, H[1]: 2 otolaryngologists, 1 paediatrician	2	400	77.0	70.0	84.0	81.0	74.0	NR	0.89
Goshtasbi 2020[27]	H[1], O	O: NR, H[1]: 2 otolaryngologists, 1 paediatrician	5	400	71.0	NR	NR	NR	NR	NR	0.91
Habib 2020[28]	O	2 otolaryngologists	2	233	76.0	76.0	76.0	76.0	76.0	NR	0.87
Khan 2020[29]	H	2 otolaryngologists	3	2484	87.0	95.0	NR	95.2	NR	95.1	0.99
Simon 2020[30]	H[1]	2 otolaryngologists, 1 paediatrician	2	956	81.4	83.6	83.8	NR	NR	NR	0.89
Viscaino 2020[31]	H[1]	1 otolaryngologist	4	720	88.1	87.8	95.9	87.7	NR	NR	1.00
Cömert 2020[32]	H[1]	2 otolaryngologists, 1 paediatrician	4	857	99.4	99.4	99.8	NR	NR	99.3	NR
Basaran 2019[33]	H[1]	2 otolaryngologists, 1 paediatrician	2	598	97.9	99.1	98.5	NR	NR	NR	NR
Basaran 2019[34]	H[1]	2 otolaryngologists, 1 paediatrician	2	223	76.1	70.8	80.1	NR	NR	NR	0.81
Cha 2019[35]	H	1 otologist, 1 physician	6	10544	94.2	93.7	96.8	NR	NR	NR	NR
Lee 2019[36]	H	2 otologists	2	1338	91.0	90.5	92.9	98.0	72.3	NR	0.92
Livingstone 2019[37]	H, O	1 otologist, 1 otolaryngology resident	14	1366	88.7	86.1	NR	90.9	NR	NR	NR

(Continues)

TABLE 1 (Continued)

Author/year	Image source	Ground truth	Classes	Total images	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F1 score	AUC
Livingstone 2019[38]	H	2 otolaryngologists	3	529	84.4	NR	NR	NR	NR	NR	NR
Seok 2019[39]	H	2 otolaryngologists	2	920	92.9	NR	NR	92.9	NR	NR	NR
Huang 2018[40]	NR	NR	3	20	70.0	NR	NR	NR	NR	NR	NR
Kasher 2018[41]	H, O	NR	2	108	82.1	NR	NR	NR	NR	NR	NR
Myburgh 2018[42]	NR	2 otologists	5	389	86.2	86.8	96.4	87.4	96.4	NR	NR
Senaras 2018[43]	H, O	NR	2	1082	84.3	86.8	81.9	NR	NR	NR	NR
Tran 2018[44]	H	NR	2	214	91.4	89.5	93.3	91.9	NR	NR	0.92
Senaras 2017[45]	H	otolaryngologists (number not reported)	2	247	84.6	87.3	81.4	NR	NR	NR	NR
Myburgh 2016[46]	H	2 otologists	5	486	80.6	80.6	94.4	81.0	94.6	NR	NR
Wang 2015[47]	H, O	NR	2	215	90.0	85.0	92.0	NR	NR	NR	NR
Shie 2014[48]	H	otolaryngologists (number not reported)	4	865	88.0	91.6	79.9	NR	NR	NR	0.91
Kuruvilla 2013[49]	H	3 otologists	3	181	89.9	NR	NR	NR	NR	NR	NR
Kuruvilla 2012[50]	H	3 otologists	3	135	84.0	NR	NR	NR	NR	NR	NR
Mironica 2011[51]	H	1 otolaryngologist	2	186	73.1	NR	NR	NR	NR	NR	NR
Vertan 2011[52]	H	1 otolaryngologist	3	100	59.9	NR	NR	NR	NR	NR	NR

Note: Classes refers to the number of diagnostic categories included in the algorithm; (1) online data set.

Abbreviations: AUC—area under the curve; H—primary care hospital database; NPV—negative predictive value; NR—not reported; O—online repository; PPV—positive predictive value.

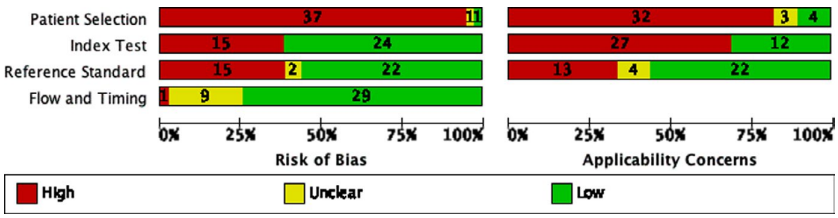


FIGURE 2 Summary of risk of bias and applicability of concerns graph using the Quality Assessment of Diagnostic Accuracy Studies tool, version 2 (QUADAS-2) [Colour figure can be viewed at wileyonlinelibrary.com]

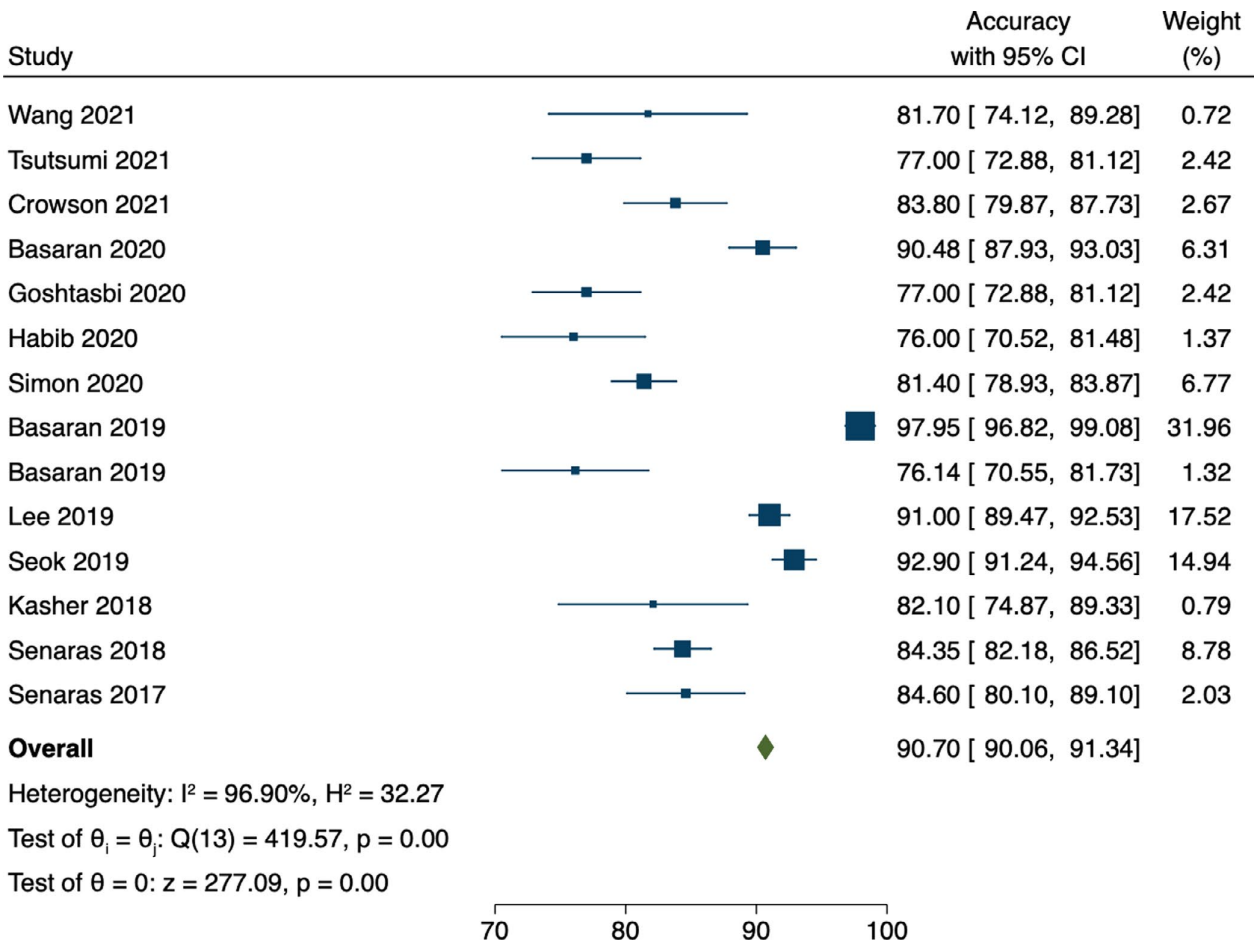


FIGURE 3 Forest plot comparing accuracy of AI algorithms to classify normal versus abnormal otoscopy images [Colour figure can be viewed at wileyonlinelibrary.com]

3.5.1 | Normal, AOM and OME

The most common multiclassification model differentiated between normal, AOM and OME in five studies[20–24] Overall, the algorithms achieved an accuracy of 97.6% (95%CI: 97.3–97.9%) with substantial heterogeneity between studies ($n = 3$ studies, $I^2 = 98.7\%$, $p = .001$, Figure S2).[20,21,24] Two studies were excluded from the meta-analysis because the number of test images was not reported.[22,23]

Wu et al.[20] used 12 203 otoscopic images and applied the Xception and MobileNets-V2 pretrained CNNs for classification. Image augmentation was used during the preprocessing phase consisting of rotation, width shift, height shift, shearing, zooming and horizontal flip. Hyperparametric tuning was used to establish the optimal classification algorithm (epochs: 100, initial learning rate: 0.001, batch size: 14 [Xception] and 64 [MobileNets-V2]). Between diagnostic categories, OME was classified less accurately (96.7%) than normal (98.3%) or AOM (98.5%).

3.6 | AI versus human classification

Five studies[25–29] compared the performance of image classification algorithms to human assessors. Of these, three studies compared the performance of the AI algorithm to human assessors using the same test set.[25,26,28] Restricting the meta-analysis to these three studies demonstrated that AI algorithms outperformed human assessors

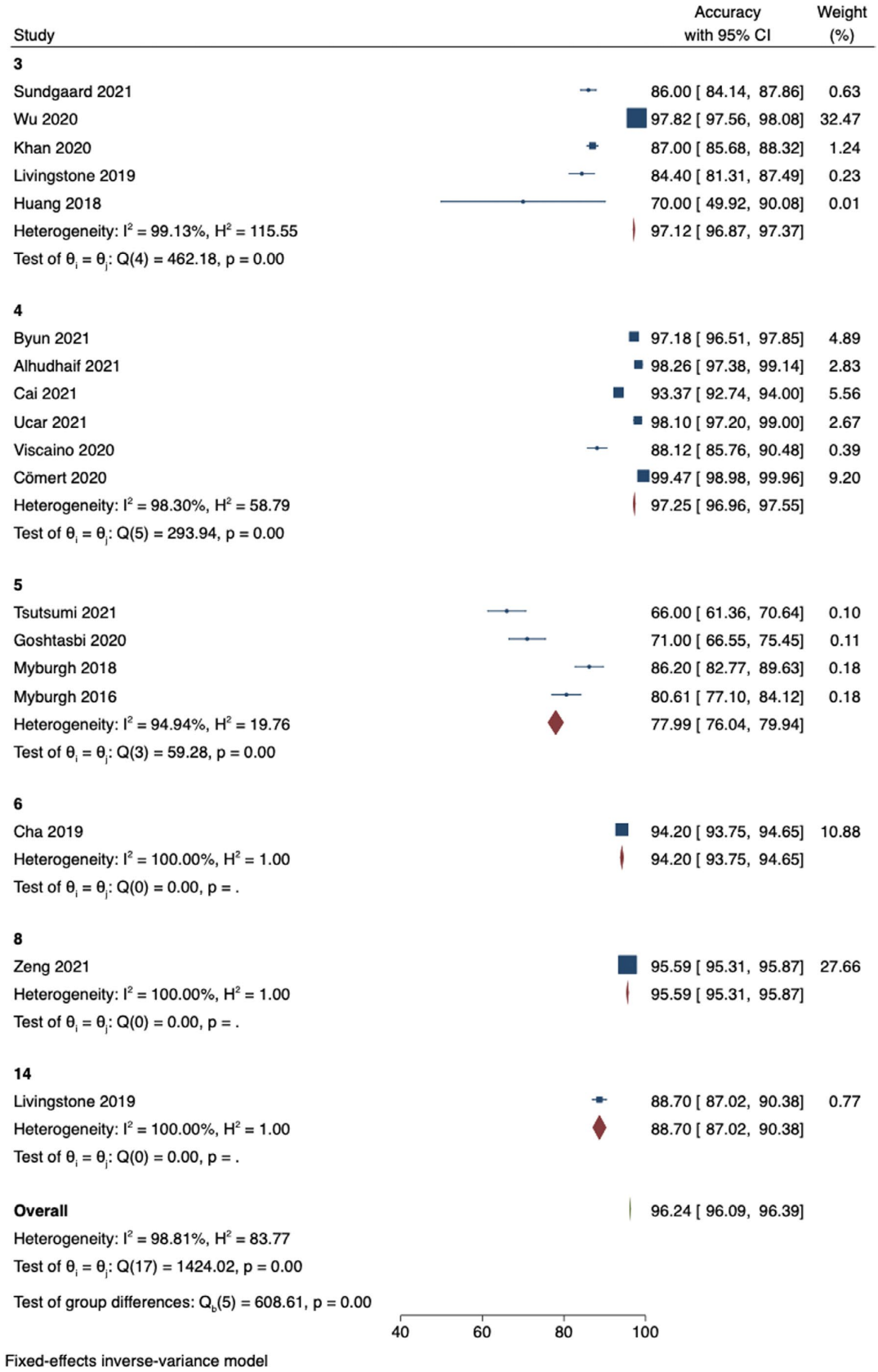


FIGURE 4 Forest plot comparing accuracy of multiclassification AI algorithms to classify ear disease from otoscopy, stratified by number diagnostic classes [Colour figure can be viewed at wileyonlinelibrary.com]

having achieved an accuracy of 93.4% (95%CI: 90.5–96.4%) versus 73.2% (95%CI: 67.9–78.5%) with substantial heterogeneity between studies (AI: $n = 3$ studies, $I^2 = 78.6\%$; human assessors: $n = 3$ studies, $I^2 = 83.7\%$, $p = .001$, Figure 5, Table S3).

Byun et al.[28] found that a 4-class image classification algorithm using ResNet-18 and a shuffle attention model outperformed 10 nonexperts (first- and second-year resident physicians) (97.1% vs. 82.9%) to classify normal, OME, COM or cholesteatoma. Otolaryngology experience of the resident physicians was not reported.

Livingstone et al.[25] recruited 10 nonexperts (general practitioners and trainees from paediatrics, emergency medicine and otolaryngology) to review a test set of 89 images for 14 diagnostic classes. Human, nonexpert assessors were summarised together, and stratification by specialty was not provided. Overall, the algorithm outperformed all the human assessors in 12 out of 14 categories (algorithm: 88.7% vs. human assessors: 58.9%). The algorithm classified cholesteatoma and otomycosis less accurately than human assessors (cholesteatoma – 50.0% vs. 55.0%, otomycosis – 0.0% vs. 40.0% respectively).

Khan et al.[26] recruited 17 human expert and nonexpert assessors (7 specialist otolaryngologists and 10 nonexperts) to review 100 test images for three diagnostic classes (normal, OME, COM). Overall, the algorithm achieved a classification accuracy of 87.0% compared to 74.0% for the pooled human expert and nonexpert assessors. Stratification of accuracy results by diagnostic class or assessor seniority was not provided.

4 | DISCUSSION

Performance in diagnosing ear disease from otoscopy varies by user training and experience.[30] This study summarises the performance of AI-based computer vision algorithms to diagnosis of ear disease from otoscopy. Thirty-nine studies were included in this review, evaluating the performance of 9 binary and 17 multiclassification algorithms. AI-based computer vision algorithms achieved 90.7% (14 studies) accuracy to differentiate between normal or abnormal otoscopy images and 97.6% (3 studies) to differentiate between normal, AOM or OME. Compared to manual classification, AI-based computer vision algorithms outperformed human assessors to classify otoscopy images (93.4% vs. 73.2% accuracy, respectively) in three studies. Substantial heterogeneity in performance was identified between studies.

AI-based computer vision algorithms with CNNs achieved greater accuracy in binary and multiclassification categories. Harnessing transfer learning by freezing the final layers of a CNN's architecture is advantageous by using established computer vision performance to evaluate new tasks of interest. CNNs may have the potential to

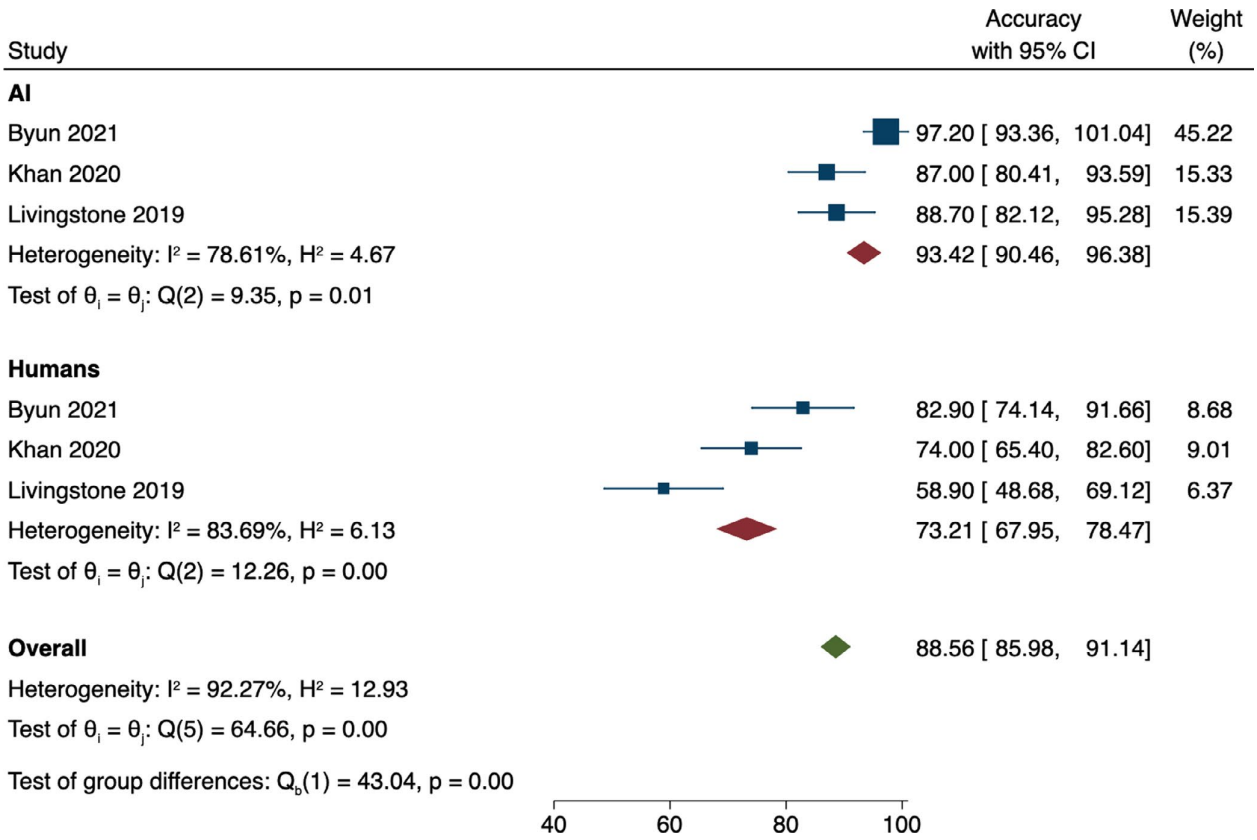


FIGURE 5 Forest plot comparing ear disease classification accuracy from otoscopy between AI algorithms and human assessors [Colour figure can be viewed at wileyonlinelibrary.com]

achieve greater accuracy than ANNs, SVMs or DTs due in part to their architecture, weighting system and features extracted in hidden layers.[20,31] However, CNNs vary in relation to the amount of computational power, speed and size of models produced. For example, Inception-V3 yielded greater accuracy than MobileNet-V2 in multiclassification but produced a larger size model.[32] This is relevant whether considering deployment of the model into an edge device (e.g. otoscope or smartphone) versus workstations with graphics processing units (GPUs) or cloud-based platforms.[32] Focus on the TM and localising anatomical or pathological characteristics with segmentation models has achieved substantial improvements in accuracy while considering various middle ear conditions.[33]

This review has several strengths. Firstly, broad inclusion criteria were used to explore articles clinically relevant to the study question. This identified algorithms with various combinations of diagnostic classes and classification approaches. Secondly, the breadth of diagnostic classes is important for clinicians to appreciate the strengths and weaknesses of computer vision algorithms and highlight potential for future exploration. Thirdly, the QUADAS-2 assessment tool was used to identify risk of biases and applicability concerns. Before image classification algorithms are introduced into routine clinical practice, rigorous efforts are required to evaluate performance and reliability.[34]

This systematic review has limitations. Firstly, this review is limited by the variability between eligible studies in terms of data sources, image capture devices and machine learning methods. Studies included used various patient selection criteria, classification categories and ground-truth definitions. As a result, this could plausibly introduce selection and measurement bias in the outcomes reported and overall conclusions drawn from the results. The heterogeneity between classification categories suggests a lack of standardisation or quality control in terms of image acquisition. It is important to acknowledge the clinical need for an AI-based computer vision algorithm and apply this technology accordingly.

AI-based computer vision algorithms depend on high-quality training data with accurate and reliable ground-truth labels. Ideally ground-truth labels should be determined by consensus of multiple independent experts and/ or by additional evidence (e.g. clinical history, histopathology or independent investigation results).[35,36] For otoscopy, ground truth may be based on expert review of images along with clinical history, tympanometry, audiometry or myringotomy results. In this review, most studies used single otolaryngologists to classify otoscopic images during routine outpatient examinations. High risk of bias was identified in patient and image selection as most sampling methods for training images were not randomised or did not use consecutive recruitment. Furthermore, ground-truth assessment was typically not validated by multiple independent experts or via other means. Crowson et al.[19] uniquely used intraoperative otoscopic images of children undergoing myringotomy for recurrent AOM or OME. The authors suggested that this approach may represent the gold standard of detecting OME. Despite this, the authors did not report whether myringotomy was performed in all children, even those with low pretest probability of disease. Furthermore, positive pressure ventilation may displace middle ear fluid immediately prior to myringotomy in up to 15% of children with OME.[37] While myringotomy results may add useful additional evidence, this approach significantly reduces the number of test images available. Only those patients proceeding to intervention can be included, effectively excluding all patients with suspected normal ears (that do not proceed to intervention) from the training dataset. Accordingly, myringotomy results cannot be used as the basis for ground-truth labelling in large training datasets.

To create large datasets for training AI-based computer vision algorithms, we recommend that ground-truth labelling for otoscopic images be based primarily on the consensus of multiple independent experts along with clinical history, and tympanometry and audiometry results. This may reflect real-world practices, where clinical suspicion from experts (e.g. otolaryngologists with subspeciality interest in otology) is used to determine which patient is appropriate for invasive interventions. Tele-otoscopy databases may be a suitable source for this research as images have been collected by clinical staff with experience in performing otoscopy and have undergone quality control, vetting by an otolaryngologist, and often review in conjunction with clinical history, tympanometry and audiometry to establish a diagnosis and treatment plan.

As identified in this review, the optimal approach to develop an AI-based computer vision algorithm can use data augmentation, CNNs (DenseNet, Xception or Inception ResNet-V2), ensemble models combining multiple classifying features, sequential multistep segmentation to localise the TM, hyperparameter tuning and cross-validation resampling to yield the greatest performance. It is an important methodological consideration to partition data into training, validation and test groups to minimise the risk of overfitting, limit training bias and consider the generalisability of model predictions to independent or heterogenous data.[38] Training a model and then testing, it on the same data would overestimate prediction performance and misrepresent its function on unseen data. However, partitioning data into discrete groups for training, validation and testing can reduce the number of samples used for model learning. To address this, cross-validation techniques can be applied to divide the training group into sets (i.e. folds) and hold-out sets to validate the model by averaging performance on a predetermined number of loops.[39,40] Most studies in this review applied cross-validation techniques. Basaran et al. (2020) demonstrated that 10-fold cross-validation yielded higher accuracy, sensitivity and specificity results than utilising 50% of the data source for training and reserving 50% for testing.[15] The collection of otoscopic images requires time, equipment, adequate storage, staff training and consideration of privacy, patient confidentiality and security. In this scenario, efforts to include a greater number of images for training and limiting bias may be important to maximise generalisability and clinical value. Attention models can be applied to visualise important areas used to establish predictions. Future algorithms may incorporate specific segmentation models using the U-Net architecture. For example, Pham et al. (2021) proposed EAR-UNet, an automatic segmentation model for TMs from video-otoscopic images integrating pretrained

CNNs (EfficientNet-B4 and ResNet), achieving 96% accuracy to localise the TM in normal, AOM, OME and COM otoscopic images.[33]

Transitioning the AI-based computer vision algorithm for otoscopy from a virtual environment to the clinical frontline will depend on real-world test performance and applicability in daily clinical practice. This technology has the potential to inform judgement, improve triage and save time and resources for batch screening. Despite this, performance of the algorithm depends on training data. Bias in patient selection, ground-truth labelling and diagnostic classes may impact generalisability. Implementation in real-world settings may depend on desirably, feasibility and viability of this technology as an adjunct to existing clinical practices. Further efforts to progress the application of AI for otoscopy may be directed at establishing a comprehensive, publicly available, open access otoscopy database with associated symptoms, tympanometry, pneumatic otoscopy and audiometry findings, validated by multiple otologists. Future studies may build on previous studies evaluating the concordance between expert and nonexpert assessments,[40] by exploring the use of AI as an adjunct to clinical decision making. Initial efforts for real-world applications may be targeted at identifying poor quality otoscopic images that limit accurate assessment (e.g. blurriness, over saturation, lack of white balance, moisture, TM not visualised and wax obstruction). Delineating the strengths and limitations of AI to autonomously classify otoscopic images is necessary for real-world applications.

5 | CONCLUSION

In this review, 39 articles explore the role of AI-based computer vision algorithms for otoscopy. AI algorithms achieved 90.7% accuracy to differentiate between normal or abnormal otoscopy images and 97.6% to differentiate between normal, AOM or OME. Compared to manual classification, AI algorithms outperformed human assessors to classify otoscopy images (93.4% versus 73.2% accuracy respectively). However, substantial heterogeneity in performance was identified between studies. A concerted effort is warranted to establish a standardised, robust, comprehensive and reliable database to develop clinically relevant computer vision algorithms for otoscopy. An AI-based computer vision algorithm for otoscopy has the potential to support health care workers and primary care practitioners with less otology experience to identify and manage ear conditions early to minimise the risk of sequelae from untreated disease.

ACKNOWLEDGEMENTS

We acknowledge support from the Garnett Passe and Rodney Williams Memorial Foundation Research Scholarship and the Avant Foundation Doctor-in-Training Research Grant. The authors would like to acknowledge Dr Mukesh Prasad for proofreading the manuscript.

CONFLICT OF INTEREST

Dr Al-Rahim Habib receives grant funding from the Garnett Passe and Rodney Williams Memorial Foundation, Microsoft AI for Humanitarian Action Grant and the Avant Foundation Doctor-in-Training Research Grant. Professor Raymond Sacks is a consultant for Medtronic. Associate Professor Narinder Singh receives grant funding from the Garnett Passe and Rodney Williams Memorial Foundation and Microsoft AI for Humanitarian Action Grant. He is a consultant for ENT Technologies, Optinose, Nasus Medical and ResMed, and a principal investigator for drug/ device trials for GSK, Lyra and Covance.

ETHICS STATEMENT

Ethics approval and patient consent were not in need for this study.

AUTHOR CONTRIBUTION

ARH, AK and NS designed the study; ARH and MK extracted and analysed the data; ARH prepared the manuscript; ARH, MK, ZH, EW, HG, CP, RS, AK, and NS involved in synthesis, revision and approval of manuscript. All authors agreed to be accountable for all aspects of the work.

[Correction added on March 21, 2022, after first online publication: Peer review history is not available for this article, so the peer review history statement has been removed.]

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Al-Rahim Habib https://orcid.org/0000-0001-6327-4648
Eugene Wong https://orcid.org/0000-0002-5799-5083

REFERENCES

1. Buchanan CM, Pothier DD. Recognition of paediatric otopathology by General Practitioners. *Int J Pediatr Otorhinolaryngol*. 2008;72(5):669-673.
2. Wormald PJ, Browning GG, Robinson K. Is otoscopy reliable? A structured teaching method to improve otoscopic accuracy in trainees. *Clin Otolaryngol Allied Sci*. 1995;20(1):63-67.
3. Oyewumi M, Brandt MG, Carrillo B, et al. Objective evaluation of otoscopy skills among family and community medicine, pediatric, and otolaryngology residents. *J Surg Educ*. 2016;73(1):129-135.
4. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
5. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-118.
6. Raumviboonsuk P, Krause J, Chotcomwongse P, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *NPJ Digit Med*. 2019;2(1):10. 10.1038/s41746-019-0099-8
7. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *J Travel Med*. 2020;27(2):1-3.
8. Seah JCY, Tang CHM, Buchlak QD, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *Lancet Digit Health*. 2021;7500(21):1-11.

9. Marom T, Kraus O, Habashi N, Tamir SO. Emerging technologies for the diagnosis of otitis media. *Otolaryngol Head Neck Surg*. 2019;160(3):447-456.
10. Smith AC, Armfield NR, Wu WI, Brown CA, Perry C. A mobile telemedicine-enabled ear screening service for Indigenous children in Queensland: activity and outcomes in the first three years. *J Telemed and Telecare*. 2012;18(8):485-489.
11. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*. 2009;6(7):e1000097.
12. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(4):529-536.
13. Higgins J & Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0 [Internet]. 2011. Accessed February 18, 2021. www.cochrane-handbook.org
14. Simon C, Caorsi V, Campillo C. Chapter 14: otitis media diagnosis model for tympanic. *Modelling and analysis of active biopotential signals in healthcare*, 2nd ed. Institute of Physics Publishing Ltd; 2020.
15. Basaran E, Comert Z, Celik Y. Convolutional neural network approach for automatic tympanic membrane detection and classification. *Biomed Signal Process Control*. 2020;56(1):1-14.
16. Goshtasbi K. Machine learning models to predict diagnosis and surgical outcomes in otolaryngology [Internet]. University of California Irvine; 2020. Accessed January 3, 2021. https://escholarship.org/uc/item/1tr0c2p0
17. Tsutsumi K, Goshtasbi K, Risbud A, et al. A web-based deep learning model for automated diagnosis of otoscopic images. *Otol Neurotol*. 2021;42(9):e1382-e1388.
18. Mironică I, Vertan C, Gheorghe DC. Automatic pediatric otitis detection by classification of global image features. 2011 E-Health and Bioengineering Conference, EHB 2011. 2011;1-4.
19. Crowson MG, Hartnick CJ, Diercks GR, et al. Machine learning for accurate intraoperative pediatric middle ear effusion diagnosis. *Pediatrics*. 2021;147(4):e2020034546.
20. Wu Z, Lin Z, Li L, et al. Deep learning for classification of pediatric otitis media. *Laryngoscope*. 2020;131:1-8.
21. Huang YK, Huang CP. A depth-first search algorithm based otoscope application for real-time otitis media image interpretation. Parallel and Distributed Computing, Applications and Technologies, PDCAT Proceedings. 2018;2017-Decem:170–5.
22. Kuruvilla A, Shaikh N, Hoberman A, Kovačević J. Automated diagnosis of otitis media: vocabulary and grammar. *Int J Biomed Imaging*. 2013;2013:1-15.
23. Kuruvilla A, Li J, Yeomans PH, Quelhas P, Instituto I, Biomedica DE. Otitis media vocabulary and grammar. *Proc Int Conf Image Proc*. 2012;2012:2845-2848.
24. Sundgaard JV, Harte J, Bray P, et al. Deep metric learning for otitis media classification. *Med Image Anal*. 2021;71:102034.
25. Livingstone D, Chau J. Otoscopic diagnosis using computer vision: an automated machine learning approach. *Laryngoscope*. 2020;130:1408-1413.
26. Khan MA, Kwon S, Choo J, et al. Automatic detection of tympanic membrane and middle ear infection from oto-endoscopic images via convolutional neural networks. *Neural Netw*. 2020;126:384-394. 10.1016/j.neunet.2020.03.023
27. Cai Y, Yu JG, Chen Y, et al. Investigating the use of a two-stage attention-aware convolutional neural network for the automated diagnosis of otitis media from tympanic membrane images: a prediction model development and validation study. *BMJ Open*. 2021;11(1):1-7.
28. Byun H, Yu S, Oh J, et al. An assistive role of a machine learning network in diagnosis of middle ear diseases. *J Clin Med*. 2021;10(15):3198.
29. Wang W, Tamhane A, Santos C, et al. Pediatric otoscopy video screening with shift contrastive anomaly detection. *Front Digit Health*. 2022;3:810427.
30. Pichichero ME, Poole MD. Comparison of performance by otolaryngologists, pediatricians, and general practioners on an otoendoscopic diagnostic video examination. *Int J Pediatr Otorhinolaryngol*. 2005;69(3):361-366.
31. Geron A, Géron A. Hands-on machine learing with scikit-learn & tensor flow [Internet]. O'Reilly Media; 2017:760. Accessed March 18, 2021. http://arxiv.org/abs/1412.3919
32. Kasher MS. Otitis media analysis-an automated feature extraction and image classification system. 2018. Accessed November 10, 2018. https://www.theseus.fi/handle/10024/144562
33. Pham VT, Tran TT, Wang PC, Chen PY, Lo MT. EAR-UNet: a deep learning-based approach for segmentation of tympanic membranes from otoscopic images. *Artif Intell Med*. 2021;115:102065.
34. Pichichero ME. Can machine learning and AI replace otoscopy for diagnosis of otitis media? *Pediatrics*. 2021;147(4):e2020049584.
35. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402.
36. Krause J, Gulshan V, Rahimy E, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology*. 2018;125(8):1264-1272.
37. Gates GA, Cooper JC. Effect of anesthetic gases on middle ear pressure in the presence of effusion. *Ann Otol Rhinol Laryngol Suppl*. 1980;89(3 Pt 2):62-64.
38. Ghojogh B, Crowley M. The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. *arXiv*. 2019;1905.12787. http://arxiv.org/abs/1905.12787
39. Hawkins DM, Basak SC, Mills D. Assessing model fit by cross-validation. *J Chem Inf Comput Sci*. 2003;579-586.
40. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv*. 2010;4:40-79.
41. Alenezi EMA, Jajko K, Reid A, et al. The reliability of video otoscopy recordings and still images in the asynchronous diagnosis of middle-ear disease. *Int J Audiology*. 2021. doi:10.1080/14992027.2021.1983217.
42. Myburgh HC, Jose S, Swanepoel DW, Laurent C. Towards low cost automated smartphone- and cloud-based otitis media diagnosis. *Biomed Signal Process Control*. 2018;39:34-52.
43. Senaras C, Moberly AC, Teknos T, Essig G, Elmaraghy C, Taj-Schaal N, et al. Detection of eardrum abnormalities using ensemble deep learning approaches. In: Mori K, Petrick N, eds. *Medical Imaging 2018: Computer-Aided Diagnosis* [Internet]. SPIE; 2018:45. Accessed February 26, 2019. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10575/2293297/Detection-of-eardrum-abnormalities-using-ensemble-deep-learning-approaches/10.1117/12.2293297.full
44. Tran TT, Fang TY, Pham VT, Lin C, Wang PC, Lo MT. Development of an automatic diagnostic algorithm for pediatric otitis media. *Otol Neurotol*. 2018;39(8):1060-1065.
45. Senaras C, Moberly AC, Teknos T, et al. Autoscope: automated otoscopy image analysis to diagnose ear pathology and use of clinically motivated eardrum features. *Med Imaging 2017: Compu-Aided Diagnosis*. 2017;10134(March 2017):101341X.
46. Myburgh HC, van Zijl WH, Swanepoel DW, Hellström S, Laurent C. Otitis media diagnosis for developing countries using tympanic membrane image-analysis. *EBioMedicine*. 2016;5:156-160.
47. Wang X, Valdez TA, Bi J. Detecting tympanostomy tubes from otoscopic images via offline and online training. *Comput Biol Med*. 2015;61:107–118. 10.1016/j.compbiomed.2015.03.025
48. Shie CK, Chang HT, Fan FC, et al. A hybrid feature-based segmentation and classification system for the computer aided self-diagnosis of otitis media. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014. 2014:4655–4658.

49. Kuruvilla A, Shaikh N, Hoberman A, Kovačević J. Automated diagnosis of otitis media: vocabulary and grammar. *Int J Biomed Imaging*. 2013;2013:1–15.
50. Kuruvilla A, Li J, Yeomans PH, Quelhas P, Instituto I, Biomedica DE. Otitis media vocabulary and grammar. *IEEE*. 2012;2845–2848.
51. Mironică I, Vertan C, Gheorghe DC. Automatic pediatric otitis detection by classification of global image features. 2011 E-Health and Bioengineering Conference, EHB 2011. 2011:1–4.
52. Vertan C, Gheorghe DC, Ionescu B, Eardrum color content analysis in video-otoscopy images for the diagnosis support of pediatric otitis. ISSCS 2011 - International Symposium on Signals, Circuits and Systems, Proceedings. 2011;129–32.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Habib A-R, Kajbafzadeh M, Hasan Z, et al. Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. *Clin Otolaryngol*. 2022;47:401–413. doi:10.1111/coa.13925

2.5 Concluding summary for this chapter and knowledge gained from this study

This systematic review and meta-analysis included 39 articles. AI-based image classification algorithms achieved 90.7% accuracy in distinguishing between normal and abnormal otoscopic images. Algorithms with multiple diagnostic classes achieved up to 97.6% accuracy depending on the combination of classes included. AI algorithms demonstrated an accuracy of 93.4% compared to 73.2% among non-expert human assessors. Among the various machine learning methods, CNNs achieved the highest classification performance. The review also reveals the adaptability of AI algorithms to handle a diverse range of diagnostic classes, from normal to specific ear conditions. However, the study identifies the heterogeneity between studies due to diverse data sources, machine learning techniques, and ground-truth definitions.

To integrate AI into clinical practice, the review emphasises the importance of establishing standardised and validated otoscopy databases for robust algorithm training. Moreover, AI's potential to address healthcare challenges in underserved settings, such as rural and remote areas, can be applied to support healthcare workers with limited otology expertise. The review also raises ethical considerations and the need to mitigate bias, promoting the responsible deployment of AI.

2.6 References

1. Habib AR, Kajbafzadeh M, Hasan Z, Wong E, Gunasekera H, Perry C, Sacks R, Kumar A, Singh N. Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. Clinical Otolaryngology. 2022 May 1;47(3):401–13.

CHAPTER THREE: Tele-otology for Aboriginal and Torres Strait Islander people living in rural and remote areas

3.1 Preface to the chapter

This chapter presents findings from a study evaluating the referral-based, tele-otology service in the Northern Territory from 2011 to 2019. The purpose of this study was to summarise the prevalence of ear disease in Aboriginal and Torres Strait Islander people living in rural and remote areas, intra- and inter-rater reliability of ear disease diagnoses established from the tele-otology service and explore the progression of ear disease in a subset of patients with two consecutive assessments. The version of the study included in this chapter has been accepted for publication and presented in manuscript format. Supplementary material accompanying the accepted manuscript is included in this chapter.

This chapter specifically addresses aim #2 of this thesis as described in Chapter 1. Dissemination of this research and author contributions for this paper are described below.

3.2 Research dissemination

Manuscript accepted for publication:

Habib AR, Crossland G, Sacks R, Singh N, Patel H. Tele-otology for Aboriginal and Torres Strait Islander people living in rural and remote areas. The Laryngoscope. 2023. Accepted for publication on 18 June 2024.

Journal impact factor: 3.3

Conference presentations:

Habib AR*, Crossland G, Sacks R, Singh N, Patel H. *Tele-otology in Aboriginal and Torres Strait Islander children:* A 10-year retrospective analysis of 13,672 ear examinations from the

Northern Territory, Australia. Australian Society for Otolaryngology – Head and Neck Surgery Annual Conference. Melbourne, Australia (virtual conference). 2021. [Oral presentation].

*Presenting author

3.3 Author attribution statement

I, Al-Rahim Habib, was responsible for designing the study, interpreting data, writing drafts of the manuscript, submitting the manuscript, responding to reviewers' comments, and coordinating submission and publication of the manuscript.

My co-authors, Graeme Crossland, Hemi Patel and Narinder Singh helped design the study. Al-Rahim Habib extracted and analysed the data. Al-Rahim Habib prepared the manuscript; Al-Rahim Habib, Graeme Crossland, Hemi Patel, and Narinder Singh were involved in synthesis, revision, and approval of the manuscript. All authors have read and approved the manuscript.

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statement above is correct.

Signature:

Associate Professor Narinder Singh

3.4 Paper in manuscript format followed by corresponding supplementary material

Title: Tele-otology for Aboriginal and Torres Strait Islander people living in rural and remote areas

Short Running Title (40 characters): Tele-otology for Aboriginal and Torres Strait Islander people

Authors:

Al-Rahim Habib[1,2,3]

Graeme Crossland[2]

Raymond Sacks[1]

Narinder Singh[1,3]

Hemi Patel[2]

Affiliations:

1. Sydney Medical School, Faculty of Medicine and Health, University of Sydney, New South Wales, Australia
2. Department of Otolaryngology - Head and Neck Surgery, Royal Darwin Hospital, Top End Health Service, Department of Health, Northern Territory, Australia
3. Department of Otolaryngology - Head and Neck Surgery, Westmead Hospital, New South Wales, Australia

Corresponding Author:

Dr Al-Rahim Habib, Faculty of Medicine and Health, University of Sydney,

(email: al-rahim.habib@sydney.edu.au)

Sources of Support: None

Disclosure of Funding:

1. Research Scholarship from the Garnett Passe and Rodney Williams Memorial Foundation (Dr Al-Rahm Habib)
2. Microsoft AI for Humanitarian Action Grant (Dr Al-Rahim Habib)
3. Avant Foundation Doctor-in-Training Research Grant (Dr Al-Rahim Habib)

Acknowledgements

We acknowledge the traditional owners of the lands on which this research project was conducted, their spiritual relationship to the land, and their elders both past, present, and emerging. We thank the Aboriginal and Torres Strait Islander people, elders, and community councils for allowing this research to be completed. We sincerely appreciate and acknowledge the assistance of Amarjit Anand, Christian Aguilar, Jade Winiata, and the many clinicians, nurses, audiologists, health workers, and administrative staff in the ENT Department at the Royal Darwin Hospital for their efforts collecting and managing the clinical data used for this study.

Ethics approval:

Ethics approval was obtained from the Menzies School of Health Research, Northern Territory, Australia, (HREC: 2019-3410).

Abstract

Objective:

To evaluate a referral-based, tele-otology service in rural and remote areas of the Northern Territory (NT), Australia.

Methods:

A retrospective observational cohort study was performed of a tele-otology service in 93 Aboriginal and Torres Strait Islander communities (2011 to 2019). Assessments included face-to-face examinations performed by Clinical Nurse Consultants (CNC) and audiologists, and asynchronous reviews performed by otolaryngologists. Multivariable logistic regression was performed to determine the likelihood of ear disease, adjusted for age and gender. Intra- and inter-rater agreement was assessed between otolaryngologists.

Results:

3,950 patients were reviewed (6,838 encounters, 13,726 ear assessments). The median age of patients was 9.8 years (inter-quartile range: 7.2 years). 62.2% of patients were identified with ear disease and 62.5% identified with hearing loss. Substantial intra- and inter-rater agreement in diagnosis was found between otolaryngologists ($\kappa = 0.71$ and $\kappa = 0.78$, respectively). The most common ear conditions identified were chronic otitis media (COM, 28.1%) and otitis media with effusion (OME, 16.5%). Topical or oral antibiotics were initiated in 14.1% of all encounters, most often for acute otitis media (AOM) or COM. Surgery was recommended in 27.7% of all encounters, most often myringoplasty, adenoidectomy, and myringotomy with insertion of tympanostomy tubes.

Conclusion:

Tele-otology is a critical component of an integrated approach to evaluating ear disease in Indigenous people living in rural and remote areas. The high prevalence of OME, COM and surgical recommendations highlights the need for community engagement, regular follow-up, and early interventions to prevent long-term hearing loss.

Keywords: otitis media, telehealth, otology, epidemiology

Introduction

Aboriginal and Torres Strait Islander people living in rural and remote areas of Australia experience a high burden of ear disease[21,121,122]. Recurrent otitis media can be associated with long-term hearing loss and can contribute to detrimental lifelong impacts on speech and language development, academic performance, behaviour, social isolation, future employment opportunities, and increased contact with the criminal justice system.[45,49,53,123] The prevalence of self-reported ear or hearing problems are greatest in the Northern Territory (NT) where geographical, cultural, and logistical factors can impact the delivery of hearing health services.[124]

Telehealth services have been developed to overcome geographical barriers to specialist access, reduce costs associated with travel, and partner with community organisations to establish culturally appropriate and sustainable care. [125,126] Real-time consultations can be conducted between the patient, family, regional clinician (referrer), and offsite otolaryngologists.[127] In rural settings, real-time video-conferencing telehealth assessments demonstrate high concordance with face-to-face assessments to establish diagnoses and surgical management decisions.[125] However, the real-time model of care relies on suitable resources, including videoconferencing equipment and facilities on-site, trained personnel on-site, and access to reliable internet service with adequate bandwidth for transmission of live video feeds.[128]

Alternative strategies such as asynchronous, “store-and-forward” telehealth approaches have been utilised.[129] One approach for community-based mobile screening services utilises a dedicated vehicle equipped with video otoscopes, audiometric testing equipment and an

Aboriginal Health Worker with advanced hearing-health training.[126] The vehicle includes a wireless broadband connection for transmission of data but uses asynchronous data transfer rather than live video feeds. This approach has contributed to an increase in referrals to otolaryngologists and reduced waiting time for specialist review.[130] Such asynchronous, "store-and-forward" telehealth strategies, known as "tele-otology" are valuable because specialists can review data when convenient, utilise data managers to organise examination results into truncated packages for expedited review, and can tailor participation relative to competing clinical commitments. In addition, complex equipment, remote transmission facilities and trained personnel can remain with the dedicated vehicle, rather than requiring duplication of equipment and staff at multiple remote locations.

A 'store and forward' tele-otology model has been used in rural and remote Indigenous communities in the NT, following its inception by authors HP and GC. Individuals identified in primary care with persistent middle ear disease are referred to the NT Ear, Nose and Throat (ENT) Service for tele-otology or specialist outreach assessment. The tele-otology service is comprised of a specialty-trained Clinical Nurse Consultant (CNC) and an Audiologist who can perform detailed face-to-face ear assessments to all major communities and homelands in the NT. The tele-otology team are equipped with mobile computer workstations to perform ear examinations (i.e. digital video otoscopy, tympanometry, audiometry). Clinical information is collected during community visits and subsequently reviewed by metropolitan otolaryngologists. This model of care provides a broad range of ear health services to individuals from rural and remote communities, and has established standardised, high-quality ear examination data for otolaryngologist review.

The aim of this study was to review the referral-based, tele-otology ear health service in the NT since its inception in 2011, and to summarise the identified burden of ear disease in Aboriginal and Torres Strait Islander people from these rural and remote areas.

Methods

Ethics

Ethics approval was obtained from the Menzies School of Health Research, Northern Territory, Australia, (HREC: 2019-3410). This research was conducted in accordance with the Helsinki Declaration and with the support and approval of the Aboriginal communities involved.[131]

Tele-otology assessments

The tele-otology ear health program was initiated by the Department of Otolaryngology at the Royal Darwin Hospital (Darwin, Northern Territory, Australia) and performed with the Northern Territory Outreach Hearing Health Program. Digital ear examinations and corresponding specialist assessments were prospectively collected from March 7, 2011 to May 2, 2019 and retrospectively reviewed.

Consultations were part of non-Medicare services funded by various Australian Government programs. These programs include the Child Health Check Initiative, the Closing the Gap initiative, the National Partnership Agreement on Stronger Futures, the National Partnership on Northern Territory Remote Aboriginal Investment, and the Healthy Ears–Better Hearing, Better Listening Program.[132]

Ear examinations were performed by CNCs with otology training and experience and accompanying audiologists, with consent provided by patients and/or carers. Examinations completed by CNCs consisted of history-taking (including collection of patient demographics, community, symptoms, previous treatments, and relevant past medical history) and still-images captured with digital video otoscopes from MedRx (Largo, Florida, USA) or Welch Allyn

(Chicago, Illinois, USA). Audiologists performed tympanometry and audiometry in onsite audiometric booths, clinic facilities, school classrooms, or other locations considered suitable to perform hearing tests. Examination results were documented on a standardised proforma (Supplementary Figure 1) and forwarded to a designated data custodian for review and storage. Where internet access was not available, examination data was electronically stored on local mobile computer workstations and uploaded to the data custodian once adequate online connectivity was re-established.

Otolaryngologists at the Royal Darwin Hospital reviewed examination data consisting of the completed pro forma clinical history, tympanometry, audiometry and accompanying otoscopic images (still images collected with digital video otoscopes) (Supplementary Figure 2). Findings from previous tele-otology assessments, audiograms, and operation reports were available for the otolaryngologist to establish a diagnosis, management recommendations, and follow-up plans. Diagnostic categories included: normal (no abnormality detected), acute otitis media (AOM) with or without perforation, otitis media with effusion (OME), chronic otitis media with actively discharging perforation (COMwAP), chronic otitis media with inactive dry perforation (COMwDP), healed perforation, cholesteatoma, eustachian tube dysfunction (ETD), otitis externa (OE), foreign body or wax obstruction, presence of tympanostomy tube (in-situ, blocked, discharging, extruded) and post-operative evaluations. Detailed evaluation of otoscopic images was completed, consisting of an assessment for perforation (location and size) and tympanostomy tube placement and position.

COMwAP is a long-term condition characterised by a persistent or recurrent perforation in the tympanic membrane, accompanied by continuous or intermittent ear discharge (i.e. otorrhea) for

more than two weeks.[133] AOMwP is a short-term condition that occurs when an acute ear infection causes a sudden perforation of the tympanic membrane, leading to sudden onset ear pain, fever, and otorrhea within two weeks.[133]

The differences between COMwAP and AOMwP lie in their duration, symptoms, and causes: COMwAP is chronic, often painless, and associated with repeated infections or Eustachian tube dysfunction, whereas AOMwP is acute, typically painful, and results from a recent bacterial or viral infection. Treatment approaches also differ, with COMwAP often requiring long-term management and possible surgical intervention, while AOMwP usually heals spontaneously with antibiotic and symptomatic treatment.[134–136]

Management recommendations included: medical treatments (topical or oral antibiotics), aural toilet, foreign body removal and surgical recommendations (examination under anaesthesia, myringotomy with insertion of tympanostomy tubes, myringotomy with insertion of tympanostomy tubes and adenoidectomy, exploration of middle ear or mastoid).

Patient cohort

Patients were screened by local primary healthcare service practitioners and referred to the mobile tele-otology service for further assessment. Patients subsequently requiring face-to-face otolaryngologist review or surgical intervention were evaluated at the Royal Darwin Hospital, NT, Australia.

Statistical analysis

Demographic characteristics (age, gender) were summarised using descriptive statistics including mean, inter-quartile range, absolute frequency, or relative proportions. Age was summarised as a continuous variable (mean years of age, inter-quartile range) and a categorical variable (≤ 3 years, 4 to 6 years, 7 to 16 years, and ≥17 years. Gender was summarised as proportion of females and males. Each ear disease condition and management recommendations were classified as binary variables (yes/no) and summarised using absolute frequencies and relative proportions.

Hearing loss (HL) was classified by type (conductive, sensorineural, mixed, none, incomplete assessment) and degree (mild [16–30 dB HL in soundproof conditions; 26–35 dB HL in non-soundproof conditions], moderate [31–60 dB HL in soundproof conditions; 36–60 dB HL in non-soundproof conditions, severe [61–90 dB HL in soundproof conditions or non-soundproof conditions], profound [91+ dB HL either in soundproof conditions or non-soundproof conditions], indeterminate).[137–139]

Associations between variables of interest were assessed using odds ratios (OR) with corresponding 95% confidence intervals (95%CI) and the chi-squared test. Logistic regression was used to determine the association between ear disease and age, gender, and year of assessment. A secondary analytic sample was established using participants that had received two tele-otology reviews throughout the observation period. The prevalence of ear disease and hearing loss was compared between the first and second assessment.

Fifty randomly selected patients who previously received a tele-otology assessment were included to assess intra- and inter-rater agreement. An a priori sample size calculation was completed using previously established inter-rater agreement statistics.[29,140] An effect size of 30% was used to determine the likelihood of agreement between two raters. A sample size of 100 ear assessments (right and left ears, 50 patients) was deemed sufficient to achieve 80% power. Raters (HP and GC) are fellowship-trained otolaryngologists with extensive experience in providing ENT outreach services for Indigenous people from the NT, Australia. Assessments were de-identified, coded, and randomised to minimise recall bias by ensuring raters were blinded to the identity of participating subjects. Assessments were provided to the raters in two sessions (Attempt 1 and Attempt 2) separated by at least 1 week. Raters were provided the 8 diagnostic options consisting of no middle ear disease (n=26), AOM without perforation (n=16), AOM with perforation (n=1), COMwAP (n=4), COmwDP (n=3), OME (n=39), ETD (n=11), foreign body or cerumen obstruction (n=0), and other (n=0). Inter-rater agreement of diagnoses between raters and the reference standard (original diagnosis) were expressed as Cohen kappa statistics (κ) to represent the proportion of agreement beyond chance, stratified by the experimental index tests of interest. Standard κ interpretations of inter-rater agreement were applied as previously proposed (<0.21 slight agreement, 0.21 – 0.40 fair agreement, 0.41 – 0.60 moderate agreement, 0.61 – 0.80 substantial agreement, >0.80 near complete agreement).[141]

Results

Demographics

A total of 3950 individuals were reviewed from 93 communities within the observation period, consisting of 13726 tele-otology ear assessments. Of all the assessments, 12385 (89.8%) included sufficient clinical information, image quality, and patency of the external acoustic canal for the otolaryngologist to establish a diagnosis and management plan. Assessments were excluded due to obstruction of the external acoustic canal from wax or foreign bodies (n=186, 1.4%) and incomplete assessments without corresponding otoscopic images, audiometry or clinical history (n=1155, 8.4%).

Table 1 provides the age distribution of the patients reviewed by the tele-otology service. Most individuals were between 7 to 16 years of age (56%) with a similar distribution of females and males. On average, 759 individuals (inter-quartile range: 402 individuals) were evaluated each year in the tele-otology service. Throughout the observation period, 2354 (59.5%) individuals received 1 assessment and 1596 (40.0%) individuals received 2 or more assessments.

Rater agreement

Substantial intra- and inter-rater agreement for ear disease diagnoses was found (overall intra-rater agreement $\kappa = 0.71$; inter-rater agreement $\kappa = 0.78$, Table 2). Raters achieved complete agreement to classify AOM with perforation, COMwAP and COMwDP ($\kappa = 1.00$). Raters achieved near complete agreement to classify normal aerated ears ($\kappa = 0.88$), substantial agreement to classify OME ($\kappa = 0.82$), and moderate agreement to classify ETD ($\kappa = 0.64$) and AOM ($\kappa = 0.56$).

Prevalence of ear disease and hearing loss

During the observation period in the tele-otology program, the proportion of individuals with ear disease was 62.2%, consisting of 34.1% with bilateral ear disease and 28.1% with unilateral ear disease. The proportion of individuals with hearing loss was 62.5%, consisting of 37.3% with bilateral hearing loss and 25.2% with unilateral hearing loss. Since inception of the tele-otology program, assessments related to post operative reviews, ETD, AOM with perforation and COMwDP increased each year of the service (Figure 1).

Ear assessments were comprised of 4442 (35.9%) normal aerated ears, 7542 (60.9%) with ear disease, 300 (2.4%) post operative reviews, and 101 (0.8%) tympanostomy tube reviews. Of cases with ear disease, the most common middle ear conditions identified were OME (16.5%), COMwAP (14.1%), and COMwDP (14.0%) (Table 3).

Hearing loss was comprised of conductive (59.8%), indeterminate (34.7%), sensorineural (2.8%), and mixed (2.7%) hearing loss.

Children ≤ 3 years of age were more likely to present with AOM with or without perforation (OR: 28.5, 95%CI: 13.1 – 62.1), OME (OR 10.7, 95%CI: 7.6 – 14.8), or ETD (OR 2.9, 95%CI: 1.8 – 4.5) than individuals ≥17 years of age. The likelihood of COM with active or dry perforation was greater among individuals ≥17 years (OR 2.1, 95%CI: 1.6 – 2.6) and 7 to 16 years (OR 1.3, 95%CI: 1.1 – 1.6) than children ≤ 3 years of age.

Management

CNCs initiated oral or topical antibiotics in 53.7% of individuals with COMwAP and AOM with perforation. Otolaryngologists recommended surgical intervention in 27.7% of all cases, consisting of myringoplasty (78.8%), adenoidectomy and myringotomy ± insertion of tympanostomy tubes (12.4%), myringotomy with insertion of tympanostomy tubes (7.9%), examination under anaesthesia (0.6%), and middle ear exploration / mastoidectomy (0.3%).

Frequency of ear disease between first and second assessments

Among the 1596 individuals who received two tele-otology assessments, 14.9% were ≤ 3 years old, 27.2% were 4-6 years old, 52.0% were 7-16 years old, and 5.9% were ≥17 years old at their first assessment. The mean time between the two assessments was 634 days (inter-quartile range: 659 days). Supplementary Table 1 summarises the distribution of ear disease across the two assessments.

At the second assessment, 39.1% had new onset middle ear disease, and among those, 41.9% had worse hearing than at their first assessment. Of those with a normal first assessment, 16.9% had OME and 8.4% had ETD at the second assessment, with 7.8% experiencing worse hearing.

Among individuals with AOM at the first assessment, 39.6% were prescribed oral antibiotics and 13.2% received topical antibiotics. At the second assessment, 50.9% had OME, 37.7% had no middle ear disease, 5.7% had ETD, and 3.8% had recurrent AOM.

For those with OME at the first assessment, 46.6% presented with OME at the second assessment, and 17.6% presented with worse hearing. Among those with recurrent OME, 36.0% received a surgical recommendation at the second assessment.

For individuals with ETD at the first assessment, 34.2% had OME at the second assessment, with 22.7% experiencing worse hearing. Among those with recurrent ETD, 16.7% had worse hearing at the second assessment.

Most individuals (87.6%) with a TM perforation identified at the first assessment had a recurrent/persistent TM perforation at the second assessment. Cases with COMwAP at the first assessment showed 51.8% with recurrent/persistent actively discharging perforation and 30.8% with a dry perforation at the second assessment. Cases with COMwDP at first assessment had 46.3% dry perforations and 35.5% progressed to COMwAP.

For suppurative OM treatment at the first assessment, 5.7% received oral antibiotics and 59.9% had topical antibiotic ear drops. At the second assessment, 55.2% experienced no change in hearing, 20.6% experienced worse hearing, and 24.1% experienced hearing improvement.

Management

CNCs initiated oral or topical antibiotics in 53.7% of individuals with COMwAP and AOM with perforation. Otolaryngologists recommended surgical intervention in 27.7% of all cases, consisting of myringoplasty (78.8%), adenoidectomy and myringotomy ± insertion of tympanostomy tubes (12.4%), myringotomy with insertion of tympanostomy tubes (7.9%),

examination under anaesthesia (0.6%), and middle ear exploration / mastoidectomy (0.3%). For individuals with TM perforations, myringoplasty was recommended in 98.1% of cases. For individuals with OME, adenoidectomy with myringotomy ± insertion of tympanostomy tubes was recommended in 54.3% of cases, and myringotomy with insertion of tympanostomy tubes alone (without adenoidectomy) was recommended in 36.3% of cases.

Discussion

Telehealth programs have been shown to improve access to tertiary services for patients in rural and remote areas, provide services for pre- and post-operative reviews, reduce travel time and associated costs, and integrate with existing primary care services.[142] However, there is a paucity of information related to the prevalence of ear disease and caseload distribution associated with long-standing telehealth services. In Australia, previous studies have reported the prevalence of ear disease to range from 15 to 91% in Aboriginal and Torres Strait Islander people, depending on sampling locations, participant age, and subtype of otitis media.[6]

In the NT, previously published studies have reported the prevalence of ear disease from longitudinal reviews of neonates, vaccination programs, community and school-based screening programs, and emergency response services.[121,143–145] From these studies, the prevalence of AOM was reported as 7% to 26%, OME was reported as 26% to 31%, and COM was reported as 15% to 36%. In comparison, the referral-based NT tele-otology service reported AOM (2.6%), OME (16.5%), and COM (28.1%).

Since inception, there was a decrease in the proportion of individuals referred each year with a normal aerated middle ear, indicating a maturation of the referral process to exclude those individuals not requiring specialist attention. Of the individuals identified with at least 1 ear condition, nearly 1 in 3 experienced hearing loss. Rates of hearing loss were highest among individuals with COM with or without discharge, and OME. CNCs initiated oral or topical antibiotics in over half of individuals with suppurative OM. Otolaryngologists recommended a

surgical intervention in 1 of every 3 individuals reviewed. These findings demonstrate that the tele-otology service captures a large referral base of chronic OM requiring point-of-care medical management and surgical interventions.

The comparison in the distribution of ear disease between first and second assessments provides valuable insight into the progression and persistence of middle ear disease in Aboriginal and Torres Strait Islander people. Over one-third of individuals with ETD at the first assessment presented with OME at the second assessment. Recurrent OME at the second assessment was identified in nearly half of individuals who were found to have OME at the first assessment. These findings demonstrate that between assessments, it is common for ear disease to progress or persist, rather than to resolve. This emphasises the importance of rigorous follow-up, surveillance, and the potential role for early surgical intervention in children with ETD and OME. For example, having a lower threshold for inserting tympanostomy tubes could be an early treatment consideration for individuals with ETD or for individuals with OME, as most will have middle ear disease at subsequent reviews. However, this needs to be balanced against the risk of introducing tube-related complications and the need for greater post-operative care in patients with tympanostomy tubes present.

Most children identified with tympanic membrane perforations presented with persistent perforations at subsequent reviews. One in 2 children with COMwAP re-presented with no improvement in ear disease status at the second assessment. For children with COMwDP, 1 in 3 developed active discharging perforations at the second review. In these cases, myringoplasty could offer a beneficial intervention by addressing the likelihood of persistent tympanic

membrane perforation, improving auditory function, and reducing the risk of new or recurrent infections. However, it is important to note that myringoplasty carries risks, such as potential graft failure, postoperative infection, and the possibility of temporary or permanent hearing loss.

The strengths of this study are demonstrated by the volume of assessments collected over nearly a decade, expansive catchment of Indigenous communities from rural and remote areas, and substantial intra- and inter-rater agreement for diagnoses. In remote settings where loss to follow-up is common, efforts to establish culturally appropriate services are important to engage patients and parents/ guardians, develop rapport, foster familiarity, and confidence. Providing primary care practitioners an avenue to refer patients for tertiary otolaryngology review is a priority to identify and manage ear disease to prevent long-term hearing loss and adverse complications.

This study has several limitations. The scope of this study was to evaluate the initial triage and prevalence estimates ascertained from the tele-otology service in the NT, Australia. This study did not track outcomes of patients referred for face-to-face specialist reviews. Future endeavours will be focused on linking tele-otology assessments to hospital admissions and surgical outcomes. Data linkage may be valuable to delineate the down-stream challenges in providing specialist services requiring face-to-face interactions and the impact of tele-otology on surgical waitlists. Although tele-otology may improve access to specialists via asynchronous assessments for diagnosis, the influx of patients requiring subsequent surgical treatment may face additional limitations in accessing these services. In addition, the reliability assessment performed in this study may limit the generalisability of the findings and underestimate agreement, given that less

common ear disease pathology could be underrepresented in the analytic cohort. Nasal steroids have shown promise in treating OME in children but were not reported in this study.[146–148] Several factors may have influenced the limited use of nasal steroids in this population and clinical context. Rural and remote locations of many communities can make accessing medications challenging due to limitations in infrastructure and access to distributors.[149] Compliance with topical medications may be lower due to the inconvenience of administration and the need for regular use. There may be differences in awareness regarding the benefits of nasal steroids for OME, impacting their usage. Effective communication about treatment plans can be challenging, further affecting compliance and proper use of medications. Moreover, topical nasal steroids may not have been a standard part of practice for the clinicians in the settings studied, possibly due to historical prescribing patterns or resource limitations. Future research should focus on understanding the barriers to using topical nasal steroids in this setting and explore strategies to improve compliance, access, and evaluate outcomes.

Asynchronous store-and-forward telehealth approaches such as these are limited by the delays between data capture, specialist interpretation, and subsequent intervention, representing a potential missed opportunity to initiate early treatment at the point of data capture. Artificial intelligence (AI) has the potential to be integrated into existing telehealth programs.[150–152] AI-based image classification algorithms could be developed to support rural and remote healthcare workers classify ear disease from digital video otoscopy and triage the appropriate clinical pathway.[153] For example, healthcare workers could be trained to perform otoscopy and upload images into an AI-based tool either directly linked with the otoscope, accessed via smartphone or desktop applications, or cloud-based models where internet access is available. [120] These

advancements would support improved triage in frontline settings where immediate access to tertiary services is limited. AI-tools could support surveillance to improve allocation of resources in high-demand areas. Otolaryngologists could review the AI-generated reports to establish prediction models that anticipate care needs and receive red flag notifications to initiate alerts for children who warrant urgent review.

Conclusion

Since inception in 2011, through to 2019, a referral-based tele-otology service in the NT, Australia has performed 13726 ear examinations in 3950 Aboriginal and Torres Strait Islander individuals from 93 communities. On average, the tele-otology service reviews 759 Aboriginal and Torres Strait Islander people per year. This service more commonly reviews individuals with CSOM than community screening programs and audiology services. Substantial intra- and inter-rater agreement was found for ear disease diagnoses using an asynchronous, 'store-and-forward' data collection workflow. The results of this review confirm that the service has made substantial progress in addressing ear disease in this critically under-served population, but key areas requiring further attention have been identified. This tele-otology service may serve as a reference for future telehealth services in rural and remote settings to optimise service delivery for ear health and hearing assessments for Aboriginal and Torres Strait Islander people.

Author statement:

The authors had full access to all the data, including statistical reports and tables, in the study.

References

1. Leach AJ. Otitis media in Australian Aboriginal children: An overview. Int J Pediatr Otorhinolaryngol. 1999;49(SUPPL. 1):S173–8.

2. 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;5:1–10.

3. World Health Organization - Child and Adolescent Health and Development Prevention of Blindness and Deafness. Chronic suppurative otitis media: Burden of Illness and Management Options [Internet]. Geneva, Switzerland; 2004. Available from: https://www.who.int/pbd/publications/Chronicsuppurativeotitis_media.pdf

4. Su JY, He VY, Guthridge S, Howard D, Leach A, Silburn S. The impact of hearing impairment on Aboriginal children's school attendance in remote Northern Territory: a data linkage study. Aust N Z J Public Health. 2019;43(6):544–50.

5. He VY, Su JY, Guthridge S, Malvaso C, Howard D, Williams T, et al. Hearing and justice: The link between hearing impairment in early childhood and youth offending in Aboriginal children living in remote communities of the Northern Territory, Australia. Health Justice. 2019;7(1).

6. Gibney KB, Morris PS, Carapetis JR, Skull SA, Smith-Vaughan HC, Stubbs E, et al. The clinical course of acute otitis media in high-risk Australian Aboriginal children: a longitudinal study. BMC Pediatr [Internet]. 2005 Dec 14 [cited 2019 Feb 19];5(1):16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15955251

7. Su JY, Guthridge S, He VY, Howard D, Leach AJ. Impact of hearing impairment on early childhood development in Australian Aboriginal children: A data linkage study. J Paediatr

Child Health [Internet]. 2020 Oct 1 [cited 2023 May 28];56(10):1597–606. Available from: https://pubmed.ncbi.nlm.nih.gov/32725651/

8. Aihw. Hearing health outreach services to Aboriginal and Torres Strait Islander children and young people in the Northern Territory: 2012–13 to 2014–15. Canberra; 2015.

9. Smith AC, Dowthwaite S, Agnew J, Wootton R. Concordance between real-time telemedicine assessments and face-to-face consultations in paediatric otolaryngology. Medical Journal of Australia. 2008;188(8):457–60.

10. 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. Telemedicine and e-Health [Internet]. 2010;16(9):950–6. Available from: http://www.liebertonline.com/doi/abs/10.1089/tmj.2010.0045

11. Ullah R, Gilliland D, Adams D. Otolaryngology consultations by real-time telemedicine. Ulster Medical Journal. 2002;71(1):26–9.

12. Smith AC. Telemedicine: Challenges and opportunities. Vol. 4, Expert Review of Medical Devices. 2007. p. 5–7.

13. Bennett RJ, Swanepoel D, Manchaiah V. Tele-audiology services in Australia: A shift in clinical practices. Audiology Now [Internet]. 2020;81:11–3. Available from: https://www.researchgate.net/publication/344664862

14. Smith AC, Armfield NR, Wu WI, Brown CA, Perry C. A mobile telemedicine-enabled ear screening service for Indigenous children in Queensland: Activity and outcomes in the first three years. J Telemed Telecare. 2012;18(8):485–9.

15. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. JAMA. 2013 Nov 27;310(20):2191–4.

16. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33(1):159–74.

17. Nash K, Macniven R, Clague L, Coates H, Fitzpatrick M, Gunasekera H, et al. Ear and hearing care programs for First Nations children: a scoping review. BMC Health Serv Res [Internet]. 2023 Dec 1 [cited 2023 Oct 16];23(1):1–12. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-023-09338-2

18. Jervis-Bardy J, Sanchez L, Carney AS. Otitis media in indigenous australian children: Review of epidemiology and risk factors. Journal of Laryngology and Otology. 2014;128(SUPPL.S1).

19. Foreman A. The Aetiology and Prevention of Otitis Media in Indigenous Children in the Northern Territory, Australia. [Darwin, Northern Territory]: University of Sydney; 1987.

20. Roberts D, Gracey M, Spargo RM. Growth and morbidity in children in a remote aboriginal community in north-west Australia. Med J Aust [Internet]. 1988 [cited 2023 Oct 17];148(2):68–71. Available from: https://pubmed.ncbi.nlm.nih.gov/3336337/

21. Australian Institute of Health and Welfare. Northern Territory Emergency Response Child Health Check Initiative: Follow- up Services for Oral and Ear Health. Canberra; 2012.

22. D'Onofrio KL, Zeng FG. Tele-Audiology: Current State and Future Directions. Vol. 3, Frontiers in Digital Health. Frontiers Media S.A.; 2022.

23. Kuziemsky C, Maeder AJ, John O, Gogia SB, Basu A, Meher S, et al. Role of Artificial Intelligence within the Telehealth Domain. Yearb Med Inform. 2019 Aug 1;28(1):35–40.

24. Schünke LC, Mello B, da Costa CA, Antunes RS, Rigo SJ, Ramos G de O, et al. A rapid review of machine learning approaches for telemedicine in the scope of COVID-19. Vol. 129, Artificial Intelligence in Medicine. Elsevier B.V.; 2022.

25. Habib AR, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, et al. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images From Australian Aboriginal and Torres Strait Islander Children. Otology and Neurotology. 2022 Apr 1;43(4):481–8.

26. Habib AR, Kajbafzadeh M, Hasan Z, Wong E, Gunasekera H, Perry C, et al. Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. Clinical Otolaryngology. 2022 May 1;47(3):401–13.

Table 1. Summary of demographic characteristics for tele-otology assessments completed at the Royal Darwin Hospital, Northern Territory, Australia from 2011 to 2019

	Total Cohort (n=3950)	
	n	%
Age		
≤ 3 years	419	10.6
4 to 6 years	564	14.3
7 to 16 years	2227	56.4
≥17 years	740	18.7
Gender		
Female	2016	51.0
Male	1934	49.0

Table 2. Comparing intra- and inter-rater agreement between reviewers for tele-otology diagnoses from the Royal Darwin Hospital, Northern Territory, Australia between 2010 to 2019.

	Overall	Standard Error
Average intra-rater agreement		
Attempt 1 and 2		
Percent agreement	83% (95%CI: 70 – 95%)	0.06
Cohen κ	0.71 (95%: 0.50 – 0.91)	0.10
Average inter-rater agreement by ground truth classification (Rater 1 vs Rater 2)		
Overall		
Percent agreement	81% (95%CI: 73 – 89%)	0.04
Cohen κ	0.78 (95%CI: 0.69 – 0.87)	0.05
Normal aerated ear / no middle ear disease		
Percent agreement	88% (95%CI: 75 –100%)	0.06
Cohen κ	0.83 (95%CI: 0.63 –1.00)	0.09
AOM without perforation		
Percent agreement	56% (95%CI: 29 – 84%)	0.13
Cohen κ	0.41 (95%CI: 0.05 – 0.78)	0.17
AOM with perforation		
Percent agreement	100%	-
Cohen κ	1.00	-
OME		
Percent agreement	82% (95%CI: 69 – 95%)	0.06
Cohen κ	0.73 (95%CI: 0.54 – 0.92)	0.09
ETD		
Percent agreement	82% (95%CI: 55 – 100%)	0.12
Cohen κ	0.64 (95%CI: 0.09 – 1.00)	0.24
COMwAP		
Percent agreement	100%	-
Cohen κ	1.00	-
COMwDP		
Percent agreement	100%	-
Cohen κ	1.00	-

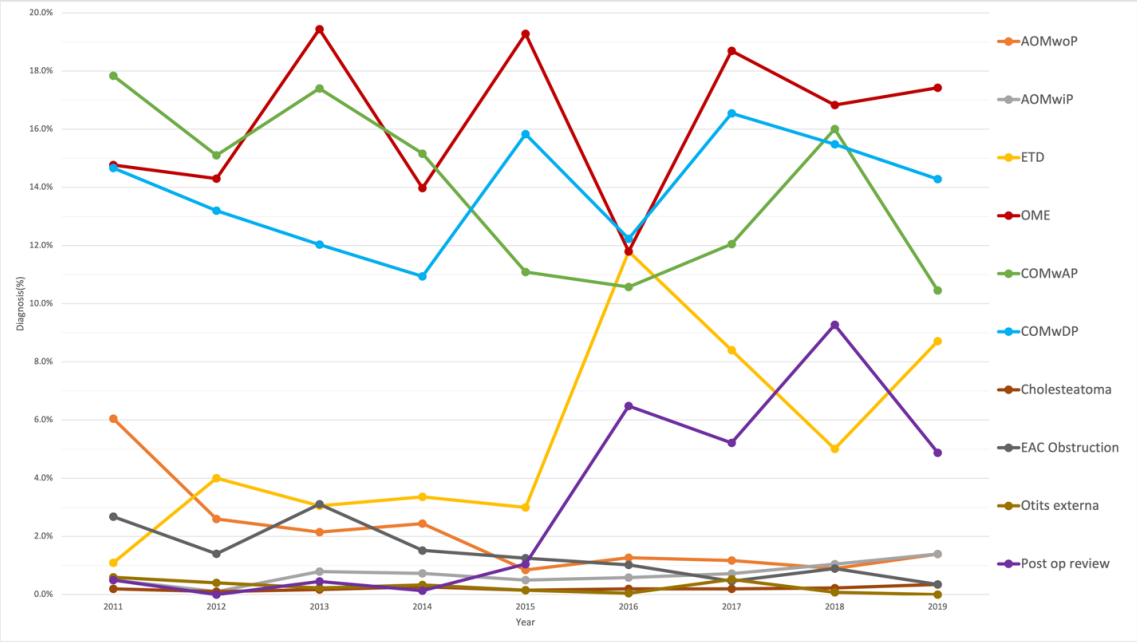
Legend: Agreement – proportion of agreement within and between raters, κ – Cohen kappa (proportion of agreement beyond chance)

Table 3. Distribution of ear disease by age group

	Total		≤3 years		4 to 6 years		7 to 16 years		≥17 years		
	n	%	n	%	n	%	n	%	n	%	Chi-squared test (p-value)
No ear disease	5394	43.6%	245	25.2%	535	33.7%	3621	46.2%	993	50.3%	<0.0001
Post operative review	300	2.4%	2	0.2%	9	0.6%	182	2.3%	107	5.4%	<0.001
AOMwoP	241	1.9%	98	10.1%	48	3.0%	93	1.2%	2	0.1%	<0.001
AOMwiP	82	0.7%	23	2.4%	18	1.1%	34	0.4%	7	0.4%	<0.001
ETD	681	5.5%	58	6.0%	121	7.6%	447	5.7%	55	2.8%	<0.001
OME	2047	16.5%	302	31.0%	462	29.1%	1209	15.4%	74	3.7%	<0.001
COMwAP	1746	14.1%	144	14.8%	214	13.5%	1058	13.5%	330	16.7%	0.008
COMwDP	1737	14.0%	90	9.2%	164	10.3%	1091	13.9%	392	19.9%	<0.001
Cholesteatoma	24	0.2%	1	0.1%	1	0.1%	11	0.1%	11	0.6%	0.001
Grommet *in situ*	94	0.8%	9	0.9%	11	0.7%	69	0.9%	5	0.3%	0.031
Grommet extruded	7	0.1%	0	0.0%	1	0.1%	6	0.1%	0	0.0%	0.519
Otitis externa	32	0.3%	1	0.1%	3	0.2%	20	0.3%	8	0.4%	0.434

Legend: Acute otitis media without perforation (AOMwoP), acute otitis media with perforation (AOMwiP), otitis media with effusion (OME), chronic suppurative otitis media with active perforation (COMwAP), chronic otitis media with dry perforation (COMwDP), Eustachian Tube dysfunction (ETD), otitis externa (OE)

Figure 1. Distribution of ear disease by year for the tele-otology service from 2011 to 2019.



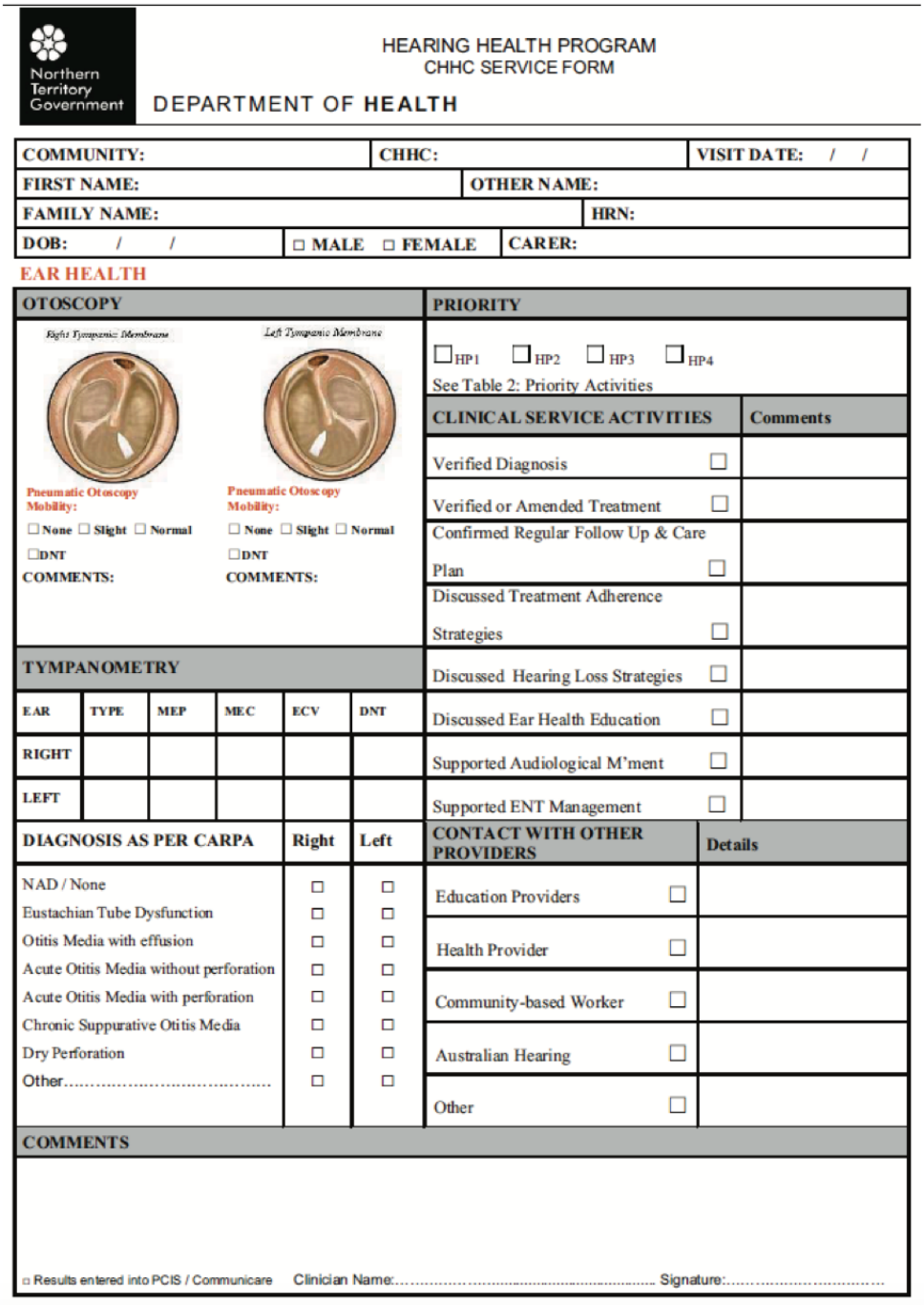
Legend: Acute otitis media without perforation (AOMwoP), acute otitis media with perforation (AOMwiP), otitis media with effusion (OME), chronic suppurative otitis media with active perforation (COMwAP), chronic otitis media with dry perforation (CSOMwDP), Eustachian Tube dysfunction (ETD), otitis externa (OE), external acoustic canal (EAC) obstruction.

Supplementary Table 1. Comparison between the first and second assessments of individuals receiving 2 tele-otology assessments

		Second Assessment										
First Assessment	Total	No middle ear disease	ETD	OME	AOMwoP	AOMwiP	CSOMwAP	CSOMwDP	Cholesteatoma	Grommet *in situ*	Grommet extruded	OE
No middle ear disease	980	591 (60.3%)	81 (8.3%)	164 (16.7%)	24 (2.4%)	8 (0.8%)	61 (6.2%)	48 (4.9%)	0	0	0	3 (0.3%)
ETD	152	62 (40.8%)	20 (13.2%)	52 (34.2%)	3 (2.0%)	0	6 (3.9%)	8 (5.3%)	0	1 (0.7%)	0	0
OME	477	149 (31.2%)	37 (7.8%)	219 (45.9%)	16 (3.4%)	7 (1.5%)	26 (5.5%)	16 (3.4%)	0	7 (1.5%)	0	0
AOMwoP	53	20 (37.7%)	3 (5.7%)	27 (50.9%)	2 (3.8%)	0	0	0	0	1 (1.9%)	0	0
AOMwiP	18	2 (11.1%)	0	7 (38.9%)	0	0	6 (33.3%)	2 (11.1%)	0	1 (5.6%)	0	0
COMwAP	423	46 (10.9%)	3 (0.7%)	20 (4.7%)	3 (0.7%)	0	220 (52.0%)	129 (30.5%)	0	1 (0.2%)	0	1 (0.2%)
COMwDP	424	51 (12.0%)	7 (1.7%)	12 (2.8%)	1 (0.2%)	0	155 (36.6%)	193 (45.5%)	5 (1.2%)	0	0	0
Cholesteatoma	5	0	0	0	0	0	0	2 (40.0%)	3 (60.0%)	0	0	0
Grommet *in situ*	28	4 (14.3%)	0	13 (46.4%)	0	0	2 (7.1%)	2 (7.1%)	0	7 (25.0%)	0	0
Grommet extruded	1	0	0	1 (100%)	0	0	0	0	0	0	0	0
OE	4	2 (50.0%)	0	2 (50.0%)	0	0	0	0	0	0	0	0

Legend: Acute otitis media without perforation (AOMwoP), acute otitis media with perforation (AOMwiP), otitis media with effusion (OME), chronic suppurative otitis media with active perforation (COMwAP), chronic otitis media with dry perforation (CSOMwDP), Eustachian Tube dysfunction (ETD), otitis externa (OE)

Supplementary Figure 1. Clinical Nurse Consultant data collection form for face-to-face assessments



Northern Territory Government

DEPARTMENT OF HEALTH

HEARING HEALTH PROGRAM
CHHC SERVICE FORM

COMMUNITY:	CHHC:	VISIT DATE: / /
FIRST NAME:	OTHER NAME:	
FAMILY NAME:	HRN:	
DOB: / /	☐ MALE ☐ FEMALE	CARER:

EAR HEALTH

OTOSCOPY

Right Tympanic Membrane

Pneumatic Otoscopy
Mobility:
☐ None ☐ Slight ☐ Normal
☐ DNT
COMMENTS:

Left Tympanic Membrane

Pneumatic Otoscopy
Mobility:
☐ None ☐ Slight ☐ Normal
☐ DNT
COMMENTS:

TYMPANOMETRY

EAR	TYPE	MEP	MEC	ECV	DNT
RIGHT					
LEFT					

PRIORITY

☐ HP1 ☐ HP2 ☐ HP3 ☐ HP4
See Table 2: Priority Activities

CLINICAL SERVICE ACTIVITIES	Comments
Verified Diagnosis ☐	
Verified or Amended Treatment ☐	
Confirmed Regular Follow Up & Care Plan ☐	
Discussed Treatment Adherence Strategies ☐	
Discussed Hearing Loss Strategies ☐	
Discussed Ear Health Education ☐	
Supported Audiological M'ment ☐	
Supported ENT Management ☐	

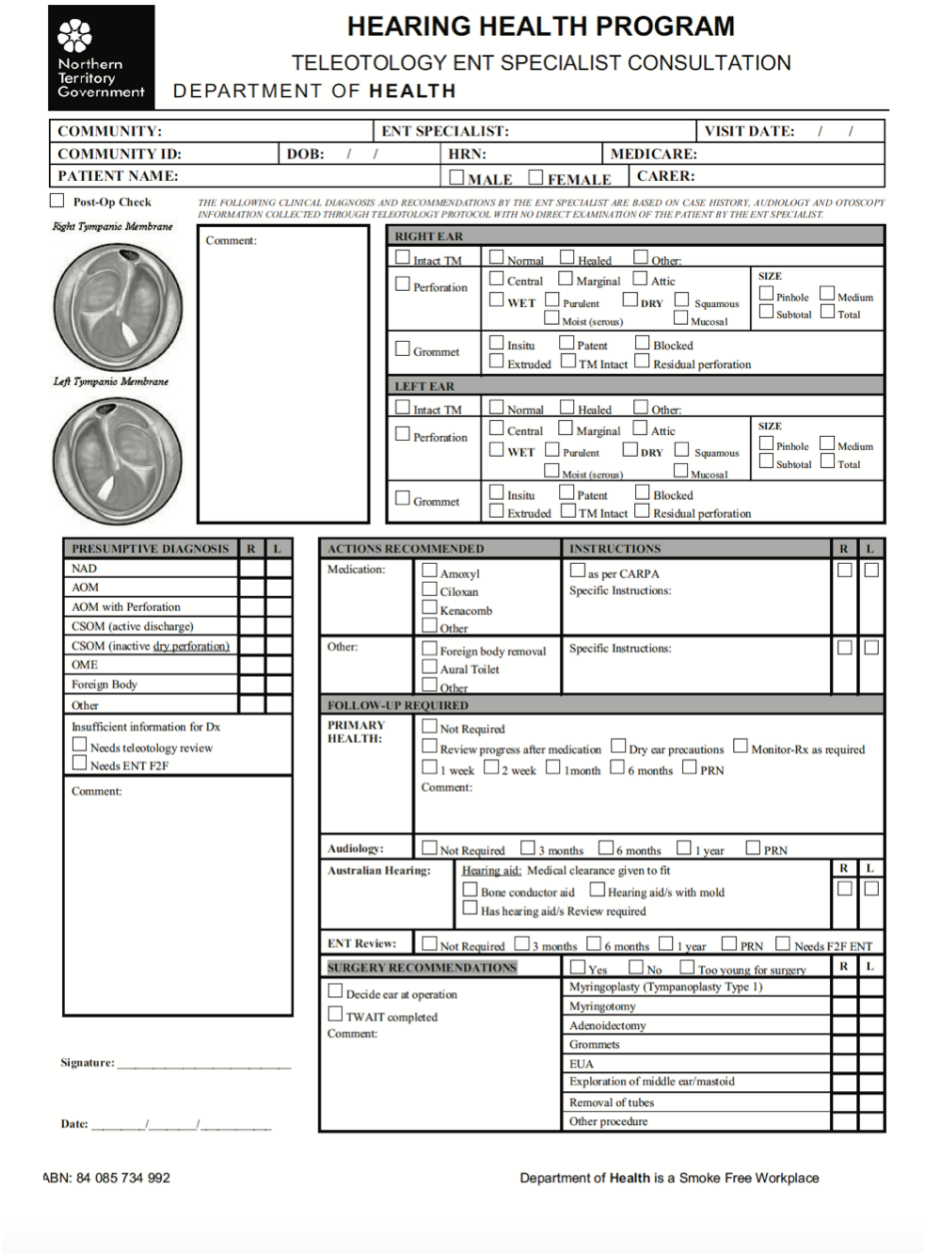
DIAGNOSIS AS PER CARPA	Right	Left
NAD / None	☐	☐
Eustachian Tube Dysfunction	☐	☐
Otitis Media with effusion	☐	☐
Acute Otitis Media without perforation	☐	☐
Acute Otitis Media with perforation	☐	☐
Chronic Suppurative Otitis Media	☐	☐
Dry Perforation	☐	☐
Other.....	☐	☐

CONTACT WITH OTHER PROVIDERS	Details
Education Providers ☐	
Health Provider ☐	
Community-based Worker ☐	
Australian Hearing ☐	
Other ☐	

COMMENTS

☐ Results entered into PCIS / Communicare Clinician Name:..... Signature:.....

Supplementary Figure 2. Otolaryngologist data collection form for tele-otology assessments



Northern Territory Government

DEPARTMENT OF **HEALTH**

HEARING HEALTH PROGRAM

TELEOTOLOGY ENT SPECIALIST CONSULTATION

COMMUNITY:	**ENT SPECIALIST:**	**VISIT DATE:** / /	
COMMUNITY ID:	**DOB:** / /	**HRN:**	**MEDICARE:**
PATIENT NAME:		☐ **MALE** ☐ **FEMALE**	**CARER:**

☐ **Post-Op Check**

THE FOLLOWING CLINICAL DIAGNOSIS AND RECOMMENDATIONS BY THE ENT SPECIALIST ARE BASED ON CASE HISTORY, AUDIOLOGY AND OTOSCOPY INFORMATION COLLECTED THROUGH TELEOTOLOGY PROTOCOL WITH NO DIRECT EXAMINATION OF THE PATIENT BY THE ENT SPECIALIST.

Right Tympanic Membrane

Left Tympanic Membrane

Comment:

RIGHT EAR		
☐ Intact TM	☐ Normal ☐ Healed ☐ Other:	
☐ Perforation	☐ Central ☐ Marginal ☐ Attic ☐ **WET** ☐ Purulent ☐ **DRY** ☐ Squamous ☐ Moist (serous) ☐ Mucosal	**SIZE** ☐ Pinhole ☐ Medium ☐ Subtotal ☐ Total
☐ Grommet	☐ Insitu ☐ Patent ☐ Blocked ☐ Extruded ☐ TM Intact ☐ Residual perforation	

LEFT EAR		
☐ Intact TM	☐ Normal ☐ Healed ☐ Other:	
☐ Perforation	☐ Central ☐ Marginal ☐ Attic ☐ **WET** ☐ Purulent ☐ **DRY** ☐ Squamous ☐ Moist (serous) ☐ Mucosal	**SIZE** ☐ Pinhole ☐ Medium ☐ Subtotal ☐ Total
☐ Grommet	☐ Insitu ☐ Patent ☐ Blocked ☐ Extruded ☐ TM Intact ☐ Residual perforation	

PRESUMPTIVE DIAGNOSIS	**R**	**L**
NAD		
AOM		
AOM with Perforation		
CSOM (active discharge)		
CSOM (inactive dry perforation)		
OME		
Foreign Body		
Other		

Insufficient information for Dx
☐ Needs teleotology review
☐ Needs ENT F2F

Comment:

ACTIONS RECOMMENDED		**INSTRUCTIONS**	**R**	**L**
Medication:	☐ Amoxyl ☐ Ciloxan ☐ Kenacomb ☐ Other	☐ as per CARPA Specific Instructions:	☐	☐
Other:	☐ Foreign body removal ☐ Aural Toilet ☐ Other	Specific Instructions:	☐	☐

FOLLOW-UP REQUIRED

PRIMARY HEALTH:	☐ Not Required ☐ Review progress after medication ☐ Dry ear precautions ☐ Monitor-Rx as required ☐ 1 week ☐ 2 week ☐ 1month ☐ 6 months ☐ PRN Comment:
Audiology:	☐ Not Required ☐ 3 months ☐ 6 months ☐ 1 year ☐ PRN
Australian Hearing:	Hearing aid: Medical clearance given to fit ☐ Bone conductor aid ☐ Hearing aid/s with mold ☐ Has hearing aid/s Review required
ENT Review:	☐ Not Required ☐ 3 months ☐ 6 months ☐ 1 year ☐ PRN ☐ Needs F2F ENT

SURGERY RECOMMENDATIONS	☐ Yes ☐ No ☐ Too young for surgery	**R**	**L**
☐ Decide ear at operation ☐ TWAIT completed Comment:	Myringoplasty (Tympanoplasty Type 1)		
	Myringotomy		
	Adenoidectomy		
	Grommets		
	EUA		
	Exploration of middle ear/mastoid		
	Removal of tubes		
	Other procedure		

Signature: _____

Date: _____/_____/_____

ABN: 84 085 734 992

Department of **Health** is a Smoke Free Workplace

3.5 Concluding summary for this chapter and knowledge gained from this study

The study provides a review of a referral-based, tele-otology service in the Northern Territory, Australia, from 2011 to 2019. The findings from this study demonstrate the tele-otology service's vital role to identify and manage ear disease within a population living in rural and remote areas, where the prevalence of hearing loss is high. This study also highlights areas that require further improvement. This study contributes evidence to direct future endeavours to improve telehealth services and offers insight into the role for AI to augment the limitations of the current 'store-and-forward' data collection workflow.

CHAPTER FOUR: Evaluate the inter-rater agreement among otolaryngologists to diagnose OM using clinical data collected during telehealth assessments

4.1 Preface to the chapter

This chapter presents findings from a published peer-reviewed manuscript that examined the inter-rater agreement among otolaryngologists to diagnose OM using various tiers of clinical data collected during telehealth assessments of Aboriginal and Torres Strait Islander children in Queensland, Australia. This chapter, consisting of the published paper[154], addresses specific aim #4 of this thesis as described in Chapter 1. Dissemination of this research and author contributions for this paper are described below.

4.2 Research dissemination

Published peer-reviewed paper:

Habib AR, Perry C, Crossland G, Patel H, Kong K, Whitfield B, North H, Walton J, Da Cruz M, Suruliraj A, Smith M, Harris R, Hasan Z, Gunaratne DA, Sacks R, Singh N. Inter-rater agreement between 13 otolaryngologists to diagnose otitis media in Aboriginal and Torres Strait Islander children using a telehealth approach. Int J Pediatr Otorhinolaryngol. 2023 May;168:111494. doi: 10.1016/j.ijporl.2023.111494. Epub 2023 Mar 13. PMID: 37003013
Journal impact factor: 1.7

Conference presentations:

Habib AR*, Perry C, Crossland G, Patel H, Kong K, Whitfield B, North H, Walton J, Da Cruz M, Suruliraj A, Smith M, Harris R, Hasan Z, Gunaratne DA, Sacks R, Singh N. Inter-rater agreement between 13 otolaryngologists to diagnose otitis media in Aboriginal and Torres Strait Islander children using a telehealth approach. Australian Society for

Otolaryngology – Head and Neck Surgery Annual Conference. Adelaide, Australia. 2022. [Oral presentation].

*Presenting author

4.3 Author attribution statement

I, Al-Rahim Habib, was responsible for designing the study, interpreting data, writing drafts of the manuscript, submitting the manuscript, responding to reviewers' comments, and coordinating submission and publication of the manuscript.

My co-authors, C Perry, G Crossland, H Patel and N Singh helped design the study. Data collection was performed by C Perry, G Crossland, H Patel, K Kong, B Whitfield, H North, J Walton, M Da Cruz, M Smith, R Harris, Z Hasan, and DA Gunaratne. My co-authors, C Perry, H Patel, G Crossland, H North, M Da Cruz, R Sacks, and N Singh assisted with the interpretation of the data analysis, manuscript revision, and approval of the final version of the manuscript submitted for publication. All authors have read and approved the manuscript.

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statement above is correct.

Signature:

Associate Professor Narinder Singh

4.4 Paper in published format

International Journal of Pediatric Otorhinolaryngology 168 (2023) 111494

Contents lists available at ScienceDirect

International Journal of Pediatric Otorhinolaryngology

journal homepage: www.elsevier.com/locate/ijporl

Inter-rater agreement between 13 otolaryngologists to diagnose otitis media in Aboriginal and Torres Strait Islander children using a telehealth approach

Al-Rahim Habib [a,b,*], Chris Perry [c], Graeme Crossland [d], Hemi Patel [d], Kelvin Kong [e], Bernard Whitfield [f], Hannah North [b], Joanna Walton [g], Melville Da Cruz [a,b], Anand Suruliraj [b], Murray Smith [b], Rhydian Harris [b], Zubair Hasan [b], Dakshika A. Gunaratne [b], Raymond Sacks [a], Narinder Singh [a,b]

[a] Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Camperdown, New South Wales, 2006, Australia
[b] Department of Otolaryngology – Head and Neck Surgery, Westmead Hospital, Westmead, New South Wales, 2145, Australia
[c] University of Queensland Medical School, St Lucia, Queensland, 4072, Australia
[d] Royal Darwin Hospital, Top End Health Service, Department of Health, Tiwi, Northern Territory, 0810, Australia
[e] School of Medicine and Public Health, University of Newcastle, Callaghan, New South Wales, 2308, Australia
[f] Griffith Medical School, Griffith University, Southport, Queensland, 4215, Australia
[g] Department of Otolaryngology – Head and Neck Surgery, The Children's Hospital at Westmead, Westmead, New South Wales, 2145, Australia

ARTICLE INFO

Keywords:
Otitis media
Telehealth
Telemedicine
Otoscopy
Agreement

ABSTRACT

Introduction: Telehealth programs are important to deliver otolaryngology services for Aboriginal and Torres Strait Islander children living in rural and remote areas, where distance and access to specialists is a critical factor.
Objective: To evaluate the inter-rater agreement and value of increasing levels of clinical data (otoscopy with or without audiometry and in-field nurse impressions) to diagnose otitis media using a telehealth approach.
Design: Blinded, inter-rater reliability study.
Setting: Ear health and hearing assessments collected from a statewide telehealth program for Indigenous children living in rural and remote areas of Queensland, Australia.
Participants: Thirteen board-certified otolaryngologists independently reviewed 80 telehealth assessments from 65 Indigenous children (mean age 5.7 ± 3.1 years, 33.8% female).
Interventions: Raters were provided increasing tiers of clinical data to assess concordance to the reference standard diagnosis: Tier A) otoscopic images alone, Tier B) otoscopic images plus tympanometry and category of hearing loss, and Tier C) as B plus static compliance, canal volume, pure-tone audiometry, and nurse impressions (otoscopic findings and presumed diagnosis). For each tier, raters were asked to determine which of the four diagnostic categories applied: normal aerated ear, acute otitis media (AOM), otitis media with effusion (OME), and chronic otitis media (COM).
Main outcome measures: Proportion of agreement to the reference standard, prevalence-and-bias adjusted κ coefficients, mean difference in accuracy estimates between each tier of clinical data.
Results: Accuracy between raters and the reference standard increased with increased provision of clinical data (Tier A: 65% (95%CI: 63–68%), $\kappa = 0.53$ (95%CI: 0.48–0.57); Tier B: 77% (95%CI: 74–79%), 0.68 (95%CI: 0.65–0.72); C: 85% (95%CI: 82–87%), 0.79 (95%CI: 0.76–0.82)). Classification accuracy significantly improved between Tier A to B (mean difference:12%, $p < 0.001$) and between Tier B to C (mean difference: 8%, $p < 0.001$). The largest improvement in classification accuracy was observed between Tier A and C (mean difference: 20%, $p < 0.001$). Inter-rater agreement similarly improved with increasing provision of clinical data.
Conclusions: There is substantial agreement between otolaryngologists to diagnose ear disease using electronically stored clinical data collected from telehealth assessments. The addition of audiometry, tympanometry and

* Corresponding author. Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Camperdown, New South Wales, 2006, Australia.
E-mail address: al-rahim.habib@sydney.edu.au (A.-R. Habib).

https://doi.org/10.1016/j.ijporl.2023.111494
Received 21 August 2022; Received in revised form 7 January 2023; Accepted 19 February 2023
Available online 13 March 2023

nurse impressions significantly improved expert accuracy and inter-rater agreement, compared to reviewing otoscopic images alone.

1. Introduction

Ear health is a major public health issue for Aboriginal and Torres Strait Islander people in Australia [1–3]. Indigenous children are three times more likely to have otitis media (OM) and twice as likely to have long-term hearing loss than non-Indigenous children [4]. The burden of OM is greatest for Indigenous children living in rural and remote areas, as these children are twice as likely to experience long-term hearing loss than non-remote Indigenous children [4]. Hearing loss in childhood can affect speech and language development, behaviour, academic performance and future employment opportunities as well as increasing the risk of contact with the criminal justice system [1]. Several socio-economic and cultural determinants of health are associated with ear disease in Aboriginal and Torres Strait Islander people including household overcrowding, exposure to second-hand smoke, poor nutrition and infrastructure, and limited access to health care services and treatment [5].

In Australia, most otolaryngologists practice in metropolitan areas [6]. Individuals living in rural and remote areas have limited access to specialists and can experience wait times longer than recommended for tertiary Ear, Nose and Throat (ENT) and audiology services [7]. Telehealth has been used as an alternative model of care to provide ear health services through partnerships with Aboriginal and Torres Strait Islander communities and state or territory-based hospital and health services [8–11]. Telehealth initiatives are asynchronous, with examinations conducted by Aboriginal Health Workers (AHW) and nurses who upload examination findings to secure online databases for off-site/metropolitan otolaryngologists to review at a later time, a method known as the "Store-and-forward" approach [8,10].

Telehealth ear health services have been shown to increase the number of children screened, reduce wait times for high-risk children requiring specialist review, and establish a model of care that is community led and culturally appropriate [10,12]. However, telehealth introduces its own limitations. The time difference between the initial assessment of a patient and subsequent implementation of specialist recommendations can delay the delivery of care [12,13] which may increase the potential for adverse or critical complications to develop and may contribute to loss of follow-up [8]. Telehealth for ear health screening relies on otolaryngologists to review serial otoscopy, tympanometry, and audiometry results, a tedious and mundane task devoid of typical patient interaction. In the existing telehealth model, children diagnosed with OME as per the national OM guidelines, are reviewed by their primary care physician prior to specialist referral [14]. High-risk children who experience speech and language, learning, behaviour, or developmental issues are referred earlier for expeditated review.

Previous research has demonstrated substantial intra-observer agreement to diagnose ear disease between on-site examinations and asynchronous, pre-recorded telehealth assessments. Otolaryngologists often agree with themselves in 81% of cases when comparing field versus pre-recorded examinations [15]. However, Eikelboom et al. (2005) demonstrated that intra-observer agreement between telehealth and on-site diagnoses was dependent on ear disease subtype [16]. The prevalence of chronic suppurative otitis media (CSOM) and cholesteatoma was comparable between field and telehealth assessments but dissimilar when classifying acute otitis media (AOM), otitis media with effusion (OME) or Eustachian tube dysfunction (ETD), despite reviewing otoscopy, clinical history, tympanometry, and audiometry [16].

The potential for intra- and inter-rater variability is an important consideration for telehealth programs where the workload of reviewing otoscopy and audiometric data is often shared. Otolaryngologists participating in telehealth may have different levels of experience at evaluating pediatric Indigenous ear disease. Variability in diagnostic accuracy may impact subsequent management decisions for a patient population where distance to health services is a critical factor and loss to follow up is common. Furthermore, the use of artificial intelligence algorithms to triage or diagnose ear disease requires accurate reference-standard data, typically achieved through multiple experts agreeing on diagnoses [17]. To date, there is a paucity of evidence evaluating the inter-rater agreement of ear disease diagnoses established from telehealth programs using asynchronous, store-and-forward approaches.

In addition, in rural and remote settings, audiology services are limited. Accordingly it is necessary to determine whether otoscopy alone, or in conjunction with increasing levels of additional clinical information (audiometry, tympanometry and nurse impressions), can yield an accurate diagnosis.

The purpose of this study was to evaluate the inter-rater agreement between otolaryngologists to diagnose OM using a telehealth program for Aboriginal and Torres Strait Islander children living in rural and remote areas of Australia. Additional aims included evaluating the value of various exam modalities (e.g. otoscopy, tympanometry, audiometry, frontline nurse impressions) on diagnostic performance.

2. Methods

2.1. Ethics

Ethics approval was obtained from the Children's Health Queensland Hospital and Health Service, Queensland, Australia (HREC: 20/QCHQ/68498). This research was conducted in accordance with the Helsinki Declaration [18]. This study was conducted in accordance with the Standard for Reporting Diagnostic Accuracy Studies (STARD) reporting guidelines [19].

2.2. Study design

Eighty ear assessments were randomly selected from 1610 telemedicine assessments from the Deadly Ears Program (Queensland's statewide ear health program for Aboriginal and Torres Strait Islander children). The assessments were dated between 2017 and 2020 and included the results of otoscopy, tympanometry, audiometry and nurse assessments performed by the "Deadly Ears Program" nurses and audiologists. Otoscopy images were captured using digital video otoscopes manufactured by Inline Systems (Warriewood, NSW, Australia) or Welch Allyn (Chicago, Illinois, USA). Video otoscopy is well tolerated, and most children are compliant, as real-time visual feedback improves engagement and interest. All clinical assessment data were electronically stored in Children's Health Queenland's (CHQ) secure data management system: QChild.

2.3. Reference standard

The reference standard was established prior to the study by board-certified otolaryngologists with expertise in Aboriginal and Torres Strait Islander pediatric telehealth assessment who participate in the Deadly Ears Program, based on the following information in QChild:

- patient demographics (age, gender, date of review)
- otoscopic images for right and left ears
- tympanometry (tympanogram type, static compliance, middle ear pressure, ear canal volume)
- pure-tone audiometry

- nurse impressions (including examination conditions, otoscopy findings, and general comments)
- treatments initiated by the nurse at the in-person clinical reviews

For each case, a sole reviewing otolaryngologist determined a diagnosis, developed a treatment and follow-up plan, and entered these into QChild for subsequent action by the Deadly Ears' nurses as well as primary care clinicians managing the childrens' health in their communities. The possible diagnostic categories were; normal aerated middle ear (no middle ear pathology), AOM with or without perforation, OME, COM (including chronic suppurative otitis media, cholesteatoma, or dry perforation), and obstructed external acoustic canal (EAC) (see Fig. 1).

2.4. Experimental index tests

To evaluate the inter-rater agreement of each reference standard diagnosis, 13 board-certified otolaryngologists (raters) were recruited to complete 3 index test experiments using tiers of increasing clinical data. The raters included 5 general otolaryngologists with Indigenous pediatric experience, 3 otologists, and 5 general otolaryngologists with limited specialist pediatric Indigenous ear disease experience. The 80 cases were randomly selected from diagnostic category subsets in the test set, to provide a mix of normal ears (n = 25), AOM (n = 14), OME (n = 23), and COM (n = 18). Cases included in this study consisted of otoscopic images where the tympanic membrane could be adequately visualised. Cases were excluded if the tympanic membrane was obstructed secondary to cerumen or foreign body. Raters were asked to independently provide the most likely diagnosis after reviewing:

Tier A) otoscopic images alone,
Tier B) otoscopic images plus tympanogram type (A, B, C, or could not test) and category of hearing loss (normal, conductive, sensorineural, mixed), or
Tier C) otoscopic images plus tympanometry (tympanogram type, static compliance, middle ear pressure, ear canal volume), pure-tone audiograms and nurse impressions (otoscopy findings and presumed diagnosis)

The same patients were used for all tiers, however the order of patients was randomised between tiers. All raters received the same data in the same order.

Identifying information was removed to protect patient privacy and confidentiality, and to ensure the raters were blinded to the patients and blinded to the reference standard diagnosis judgements provided by the original raters.

2.5. Statistical analysis

Performance between the experimental index tests was assessed by accuracy (proportion of correct diagnoses compared to the reference standard). Inter-rater agreement of diagnoses between raters and the reference standard were expressed as weighted kappa statistics (κ) to represent the proportion of agreement beyond chance, stratified by the experimental index tests of interest. An additional analysis was completed to determine the inter-rater agreement within the test group, considered as the proportion of agreement among the 13 raters (excluding the reference standard results), stratified by the provision of increasing levels of clinical data. Standard κ interpretations of inter-rater agreement were applied as previously proposed (<0.21 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, >0.80 almost perfect agreement) [20].

A sample size of 80 cases was required to test the null hypothesis that inter-rater agreement did not significantly differ between Tier A versus Tier C by a pre-determined clinically important threshold of 15%, to achieve a type 2 error rate of 80%, type error 1 error rate of 5%, and adjusting for potential missing data (30%).

As described previously, the Stuart-Maxwell test of marginal homogeneity was applied to evaluate systematic differences between the reference standard and raters' diagnoses [21,22]. The null hypothesis was defined as the marginal proportion of diagnoses established by the reference standard was equal to the marginal proportion of diagnoses provided by the raters. The student's t-test was used to compare rater-to-reference standard, inter-rater, and intra-rater accuracy and agreement, stratified by index experiment tier (e.g. Tier A, B, C). Results were considered significant with $p < 0.05$.

Statistical analysis was performed using Stata 17 (StataCorp LLC, College Station, Texas, USA, 2021).

3. Results

The study sample consisted of 80 ear assessments (40 right, 40 left) from 65 children (mean age 5.7 ± 3.1 years, 33.8% female). A summary of the demographic characteristics and ear disease distribution of the study sample is provided in Table 1.

13 board-certified otolaryngologists (raters) were recruited from

Table 1
Summary of demographic and clinical characteristics between population and study sample.

	Total cohort	Study cohort
Demographics		
Number of children	639	65
Age (years, mean ± SD)	6.5 ± 3.4	5.7 ± 3.1
Females	301 (47%)	22 (34%)
Males	338 (53%)	43 (66%)
Ear assessments		
Total	1610	80
Ear disease distribution		
Normal	756 (47%)	25 (31%)
Acute Otitis Media (with or without perforation)	48 (3%)	14 (18%)
Otitis Media with Effusion	238 (15%)	23 (29%)
Chronic Otitis Media	250 (16%)	18 (22%)
Chronic Suppurative Otitis Media	*137*	*12*
Dry Perforation	*104*	*5*
Cholesteatoma	*9*	*1*
Wax obstruction	177 (11%)	0
Not reported	141 (8%)	0

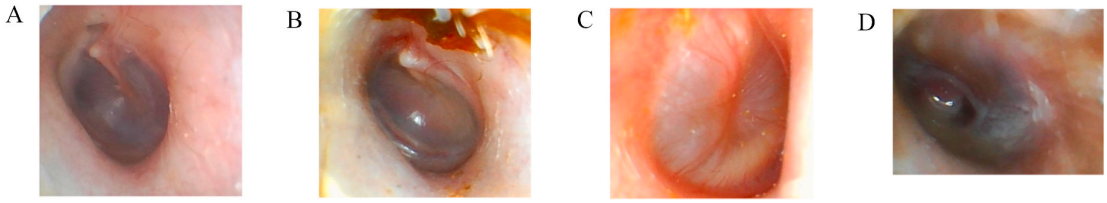


Fig. 1. Examples of otoscopic images obtained from otoscopy during telehealth assessments as per the Deadly Ears Program, Queensland, Australia. A. Normal Aerated Ear. B. Otitis Media with Effusion. C. Acute Otitis Media. D. Chronic Suppurative Otitis Media.

New South Wales, Queensland, and the Northern Territory, Australia. The raters included general otolaryngologists with limited pediatric Indigenous ear disease experience.

3.1. Agreement between raters and reference standard

Using Tier A clinical data (otoscopy images alone), raters achieved 65% accuracy and moderate agreement (κ = 0.53) to the reference standard (Table 2). The marginal proportions of diagnoses established by raters were significantly different than the reference standard (χ^2 = 26.4, p < 0.001). Accuracy achieved by all raters reviewing otoscopy alone by diagnostic subtype was 64% for normal aerated ears, 50% for AOM, 55% for OME, and 86% for COM (Supplementary Table 1). The most common errors were AOM misclassified as OME (35%), normal misclassified as OME (26%), and OME misclassified as normal (23%).

When raters were provided Tier B clinical data there was 77% accuracy and substantial agreement (κ = 0.68) between all raters and the reference standard (Table 2). Proportion of classifications were significantly different than the reference standard (χ^2 = 10.5, p = 0.014). Accuracy achieved by all raters by diagnostic subtype was 88% for normal, 47% for AOM, 70% for OME, and 84% for COM (Supplementary Table 1). The most common error was AOM misclassified as OME (38%), of which most were found to have Type B tympanograms (98%).

When raters were provided Tier C clinical data there was 85% accuracy and substantial agreement (κ = 0.79) between all raters and the reference standard (Table 2). Proportion of classifications were significantly different than the reference standard (χ^2 = 11.8, p = 0.008). Accuracy achieved by all raters by diagnostic subtype was 93% for normal, 69% for AOM, 78% for OME, and 89% for COM. The most common error was AOM misclassified as OME (19%). In these misclassification errors (Supplementary Table 2), all were found to have Type B tympanograms (100%) and conductive hearing loss (81%).

Classification accuracy improved in all raters between Tier A to B (mean difference: 12%, p < 0.001) and between Tier B to C (mean difference: 8%, p < 0.001) (Table 2). The largest improvement was observed in classification accuracy between raters reviewing Tier A versus Tier C (mean difference: 20%, range: 18%–22%, p < 0.001).

3.2. Inter-rater agreement within test group

The overall inter-rater agreement between the 13 otolaryngologists tested (excluding the reference standard results) improved between Tier A (57%, κ = 0.42), Tier B (74%, κ = 0.64), and Tier C (82%, κ = 0.75) (Table 3). Inter-rater agreement significantly improved between Tier A to B (mean difference: 17%, p < 0.001) and Tier A to C (mean difference: 25%, p < 0.001).

Supplementary Table 3 summarises inter-rater agreement by diagnostic category across each Tier of clinical data. Inter-rater agreement increased from Tier A to Tier B, to Tier C. Using Tier C data, near perfect inter-rater agreement was found to diagnose COM (κ = 0.86) and normal aerated ears (κ = 0.86). Moderate inter-rater agreement was found to classify AOM (κ = 0.43) and OME (κ = 0.57).

4. Discussion

The use of otoscopy as a single modality yielded 'fair' agreement with congruent classifications in nearly half of the assessments. However, performance varied by diagnostic category. Based on otoscopy alone, the inter-rater agreement was 'substantial' for COM and 'fair' for normal, AOM and OME, relative to the previously established reference standard [20]. Inter-rater agreement significantly improved when raters were provided otoscopy with additional clinical and audiometric findings. Between Tier A to C, inter-rater agreement to classify normal, aerated ears improved from 'fair' to 'near-perfect' agreement, AOM and OME improved from 'fair' to 'moderate', and COM improved from 'substantial' to 'near-perfect' agreement.

To diagnose OME, accuracy among raters improved from 55% using otoscopy alone, to 70% with the inclusion of tympanogram type and category of hearing loss. Further improvements to diagnose OME were found as raters achieved 78% accuracy and 'substantial' concordance with the reference standard when additionally provided audiometry and nurse impressions. Raters correctly identified normal ears in 64% of cases if using otoscopy alone, achieving 'moderate' agreement with the reference standard. Although, concordance improved substantially with the addition of tympanometry and audiometry (Table 2). It is important to note that raters achieved 'substantial' accuracy to diagnose COM when reviewing otoscopy with or without audiometry and nurse findings.

As demonstrated earlier, accuracy improved when raters were provided otoscopy with clinical and audiometric findings. This suggests that a comprehensive ear health and hearing assessment improves the accuracy of OM diagnoses. Although intuitive, the availability of audiological resources in rural and remote settings may impact the likelihood of otolaryngologists receiving complete ear examinations. For example, audiologist services may not be readily available due to logistical or staffing limitations and therefore ear assessments may not include pure-tone audiograms. In these circumstances, otolaryngologists may only receive otoscopic images, potentially contributing to a less accurate OM diagnosis. Findings from this study demonstrate that raters using otoscopy alone are more likely to identify COM with accuracy comparable to scenarios with complete ear assessments. This may be due to overt pathology, such as large TM perforations, with or without active discharge, or cholesteatomas, which tend to be identified more readily with otoscopy than audiometry findings.

However, this may not be relevant to differentiating normal aerated ears from AOM or OME. Using otoscopy alone, raters correctly identified normal, aerated ears in approximately 2 out of 3 assessments and AOM or OME in approximately 1 out of 2 assessments. These findings highlight the value of audiometry and on-site assessments when differentiating between OM diagnoses that appear phenotypically similar (as

Table 2
Comparison of congruence between raters and the reference standard by index tier.

	Tier A[1]		Tier B[2]		Tier C[3]		Tier A vs. B		Tier A vs C		Tier B vs. C	
	Accuracy (%)	Adjusted κ	Accuracy (%)	Adjusted κ	Accuracy (%)	Adjusted κ	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
All Raters (n = 13) vs reference standard	65% (95% CI: 63–68%)	0.53 (95% CI: 0.48–0.57)	77% (95% CI: 74–79%)	0.68 (95% CI: 0.65–0.72)	85% (95% CI: 82–87%)	0.79 (95% CI: 0.76–0.82)	12% (95% CI: 8–15%)	<0.001	20% (95% CI: 16–23%)	<0.001	8% (95% CI: 5–11%)	<0.001

Legend: Agreement – proportion of agreement between raters and reference standard, Adjusted κ – prevalence-and-bias-adjusted linearly weighted.
Tier A - Raters review otoscopic images only.
Tier B - Raters review otoscopic images with tympanogram type and category of hearing loss.
Tier C - Raters review otoscopic images with tympanometry (tympanogram type, static compliance, middle ear pressure, ear canal volume), pure-tone audiograms and nurse impressions (otoscopy findings and presumed diagnosis).

Table 3
Comparison of inter-rater agreement of 13 rates by index tier.

	Tier A[1]		Tier B[2]		Tier C[3]		Tier A vs. B		Tier A vs C		Tier B vs. C	
	Agreement (%)	Adjusted κ	Agreement (%)	Adjusted κ	Agreement (%)	Adjusted κ	Mean difference	p-value	Mean difference	p-value	Mean difference	p-value
All Raters (n = 13)	57% (95%CI: 50–63%)	0.42 (95%CI: 0.33–0.50)	74% (95%CI: 67–80%)	0.64 (95%CI: 0.56–0.73)	82% (95%CI: 75–88%)	0.75 (95%CI: 0.66–0.84)	17% (95%CI: 11–23%)	<0.001	25% (95%CI: 19–31%)	<0.001	8% (95%CI: 4–13%)	0.001

Legend: Agreement – proportion of agreement between 13 raters, Adjusted κ – prevalence-and-bias-adjusted linearly weighted.
Tier A - Raters review otoscopic images only.
Tier B - Raters review otoscopic images with tympanogram type and category of hearing loss.
Tier C - Raters review otoscopic images with tympanometry (tympanogram type, static compliance, middle ear pressure, ear canal volume), pure-tone audiograms and nurse impressions (otoscopy findings and presumed diagnosis).

demonstrated by the frequent misclassification error between AOM and OME). Without the adjunct of audiometry and nurse findings, otolaryngologists may be less likely to achieve an accurate and reliable diagnosis for non-chronic OM. The difference in performance may reflect the value of frontline, on-site assessments to accurately recognise AOM. The presumptive AOM diagnoses established by the nurses conducting the ear assessment may consider a child's disposition during the ear exam and clinical history reported by their carer. The telemedicine model of care used in this study did not record symptoms as discreet categories. However, this may have been intrinsically considered in the nurses' presumptive diagnosis and in-person impressions. In these settings, a reliable clinical history and assessment is an important component of diagnosing AOM in telehealth programs. Under-recognising AOM at the initial point-of-contact may have downstream impact on the final diagnosis established by otolaryngologists. Without context of the child's symptoms and disposition, otolaryngologists may over emphasise the value of tympanometry to favour a diagnosis of OME rather than AOM. Raters were more likely to misclassify OME as normal, healthy ears and vice versa, when reviewing otoscopy alone. With the addition of tympanometry and audiometry, the misclassification rates substantially reduced.

This study has limitations. Firstly, the methods used to validate the reference standard for ear disease diagnoses from telemedicine reflected the current standard of care. In this study, demographic characteristics such as age, gender, and community, were excluded to maintain privacy and confidentiality. It is plausible that by excluding these factors, raters may have been disadvantaged at establishing a diagnosis. Age is an important diagnostic consideration, as 1 in 2 Aboriginal and Torres Strait Islander children under the age of 30 months are likely to present with OM [23]. Furthermore, the reference standard was based on the telemedicine model of care with a single specialist review and did not include independent, in-person review by otolaryngologists. It is possible that on-site specialist consultations may contribute to a different diagnostic outcome. Smith et al. (2006) demonstrated substantial agreement between on-site assessments and a consensus panel of otolaryngologists reviewing pre-recorded ear examinations [15]. Of disagreements, 12% were related to differences in the interpretation of clinical findings and 10% were related to differences in the reported history [15]. OME is a challenging condition to detect, and pneumatic otoscopy is an important method to assess TM mobility [24,25]. The addition of pneumatic otoscopy could enhance diagnostic accuracy. Pichichero et al. (2002 and 2005) demonstrated that otolaryngologists achieved 35–88% accuracy to correctly identify OME based on still frames and a 10-s video of pneumatic otoscopy, respectively [24,26]. However, the authors did not include clinical evaluations or audiometry results. For the current study it is important to note that pneumatic otoscopy or video recordings were not routinely collected in the current telemedicine model of care.

Agreement between the 13 raters alone, excluding the original reference standard results, was similarly significant and improved with the increasing provision of clinical information (Table 3). The low numbers within the 3 raters subgroups (otolaryngologists with and without Indigenous pediatric experience and otologists) prevented subgroup analysis and this is worth of further investigation in future studies.

The two key findings of this study are that, for telehealth assessment in Indigenous children, there is significant concordance between raters and that otoscopy alone is an inferior modality for diagnosing AOM and OME, important conditions that require urgent and semi-urgent intervention, respectively. The addition of audiometry, tympanometry and in-person assessment significantly improves accuracy and inter-rater agreement. This has important clinical implications in the development of telehealth programs and resourcing telehealth facilities. The high concordance found implies that otolaryngologists can establish congruent diagnoses using asynchronously performed telehealth assessments. An increased pool of otolaryngologists could significantly

reduce delays in the telehealth process, resulting in marked improvements in patient outcomes. Future research is warranted to explore the potential of enhancing audiological services in rural and remote settings, given their demonstrated value in improving diagnostic accuracy. The evaluation of alterative audiological devices (e.g. tablet- or mobile-based audiometry) compared to formal pure-tone audiograms may be a valuable addition to the capabilities of frontline practitioners to investigate ear health and hearing in resource limited settings. Augmenting otoscopy with artificial intelligence-based algorithms may provide added benefits at reducing inter-rater variability, resource requirements and time commitments associated with telehealth screening [17,27,28].

5. Conclusion

Telehealth is an important component of ear health assessments for Aboriginal and Torres Strait Islander children living in rural and remote areas of Australia. The accurate diagnosis of AOM and OME, two critical conditions in this patient cohort, is significantly enhanced with additional data. In rural and remote settings where otolaryngology services are limited, accurate expert-level diagnoses to identify non-chronic forms of OM requires a multi-disciplinary, team-based approach. The addition of audiometry, tympanometry, clinical context, and patient disposition is critical to achieve an accurate diagnosis of OM.

Meeting information

Verbal presentation at the 2022 Australian Society of Otolaryngology – Head and Neck Surgery Annual Scientific Meeting on Saturday, June 11, 2022.

Contributors

A-RH is the first author. A-RH, CP, GC, HP, HN, and NS conceived the study idea. CP, GC, HP, BW, KK, HN, JW, MDC, AS, MS, RH, ZH, DG were involved in data collection. A-RH extracted data and performed the data analysis. A-RH wrote the first draft of the manuscript. A-RH, CP, HP, GC, HN, MDC, RS, and NS interpreted the data analysis and critically revised the manuscript. A-RH and NS are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. All authors read and approved the final manuscript.

Funding

This project was funded by the Garnett Passe and Rodney Williams Memorial Foundation (GPRWMF) and the Avant Foundation Doctor-in-Training research grant. The views expressed in this work are those of the authors and do not represent the views of the GPRWMF and Avant Mutual.

Ethical approval

Ethics approval was obtained from the Children's Health Queensland Hospital and Health Service, Queensland, Australia (HREC: 20/QCHQ/68498).

Data sharing

Anonymised rater-level data that support the findings of this study are available from the corresponding author on reasonable request.

Transparency statement

A-RH and NS affirm the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned (and, if relevant, registered) have been explained.

Declaration of competing interest

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf. A-RH discloses research funding from the Garnett Passe and Rodney Williams Memorial Foundation, Microsoft AI for Humanitarian Action grant, and the Avant Mutual Foundation Doctor-in-Training research grant. CP discloses conflict of interest from Australian Medical Association (Queensland) and provides consulting services for the Deadly Ears Program. MDC discloses conflict of interest from Cochlear Limited. RS is a consultant for Medtronic. NS discloses research funding from Microsoft's AI for Humanitarian Action grant, consultant for ResMed, Optinose, Nasus, GSK, and ENT Technologies, and has received conference funding from Medtronic, Karl Storz, and NeilMed. GC, HP, KK, BW, HN, JW, AS, MS, RH, ZH, DG have no disclosures or conflicts of interest to report.

Acknowledgements

We acknowledge the traditional owners of the lands on which this research project was conducted, their spiritual relationship to the land, and their elders both past, present, and emerging. We thank the Aboriginal and Torres Strait Islander people, elders, and community councils for allowing this research to be completed. We sincerely appreciate and acknowledge the assistance of Matthew Brown, Jonathan Wood, Gavin Bott, and the many clinicians, nurses, audiologists, and health workers of the Deadly Ears Program for their efforts collecting otoscopic images and managing the clinical data used for this study. We acknowledge the Garnett Passe and Rodney Williams Memorial Foundation, the Avant Foundation, and the Microsoft AI for Good Lab.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijporl.2023.111494.

References

[1] A.J. Leach, Otitis media in Australian Aboriginal children: an overview, Int. J. Pediatr. Otorhinolaryngol. 49 (1999), https://doi.org/10.1016/S0165-5876(99)00156-1. S173–S178.
[2] A. Leach, P. Morris, Otitis Media and Hearing Loss Among Aboriginal and Torres Strait Islander Children : a Research Summary Australian Parliament ' S Standing Committee on Health , Aged Care and Sport Public Hearing in Reference to the Inquiry into the Hearing Health and Wel, 2017, pp. 1–18.
[3] World Health Organization - Child and Adolescent Health and Development Prevention of Blindness and Deafness, Chronic suppurative otitis media: burden of illness and management options, Geneva, Switzerland, https://www.who.int/pbd/publications/Chronicsuppurativeotitis_media.pdf, 2004.
[4] Australian Institute of Health and Welfare, Indigenous hearing health, Canberra, Australia, https://www.aihw.gov.au/reports/australias-health/indigenous-hearing-health, 2020.
[5] J. Jervis-Bardy, L. Sanchez, A.S. Carney, Otitis media in indigenous australian children: review of epidemiology and risk factors, J. Laryngol. Otol. 128 (2014), https://doi.org/10.1017/S0022215113003083.
[6] G. Shein, K. Hunter, K. Gwynne, S. Tobin, K. Kong, M. Lincoln, T. Lee, J. Jr, J. Caswell, A.J. Saxby, The O . P . E . N . Survey : outreach projects in ear , nose and throat (ENT) in New South Wales, Aust. J. Oto Laryngol. 2 (2019), https://doi.org/10.21037/ajo.2019.04.01.
[7] H. Gunasekera, P.S. Morris, J. Daniels, S. Couzos, J.C. Craig, 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 45 (2009) 425–430. http://www.ncbi.nlm.nih.gov/pubmed/19722295. (Accessed 20 February 2019).
[8] G. Elliott, A.C. Smith, M.E. Bensink, C. Brown, C. Stewart, C. Perry, P. Scuffham, The feasibility of a community-based mobile telehealth screening service for aboriginal and Torres Strait Islander children in Australia, Telemedicine and E-Health 16 (2010) 950–956, https://doi.org/10.1089/tmj.2010.0045.
[9] A.C. Smith, N.R. Armfield, W.I. Wu, C.A. Brown, C. Perry, A mobile telemedicine-enabled ear screening service for Indigenous children in Queensland: activity and

outcomes in the first three years, J. Telemed. Telecare 18 (2012) 485–489, https://doi.org/10.1258/jtt.2012.GTH114.
[10] A.C. Smith, C. Brown, N. Bradford, L.J. Caffery, C. Perry, N.R. Armfield, Monitoring ear health through a telemedicine-supported health screening service in Queensland, J. Telemed. Telecare 21 (2015) 427–430, https://doi.org/10.1177/1357633X15605407.
[11] A.C. Smith, N.R. Armfield, M.G. Coulthard, M.L. Williams, L.J. Caffery, Queensland telepaediatric service: a review of the first 15 Years of service, Front Digit Health 2 (2020), https://doi.org/10.3389/fdgth.2020.587452.
[12] A.G. Graham, R. Macniven, K. Kong, K. Gwynne, Effectiveness of ear , nose and throat outreach programmes for Aboriginal and Torres Strait Islander Australians : a systematic review, BMJ Open 10 (2020), e038273, https://doi.org/10.1136/bmjopen-2020-038273.
[13] A.C. Smith, Telemedicine: challenges and opportunities, Expet Rev. Med. Dev. 4 (2007) 5–7, https://doi.org/10.1586/17434440.4.1.5.
[14] A.J. Leach, P.S. Morris, H.L.C. Coates, S. Nelson, S.J. O'Leary, P.C. Richmond, H. Gunasekera, S. Harkus, K. Kong, C.G. Brennan-Jones, S. Brophy-Williams, K. Currie, S.K. Das, D. Isaacs, K. Jarosz, D. Lehmann, J. Pak, H. Patel, C. Perry, J. S. Reath, J. Sommer, P.J. Torzillo, Otitis media guidelines for Australian Aboriginal and Torres Strait Islander children: summary of recommendations, Med. J. Aust. 214 (2021) 228–233, https://doi.org/10.5694/mja2.50953.
[15] A.C. Smith, C. Perry, J. Agnew, R. Wootton, Accuracy of pre-recorded video images for the assessment of rural indigenous children with ear, nose and throat conditions, J. Telemed. Telecare 12 (2006) 76–80, https://doi.org/10.1258/135763306779380138.
[16] R.H. Eikelboom, M.N. Mbao, H.L. Coates, M.D. Atlas, M.A. Gallop, Validation of tele-otology to diagnose ear disease in children, Int. J. Pediatr. Otorhinolaryngol. 69 (2005) 739–744, https://doi.org/10.1016/j.ijporl.2004.12.008.
[17] A.R. Habib, M. Kajbafzadeh, Z. Hasan, E. Wong, H. Gunasekera, C. Perry, R. Sacks, A. Kumar, N. Singh, Artificial intelligence to classify ear disease from otoscopy: a systematic review and meta-analysis, Clin. Otolaryngol. 47 (2022) 401–413, https://doi.org/10.1111/coa.13925.
[18] World Medical Association, World medical association declaration of Helsinki: ethical principles for medical research involving human subjects, JAMA 310 (2013) 2191–2194, https://doi.org/10.1001/jama.2013.281053.
[19] J.F. Cohen, D.A. Korevaar, D.G. Altman, D.E. Bruns, C.A. Gatsonis, L. Hooft, L. Irwig, D. Levine, J.B. Reitsma, H.C.W. de Vet, P.M.M. Bossuyt, STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration, BMJ Open 6 (2016) 1–17, https://doi.org/10.1136/bmjopen-2016-012799.
[20] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159–174.
[21] H. Gunasekera, H.M. Miller, L. Burgess, S. Chando, S.L. Sheriff, J.D. Tsembis, K. M. Kong, H.L.C. Coates, J. Curotta, K. Falster, P.B. McIntyre, E. Banks, N.J. Peter, J. C. Craig, Agreement between diagnoses of otitis media by audiologists and otolaryngologists in Aboriginal Australian children, Med. J. Aust. 209 (2018) 29–35, https://doi.org/10.5694/mja18.00249.
[22] A. Stuart, A test for homogeneity of the marginal distributions in a two-way classification, Biometrika 42 (1955) 412–416.
[23] P.S. Morris, A.J. Leach, P. Silberberg, G. Mellon, C. Wilson, E. Hamilton, J. Beissbarth, Otitis media in young Aboriginal children from remote communities in Northern and Central Australia: a cross-sectional survey, BMC Pediatr. 5 (2005) 1–10, https://doi.org/10.1186/1471-2431-5-27.
[24] M.E. Pichichero, M.D. Poole, Comparison of performance by otolaryngologists, pediatricians, and general practioners on an otoendoscopic diagnostic video examination, Int. J. Pediatr. Otorhinolaryngol. 69 (2005) 361–366, https://doi.org/10.1016/j.ijporl.2004.10.013.
[25] R.M. Rosenfeld, J.J. Shin, S.R. Schwartz, R. Coggins, L. Gagnon, J.M. Hackell, D. Hoelting, L.L. Hunter, A.W. Kummer, S.C. Payne, D.S. Poe, M. Veling, P.M. Vila, S.A. Walsh, M.D. Corrigan, Clinical Practice Guideline: Otitis Media with Effusion (Update), vol. 154, Otolaryngology - Head and Neck Surgery, United States), 2016, pp. S1–S41, https://doi.org/10.1177/0194599815623467.
[26] M.E. Pichichero, Assessing diagnostic accuracy and tympanocentesis skills of South African physicians in management of otitis media [7], S. Afr. Med. J. 92 (2002) 137–138.
[27] A.R. Habib, G. Crossland, H. Patel, E. Wong, K. Kong, H. Gunasekera, B. Richards, L. Caffery, C. Perry, R. Sacks, A. Kumar, N. Singh, An artificial intelligence computer-vision algorithm to triage otoscopic images from Australian aboriginal and Torres Strait Islander children, Otol. Neurotol. 43 (2022) 481–488, https://doi.org/10.1097/MAO.0000000000003484.
[28] A. Mothershaw, A.C. Smith, C.F. Perry, C. Brown, L.J. Caffery, Does artificial intelligence have a role in telehealth screening of ear disease in Indigenous children in Australia? Aust. J. Oto Laryngol. 4 (2021) https://doi.org/10.21037/ajo-21-14.

4.5 Concluding summary for this chapter and knowledge gained from this study

This study evaluated the diagnostic accuracy and inter-rater agreement among otolaryngologists performing telehealth ear assessments for Aboriginal and Torres Strait Islander children living in rural and remote areas. The combination of clinical, otoscopic and audiometric data significantly improved diagnostic accuracy and inter-rater agreement compared to otoscopy alone. These findings highlight the importance of a multidisciplinary approach to collect clinical history and hearing test results, in addition to otoscopic images for otolaryngologists to achieve an accurate diagnosis of middle ear disease, including acute otitis media and otitis media with effusion.

Achieving precision in telehealth ear assessments requires a multidisciplinary approach, involving the collaboration of nurses and audiologists to conduct face-to-face evaluations. As demonstrated by this study, performance among otolaryngologists improves if they are provided a comprehensive array of clinical data, encompassing patient medical histories, audiometric evaluations, and otoscopic images. This is especially important to differentiate between cases such as a normal middle ear, acute otitis media and otitis media with effusion, where otoscopic images are more commonly misclassified.

Integrating AI into the telehealth clinical workflow offers the potential to optimise the synthesis of clinical data to support nurses and audiologists in face-to-face settings. This has the potential to enhance interpretation and triage decisions where otolaryngologists may not be readily available for immediate point-of-care assessments. This may have a greater impact on differentiating between middle ear conditions where the tympanic membrane is intact, versus identifying chronic forms of otitis media with tympanic membrane perforations.

4.6 References

1. Habib AR, Perry C, Crossland G, Patel H, Kong K, Whitfield B, North H, Walton J, Da Cruz M, Suruliraj A, Smith M, Harris R, Hasan Z, Gunaratne DA, Sacks R, Singh N. Inter-rater agreement between 13 otolaryngologists to diagnose otitis media in aboriginal and Torres Strait islander children using a telehealth approach. Int J Pediatr Otorhinolaryngol. 2023 Mar;111494.

CHAPTER FIVE: Investigate the generalisability and performance of deep learning algorithms in detecting middle ear disease using otoscopy

5.1 Preface to the chapter

This chapter presents findings from a published peer-reviewed manuscript that evaluated the generalisability and performance of image classification algorithms to identify middle ear disease using otoscopic images. Deep learning methods were used to train models to differentiate normal versus abnormal otoscopic images using data retrieved from 3 open-access data sources. Internal performance was evaluated within each data source and external performance was evaluated across the data sources to explore generalisability. The published paper forms this chapter[155] and addresses specific aim #4 of this thesis. Dissemination of this research and author contributions for this paper are described below.

5.2 Research dissemination

Published peer-reviewed paper

Habib AR, Xu Y, Bock K, Mohanty S, Sederholm T, Weeks WB, Dodhia R, Ferres JL, Perry C, Sacks R, Singh N. Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. Sci Rep. 2023 Apr 1;13(1):5368. doi: 10.1038/s41598-023-31921-0. PMID: 37005441; PMCID: PMC10067817.

Journal impact factor: 4.6

Conference presentations:

Habib AR*, Xu Y, Bock K, Mohanty S, Sederholm T, Weeks WB, Dodhia R, Ferres JL, Perry C, Sacks R, Singh N. Artificial intelligence for otoscopy. Australian Society for Otolaryngology – Head and Neck Surgery Annual Conference. Brisbane, Australia. 2023. [Oral presentation].

Habib AR*, Xu Y, Bock K, Mohanty S, Sederholm T, Weeks WB, Dodhia R, Ferres JL, Perry C, Sacks R, Singh N. Artificial intelligence for otoscopy. Indigenous Ear Workshop at the Australian Society for Otolaryngology – Head and Neck Surgery Annual Conference. Adelaide, Australia. 2022. [Oral presentation].

Habib AR*, Xu Y, Bock K, Mohanty S, Sederholm T, Weeks WB, Dodhia R, Ferres JL, Perry C, Sacks R, Singh N. Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. Australian Society for Otolaryngology – Head and Neck Surgery Annual Conference. Adelaide, Australia. 2022. [Poster presentation]. Awarded winner of poster competition prize.

Habib AR, Xu Y*, Bock K, Mohanty S, Sederholm T, Weeks WB, Dodhia R, Ferres JL, Perry C, Sacks R, Singh N. Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. Artificial Intelligence in Medicine (AIMed) Global Summit. San Francisco, USA. 2022. [Oral presentation]. Awarded oral presentation prize.

*Presenting author

5.3 Author attribution statement

I, Al-Rahim Habib, was responsible for designing the study, interpreting data, writing drafts of the manuscript, submitting the manuscript, responding to reviewers' comments, and coordinating submission and publication of the manuscript.

My co-authors, Y Xu, K Bock, S Mohanty, T Sederholm, R Dodhia, JL Ferres, C Perry, R Sacks and N Singh conceived the idea. Myself and my co-authors Y Xu, K Bock, and S Mohanty were involved in data collection. Myself and Y Xu extracted the data and performed the data analysis. My co-authors Y Xu, K Bock, S Mohanty, T Sederholm, WB Weeks, R Dodhia, JL Ferres, C Perry, R Sacks and N Singh assisted with the interpretation of the data analysis, manuscript revision, and approval of the final version of the manuscript submitted for publication. All authors have read and approved the manuscript.

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statement above is correct.

Signature:

Associate Professor Narinder Singh

5.4 Paper in published format

scientific reports

Check for updates

OPEN

Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy

Al-Rahim Habib[1,2]✉, Yixi Xu[3], Kris Bock[4], Shrestha Mohanty[6], Tina Sederholm[3], William B. Weeks[3], Rahul Dodhia[3], Juan Lavista Ferres[3], Chris Perry[5], Raymond Sacks[1] & Narinder Singh[1,2]

To evaluate the generalizability of artificial intelligence (AI) algorithms that use deep learning methods to identify middle ear disease from otoscopic images, between internal to external performance. 1842 otoscopic images were collected from three independent sources: (a) Van, Turkey, (b) Santiago, Chile, and (c) Ohio, USA. Diagnostic categories consisted of (i) normal or (ii) abnormal. Deep learning methods were used to develop models to evaluate internal and external performance, using area under the curve (AUC) estimates. A pooled assessment was performed by combining all cohorts together with fivefold cross validation. AI-otoscopy algorithms achieved high internal performance (mean AUC: 0.95, 95%CI: 0.80–1.00). However, performance was reduced when tested on external otoscopic images not used for training (mean AUC: 0.76, 95%CI: 0.61–0.91). Overall, external performance was significantly lower than internal performance (mean difference in AUC: −0.19, p ≤ 0.04). Combining cohorts achieved a substantial pooled performance (AUC: 0.96, standard error: 0.01). Internally applied algorithms for otoscopy performed well to identify middle ear disease from otoscopy images. However, external performance was reduced when applied to new test cohorts. Further efforts are required to explore data augmentation and pre-processing techniques that might improve external performance and develop a robust, generalizable algorithm for real-world clinical applications.

It estimated that over 1.5 billion people in the world live with hearing loss, representing nearly 20% of the global population[1]. The prevalence of hearing loss is expected to rise to over 2.5 billion by 2050[1]. In the US, it is estimated that the overall cost of deafness and hearing loss amounts to US$980 billion annually[2]. Direct and indirect costs related to hearing loss are comprised of health sector costs, educational support, loss of productivity, and societal costs[3,4]. In the global pediatric population, 34 million children experience hearing loss, of which, 60% can be attributed to preventable causes[5]. The WHO estimates that over one-third of preventable childhood hearing loss is attributed to infections including mumps, rubella, meningitis, measles and middle ear infections[5].

Otoscopy is an important component of ear health and hearing assessments to identify ear infections. Otoscopy is performed routinely by multiple healthcare workers including medical students, nurses, general practitioners, emergency medicine physicians, paediatricians, audiologists, and otolaryngologists. However, the ability to accurately interpret otoscopic findings varies by user experience[6–10]. Pichichero et al. (2005) demonstrated differences between general practitioners, paediatricians, and otolaryngologists to recognise tympanic membrane (TM) abnormalities[8]. Significant performance differences were found when non-otolaryngologists were asked to differentiate between acute otitis media (AOM), otitis media with effusion (OME), or retracted TMs[8]. These findings are consistent with recent studies that demonstrate that diagnostic accuracy is reduced when non-experts must differentiate between multiple ear disease sub-types[6,11]. Otolaryngologists significantly outperformed non-otolaryngologists to identify ear disease sub-types and important pathological characteristics (e.g. tympanic membrane perforations, attic retraction, or myringitis)[6,8].

[1]Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia. [2]Department of Otolaryngology, Head and Neck Surgery, Westmead Hospital, Sydney, NSW, Australia. [3]AI for Good Lab, Microsoft, Redmond, WA, USA. [4]Azure FastTrack Engineering, Brisbane, QLD, Australia. [5]University of Queensland Medical School, Brisbane, QLD, Australia. [6]Microsoft, Redmond, WA, USA. ✉email: al-rahim.habib@sydney.edu.au

Accurately recognising middle ear diseases is important for early management, prevention of long-term hearing loss, can improve speech and language development and reduce healthcare costs[12]. However, the subjective and often inconsistent nature of traditional otoscopy makes accurate interpretation difficult. To address this challenge, deep learning -based image classification algorithms have been developed using machine learning techniques[13–35]. Traditional machine learning algorithms rely on statistical models and pre-defined features, whereas deep learning algorithms use neural networks to learn features and extract patterns directly from raw data[13,36,37]. Deep learning methods have the potential to improve early ear disease detection by bridging the gap between expert and non-expert performance and enhancing the performance of previously developed machine learning models for otoscopy[38].

Previous studies have explored image classification algorithms for otoscopy using deep learning approaches[13–35]. In a recent systematic review and meta-analysis, a pooled analysis of 14 studies achieved 91% accuracy in differentiating between normal versus abnormal otoscopic images[39]. From a pooled analysis of three studies, multi-classification algorithms achieved 98% to differentiate AOM, OME, and normal aerated ears without middle ear pathology[39]. Despite high levels of accuracy reported, the binary and multi-classification analyses yielded substantial analyses heterogeneity between studies[39]. Variability between models to classify ear disease subtypes suggests performance may not be generalisable to new test environments.

The inability to generalize to environments beyond those used for training could limit the clinical utility and sustainability of deep learning algorithms for otoscopy. Variability in data source, capture devices, ear disease sub-type, and deep learning methods may impact classification performance in clinical encounters different to those used for training. To date, an exploration of the generalisability of deep learning algorithms for otoscopy has not been evaluated.

Deep learning techniques have been shown to excel in tasks such as image classification, with the ability to classify images with high accuracy and to generalize to new data[40,41]. Unlike traditional machine learning techniques, deep learning algorithms can automatically extract relevant features from images, reducing the need for manual feature engineering. This makes deep learning algorithms less sensitive to variations in data quality and pre-processing steps, and more robust to changes in data source, capture devices, and ear disease subtypes.

The purpose of this study was to construct a deep learning-based image classifier capable of differentiating between normal and abnormal otoscopic images using three independently collected otoscopy databases, evaluate the generalisability of the algorithm on images not used for training, and explore the optimal convolutional neural network (CNN) and deep learning approach to optimise performance.

Methods

Ethics. This research study was conducted in accordance with the Helsinki Declaration, the Standard for Reporting Diagnostic Accuracy Studies (STARD) and the Consolidated Standards of Reporting Trials for interventions involving artificial intelligence (CONSORT-AI) reporting guidelines[42,43]. This study utilised three open-access otoscopic image databases that were collected after obtaining Institutional Review Board (IRB) approval from each of the respective human research ethics committees and published online for public access, as indicated by each of the authorship groups (Bitlis Eren University Ethics Committee[34,44], University of Chile Scientific and Research Ethics Committee[18], Ohio State University Institutional Review Board[45]).

Data sources. Otoscopic images ascertained for the purposes of this study can be found elsewhere for online public access, as per the authorship groups[18,34,45]. In Turkey (Cohort A), 848 otoscopic images were collected at the Özel Van Akdamar Hospital using digital video otoscopes (unspecified brand and model) with ground-truth classified by two otolaryngologists and one paediatrician as: normal, acute otitis media (AOM), and chronic otitis media (COM) consisting of chronic suppurative otitis media[34]. In Chile (Cohort B), 540 otoscopic images were collected at the University of Chile Clinical Hospital using the Firefly DE500 digital video otoscope (Firefly Global, Belmont, MA, US. 2022) with ground-truth classified by 1 otolaryngologist as: normal or COM[18]. In the United States (Cohort C), 454 otoscopic images were collected at Nationwide Children's Hospital and Ohio State University using the Jedmed Horus+ (Jedmed, St Louis, MO, US, 2022) with ground-truth classified by 1 otolaryngologist as: normal or otitis media with effusion (OME)[45]. Otoscopic images where the TM could not be visualised were excluded (e.g., obstruction with wax or foreign body).

Algorithm development. Deep learning-based binary class algorithms were developed to classify normal or abnormal otoscopic images using Cohort A, B and C independently. Abnormal sub-classes included AOM, OME, and COM. Pre-existing ground-truth labels were available for each otoscopic image from the online data sources[18,34,45]. We used four different neural net architectures including ResNet-50 (comprised of 50 layers and 23 million parameters)[46], VGGNet-16 (comprised of 16 layers and 134 million parameters)[47], DenseNet-161 (comprised of 16 layers and 26 million parameters)[48], and Vision Transformer (comprised of 12 layers and 85 million parameters)[49]. Images were resized to 224 by 224 pixels. Each model was trained for 200 epochs using the Stochastic Gradient Descent (SGD) with a learning rate = 0.003 as the optimizer. The loss function is cross entropy. Training was completed using the Microsoft Azure Machine Learning Studio (Redmond, Washington, USA). All the models used the weights of models trained on the ImageNet dataset as their initial weights. Raw data was used for the purposes of developing the model and validation.

To assess the generalizability of a model trained on a single-source dataset, we first randomly split the dataset from one single source into training and validation and testing sets, allocating 70% for training, 10% for validation and 20% for testing (internal validation). We trained the model using the training set and selected the final model as the one at the best-performing epoch minimizing the loss on the validation set. Next, we tested the final model on the testing (internal validation) set. Then, we tested the final model on the two external validation

cohorts. To ensure the robustness and reliability of the results, we ran each experiment for three times, and the mean and standard deviation of accuracy, area under the curve (AUC), sensitivity, and specificity were computed, as shown in Table 3. This was done to account for any potential variability or instability in the results due to the stochastic nature of deep learning algorithms. The mean and standard deviation values provided a more comprehensive understanding of the performance of the models, as well as the variability of their performance across multiple runs.

To evaluate the pooled performance, we first combined all the data from the three cohorts, and then used fivefold cross-validation as follows: (1) The data was randomly split into fivefolds; (2) a model was trained using the training set consisting of three of the five folds. We selected the final model as the one at the best-performing epoch minimizing the loss on the validation set consisting of one of the five folds. We tested the final model on the remaining fold. (3) Step 2 was repeated five times until each fold had been used as a test fold exactly once. All the metrics were averaged to estimate the pooled performance.

Statistical analysis. To compare the algorithm to the reference standard for Cohorts A, B, and C, contingency tables were generated to summarise Area Under the Curve (AUC), accuracy, sensitivity, and specificity. To evaluate internal performance, models were trained and tested using otoscopic images from the same cohort (e.g., split Cohort A otoscopic images into training and test sets). To evaluate external validation performance and generalizability, models were validated on external cohorts (e.g., trained using Cohort A and validated on Cohort B and Cohort C, Fig. 1)[50]. Mean difference in AUC were calculated to compare internal and external performance. We used bootstrapping to compare the Receiver Operating Characteristics (ROC) curves, and the null hypothesis is that the external validation AUC is greater or equal to the internal validation AUC[50]. Pooled assessment was performed by combining Cohort A, B, and C and fivefold cross-validation was applied (Fig. 1).

Results

In total, 1842 otoscopic images were collected from three online resources, comprised of 1117 normal and 725 abnormal images. Table 1 summarises the distribution of normal and abnormal otoscopic images by Cohort A, B and C.

Internal performance. Table 2 summarises internal performance for each cohort. Sensitivity and specificity results are provided in the Supplementary Appendix. Models trained using Cohort A otoscopic images achieved AUC between 0.82 and 0.86 (accuracy: 80 to 84%; sensitivity: 57 to 70%; specificity 90 to 91%). Models trained using Cohort B otoscopic images achieved an AUC of 1.00 (100% accuracy, 100% sensitivity, and 100%

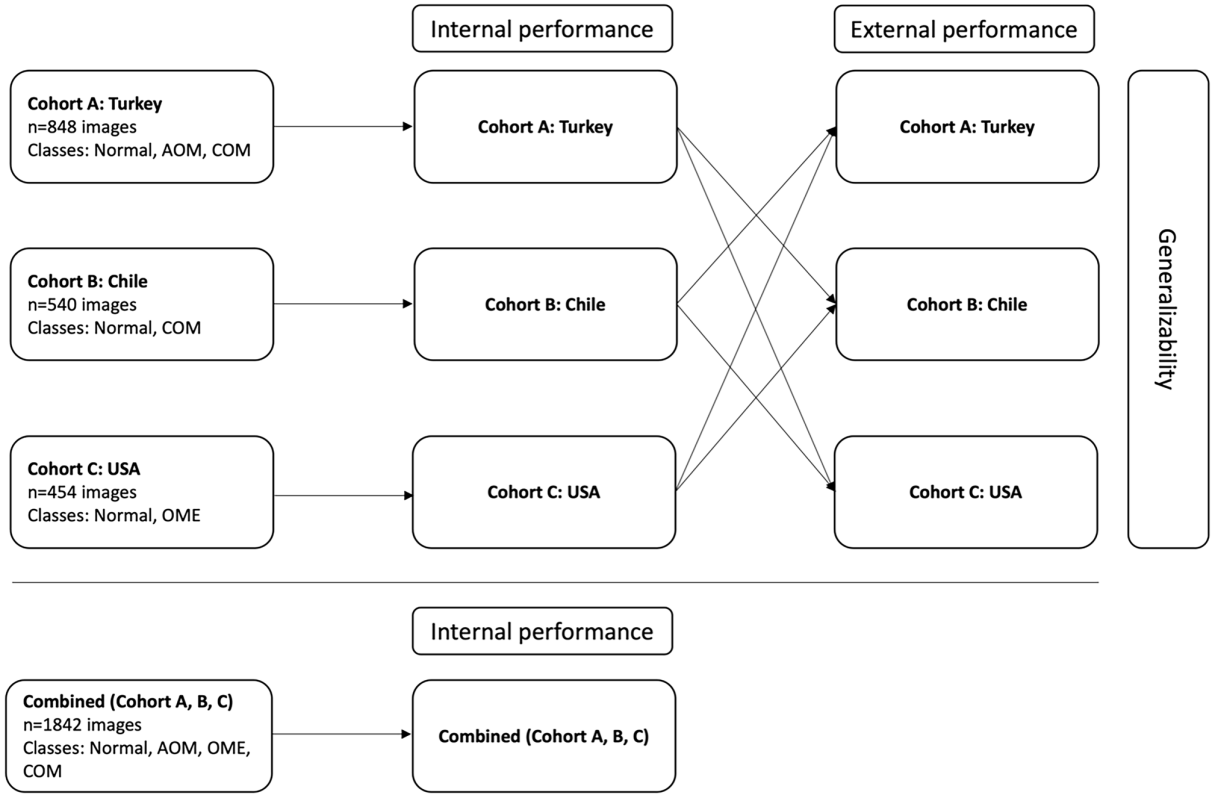


Figure 1. Flow chart of proposed study design.

	Otoscope	Image resolution (pixels)	Ground-truth	Total images	Normal images	Abnormal images	Ear disease diagnoses	Published accuracy
Cohort A (Özel Van Akdamar Hospital, Van, Turkey)	Unspecified brand/model	500 × 500 (DPI: 72)	2 otolaryngologists, 1 paediatrician	848	578	270	Normal, AOM, COM	99.4% (857 training images including AOM, OME, COM, Wax obstruction)
Cohort B (University of Chile's Hospital, Santiago, Chile)	Firefly DE500 digital video otoscope	420 × 380 (DPI: 72)	1 otolaryngologist	540	360	180	Normal, COM	93.9%
Cohort C (Nationwide Children's Hospital, Columbus, Ohio, US)	JEDMED HORUS + digital video otoscope	1397 × 982 to 441 × 355 (DPI: 72)	1 otolaryngologist	454	179	275	Normal, OME	80.6%
Combined (Cohort A, B, C)	–	–	–	1842	1117	725	Normal, AOM, OME, COM	–

Table 1. Summary of data sources. *AOM* acute otitis media, *COM* chronic otitis media, *OME* otitis media with effusion, *DPI* dots per inch.

Experiments		ResNet-50		DenseNet-161		VGG16		Vision Transformer	
Train	Validate	Acc % (SD)	AUC	Acc % (SD)	AUC	Acc % (SD)	AUC	Acc % (SD)	AUC
Cohort A	Cohort A (internal)	78 (0.02)	0.83 (0.01)	83 (0.01)	0.86 (0.01)	82 (0.01)	0.84 (0.0)	82 (0.0)	0.85 (0.02)
	Cohort B (external)	75 (0.05)	0.88 (0.01)	65 (0.04)	0.84 (0.03)	80 (0.03)	0.89 (0.02)	62 (0.06)	0.86 (0.02)
	Cohort C (external)	75 (0.01)	0.81 (0.02)	67 (0.03)	0.71 (0.05)	73 (0.03)	0.78 (0.05)	65 (0.02)	0.64 (0.06)
Cohort B	Cohort B (internal)	100 (0.0)	1.0 (0.0)	100 (0.0)	1.0 (0.0)	100 (0.0)	1.0 (0.0)	100 (0.0)	1.0 (0.0)
	Cohort A (external)	59 (0.04)	0.62 (0.02)	56 (0.07)	0.67 (0.05)	64 (0.02)	0.63 (0.01)	67 (0.03)	0.62 (0.05)
	Cohort C (external)	76 (0.02)	0.87 (0.03)	81 (0.01)	0.89 (0.01)	77 (0.01)	0.9 (0.01)	67 (0.03)	0.82 (0.04)
Cohort C	Cohort C (internal)	92 (0.01)	0.99 (0.0)	92 (0.01)	0.98 (0.0)	94 (0.01)	0.99 (0.0)	96 (0.01)	0.99 (0.0)
	Cohort A (external)	53 (0.12)	0.69 (0.01)	52 (0.04)	0.66 (0.03)	66 (0.01)	0.58 (0.02)	53 (0.11)	0.57 (0.03)
	Cohort B (external)	37 (0.02)	0.86 (0.01)	36 (0.01)	0.92 (0.01)	41 (0.03)	0.92 (0.02)	38 (0.02)	0.9 (0.01)

Table 2. Summary of internal and external performance between Cohorts to differentiate normal versus abnormal otoscopic images trained using ResNet-50, DenseNet-161, VGG16, vision transformer. *Acc* accuracy, *AUC* area under the receiver operating characteristics curve.

specificity). Models trained using Cohort C otoscopic images achieved AUC between 0.98 to 0.99 (accuracy: 91 to 95%; sensitivity: 91 to 96%; specificity: 89 to 96%).

Table 2 summarises external validation performance for each cohort. Models trained using Cohort A otoscopic images and tested on Cohort B achieved AUC between 0.80 and 0.87 (accuracy: 62 to 77%; sensitivity: 76 to 96%; specificity: 47 to 78%), while those tested on Cohort C achieved AUC between 0.61 and 0.82 (accuracy: 65 to 76%; sensitivity: 80 to 92%; specificity: 25 to 69%).

Models trained using Cohort B otoscopic images and tested on Cohort A achieved AUC between 0.60 to 0.67 (accuracy: 59 to 65%; sensitivity: 4 to 54%; specificity: 61 to 99%), while those tested on Cohort C achieved AUC between 0.87 and 0.91 (accuracy: 72 to 80%; sensitivity: 62 to 73%; specificity: 88 to 95%) (Table 2).

Models trained using Cohort C otoscopic images and tested on Cohort A achieved AUC: 0.54 to 0.68 (accuracy: 44 to 70%; sensitivity: 14 to 82%; specificity: 26 to 96%), while those tested on Cohort B achieved AUC: 0.85 to 0.94 (accuracy: 34 to 45%; sensitivity: 100%; specificity: 1 to 17%) (Table 2).

On average, external AUC was significantly lower than internal AUC (range of mean difference in AUC for DenseNet-161: 0.07–0.39). As shown in Table 3, Cohort A had the smallest difference between internal and external AUC between architectures (range of mean differences: −0.05 to 0.07).

Pooled performance. Models developed using a combination dataset of all cohorts achieved 90 to 91% accuracy (AUC: 0.96; sensitivity: 84 to 87%; specificity: 93 to 95%) (Table 4). DenseNet-161 had the highest AUC and smallest standard deviation for fivefold cross validation.

	Train	External	Mean difference in AUC (internal – external)	p-value
ResNet-50	Cohort A	Cohort B	−0.05	0.87
	Cohort A	Cohort C	0.01	0.42
	Cohort B	Cohort A	0.40	< 0.01
	Cohort B	Cohort C	0.09	< 0.01
	Cohort C	Cohort A	0.31	< 0.01
	Cohort C	Cohort B	0.14	< 0.01
DenseNet-161	Cohort A	Cohort B	0.07	0.04
	Cohort A	Cohort C	0.14	< 0.01
	Cohort B	Cohort A	0.39	< 0.01
	Cohort B	Cohort C	0.10	< 0.01
	Cohort C	Cohort A	0.36	< 0.01
	Cohort C	Cohort B	0.07	< 0.01
VGG16	Cohort A	Cohort B	−0.03	0.79
	Cohort A	Cohort C	0.12	< 0.01
	Cohort B	Cohort A	0.37	< 0.01
	Cohort B	Cohort C	0.09	< 0.01
	Cohort C	Cohort A	0.43	< 0.01
	Cohort C	Cohort B	0.05	< 0.01
Vision transformer	Cohort A	Cohort B	−0.04	0.85
	Cohort A	Cohort C	0.24	< 0.01
	Cohort B	Cohort A	0.33	< 0.01
	Cohort B	Cohort C	0.13	< 0.01
	Cohort C	Cohort A	0.45	< 0.01
	Cohort C	Cohort B	0.07	< 0.01

Table 3. Summary of mean difference in area under the curve (AUC) estimates between internal and external performance to differentiate normal versus abnormal otoscopic images trained using DenseNet-161.

Model architecture	Acc (SD)	AUC (SD)	Sen (SD)	Spec (SD)
ResNet-50	0.91 (0.02)	0.96 (0.01)	0.87 (0.04)	0.93 (0.01)
DenseNet-161	0.91 (0.02)	0.96 (0.01)	0.86 (0.04)	0.94 (0.01)
VGG16	0.90 (0.02)	0.96 (0.01)	0.84 (0.02)	0.94 (0.02)
Vision transformer	0.92 (0.01)	0.96 (0.01)	0.86 (0.01)	0.95 (0.01)

Table 4. Summary of pooled performance to differentiate normal versus abnormal otoscopic images using combined dataset (cohort A, B, and C) by pre-trained CNN architecture. *Acc* accuracy, *AUC* area under the receiver operating characteristics curve, *Sen* sensitivity, *Spec* specificity, *SD* standard deviation.

Discussion

Otoscopy is a common clinical task performed by multiple healthcare workers and practitioners with varying levels of expertise. Identifying normal anatomical landmarks and abnormal pathological processes is important in evaluating ear health. The diagnosis of ear disease relies on a combination of presenting symptoms, clinical history, tympanometry, audiometry, and otoscopic findings. AI-based algorithms have been explored in healthcare to augment existing clinical practices in order to enhance judgement and decision-making[51]. AI-based algorithms for otoscopy have achieved substantial accuracy to differentiate between normal or abnormal otoscopic images utilising pre-trained CNN architectures[52–56]. Despite achieving high performance in training and internal validation groups, substantial heterogeneity has been found between model results[39]. The differences may, in part, be attributed to the variability in image capture methods or devices, definition of diagnoses, ground truth determination, sample populations, and methodology used for algorithm development[39]. Translating performance achievements from training environments to new test cohorts is an important aspect for an AI-based model to be generalizable for routine clinical practice. Without achieving generalizable performance and real-world utility[57], it is unclear if applying AI to otoscopy practices will contribute to improving clinical practice. The purpose of this study was to evaluate the generalisability of a binary classification AI-otoscopy algorithm to classify normal or abnormal otoscopic images, using three independently collected databases and trained using standardised deep learning methods.

This study demonstrates substantial internal performance within each cohort to differentiate between normal or abnormal otoscopic images. Between CNN architectures, models developed using DenseNet-161 achieved the highest internal performance for the pooled assessment. However, when the models were applied externally

(i.e., to new validation cohorts not used for training), overall model performance was reduced. The extent of performance variability was dependent on the outcome measure assessed (e.g., accuracy, AUC, sensitivity, and specificity). Performance was reduced in most external comparisons, although the mean difference varied by outcome measure. Cohort A (Özel Van Akdamar Hospital, Van, Turkey) was found to have the smallest external performance drop in accuracy, AUC, and sensitivity. On the other hand, Cohort B (University of Chile's Hospital, Santiago, Chile) was found to have the smallest performance drop in specificity, but the greatest performance drop in AUC and sensitivity. These findings demonstrate that measuring external model performance and evaluating generalizability to new cohorts need to be interpreted relative to the outcome measure of interest. The outcome variable 'accuracy' evaluated final predictions as discrete categories, where predictions could be considered either correct or incorrect. However, the outcome variable 'AUC' measured the degree of separability and as previously demonstrated, is a better measure of algorithm performance than accuracy[58,59].

External validation performance differences between cohorts also may be attributed to intrinsic differences in the number of otoscopic images available for training, ear disease categories, and image capture devices. Cohort A had the lowest internal performance despite having both the greatest number of otoscopic images available for training and number of ear disease categories within overarching 'abnormal' classification labels. Although these factors may have contributed to less accurate internal performance due to greater heterogeneity in model learning across each category, these deficiencies were not translated to the model's ability to generalize compared to Cohort B and C.

Interestingly, models naïve to OME during training were able to achieve an AUC between 0.72 to 0.90 and to differentiate OME from normal TMs in validation cohorts. This achievement may reflect the CNN's architectural design and the feature extraction layers that were used to extract pertinent features common to normal TMs, ones that are translatable to new cohorts. However, the addition of AOM in Cohort A may have negatively impacted the CNN's ability to extract pertinent features as, often, AOM, OME, and normal TMs are phenotypically comparable and challenging to differentiate, even among experts[8,9,60,61].

In a pooled analysis of all three cohorts combined, the DenseNet-161 architecture achieved substantial performance (AUC 0.96) and exceeded the external performance from each individual cohort, respectively. This finding demonstrated that the reduction in performance between internal and external validation, could potentially and in part, be overcome by coming the data into a pooled cohort by sampling a broad collection of otoscopic images from multiple sources.

This study represents the first attempt to evaluate the generalisability of AI-otoscopy algorithms trained and tested on three unique and independently collected databases. Although the image capture devices and data collection protocols varied by each cohort, the deep learning methodology was standardised and uniform between groups. This approach revealed the potential for pre-trained CNNs to extract important features from otoscopic images used during training that could be applied universally to new images. Comparable to a healthcare worker's evolving knowledge base that improves with otoscopy experience, CNNs demonstrate the potential to identify useful patterns in otoscopic images despite differences in otoscopes, image quality, and underlying ear disease diagnoses. The patterns and features extracted by pre-trained CNNs enable broad differentiation between normal versus abnormal TMs. However, the choice of architecture can impact model performance, with more complex architectures potentially achieving better performance to capture complex features in the otoscopic images. This was found as DenseNet-161, which has the largest number of layers compared to other architectures used, achieved the highest performance to generalise between cohorts. However, more complex architectures may also require more data and computing resources to train and may be more prone to overfitting. As these architectures were pre-trained on large datasets, they can extract features from images that are relevant for many different image classification tasks.

This study has several limitations. First, this study utilises secondary data published online for open-access use. As a result, the data collection protocols, image capture devices, ground-truth definitions, and labelling formats were not standardised across cohorts. Although the models developed for this study evaluated normal versus abnormal classifications, it is plausible that differences in the ear disease definitions and potential misclassifications could influence the models' performance and overestimate its accuracy. Second, it is unclear if the ear disease diagnoses were based on otoscopic images alone or in concert with tympanometry, audiometry, pneumatic otoscopy, and clinical history. Inter-cohort variability and interpretation bias may exist in the ground-truth labels, potentially contributing to external measurement errors and misrepresenting the algorithm's generalisability. Third, ground-truth labels were not established by expert consensus. Inter-rater variability and rater bias could impact the overall reliability of ground-truth labels and adversely impact external validation.

The potential for AI-based algorithms to augment otoscopy is promising. Autonomous AI-otoscopy tools have the potential to improve triage and clinical decision-making. In settings where the prevalence of ear disease is high and access to healthcare services is limited (e.g., rural and remote areas; Indigenous communities in North America, Australia, and New Zealand; under-resourced or marginalised populations in sub-Saharan Africa, India, south-east Asia)[12,62,63], AI-otoscopy algorithms may enhance telehealth programs and empower local healthcare workers to perform otoscopy with more certainty. However, for this to be achieved, AI-otoscopy models must be accurate, reliable, generalizable, accessible in online/offline settings, sustainable, and ethnically constructed and deployed. Early studies have demonstrated that CNNs can achieve substantial performance to interpret otoscopy, although, as shown in this study, external validation performance and generalisability to new settings is limited. Further efforts are required to expand otoscopic image databases to include images from multiple countries, rural versus urban settings, various image capture devices, and include users from various experience levels to account for variability in image capture quality. Establishing standardised diagnostic definitions or consensus from expert panels is required to ensure the ground-truth labels are valid and subsequent models are trained with minimal classification bias. The consideration of data augmentation and pre-processing techniques may be explored to enhance external performance and overcome performance limitations. Otoscopy practices will

inevitably vary between clinical settings and a robust, comprehensive, and generalizable AI-otoscopy algorithm is necessary for real-world applications.

Conclusion

Internal and pooled performance results are comparable to previously reported findings. However, external performance was reduced when applied to new test cohorts. Further efforts are required to explore data augmentation and pre-processing techniques to improve external performance and develop a robust, generalizable algorithm for real-world clinical applications.

Data availability

Data that support the findings of this study are available from the corresponding author on reasonable request.

Received: 30 August 2022; Accepted: 20 March 2023
Published online: 01 April 2023

References

1. World Health Organisation. *World Report on Hearing [Internet]. Geneva; 2021.* https://www.who.int/publications/i/item/world-report-on-hearing. Accessed 28 Aug 2022 (2021).
2. World Health Organization. *Global Costs of Unaddressed Hearing Loss and Cost-Effectiveness of Interventions: A WHO Report [Internet]. Geneva; 2017.* https://apps.who.int/iris/bitstream/handle/10665/254659/9789241512046-eng.pdf. Accessed 28 Aug 2022 (2017).
3. Deloitte Access Economics. *The Social and Economic Costs of Hearing Loss in Australia [Internet].* https://apo.org.au/node/102776. Accessed 28 Aug 2022 (2017)
4. Shield, B. *Evaluation of the Social and Economic Costs of Hearing Impairment: A Report for Hear-It [Internet].* https://www.hear-it.org/sites/default/files/multimedia/documents/Hear_It_Report_October_2006.pdf. Accessed 28 Aug 2022 (2006).
5. World Health Organization. *Childhood Hearing Loss—Act Now, Here's How! [Internet]. Geneva; 2016.* https://apps.who.int/iris/handle/10665/204507. Accessed 28 Aug 2022 (2016).
6. Cha, D. *et al.* Differential biases and variabilities of deep learning-based artificial intelligence and human experts in clinical diagnosis: Retrospective cohort and survey study. *JMIR Med. Inform.* **9**(12), 33049 (2021).
7. Kleinman, K. *et al.* Evaluation of digital otoscopy in pediatric patients: A prospective randomized controlled clinical trial. *Am. J. Emerg. Med.* **1**(46), 150–155 (2021).
8. Pichichero, M. E. & Poole, M. D. Comparison of performance by otolaryngologists, pediatricians, and general practioners on an otoendoscopic diagnostic video examination. *Int. J. Pediatr. Otorhinolaryngol.* **69**(3), 361–366 (2005).
9. Pichichero, M. E. Assessing diagnostic accuracy and tympanocentesis skills of South African physicians in management of otitis media [7]. *S. Afr. Med. J.* **92**(2), 137–138 (2002).
10. Wormald, P. J., Browning, G. G. & Robinson, K. Is otoscopy reliable? A structured teaching method to improve otoscopic accuracy in trainees. *Clin. Otolaryngol. Allied Sci.* **20**(1), 63–67 (1995).
11. Khan, M. A. *et al.* Automatic detection of tympanic membrane and middle ear infection from oto-endoscopic images via convolutional neural networks. *Neural Netw. [Internet].* **126**, 384–394. https://doi.org/10.1016/j.neunet.2020.03.023 (2020).
12. Graydon, K., Waterworth, C., Miller, H. & Gunasekera, H. Global burden of hearing impairment and ear disease. *J. Laryngol. Otol.* **133**(1), 18–25 (2019).
13. Zhang, J., Li, C., Yin, Y., Zhang, J. & Grzegorzek, M. Applications of artificial neural networks in microorganism image analysis: A comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. *Artif. Intell. Rev.* **1**, 1–58 (2022).
14. Sandström, J., Myburgh, H., Laurent, C., Swanepoel, D. W. & Lundberg, T. A machine learning approach to screen for otitis media using digital otoscope images labelled by an expert panel. *Diagnostics* **12**(6), 1318 (2022).
15. Crowson, M. G. *et al.* Machine learning for accurate intraoperative pediatric middle ear effusion diagnosis. *Pediatrics* **147**(4), e2020034546 (2021).
16. Byun, H. *et al.* An assistive role of a machine learning network in diagnosis of middle ear diseases. *J. Clin. Med.* **10**(15), 3198 (2021).
17. Habib, A. R., Wong, E., Sacks, R. & Singh, N. Artificial intelligence to detect tympanic membrane perforations. *J. Laryngol. Otol.* **134**(4), 311–315 (2020).
18. Viscaino, M. *et al.* Computer-aided diagnosis of external and middle ear conditions: A machine learning approach. *PLoS One [Internet].* **15**(3), 1–18. https://doi.org/10.1371/journal.pone.0229226 (2020).
19. Goshtasbi, K. *Machine Learning Models to Predict Diagnosis and Surgical Outcomes in Otolaryngology [Internet].* https://escholarship.org/uc/item/1tr0c2p0 (University of California Irvine, 2020).
20. Livingstone, D., Talai, A. S., Chau, J. & Forkert, N. D. Building an otoscopic screening prototype tool using deep learning. *J. Otolaryngol. Head Neck Surg.* **48**(66), 1–5 (2019).
21. Livingstone, D. & Chau, J. Otoscopic diagnosis using computer vision: An automated machine learning approach. *Laryngoscope* **13**, 1–6 (2019).
22. 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 [Internet].* **39**, 34–52. https://doi.org/10.1016/j.bspc.2017.07.015 (2018).
23. 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 [Internet]* **5**, 156–160. https://doi.org/10.1016/j.ebiom.2016.02.017 (2016).
24. Başaran, E., Cömert, Z. & Çelik, Y. Neighbourhood component analysis and deep feature-based diagnosis model for middle ear otoscope images. *Neural Comput. Appl.* **34**(8), 6027–6038 (2022).
25. Binol, H. *et al.* OtoXNet—Automated identification of eardrum diseases from otoscope videos: A deep learning study for video-representing images. *medRxiv (Internet)* https://doi.org/10.1101/2021.08.05.21261672 (2021).
26. Binol, H., Niazi, M.K.K., Elmaraghy, C., Moberly, A.C., & Gurcan, M.N. Automated video summarization and label assignment for otoscopy videos using deep learning and natural language processing. in *Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications [Internet]* (Park, B.J., Deserno, T.M. Eds.). 28. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11601/2582009/Automated-video-summarization-and-label-assignment-for-otoscopy-videos-using/https://doi.org/10.1117/12.2582009.full. Accessed 17 Mar 2021 (SPIE, 2021).
27. Binol, H. *et al.* SelectStitch: Automated frame segmentation and stitching to create composite images from otoscope video clips. *Appl. Sci. (Switzerland)* **10**(17), 1–13 (2020).
28. Binol, H. *et al.* Digital otoscopy videos versus composite images: A reader study to compare the accuracy of ENT physicians. *Laryngoscope* **131**(5), E1668–E1676 (2021).

29. Wang, W., Tamhane, A., Rzasa, J., Clark, J., Canares, T., & Unberath, M. Otoscopy video screening with deep anomaly detection. in (Drukker, K., Mazurowski, M.A. eds.) *Medical Imaging 2021: Computer-Aided Diagnosis [Internet]*. 50. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11597/2581902/Otoscopy-video-screening-with-deep-anomaly-detection/https://doi.org/10.1117/12.2581902.full. Accessed 17 Mar 2021. (SPIE, 2021)
30. Wang, W., Tamhane, A., Santos, C., Rzasa, J.R., Clark, J.H., Canares, T.L. *et al. Pediatric Otoscopy Video Screening with Shift Contrastive Anomaly Detection*. http://arxiv.org/abs/2110.13254. Accessed 25 Oct 2021 (2021).
31. Uçar, M., Akyol, K. & AtilaUçar, E. Classification of different tympanic membrane conditions using fused deep hypercolumn features and bidirectional LSTM. *Irbm (Internet)* **1**, 1–11. https://doi.org/10.1016/j.irbm.2021.01.001 (2021).
32. Alhudhaif, A., Cömert, Z. & Polat, K. Otitis media detection using tympanic membrane images with a novel multi-class machine learning algorithm. *PeerJ Comput. Sci.* **7**, e405 (2021).
33. Sundgaard, J. V. *et al.* Deep metric learning for otitis media classification. *Med. Image Anal.* **1**, 71 (2021).
34. Cömert, Z. Fusing fine-tuned deep features for recognizing different tympanic membranes. *Biocybern. Biomed. Eng. (Internet)* **40**(1), 40–51. https://doi.org/10.1016/j.bbe.2019.11.001 (2020).
35. Wu, Z. *et al.* Deep learning for classification of pediatric otitis media. *Laryngoscope.* **131**, 1–8 (2020).
36. Liu, W. *et al.* Is the aspect ratio of cells important in deep learning? A robust comparison of deep learning methods for multi-scale cytopathology cell image classification: From convolutional neural networks to visual transformers. *Comput. Biol. Med.* **1**, 141 (2022).
37. Li, X. *et al.* A comprehensive review of computer-aided whole-slide image analysis: From datasets to feature extraction, segmentation, classification and detection approaches. *Artif. Intell. Rev.* **55**(6), 4809–4878 (2022).
38. Habib, A. R. *et al.* An artificial intelligence computer-vision algorithm to triage otoscopic images from Australian Aboriginal and Torres Strait Islander children. *Otol. Neurotol.* **43**(4), 481–488 (2022).
39. Habib, A. R. *et al.* Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. *Clin. Otolaryngol.* **47**(3), 401–413 (2022).
40. Chen, H. *et al.* IL-MCAM: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach. *Comput. Biol. Med.* **1**, 143 (2022).
41. Chen, H., Li, C., Wang, G., Li, X., Rahaman, M., Sun, H. *et al. GasHis-Transformer: A Multi-scale Visual Transformer Approach for Gastric Histopathological Image Detection*. http://arxiv.org/abs/2104.14528. Accessed 29 Apr 2021 (2021).
42. Cohen, J. F. *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open* **6**(11), 1–17 (2016).
43. Liu, X. *et al.* CONSORT-AI extension. *Nat Med.* **26**, 1364–1374 (2020).
44. Başaran, E., Cömert, Z., Celik, Y., Velappan, S. & Togacar, M. Determination of tympanic membrane region in the middle ear otoscope images with convolutional neural network based YOLO method. *Deu Muhendislik Fakultesi Fen ve Muhendislik.* **22**(66), 919–928 (2020).
45. Camalan, S. *et al.* OtoMatch: Content-based eardrum image retrieval using deep learning. *PLoS One [Internet].* **15**(5), 1–16. https://doi.org/10.1371/journal.pone.0232776 (2020).
46. He K, Zhang X, Ren S, Sun J. *Deep Residual Learning for Image Recognition*. http://arxiv.org/abs/1512.03385. Accessed 10 Dec 2015 (2015).
47. Simonyan, K., & Zisserman, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. http://arxiv.org/abs/1409.1556. Accessed 4 Sep 2014 (2014).
48. Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K.Q. *Densely Connected Convolutional Networks*. http://arxiv.org/abs/1608.06993. Accessed 24 Aug 2016 (2016).
49. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T. *et al. An Image is Worth 16 x 16 Words: Transformers for Image Recognition at Scale. arXiv [Internet].* 11929. http://arxiv.org/abs/2010.11929 (2020).
50. Ramspek, C. L., Jager, K. J., Dekker, F. W., Zoccali, C. & van Diepen, M. External validation of prognostic models: What, why, how, when and where?. *Clin. Kidney J. (Oxford University Press)* **14**, 49–58 (2021).
51. Yin, J., Ngiam, K. Y. & Teo, H. H. Role of artificial intelligence applications in real-life clinical practice: Systematic review. *J. Med. Internet Res. (JMIR Publications Inc.)* **23**, 25759 (2021).
52. Basaran, E., Comert, Z. & Celik, Y. Convolutional neural network approach for automatic tympanic membrane detection and classification. *Biomed. Signal Process Control* **56**(1), 1–14 (2020).
53. Basaran, E., Sengur, A., Comert, Z., Budak, U., Celik, Y., & Velappan, S. Normal and acute tympanic membrane diagnosis based on gray level co-occurrence matrix and artificial neural networks. in *2019 International Conference on Artificial Intelligence and Data Processing Symposium, IDAP 2019*. 5–10 (2019).
54. Basaran, E., Comert, Z., Sengur, A., Budak, U., Celik, Y., & Togacar, M. Chronic tympanic membrane diagnosis based on deep convolutional neural network. in *UBMK 2019—Proceedings, 4th International Conference on Computer Science and Engineering*. 635–638 (2019).
55. Lee, J. Y., Choi, S. H. & Chung, J. W. Automated classification of the tympanic membrane using a convolutional neural network. *Appl. Sci. (Switzerland).* **9**(9), 1827 (2019).
56. Seok, J., Song, J.J., Koo, J.W., Kin, H.C., & Choi, B.Y. The semantic segmentation approach for normal and pathologic tympanic membrane using deep learning (internet). *BioRxiv*. https://doi.org/10.1101/515007v2.full. Accessed 3 Mar 2019 (2019).
57. Futoma, J., Simons, M., Panch, T., Doshi-Velez, F. & Celi, L. A. The myth of generalisability in clinical research and machine learning in health care. *Lancet* **1**(2), e489–e492 (2020).
58. Bradley, A. P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Patter Recognit.* **30**(7), 1145–1159 (1997).
59. Huang, J. & Ling, C. X. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* **17**(3), 299–310 (2005).
60. Alenezi, E. M. A. *et al.* The reliability of video otoscopy recordings and still images in the asynchronous diagnosis of middle-ear disease. *Int. J. Audiol.* **61**, 921 (2021).
61. Alenezi, E. M. A. *et al.* Clinician-rated quality of video otoscopy recordings and still images for the asynchronous assessment of middle-ear disease. *J. Telemed. Telecare* **26**, 1357633X20987783 (2021).
62. Wahl, B., Cossy-Gantner, A., Germann, S. & Schwalbe, N. R. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings?. *BMJ Glob Health (Internet)* **3**(4), e000798. https://doi.org/10.1136/bmjgh-2018-000798 (2018).
63. Alami, H. *et al.* Artificial intelligence in health care: Laying the foundation for responsible, sustainable, and inclusive innovation in low- and middle-income countries. *Global Health.* **16**(1), 1–6 (2020).

Acknowledgements

We acknowledge Microsoft's AI for Humanitarian Action Grant, the Garnett Passe and Rodney Williams Memorial Foundation, the AI for Good Research Lab, and the Avant Foundation for providing research funding to support this project.

Author contributions

A.-R.H., Y.X., K.B., S.M., T.S., R.D., J.L.F., C.P., R.S., N.S. conceived the study idea. A.-R.H., Y.X., K.B., S.M. were involved in data collection. A.-R.H., Y.X. extracted data and performed the data analysis. A-RH wrote the first draft of the manuscript. A.-R.H., Y.X., K.B., S.M., T.S., W.B.W., R.D., J.L.F., C.P., R.S., interpreted the data analysis and critically revised the manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. All authors read and approved the final manuscript.

Funding

Microsoft AI for Humanitarian Action Program providing access to Azure and in-kind Data Science resources. (1) Microsoft AI for Humanitarian Action Grant, (2) Research Scholarship from the Garnett Passe and Rodney Williams Memorial Foundation, (3) Avant Foundation Doctor-in-Training and Early Career Research Grants.

Competing interests

A-RH discloses research funding from the Garnett Passe and Rodney Williams Memorial Foundation, Microsoft's AI for Humanitarian Action grant, and the Avant Mutual Foundation Doctor-in-Training and Early Career research grants. CP discloses conflict of interest from Australian Medical Association (Queensland) and provides consulting services for the Deadly Ears Program. RS is a consultant for Medtronic. NS discloses research funding from Microsoft's AI for Humanitarian Action grant, consultant for ResMed, Optinose, Nasus, GSK, and ENT Technologies, and has received conference funding from Medtronic, Karl Storz, and NeilMed. YX, KB, SM, TS, RD, and JLF are affiliated with the AI for Good Research Lab.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-31921-0.

Correspondence and requests for materials should be addressed to A.-R.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

5.5 Concluding summary for this chapter and knowledge gained from this study

This study investigates the applicability of AI-based otoscopy algorithms in enhancing ear health assessments, emphasising the significance of generalisability for real-world applications. It highlights the issue of overfitting and overtraining within internal settings, where the observed performance may not extrapolate to external scenarios. To address these challenges, potential strategies include diversifying the dataset sources, employing data augmentation techniques, and optimising hyperparameters.

The research demonstrates the robust internal performance of AI-otoscopy algorithms in effectively distinguishing normal from abnormal otoscopic images. However, when subjected to new validation cohorts, there is a notable decline in overall model performance. This performance gap is attributed to discrepancies in several factors, including variations in training data, categories of ear diseases, and the devices used for image capture.

It is worth noting that even models not specifically trained for conditions, such as otitis media with effusion, display the ability to differentiate these conditions from normal cases in validation cohorts. This highlights the adaptability of pre-trained CNNs and their capacity to extrapolate patterns across different datasets. Pooling data from diverse sources emerges as a promising approach to mitigate the decline in external performance, as exemplified by the substantial performance achieved by the DenseNet-161 architecture in a comprehensive analysis of all three cohorts.

Although AI-otoscopy algorithms offer potential for advancing ear health assessments, maintaining consistent external performance across diverse cohorts is a challenge. To fully leverage the benefits of AI in clinical otoscopy, future endeavours should prioritise data diversity, standardised diagnostic criteria, and advanced processing techniques to develop a comprehensive and universally applicable algorithm suitable for practical clinical applications in various settings.

5.6 References

1. Habib AR, Xu Y, Bock K, Mohanty S, Sederholm T, Weeks WB, Dodhia R, Lavista Ferres J, Perry C, Sacks R, Singh N. Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. Nature Scientific Reports. 2023;13(5368).

CHAPTER SIX: Develop and evaluate a supervised computer-vision algorithm to triage otoscopic images for Aboriginal and Torres Strait Islander people living in rural and remote areas

6.1 Preface to the chapter

This chapter presents findings from a published peer-reviewed manuscript that explores an image classification algorithm to classify otoscopic images collected from Aboriginal and Torres Strait Islander people living in rural and remote areas. The dataset used to train, validate, and test the algorithm were ascertained from the tele-otology program described in Chapter 3. The published manuscript forms this chapter[153] and addresses specific aim #6 of this thesis. Dissemination of this research and author contributions for this paper are described below.

6.2 Research dissemination

Published peer-reviewed paper:

Habib AR, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images From Australian Aboriginal and Torres Strait Islander Children. Otol Neurotol. 2022 Apr 1;43(4):481-488. doi: 10.1097/MAO.0000000000003484. PMID: 35239622.

Journal impact factor: 2.6

Conference presentations:

Habib AR*, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. DrumBeat.ai: Artificial intelligence to triage ear disease in rural and remote areas. University of Sydney Faculty of Medicine and Health Higher Degree by Research Conference. Westmead, Australia. 2023. [Oral Presentation].

Awarded prize for oral presentation in category of ‘Cancer – Public Health and Clinical Research'.

Habib AR*, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. DrumBeat.ai: Artificial intelligence to triage ear disease in rural and remote areas. 4th Annual Virtual Conference on Computational Audiology. Virtual conference. 2023. [Oral Presentation].

Habib AR*, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. DrumBeat.ai: Artificial intelligence to triage ear disease in rural and remote areas. Westmead Association Hospital Week. Westmead, Australia. 2023. [Poster Presentation].

Habib AR*, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. DrumBeat.ai: Artificial intelligence to triage ear disease in rural and remote areas. International Federation of Otorhinolaryngology Societies (IFOS) Annual Scientific Conference. Dubai, UAE. 2023. [Oral Presentation].

Habib AR*, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images from Australian Aboriginal and Torres Strait Islander Children. Australian Artificial Intelligence in Medicine, Surgery and Healthcare (AMSAH) Conference. Brisbane, Australia. 2023. [Oral Presentation].

Habib AR*, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images from Australian Aboriginal and Torres Strait Islander Children. Ramsay Health Research Foundation Poster Competition. Sydney, Australia. 2022. [Poster Presentation]. Awarded poster competition research prize.

Habib AR*, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images from Australian Aboriginal and Torres Strait Islander Children. Academy of Child and Adolescent Health (ACAH) Annual Conference. Sydney, Australia. 2022. [Oral Presentation]. Awarded conference research prize.

Habib AR*, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images from Australian Aboriginal and Torres Strait Islander Children. Royal Australasian College of Surgeons (RACS) Annual Conference. Brisbane, Australia. 2022. [Oral Presentation]. Awarded Mi-Tec scientific research prize for oral presentation.

Habib AR*, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images from Australian Aboriginal and Torres Strait Islander Children. Royal Australasian College of Surgeons (RACS) Annual Queensland State Meeting. Noosa Heads, Australia. 2021. [Oral Presentation]. Awarded Neville Davis Prize for oral presentation.

*Presenting author

6.3 Author attribution statement

I, Al-Rahim Habib, was responsible for designing the study, interpreting data, writing drafts of the manuscript, submitting the manuscript, responding to reviewers' comments, and coordinating submission and publication of the manuscript.

My co-authors G Crossland, H Patel, and N Singh conceived the idea. Myself and my co-authors G Crossland and H Patel were involved in data collection. I extracted the data and performed the data analysis. My co-authors G Crossland, H Patel, E Wong, K Kong, H Gunasekera, B Richards, L Caffery, C Perry, R Sacks, A Kumar, and N Singh assisted with the interpretation of the data analysis, manuscript revision, and approval of the final version of the manuscript submitted for publication. All authors have read and approved the manuscript.

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statement above is correct.

Signature:

Associate Professor Narinder Singh

6.4 Paper in published format

Original Study

An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images From Australian Aboriginal and Torres Strait Islander Children

*†‡Al-Rahim Habib, §Graeme Crossland, §Hemi Patel, *‡Eugene Wong, ||¶#Kelvin Kong, ***Hasantha Gunasekera, †††‡‡‡Brent Richards, §§Liam Caffery, ||||Chris Perry, *Raymond Sacks, ¶¶Ashnil Kumar, and *‡Narinder Singh

*Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Camperdown, New South Wales, Australia; †Department of Otolaryngology–Head and Neck Surgery, Princess Alexandra Hospital, Brisbane, Queensland, Australia; ‡Department of Otolaryngology – Head and Neck Surgery, Westmead Hospital, Sydney, New South Wales, Australia; §Department of Otolaryngology – Head and Neck Surgery, Royal Darwin Hospital, Darwin, Northern Territory, Australia; ||School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia; ¶Department of Linguistics, Faculty of Medicine, Macquarie University, Sydney, New South Wales, Australia; #School of Population Health, Faculty of Medicine, University of New South Wales, Sydney, Australia; **The Children's Hospital at Westmead, Sydney, New South Wales, Australia; ††Division of Medical Services, Gold Coast University Hospital, Gold Coast, Queensland, Australia; ‡‡Griffith Health, Griffith University Queensland, Australia; §§Centre for Online Health, University of Queensland, Australia; ||||Centre for Online Health, University of Queensland, Australia; and ¶¶School of Biomedical Engineering, Faculty of Engineering, University of Sydney, Camperdown, New South Wales, Australia

Objective: To develop an artificial intelligence image classification algorithm to triage otoscopic images from rural and remote Australian Aboriginal and Torres Strait Islander children.

Study Design: Retrospective observational study.

Setting: Tertiary referral center.

Patients: Rural and remote Aboriginal and Torres Strait Islander children who underwent tele-otology ear health screening in the Northern Territory, Australia between 2010 and 2018.

Intervention(s): Otoscopic images were labeled by otolaryngologists to classify the ground truth. Deep and transfer learning methods were used to develop an image classification algorithm.

Main Outcome Measures: Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, area under the curve (AUC) of the resultant algorithm compared with the ground truth.

Results: Six thousand five hundred twenty seven images were used (5927 images for training and 600 for testing). The algorithm achieved an accuracy of 99.3% for acute otitis media, 96.3% for chronic otitis media, 77.8% for otitis media with effusion (OME), and 98.2% to classify wax/obstructed canal. To differentiate between multiple diagnoses, the algorithm achieved 74.4 to 92.8% accuracy and an AUC of 0.963 to 0.997. The most common incorrect classification pattern was OME misclassified as normal tympanic membranes.

Conclusions: The paucity of access to tertiary otolaryngology care for rural and remote Aboriginal and Torres Strait Islander communities may contribute to an under-identification of ear disease. Computer vision image classification algorithms can accurately classify ear disease from otoscopic images of Indigenous Australian children. In the future, a validated algorithm may integrate with existing telemedicine initiatives to support effective triage and facilitate early treatment and referral. **Key Words:** Artificial intelligence—Computer-vision—Deep learning—Image classification—Machine learning—Otitis media—Otoscopy—Triage.

Otol Neurotol **43:**481–488, 2022

Aboriginal and Torres Strait Islander children living in rural and remote areas of Australia have the highest rates of ear disease in the world (1). Such Indigenous children are three times more likely to have otitis media and twice as likely to have long-term hearing loss than non-Indigenous children in Australia (2). Remote Indigenous children are nearly twice as likely to experience long-term hearing loss than non-remote Indigenous counterparts (2). The World Health Organization has declared this an Australian ''Public Health Crisis'' requiring urgent and innovative solutions (3). The sequelae of long-term hearing loss in childhood, a critical stage of development, are well established; impacting speech,

Address correspondence and reprint requests to Al-Rahim Habib, M.D., M.Sc., Anderson Stuart Building, Faculty of Medicine and Health, The University of Sydney, Camperdown NSW 2050, Australia; E-mail: ahab1907@uni.sydney.edu.au

Sources of Support: None.

Disclosure of Funding: Research scholarship from The Garnett Passe and Rodney Williams Memorial Foundation.

The authors disclose no conflicts of interest.

DOI: 10.1097/MAO.0000000000003484

academic performance, behavior, social skills, and future employment opportunities (4).

Access to otolaryngologists remains limited in rural and remote communities (5). The SARS-CoV-2 pandemic further restricted specialist access in rural and remote Indigenous communities, given uncontrolled outbreak concerns, which would devastate such communities (6). In rural areas, most ear health screening is performed by nurses or community health workers (HW) in primary care, who are often less experienced than tertiary otolaryngologists to accurately diagnose ear disease (7).

Telemedicine initiatives, such as tele-otology, are one of the currently utilized methods for ear health screening in rural and remote communities (8). The workflow typically being: images are digitally captured by local HW; electronically stored; then uploaded for otolaryngologists to review from remote access points. While tele-otology is often a feasible strategy for remote ear assessments, clinical decision-making and communication of results is typically delayed and potentially increases the risk for adverse or critical complications to develop, along with potential loss of follow-up (8).

Outreach services are an important aspect of ear health management for children in rural and remote areas. The availability of outreach clinics depends on existing partnerships between specialists and communities, time commitments from visiting surgeons, accessibility for fly-in-fly-out services, equipment, and funding (9). For Aboriginal and Torres Strait Islander children living in rural and remote areas, access to otolaryngologist services can be costly, infrequent, require long-distance travel, and difficult to navigate for families that are unfamiliar with the existing healthcare system (10).

Advances in modern computing and artificial intelligence (AI) may provide a novel approach to overcome the delays in diagnosis and improve the triage of high-risk children (11). Machine learning is a fundamental component of AI allowing computers to learn by recognizing patterns in training data and make predictions based on these learned patterns on previously unseen (test) data. Convolutional neural networks (CNNs) are machine learning techniques that have been used to develop computer vision algorithms for clinical practice, such as for detecting diabetic retinopathy from retinal fundus photographs (12). Standard retinal photographs have been evaluated by an AI screening tool to indicate whether a patient requires specialist referral and treatment (12). HW have been trained to effectively and reliably capture images of a patient's retina and upload it into the AI screening tool to acquire an instantaneous assessment (13). The application of AI for eye disease highlights the potential for improving and modernizing screening practices for ear disease, particularly in under-resourced communities.

Simple, binary image classification algorithms using CNNs have achieved an accuracy of up to 82% to classify acute otitis media (AOM), up to 98% to classify chronic otitis media (COM), and 77 to 93% to classify tympanic membrane perforations from normal images (14–17). More sophisticated algorithms have been able to differentiate between normal, AOM, otitis media with effusion (OME), COM, and wax/obstruction with 81 to 86% accuracy using decision trees and artificial neural networks (18,19). Augmenting current screening programs with such computer vision algorithms may assist frontline HW triage ear disease. Computer-assisted classification systems, binary or multiclass, integrated into smartphone applications or even into otoscopes themselves may have the potential to support accurate and timely triage of otoscopic images.

Initial efforts to explore image classification models classified images in a binary model as ''intact'' versus ''perforated'' using images collected from online repositories as the training source (16). Building on this, the aim of the current study was to develop a multiclassification model to triage otoscopic images for Aboriginal and Torres Strait Islander children living in rural and remote areas, where the prevalence of ear disease is high. These data may be tested for generalizability in other populations in future work.

METHODS

Ethics

Approval was obtained from the Menzies School of Health Research, Northern Territory, Australia, (HREC: 2019-3410). The research was undertaken in compliance with the Helsinki Declaration (20).

Study Design, Data Source, and Collection

Otoscopic images were retrospectively collected from a database of routine ear health examinations performed as part of the Northern Territory Outreach Hearing Health Program from 2010 to 2018 in 93 rural and remote areas of the Northern Territory, Australia. Patient assessment was performed by otolaryngology Clinical Nurse Specialists and comprised of focused ear-related history-taking, audiometry, tympanometry, and otoscopy with digital image capture using MedRx (Largo, FL) or Welch Allyn (Chicago, IL) video otoscopes.

The ear-related history consisted of specific questioning related to the presence or absence of pain, discharge, hearing loss, use of hearing aids, and relevant surgical history. Results were recorded on a standardized data-collection form. Images were electronically stored and submitted to trained otolaryngologists for review. The ground-truth was defined by either of two board-certified otolaryngologists with extensive experience in pediatric Indigenous ear disease (authors: G.C. and H.P.). Each otolaryngologist independently classified otoscopic images using patient history, tympanometry, audiometry, and otoscopy findings. Diagnostic categories were normal external acoustic canal (EAC) and intact tympanic membrane (TM); obstructed external acoustic canal with wax or foreign body; acute otitis media (AOM); otitis media with effusion (OME); and chronic otitis media (COM), which included cholesteatoma, dry perforation, and discharging perforation. All otoscopic images included in the analytical dataset were de-identified and replaced with a unique identification code to maintain confidentiality and blinding of the participant to the reviewer. Text was removed from each image for patient privacy and to establish homogeneity of the dataset. A secondary, independent reviewer (author: A-R.H) reviewed each image and corresponding ground-truth label before entry into the algorithm. Any discrepancies were discussed between two board-certified otolaryngologists with extensive experience managing

Indigenous ear disease (authors: G.C. and H.P.) and consensus established. Where consensus could not be clearly established, the image was removed from the dataset.

Algorithm Generation and Performance Evaluation

The image classification algorithm was generated using deep and transfer learning with Microsoft Azure Custom Vision (Redmond, WA). The Custom Vision Application Programming Interface (API) used CNNs (ResNet backbone) to establish a supervised machine learning classification model from input image labels established before training. As described previously, transfer learning was used to harness previous knowledge gained by CNNs from previous classification tasks (21,22). The final classification layer of the CNNs was retrained to yield a custom model developed specifically for otoscopic images. Binary (normal versus wax/obstructed EAC or AOM or OME or COM) and multi-classification models were developed using the diagnostic categories. Diagnostic categories were combined to establish a multi-classification model. The algorithm was trained in the Custom Vision API for 24 hours using various image augmentation techniques (gaussian blur, rotation, horizontal flip, contrast and brightness adjustments, scale/zoom, and shear transformations) and hyperparameter tuning to optimize training performance. Test images were randomly selected from the overall dataset. These test images were excluded from the training cohort and unknown to the algorithm.

Statistical Analysis

Performance of the binary and multi-classification models was assessed for accuracy compared with the ground truth (otolaryngologist assessment) including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The receiver operating characteristics (ROC) curve was constructed to evaluate the algorithm's performance. Confusion matrices were constructed to summarize performance between diagnostic classes. Output probabilities of each test image were used to construct a ROC curve. The area under the curve (AUC) was used to evaluate the predictive capacity of the algorithm with corresponding 95% confidence intervals. Statistical analysis was performed using SPSS (Version 24, IBM Corporation, Armonk, NY, 2016).

RESULTS

A total of 6,527 otoscopic images were included in the analytic sample. The image dataset consisted of normal (3,102), wax/obstructed EAC (332), AOM (192), COM (1,499), and OME (1,402) otoscopic images. COM included cholesteatoma (18), dry perforations (739), and discharging perforations (742). Figure 1 depicts a selection of otoscopic image samples for each diagnostic category.

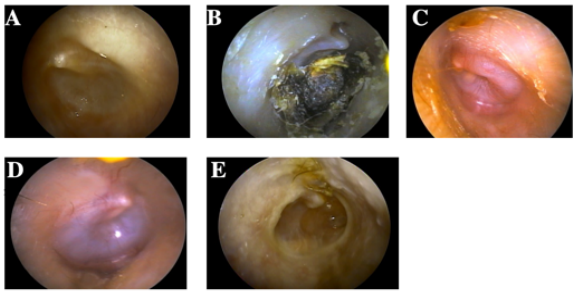


FIG. 1. Selection of otoscopic images used for training. *A*, Normal; *B*, wax/obstructed external acoustic canal; *C*, acute otitis media; *D*, otitis media with effusion; *E*, chronic otitis media.

The training image set consisted of 5,927 otoscopic images and the test image set consisted of 600 randomly selected otoscopic images not used in training.

Binary Classification

A summary of binary classification models is provided in Table 1 and a confusion matrix is provided in Table 2. Classification results were obtained for both the training and test data to assess whether the model was overfit (memorized the training cohort at the expense of generalizing to the unseen test cohort).

The binary classification models achieved more than or equal to 96.3% accuracy, more than or equal to 90.7% sensitivity, more than or equal to 98% specificity, more than or equal to 85.7% positive predictive value, and more than or equal to 94.7% negative predictive value on the test datasets for the classification tasks (obstructed EAC versus normal EAC, AOM versus normal TM, and COM versus normal TM). In the remaining task of OME versus normal TM, the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were lower. There was a higher degree of false positives and false negatives when classifying between OME versus normal TM, as shown in Table 2.

TABLE 1. *Summary of performance for binary image classification models to differentiate normal versus wax/obstructed external acoustic canal, acute otitis media, chronic otitis media, or otitis media with effusion*

Category	Total Images (n)	Training Images (n)	Test Images (n)	Test Accuracy (%)	Test Sensitivity (%)	Test Specificity (%)	Test PPV (%)	Test NPV (%)
Wax/Obstructed EAC versus normal	3434	W/O—302 N—2852	W/O—30 N—250	98.2	100	98.0	85.7	100
AOM versus normal	3294	AOM—172 N—2852	AOM—20 N—250	99.3	100	99.2	90.9	100
COM versus normal	4601	COM—1349 N—2852	COM—150 N—250	96.3	90.7	99.6	99.3	94.7
OME versus normal	4504	OME—1252 N—2852	OME—150 N—250	77.8	59.3	88.8	76.1	78.5

AOM indicates acute otitis media; COM, chronic otitis media; N, normal; OME, otitis media with effusion; W/O, wax/obstructed external acoustic canal.

TABLE 2. *Confusion matrix to compare the performance of binary classification models in the test dataset*

Wax/Obstruction Versus Normal				
		Ground Truth		
		Wax/Obstructed EAC	Normal	Total
Prediction	Wax/Obstructed EAC	30	5	35
	Normal	0	245	245
	Total	30	250	280
AOM Versus Normal				
		Ground Truth		
		AOM	Normal	Total
Prediction	AOM	20	2	22
	Normal	0	248	248
	Total	20	250	270
COM Versus Normal				
		Ground Truth		
		COM	Normal	Total
Prediction	COM	136	1	137
	Normal	14	249	263
	Total	150	250	400
OME Versus Normal				
		Ground Truth		
		OME	Normal	Total
Prediction	OME	89	28	117
	Normal	61	222	283
	Total	150	250	400

AOM indicates acute otitis media; COM, chronic otitis media; N, normal; OME, otitis media with effusion; W/O, wax/obstructed external acoustic canal.

Multi-classification

A summary of the multi-classification models is provided in Table 3 and a confusion matrix is provided in Table 4.

For all diagnoses combined (five classes), the algorithm achieved 74.5% accuracy and an AUC of 0.963 (95% CI: 0.941–0.986). The most common incorrect classification patterns were OME misclassified as normal, normal misclassified as OME, and COM misclassified as normal.

Excluding OME alone from the combined multi-classification model achieved 92.8% accuracy and an AUC of 0.997 (95% CI: 0.994–1.000). As shown in Table 4, this model improved classification accuracies for normal, AOM, and COM. There were fewer misclassifications of COM predicted as normal, although more misclassifications of normal predicted as COM, obstructed EAC predicted as COM, and AOM predicted as normal.

Excluding obstructed EAC alone from the combined multi-classification model achieved 74.4% accuracy and an AUC of 0.972 (95% CI: 0.965–0.989). Compared with the combined model (five classes), this model improved the classification accuracies for OME and COM (Table 4).

DISCUSSION

This study has shown that a multi-classification model based on deep and transfer learning can achieve substantial accuracy in classifying specific ear disease from otoscopic images. This represents a novel exploration of computer vision to address a significant, urgent public health need.

Managing ear health in rural and remote Aboriginal and Torres Strait Islander children is a complex, multifactorial public health issue with several contributing socio-economic, cultural, geographic, and health inequity factors (23). The ''Close the Gap'' campaign is a collaborative effort between Indigenous and non-Indigenous health organizations, non-governmental organizations, and human rights groups that aims to improve the health equality and life expectancy differences between Aboriginal and Torres Strait Islander people and non-Indigenous Australians within a generation (24). Efforts to ''close the gap'' in the burden of ear disease have included telemedicine outreach programs consisting of tele-otology ear assessments (8,25).

TABLE 3. *Summary of performance for multi-classification models*

Model	Categories	Classes	Total Images (n)	Training Images (n)	Test Images (n)	Test Accuracy (%)	AUC
1	Normal, OME, AOM, COM, Wax/obstructed EAC	5	6527	5927	600	74.5	0.963 (95%CI: 0.941–0.986)
2	Normal, AOM, COM, Wax/obstructed EAC	4	5125	4675	450	92.8	0.997 (95%CI: 0.994–1.000)
3	Normal, OME, AOM, COM	4	6195	5625	570	74.4	0.972 (95%CI: 0.965–0.989)

AOM indicates acute otitis media; COM, chronic otitis media; N, normal; OME, otitis media with effusion; W/O, wax/obstructed external acoustic canal.

Telemedicine is recognized as an important component of modern health service delivery and has revolutionized the way in which clinicians can remotely evaluate patients. Notably, the SARS-CoV-2 pandemic has highlighted the strength of telemedicine to triage new patients, streamline referral pathways, reduce patient travel expenses, and decrease wait times for specialist review (26,27). Tele-otology strategies and ''store-and-forward'' models have harnessed local HW to obtain patient histories and examination findings and relay the results to otolaryngologists for review (25,28–30). However, tele-otology relies on otolaryngologists to manually review each otoscopic image and provide a corresponding interpretation. This leads to several areas of significant weakness in the tele-otology model. Importantly, there are multiple points of delay between data collection, data forwarding, otolaryngologist review, communication of results, and actions based on the results. This leads to a missed opportunity for immediate action at the time of data acquisition. This can be particularly problematic in rural and remote Aboriginal and Torres Strait Islander children where predictable follow-up cannot always be assured. Such delays may put children at risk of developing serious complications and hearing impairment. In addition, continuous review of mostly normal TM images is not a productive use of otolaryngologist resources.

The advancements and accessibility of AI tools has the potential to transition tele-otology from manual assessments to the incorporation of AI for assistance or augmentation of clinical reviews. Augmenting ear health screening can save time and resources, as a greater volume of patients can be captured and assessed, and efforts can be focused on high-risk cases rather than screening normal images. Cha et al. (31) demonstrated the potential of an AI-enabled, automated system to

TABLE 4. *Confusion matrix to compare the performance of multi-classification models in the test dataset*

		Ground Truth					
Model 1		Normal	OME	W/O	AOM	COM	Total
Prediction	Normal	207	54	1	1	12	250
	OME	38	73	0	5	0	150
	W/O	2	12	27	0	7	30
	AOM	1	5	0	11	3	20
	COM	2	6	2	3	128	150
	Total	250	150	30	20	150	600
Model 2		Normal	OME	W/O	AOM	COM	Total
Prediction	Normal	244	–	0	3	8	250
	OME	–	–	–	–	–	–
	W/O	1	–	26	0	6	30
	AOM	1	–	0	14	2	20
	COM	4	–	4	3	134	150
	Total	250	–	30	20	150	450
Model 3		Normal	OME	W/O	AOM	COM	Total
Prediction	Normal	207	60	–	1	16	250
	OME	42	79	–	8	4	150
	W/O	–	–	–	–	–	–
	AOM	0	0	–	8	0	20
	COM	1	11	–	3	130	150
	Total	250	150	–	20	150	450

AOM indicates acute otitis media; COM, chronic otitis media; N, normal; OME, otitis media with effusion; W/O, wax/obstructed external acoustic canal.

classify ear disease by evaluating more patients in less time than an asynchronous, human-based telemedicine structure, or face-to-face evaluations. Despite Cha et al. (31) showing that an automated system yielded a lower accuracy than manual telemedicine assessments (69.5% versus 77.4%, respectively), the potential of the AI approach lies in its speed and the volume of assessments that can be completed. Although further efforts to validate this approach are necessary, the transition from teleotology performed by humans to a model that includes assistance with AI will represent a leap forwards to the modern otolaryngology telemedicine delivery model.

Several studies have evaluated the use of supervised machine learning algorithms to classify otoscopic images. Viscaino et al. (32) demonstrated that support vector machines could outperform classification methods such as k-nearest neighbors or decision trees to differentiate wax impaction, myringosclerosis, COM, and normal TMs with 93.9% accuracy. Khan et al. (33) demonstrated that deep learning with CNNs combined with image augmentation techniques achieved an accuracy of 95.0% to classify otoscopic images as either normal, COM with perforation or OME. Further efforts to delineate the optimal CNN were proposed by Wu et al. (21) who showed that the pre-trained Xception CNN could achieve 97.8% accuracy to classify normal, AOM, or OME. Although the performance yielded by CNNs exceed other classification methods, further improvements to these single stage models were made with the addition of attentional models and ensemble algorithms. Alhudhaif et al. (34) showed that multi-stage algorithms using global classifiers (i.e., considering the entire otoscopic image) followed by secondary focal/localized attention models (i.e., focusing on the TM within an otoscopic image) can yield even further accuracy improvements. However, the application of sophisticated multi-stage algorithms using bespoke methodology requires specialized expertise and significant computational power.

Alternative resources such as cloud-based APIs are becoming increasingly popular due to their simplified and user-friendly architecture. Computer vision services provided by Microsoft Azure Custom Vision (Redmond, WA), Google's Auto ML (Moutain View, CA), and Amazon's SageMaker (Bellevue, WA) have the potential to provide clinicians with limited data science expertize the capability to develop algorithms and address immediate clinical needs. Livingstone and Chau (22) used Google's Auto ML to classify 1,366 otoscopic images collected from Google Images and outpatient reviews into 14 diagnostic categories, yielding an average classification accuracy of 88.7%. Compared with a group of GPs and various trainees (pediatrics, emergency medicine, and otolaryngology), the algorithm outperformed human assessors in 12 out of 14 diagnostic categories (22). However, a direct and comprehensive comparison between the current study and Livingstone and Chau (22) cannot be performed because the patient populations used for training are heterogenous, the number and type of diagnostic classes are dissimilar, and the underlying architecture to establish predictions may be different. This highlights the importance of the effect of training data on prediction bias and generalizability. Image classification models may perform differently in various test environments and contribute to bias in-favor of the cohort used for training. In the current study, images were sampled from Aboriginal and Torres Strait Islander children from rural and remote areas where the prevalence of ear disease is significantly higher than non-Indigenous children (1,4). Most Aboriginal and Torres Strait Islander children have had some form of otitis media and often their TMs bear evidence of previous scarring, perforation, and refractory infection (35). Differentiating between diagnostic classes in the context of previous TM pathology may require more training data and bespoke classification algorithms. It is unclear if previously published algorithms will perform similarly between Indigenous and non-Indigenous populations and result in prediction bias when used for patient cohorts not included in training samples.

The current study demonstrates both the power of AI algorithms and their weaknesses. In particular, the algorithm demonstrated reduced accuracy in distinguishing OME from normal TMs. An improvement in performance between multi-classification models was observed when OME was excluded from the dataset. Congruent with human assessors, OME is a challenging diagnosis as the visual differences with other diagnoses are relatively subtle. OME is the most common ear pathology in rural Indigenous children but easily misdiagnosed (36).

A unique strength of this study is that the dataset is specific to the population being treated—a population with unique clinical challenges. While the long-term aim of this research is to eventually translate the benefits of such technology to a wider global population, a potential limitation of this study is the single data source with an unequal distribution of ear pathology in the data source. Imbalance in training data is a common scenario in real-world machine learning applications where an ideal distribution is rare (37).

In the current study, classification was performed using the entire image including the surrounding external acoustic canal. The method is simple, easily performed, and gathers information from the entire TM and external acoustic canal to establish the prediction algorithm. However, the disadvantage of this method can be appreciated from the binary algorithm's performance to differentiate between normal and OME. The algorithm was trained to evaluate the entire otoscopic image and not specific sections or features (e.g., fluid meniscus, bubbles, focal hypervascularity, TM opacity/color) (38). In manual assessments, the proportion of misdiagnosed OME is greater among pediatricians than otolaryngologists (36). This may occur because pediatricians are less familiar at recognizing key TM features of OME when compared with otolaryngologists (36). The algorithm in this study was not taught to focus on the TM specifically and this may have contributed to the proportion of OME misclassified as normal. Applying segmentation or Class Activation Maps to highlight the TM as a region of interest may be used to outperform pre-trained CNNs alone (39,40).

Even among otolaryngologists, the use of otoscopy to classify middle ear pathology can be challenging (36,41). The source for ground-truth in this study was specialist otolaryngologists evaluating digital otoscopic images captured by Clinical Nurse Specialists, along with clinical history on a standardized pro-forma, tympanometry, and audiometry results. Part of the reason for the algorithm's limited performance in classifying OME may be reflected in the choice of ground-truth. It has been suggested that a more accurate ground-truth can be established from intraoperative findings or myringotomy (42). However, preoperatively, middle ear fluid may be displaced by positive pressure ventilation in up to 15% of children with OME (43). In real-world settings it may not be appropriate or feasible to administer undue risk to definitively classify middle ear pathology using invasive procedures. Therefore, the current approach of using expert opinion to interpret otoscopic images appears to be a reasonable compromise, with the consequence of reduced algorithm performance. Despite these limitations, this study uses real-world data to establish a triage tool using the existing standard of care for a priority population.

The future of autonomous otoscopic interpretation for ear health screening and triage is promising. The performance achieved by the algorithm developed in the current study to classify normal, AOM, or wax/obstructed EACs in this study exceeds reported accuracy estimates from trainees (pediatrics, emergency medicine, and otolaryngology), GPs, and pediatricians (22,36). Future performance may be improved by adding symptoms, socio-demographic characteristics, tympanometry, audiometry, and dynamic video clips from pneumatic otoscopy to the classification model (40,44). Cholesteatoma could be incorporated into future models as a discrete category within COM in view of its potentially serious sequalae. Identification of this classification would prompt users to initiate urgent specialist referral for review and management.

The implications of this research are that existing teleotology services may have the potential to be enhanced with autonomous otoscopic image classification. Frontline rural and remote HW may be able to harness these tools to improve triage and decision-making and make more informed judgements when screening Aboriginal and Torres Strait Islander children for ear disease. Similarly, the workload for otolaryngologists providing online screening services could be reduced by using AI to augment batch screening. AI-based tools could be used to flag cases that are diagnostically uncertain or screen through normal images to rapidly identify concerning pathology and recommend treatment or surveillance. Reducing delays between frontline ear examinations and specialist interpretation may provide a valuable step towards the identification of early ear disease, initiation of necessary treatment, and prompt specialist referral, as required.

CONCLUSION

The current study established an autonomous computer vision algorithm to triage otoscopic images from Australian Aboriginal and Torres Strait Islander children living in rural and remote areas. Reducing delays between ear examinations and tele-otology specialist reviews may have the potential to reduce the risk of long-term hearing loss. Automating otoscopic image screening may also save specialists' time and resources and support prompt identification of high-risk cases warranting immediate review. This study represents an essential first step in developing AI tools to support health workers to evaluate ear health in rural and remote Australian Aboriginal and Torres Strait Islander children and then, potentially, Indigenous, and non-indigenous children globally.

Acknowledgments: The authors acknowledge the traditional owners of the land on which this research project was conducted, their spiritual relationship to the land, and their elders both past and present. They thank the Aboriginal and Torres Strait Islander people, elders, and community councils for allowing this research to be completed. They acknowledge the Garnett Passe and Rodney Williams Memorial Foundation for providing a research scholarship to support this project. They sincerely appreciate and acknowledge Professor Robyn Aitken, Amarjit Anand, Christian Aguilar, the ENT Department at the Royal Darwin Hospital, and the assistance of many clinical and administrative staff in the Top End Health Service, Northern Territory for their efforts collecting, and managing clinical data used for this study.

REFERENCES

1. Jervis-Bardy J, Sanchez L, Carney AS. Otitis media in indigenous australian children: review of epidemiology and risk factors. *J Laryngol Otol* 2014;128 (suppl):S16–27.
2. Australian Institute of Health and Welfare. Indigenous Hearing Health; 2020. Available at: https://www.aihw.gov.au/reports/australias-health/indigenous-hearing-health. Accessed February 1, 2020.
3. World Health Organization. Australia's disturbing health disparities set Aboriginals apart. *Bull World Health Organ* 2008;86:241–320.
4. Leach AJ. Otitis media in Australian Aboriginal children: an overview. *Int J Pediatr Otorhinolaryngol* 1999;49 (suppl):S173–8.
5. 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.
6. Kaplan HS, Trumble BC, Stieglitz J, et al. Voluntary collective isolation as a best response to COVID-19 for indigenous populations? A case study and protocol from the Bolivian Amazon. *Lancet* 2020;395:1727–34.
7. Australian Medical Association. Report Card on Indigenous Health: A National Strategic Approach to Ending Chronic Otitis Media and its Life Long Impacts in Indigenous Communities; 2017.
8. Elliott G, Smith AC, Bensink ME, et al. The feasibility of a community-based mobile telehealth screening service for Aboriginal and Torres Strait Islander children in Australia. *Telemed e-Health* 2010;16:950–6.
9. Shein G, Hunter K, Gwynne K, et al. The O. P. E. N. Survey: outreach projects in ear, nose and throat (ENT) in New South Wales. *Aust J Otolaryngol* 2019;2:14.
10. Graham AG, Macniven R, Kong K, Gwynne K. Effectiveness of ear, nose and throat outreach programmes for Aboriginal and Torres Strait Islander Australians: a systematic review. *BMJ Open* 2020;10:e038273.
11. Goggin LS, Eikelboom RH, Atlas MD. Clinical decision support systems and computer-aided diagnosis in otology. *Otolaryngol Head Neck Surg* 2007;136 (4 suppl):21–6.
12. Centre for Eye Research Australia. Artificial intelligence: the future of eye screening | CERA. Available at: https://www.cera.org.au/2019/07/artificial-intelligence-the-future-of-eye-screening/. Accessed May 22, 2021.

13. Centre for Eye Research Australia. Using AI to help close the eye health care gap | CERA. Available at: https://www.cera.org.au/2019/07/using-ai-to-help-close-the-eye-health-care-gap/. Accessed May 22, 2021.
14. Kasher MS. Otitis Media Analysis-An Automated Feature Extraction and Image Classification System; 2018. Available at: https://www.theseus.fi/handle/10024/144562. Accessed March 5, 2019.
15. Basaran E, Comert Z, Sengur A, Budak U, Celik Y, Togacar M. Chronic tympanic membrane diagnosis based on deep convolutional neural network. UBMK 2019 - Proceedings, 4th Int Conf Comput Sci Eng., 2019:635–638. doi:10.1109/UBMK.2019.8907070. Accessed September 25, 2020.
16. Habib A-R, Wong E, Sacks R, Singh N. Artificial intelligence to detect tympanic membrane perforations. *J Laryngol Otol* 2020;134:311–5.
17. Seok J, Song J-J, Koo J-W, et al. The semantic segmentation approach for normal and pathologic tympanic membrane using deep learning. 2019. Available at: https://www.biorxiv.org/content/10.1101/515007v2.full. Accessed March 3, 2019.
18. 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.
19. Myburgh HC, Jose S, Swanepoel DW, Laurent C. Towards low cost automated smartphone- and cloud-based otitis media diagnosis. *Biomed Signal Process Control* 2018;39:34–52.
20. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310:2191–4.
21. Wu Z, Lin Z, Li L, et al. Deep learning for classification of pediatric otitis media. *Laryngoscope* 2021;131:E2344–51.
22. Livingstone D, Chau J. Otoscopic diagnosis using computer vision: an automated machine learning approach. *Laryngoscope* 2019; 130:1408–13.
23. Delacy J, Dune T, MacDonald JJ. The social determinants of otitis media in aboriginal children in Australia: are we addressing the primary causes? A systematic content review. *BMC Public Health* 2020;20:1–9.
24. Australian Human Rights Commission. Close the Gap: Indigenous Health Campaign | Australian Human Rights Commission. Available at: https://humanrights.gov.au/our-work/aboriginal-and-torres-strait-islander-social-justice/projects/close-gap-indigenous-health. Accessed April 29, 2021.
25. Smith AC, Brown C, Bradford N, Caffery LJ, Perry C, Armfield NR. Monitoring ear health through a telemedicine-supported health screening service in Queensland. *J Telemed Telecare* 2015;21:427–30.
26. Bhaskar S, Bradley S, Sakhamuri S, et al. Designing futuristic telemedicine using artificial intelligence and robotics in the COVID-19 era. *Front Public Heal* 2020;8:556789.
27. Singh AK, Kasle DA, Jiang R, et al. A review of telemedicine applications in otorhinolaryngology: considerations during the coronavirus disease of 2019 pandemic. *Laryngoscope* 2021;131:744–59.
28. Smith AC, Perry C, Agnew J, Wootton R. Accuracy of pre-recorded video images for the assessment of rural indigenous children with ear, nose and throat conditions. *J Telemed Telecare* 2006;12 (suppl):76–80.
29. Kokesh J, Ferguson AS, Patricoski C. Preoperative planning for ear surgery using store-and-forward telemedicine. *Otolaryngol Head Neck Surg* 2010;143:253–7.
30. Kokesh J, Ferguson AS, Patricoski C, LeMaster B. Traveling an audiologist to provide otolaryngology care using store-and-forward telemedicine. *Telemed J E Health* 2009;15:758–63.
31. Cha D, Shin SH, Kim J, et al. Feasibility of asynchronous and automated telemedicine in otolaryngology: prospective cross-sectional study. *JMIR Med Inform* 2020;8:e23680.
32. Viscaino M, Maass JC, Delano PH, Torrente M, Stott C, Auat Cheein F. Computer-aided diagnosis of external and middle ear conditions: a machine learning approach. *PLoS One* 2020;15:1–18.
33. Khan MA, Kwon S, Choo J, et al. Automatic detection of tympanic membrane and middle ear infection from oto-endoscopic images via convolutional neural networks. *Neural Networks* 2020;126:384–94.
34. Alhudhaif A, Cömert Z, Polat K. Otitis media detection using tympanic membrane images with a novel multi-class machine learning algorithm. *PeerJ Comput Sci* 2021;7:e405.
35. Morris PS, Leach AJ, Silberberg P, et al. Otitis media in young Aboriginal children from remote communities in Northern and Central Australia: a cross-sectional survey. *BMC Pediatr* 2005; 5:1–10.
36. Pichichero ME. Assessing diagnostic accuracy and tympanocentesis skills of South African physicians in management of otitis media. *South African Med J* 2002;92:137–8.
37. Kaur H, Pannu HS, Malhi AK. A systematic review on imbalanced data challenges in machine learning. *ACM Comput Surv* 2019; 52:1–36.
38. Vanneste P, Page C. Otitis media with effusion in children: pathophysiology, diagnosis, and treatment. A review. *J Otol* 2019;14:33–9.
39. Basaran E, Comert Z, Celik Y. Convolutional neural network approach for automatic tympanic membrane detection and classification. *Biomed Signal Process Control* 2020;56:1–14.
40. Cai Y, Yu JG, Chen Y, et al. Investigating the use of a two-stage attention-aware convolutional neural network for the automated diagnosis of otitis media from tympanic membrane images: a prediction model development and validation study. *BMJ Open* 2021;11:1–7.
41. Pichichero ME, Poole MD. Comparison of performance by otolaryngologists, pediatricians, and general practioners on an otoendoscopic diagnostic video examination. *Int J Pediatr Otorhinolaryngol* 2005;69:361–6.
42. Crowson MG, Hartnick CJ, Diercks GR, et al. Machine learning for accurate intraoperative pediatric middle ear effusion diagnosis. *Pediatrics* 2021;147:e2020034546.
43. Gates GA, Cooper JC. Effect of anesthetic gases on middle ear pressure in the presence of effusion. *Ann Otol Rhinol Laryngol Suppl* 1980;89 (3 pt 2):62–4.
44. Takata GS, Chan LS, Morphew T, Mangione-Smith R, Morton SC, Shekelle P. Evidence assessment of the accuracy of methods of diagnosing middle ear effusion in children with otitis media with effusion. *Pediatrics* 2003;112:1379–87.

6.5 Concluding summary for this chapter and knowledge gained from this study

To our knowledge, this was the first study to explore an AI-based image classification algorithm to classify ear disease using otoscopic images collected from Aboriginal and Torres Strait Islander people. Binary algorithms demonstrated notable accuracy, sensitivity, and specificity to distinguish between otoscopic images of normal middle ears versus those of EAC obstruction, AOM, and COM. In multi-classification algorithms, where the model predictions were selected from multiple options, performance was dependent on the number of classes and type of ear condition. The most common incorrect classifications patterns were between normal aerated ears and OME. Excluding OME from the training dataset improved multi-classification performance.

The study identifies the importance of including a dataset specific to the population being studied within the training data. OME is common in Aboriginal and Torres Strait Islander children and often associated with hearing loss. However, OME can be missed during ear assessments performed by practitioners with limited experience recognising relevant macroscopic characteristics and variations in presentation. In these settings, AI-based algorithms have the potential to support triage and decision-making for practitioners with limited otology experience. Additional research is needed to evaluate the external performance and generalisability of the algorithm across multiple Aboriginal and Torres Strait Islander populations. Additional CNN architectures, data augmentation, and hyperparameter tuning could be explored in the future to explore methods to improve performance.

6.6 References

1. Habib AR, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, Richards B, Caffery L, Perry C, Sacks R, Kumar A, Singh N. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images From Australian Aboriginal and Torres Strait Islander Children. Otology and Neurotology. 2022 Apr 1;43(4):481–8.

CHAPTER SEVEN: Discussion

This thesis explored the use of AI to address ear disease in Aboriginal and Torres Strait Islander people living in rural and remote areas. This thesis includes a systematic literature review of previously studied machine learning techniques (Chapter 2)[120], an analysis of epidemiological data from tele-otology assessments performed in rural and remote areas (Chapter 3), an evaluation of intra- and inter-rater agreement among otolaryngologists to determine the reliability of tele-otology diagnostic labels (Chapter 4)[154], an investigation of the generalisability of deep learning image classification algorithms (Chapter 5)[155], and the development of a deep learning-based otoscopic image classifier to identify ear disease in Aboriginal and Torres Strait Islander people (Chapter 6)[153].

This chapter (Chapter 7) discusses the overall significance of the findings, implications for clinical practice and policy, the strengths and limitations of this research, and directions for future research.

7.1 Significance of findings

The implications of this research are relevant for ear disease screening and triage particularly among high-risk populations such as Aboriginal and Torres Strait Islander people. Telehealth is an important strategy to address limited access to specialist care, geographical remoteness, and cultural factors associated with health literacy, engagement, and follow-up. However, the asynchronous store-and-forward telehealth model, while beneficial in rural and remote settings, can contribute to delays between assessment and subsequent action, diagnosis, or management decisions. These delays could have notable implications, especially for children with ear disease, for recurrent ear disease and long-term hearing loss.

The application of AI, particularly the use of machine learning techniques, to automate the screening of otoscopic images has the potential to advance the telehealth service model. Advancing existing telehealth services could support practitioners with limited otology experience to identify ear disease with improved accuracy, provide a preliminary screening of images to reduce the workload on specialists, and support patients by providing timely information to support referrals and treatment pathways.

In Chapter 2, the systematic review and meta-analysis reviewed published studies exploring the application of machine learning to classify otoscopic images. The findings from Chapter 2 identified that CNNs achieve the greatest performance to classify otoscopic images compared to other machine learning methods. Models developed with CNNs achieve substantial internal accuracy, but model performance depends on the number of prediction classes, and combination of ear disease subtype. Substantial heterogeneity was found between studies, demonstrating variability between data sources, ground-truth classification, model architecture, training parameters, and ear disease subtypes. Deep learning-based models achieve greater accuracy than human, non-expert assessors. The literature review also highlights the importance of high-quality training data and emphasises the importance of standardised data collection.

In Chapter 3, the retrospective observational cohort study provides valuable insights into the prevalence of ear conditions among Indigenous people in rural and remote regions in the NT, Australia. Ear disease is found in most tele-otology assessments and nearly two-thirds of individuals present with hearing loss. Age is significantly associated with the type of ear disease identified. Most individuals with TM perforations are recommended for surgery and individuals with suppurative OM often require topical or oral antibiotics. Recurrent ear disease is prominent, especially among individuals with CSOM and OME. Otolaryngologists achieve substantial inter-rater agreement, although performance varies based on the type of

ear disease. Otolaryngologists are more likely to disagree between middle ear pathology without TM perforations. The tele-otology service in the NT demonstrates the importance of regular follow-up and timely interventions to mitigate long-term hearing loss. Most of the clinical data collected by CNCs and audiologists can be used by otolaryngologists to establish diagnosis and management plans. Furthermore, face-to-face assessments by CNCs provide a substantial opportunity to initiate frontline treatment.

Findings from Chapter 4 emphasise the importance of comprehensive clinical data in achieving accurate ear disease diagnoses. Otoscopy combined with audiometry, tympanometry, and CNC impressions significantly enhance an otolaryngologist's performance to diagnose ear disease compared with the use of otoscopy alone. The improvement in inter-rater agreement is most notable between the various tiers of clinical data to evaluate normal aerated ears, AOM and OME. These findings were congruent with previous results from Chapter 3 which demonstrate that the most challenging ear conditions to classify are between normal aerated ears versus subtypes of abnormal ears without tympanic membrane perforations. Chapter 4 also highlights the implications of the ground-truth diagnosis label. The difference in accuracy between an otolaryngologist evaluating an otoscopic image alone versus in combination with additional clinical data reveals the risk of misclassification, observer variability, and subsequent anchoring bias when relying primarily on one modality for diagnosis. The value of the information ascertained from an otoscopic image may be weighted based on the type of ear disease subtype. In cases with COM, the accuracy between Tier A to C was less pronounced than in normal aerated ears, AOM or OME. This may occur because recognising a TM perforation with or without discharge relies more on the characteristics of the otoscopic image, than the tympanometry and audiometry results. In contrast, differentiating between middle ear pathology where the TM is intact may require the addition of audiometry and tympanometry, in conjunction with otoscopy,

compared to otoscopy alone. These findings highlight that training image classification models on ground-truth labels extracted from otoscopic images alone, could be less accurate than those trained using ground-truth labels constructed with otoscopy, tympanometry, audiometry and the impressions of the HCW collecting the data.

Chapter 5 identified that a deep learning-based otoscopic image classifier can achieve substantial internal performance although generalisation to external cohorts is a challenge. Factors such as data source variability, variations in image capture devices, and ground-truth labelling discrepancies significantly impact algorithm performance. Potential options to overcome the difference between internal and external performance include; broadening the data source (i.e., pooling databases together) to include diverse training examples to improve generalisability; or using CNN architectures with a larger number of layers and feature extraction parameters.

In Chapter 6, an image classifier was developed using otoscopic images collected from the tele-otology program discussed in Chapter 3. The otoscopic image classifier was developed to address the limitations of asynchronous store-and-forward telehealth approaches as identified in earlier chapters of this thesis. The otoscopic image classifier demonstrated high levels of accuracy to differentiate between normal aerated ears and EAC obstruction, AOM and COM. However, congruent with expert performance as found in Chapter 4, the classifier performed less accurately to identify OME using an otoscopic image alone. These findings reveal that deep learning models can identify patterns and limitations present in training data that coincide with human performance. For example, when experts reviewed an otoscopic image, greater performance was found to identify gross characteristic differences (e.g., intact versus perforated TM) but performance was reduced in identifying subtle differences between intact TMs where middle ear pathology was present.

Four of the five studies presented in this thesis have been published (Chapters 2, 4-6) and one is currently under review for publication (Chapter 3). Many of these studies were among the first published worldwide to; systematically evaluate the literature related to machine learning for otoscopic image classifiers (Chapter 2); evaluate expert performance using varying tiers of clinical data (Chapter 4); explore the generalisability of otoscopic image classifiers (Chapter 5); and develop an otoscopic image classifier to identify ear disease in Aboriginal and Torres Strait Islander people (Chapter 6).

Overall, this thesis provides important public health findings of significance for Aboriginal and Torres Strait Islander people, nurses and healthcare workers performing ear disease screening and triage in rural and remote settings, primary care clinicians involved in ear disease management, otolaryngologists engaged in telehealth programs and outreach services, data scientists exploring machine learning applications in healthcare, and policy makers evaluating innovative solutions to optimise healthcare delivery and service delivery in under-resourced settings.

7.2 Implications of this research

The collection of studies presented in this thesis carries important implications for clinical practice and public health programs, particularly in the field of ear disease detection. The development of an AI-based otoscopic image classifier represents a novel advancement that holds the potential to transform healthcare delivery in rural and remote areas.

Validating the otoscopic image classifier in real-world clinical environments, beyond the control of virtual test settings, is required before implementation into clinical practice can be conducted. Rigorous field trials are essential to assess the classifier's performance and potential biases across diverse patient populations, primary end users, and otoscope manufacturers. Benchmarking against established telehealth protocols, expert assessors, and

comparing results with previously reported prevalence estimates is important to validate the classifier's effectiveness and reliability in real-world, clinical applications.

If such field-trials demonstrate the process to be safe and effective, the classifier may then be integrated into existing telehealth programs, offering a tool to augment existing capacity. Given the importance of telehealth to overcome geographical barriers and provide healthcare access to rural and remote communities, seamless integration is critical for successful implementation. Deployment options include both cloud-based and edge-based solutions, catering to a variety of settings, primary end-users, and availability of resources.

Cloud-based solutions offer scalability, making it possible for healthcare providers in remote areas to access the tool from a wide range of locations. Scalability is important in regions with dispersed populations, where centralised processing and storage can efficiently serve multiple locations. Cloud-based deployment may also reduce the need for local computing infrastructure. In rural and remote areas, costs associated with procuring, maintaining, and upgrading on-site servers and computing equipment could be avoided. For example, a remote healthcare clinic in a rural area can sidestep the need for a data centre and instead rely on cloud-based servers hosted by an off-site by provider. This could save significant expenditure but also relieve the clinic of the ongoing operational expenses linked to hardware maintenance, including expenses for IT personnel, repairs, and upgrades. In remote regions where resources may be limited, these cost savings may be transformative, allowing healthcare facilities to allocate their budgets more effectively and redirect funds toward other critical needs, such as patient care and staff training.

Cloud-based deployment may also support collaboration within the healthcare system. For example, a healthcare practitioner working in a remote clinic encountering a child with hearing loss or relevant symptoms can use a digital video otoscope to capture an otoscopic image. The image can be securely transmitted to the cloud-based AI system, where the

classifier can analyse the image and provide additional information to guide triage. Meanwhile, an otolaryngologist based in a distant location can access the cloud-based system to remotely review the cases and evaluate the classifier's predictions.

Alternatively, edge-based deployment can support local processing to reduce the dependence on high-speed internet access. The versatility of an edge-based option can ensure that the tool remains practical and accessible, even in regions where internet connectivity may be intermittent or restricted. For example, a remote clinic using an edge-based otoscopic image classifier can process the image locally on a dedicated computer. This option would avoid the need for continuous high-speed internet connectivity, reduce operational costs and potential disruptions to patient care. The reduced latency could also support the practitioner to make timely decisions about triage and referral. This would be critical in situations where expedited assessment is required (e.g., AOM or suppurative OM) or follow up is uncertain.

Regarding public health policy, supporting infrastructure and data repositories can be expanded to support refinement and focus AI-based applications to address the needs and challenges of Aboriginal and Torres Strait Islander people. The creation of a centralised tele-otology database to include a broad spectrum of patient demographics and communities, otoscopic images, audiometry, and relevant clinical history could be a valuable resource to refine and enhance the classifier. Policymakers can facilitate collaborations between research institutions, local health districts, and technology-focused organisations to tailor and support culturally sensitive applications of the classifier. In addition, policymakers could incentivise use of the classifier with a reimbursement model to expand the network of health districts participating in tele-otology services. Additional training and capacity building can be allocated to Aboriginal and Torres Strait Islander healthcare workers, nurses, and primary care practitioners to effectively use the classifier to support ongoing data collection and sustainability of the service for ongoing use.

It is important to acknowledge that while the otoscopic image classifier may improve ear disease detection, its impact is limited without addressing the social determinants of health that contribute to ear disease in Aboriginal and Torres Strait Islander communities. For example, many rural and remote communities experience delays in accessing specialist services and wait times exceed recommended guidelines.[58] Although the classifier has the potential to expedite triage and support referral pathways, patients may experience ongoing delays to receive essential care in timely manner. Furthermore, reducing the time between face-to-face assessments to interpretation, with use of the classifier, may overwhelm an existing healthcare system ill-equipped to manage an additional influx of cases. Primary care practitioners and specialists with time and resource constraints, may not be able to meet the demands of their referral network. As a result, patients where the classifier has identified ear disease may be unable to receive timely care. Patients, families, practitioners, and policy makers must consider these factors in conjunction with widespread adoption of an AI-based clinical workflow.

Informed consent, data privacy, security, and confidentiality are important considerations to highlight when considering the implications of this research on clinical practice. Patients and families should be explicitly informed about the data collection process, implications of participating in an AI-based clinical workflow, and the responsibility of their healthcare provider to act or not act on the classifier's predictions. Information should be provided in a clear, easy-to-understand, and culturally sensitive manner. To protect privacy, all information that could be used to identify patients should be removed during subsequent training and refinement of existing models. This will be critical to maintain privacy and confidentiality should cloud-based solutions be provided or collaboration with external organisations explored. Secure data storage can include encrypted databases, protected servers, and restricted access to authorised personnel. Further regulatory compliance with

existing guidelines will be required to ensure the augmentation of clinical workflows with AI are ethical, safe, secure, transparent, culturally sensitive, and sustainable.

7.3 Methodological strengths and limitations

One of the main strengths of this thesis is that it demonstrates a methodological workflow to apply AI in healthcare. This framework started with exploring previously published approaches, evaluating the burden of ear disease in rural and remote communities, identifying limitations of the current standard of care, examining a telehealth database, evaluating the ground-truth label for model training, exploring the importance of generalisability, and developing a clinically relevant machine learning model to address a specific healthcare need. This framework may serve as a template for future researchers.

In Chapter 2, a comprehensive systematic review was completed using established guidelines to evaluate previously published approaches. Each study included in the review was critically appraised yielding gaps in the existing knowledge base, which were explored in subsequent chapters. Chapter 3 provided a detailed assessment of a state-wide telehealth program to provide an overview of the burden of ear disease in rural and remote communities. The advantages and limitations of the telehealth database were explored and critically evaluated as a training source for algorithm development. Chapter 4 examined factors associated with accuracy of expert-level diagnoses using various tiers of clinical data. This exploration was not discussed in previously published studies describing otoscopic image classifiers and represented novel findings in field. Furthermore, the analysis conducted in Chapter 4 has implications on any future efforts to develop otoscopic image classifiers, as an accurate ground-truth requires the use of otoscopy with audiometry, tympanometry, and clinical history. Chapter 5 identified important limitations of machine learning models when applied to new test cohorts. These findings are of critical importance when considering

widespread use of the classifier in settings not included in the training cohort. Chapter 6 described the development of a clinically relevant machine learning model, tailored to address the specific needs of the telehealth program. The performance of the model was critically evaluated to identify strengths and limitations which can be refined in future iterations.

There are several methodological limitations. The use of tele-otology assessments, though valuable in rural and remote settings, introduces potential challenges associated with image quality and variability. Differences in image capture devices, lighting conditions, image resolution, and EAC obstruction could impact the robustness and generalisability of the otoscopic image classifier. Secondly, the ground-truth labels used for model training were based on a single otolaryngologist reviewing the otoscopy, audiometry, tympanometry, and clinical history. Although this is the standard of care, the ideal training database would include ground-truth labels from multiple reviewers or intraoperative findings. For example, Crowson et al. (2021) developed an otoscopic image classification model to detect middle ear effusions using images collected intraoperatively in children with the intent of performing myringotomy and insertion of tympanostomy tubes.[156] The use of intraoperative findings in the experiments performed in this thesis could reduce the likelihood of misclassifying middle ear effusions. Although, positive pressure ventilation during intraoperative assessments can displace middle ear effusions.[157] Thirdly, while the deep learning-based otoscopic image classifier demonstrated high accuracy in distinguishing certain ear conditions such as COM, it exhibited limitations in identifying OME, aligning with the challenges observed in human expert performance. This limitation suggests the need for further refinement in capturing subtle features within otoscopic images.

A critical limitation to acknowledge is the use of still otoscopic images and the exclusion of video of pneumatic otoscopy in the ground-truth assessment. Consequently, the classifier was

not trained with data incorporating this modality, which is useful to assess the presence of middle ear fluid. Incorporating pneumatic otoscopy could have improved the ground-truth classification, potentially enhancing the classifier's ability to distinguish between normal ear conditions and those with middle ear fluid, resulting in reduced TM mobility. Even without video of pneumatic otoscopy, the utility of digital video otoscopy as an assessment tool was not fully explored. Video otoscopy could have provided a more comprehensive dataset, offered raters additional visual cues and contributed greater diversity to the training database. For instance, video clips could be segmented into multiple still images or used in their entirety for training, encompassing a broader range of visual characteristics such as colour variation, illumination, movement, angulation, and reflection. The reliance on still images may inadvertently introduce user bias, as these images are subject to the user's interpretation and selection of what constitutes an optimal view of the TM. This variability could be influenced by the user's level of experience, potentially impacting the classifier’s training and its subsequent generalisability.

Finally, it is essential to recognise that the impact of an AI-based classifier is contingent on addressing broader social determinants of health and healthcare system capacity. Delays in accessing specialist services, potential resource constraints, and the ability of healthcare providers to manage an increased caseload due to AI-based triage must be considered when implementing this technology. Therefore, while an AI-based classifier offers promise, its successful integration into clinical practice requires a comprehensive understanding of its methodological limitations and the broader healthcare landscape.

7.4 Future directions

The exploration of AI-based otoscopic image classification and its potential applications in improving ear health for Aboriginal and Torres Strait Islander communities is promising. To advance this field and maximise its impact on healthcare delivery, there are several areas which may be explored in the future.

To ensure AI algorithms truly meet the needs of Aboriginal and Torres Strait Islander communities, future research should emphasise community-centred co-design processes. Engaging community members, healthcare workers, and local leaders in the design of the user-interface, including language, outputs, and features, will ensure that the technology is culturally sensitive, clinically relevant, and addresses community-specific healthcare challenges.[158–161] This participatory approach fosters trust and ownership, making the technology more readily embraced by the target population. For example, future endeavours may include organising community workshops where community members actively participate in shaping the design of AI-driven educational resources and healthcare solutions. This may also facilitate discussion regarding data privacy concerns to establish transparent data usage agreements with clear opt-in/opt-out mechanisms.[158] Guidance can be sought from local Aboriginal and Torres Strait Islander elders to ensure cultural sensitivity and alignment with community values and preferences.[159]

Further areas can be explored to improve the classifier's performance. One promising approach is the use of data augmentation parameters to bolster the model's robustness by diversifying the training dataset through various transformative processes[162,163]. This could include flipping, rotation, cropping, scaling, photometric and colour transformations, and noise injections.[164,165] Additional otoscopic variations could include varying lighting conditions, otoscopic light reflected off the TM, and anatomical variations specific to the ear canal and TM. Another intriguing avenue is the use of Generative Adversarial Networks (GANs).[165–167] GANs possess the capability to synthesise realistic yet diverse otoscopic

images, effectively enriching the training dataset.[168] These synthetic data closely mimic real-world variations, mitigating overfitting and enhancing a model's robustness in handling the diversity of otoscopic images from multiple sources. Equally important is the balancing of the dataset in terms of different ear conditions.[169] For example, AOM was underrepresented in the training data set, as it is less prevalent than other ear conditions (Chapter 3). Further efforts to include more relevant cases could improve the distribution of case mix used for training. This ensures that the model does not exhibit bias towards the majority class, a fundamental consideration for accurate diagnosis.[170]

Ensemble learning strategies may also provide a valuable area of exploration in the future. Ensemble models are comprised of multiple classifiers, each exposed to distinct data augmentations or architectural configurations.[171] By amalgamating the collective intelligence of multiple models, it is plausible that additional improvements in predictive accuracy may be achieved.[172] For instance, creating an ensemble of CNNs with varying filter sizes and depths, along with techniques such as averaging and voting can bolster the classifier's overall performance.[171,173,174]

Embedding interpretability into the AI model is paramount for its clinical utility.[175] Interpretability techniques such as Grad-CAM, which highlights influential regions in otoscopic images offering insights into feature contributions, empower healthcare practitioners to understand the model's decision-making process.[176] The interpretability provided by these measures may support clinicians to understand factors contributing to the classifier's prediction. For example, a hierarchical multi-stage model can be explored that provides prediction outputs at each level of the analysis. This may consist of multiple classifiers that evaluate image quality, EAC obstruction, presence or absence of TM perforations, key pathological or anatomical landmarks such as active discharge, erythema, retraction pockets, cholesteatoma, or evidence of middle ear effusions. Embedding a degree

of transparency into the classifier's functionality may also support clinicians to cross-reference the AI-based predictions with their own domain expertise. This can provide more information on areas in which the classifier surpasses or falls short compared to human performance.

While the thesis has demonstrated promising results in controlled settings, future research should focus on the real-world implementation and evaluation of AI-based triage tools. Prospective clinical trials will provide valuable insights into the technology's effectiveness, feasibility, and impact on clinical outcomes. Additionally, assessing the cost-effectiveness and scalability of these solutions will be important for policymakers and healthcare providers when considering wider adoption.

A practical application of this may include a telehealth system with integrated AI-powered otoscopy. For example, a rural and remote healthcare clinic situated in an Aboriginal and Torres Strait Islander community may be provided an AI-enhanced otoscopy device, equipped with a digital video otoscope capable of capturing high-quality otoscopic images and videos. Concurrently, a user-friendly telehealth platform can be coupled with the otoscope to empower healthcare providers to capture otoscopic images during live consultations and harness the AI model for instantaneous analysis. To determine the system's effectiveness, a comprehensive evaluation would be required, focusing on patient outcomes, diagnostic accuracy, and efficiency to deliver otologic services in conjunction with the existing standard of care. Evaluating the system's functionality as a clinical tool will provide insight into its value and utility in real-world settings.

Future research could explore the potential of AI in longitudinal monitoring and predictive analytics for ear health and hearing. AI algorithms could be developed to identify hearing loss, predict disease progression, identify at-risk individuals, and inform targeted

interventions, ultimately promoting preventive healthcare and reducing the burden of chronic ear disease.

The future of AI-based ear disease triage in Aboriginal and Torres Strait Islander people holds immense potential. By centring research on practical clinical applications, conducting real-world evaluations, developing a robust infrastructure for model refinement, and fostering global collaboration, researchers can push this field forward and make significant strides in improving ear health outcomes and reducing disparities in healthcare access and delivery for underserved populations.

7.5 References

1. Habib AR, Kajbafzadeh M, Hasan Z, Wong E, Gunasekera H, Perry C, et al. Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. Clinical Otolaryngology. 2022 May 1;47(3):401–13.

2. Habib AR, Perry C, Crossland G, Patel H, Kong K, Whitfield B, et al. Inter-rater agreement between 13 otolaryngologists to diagnose otitis media in aboriginal and Torres Strait islander children using a telehealth approach. Int J Pediatr Otorhinolaryngol. 2023 Mar;111494.

3. Habib AR, Xu Y, Bock K, Mohanty S, Sederholm T, Weeks WB, et al. Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. Nature Scientific Reports. 2023;13(5368).

4. Habib AR, Crossland G, Patel H, Wong E, Kong K, Gunasekera H, et al. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images From Australian Aboriginal and Torres Strait Islander Children. Otology and Neurotology. 2022 Apr 1;43(4):481–8.

5. 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 [Internet]. 2009 [cited 2019 Feb 20];45(7–8):425–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19722295

6. Crowson MG, Hartnick CJ, Diercks GR, Gallagher TQ, Fracchia MS, Setlur J, et al. Machine Learning for Accurate Intraoperative Pediatric Middle Ear Effusion Diagnosis. Pediatrics. 2021;147(4):e2020034546.

7. Gates GA, Cooper JC. Effect of anesthetic gases on middle ear pressure in the presence of effusion. Ann Otol Rhinol Laryngol Suppl [Internet]. 1980 [cited 2021 Aug 17];89(3 Pt 2):62–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6778352

8. McCalman J, Bainbridge R, James YC, Bailie R, Tsey K, Matthews V, et al. Systems integration to promote the mental health of Aboriginal and Torres Strait Islander children: protocol for a community-driven continuous quality improvement approach. BMC Public Health [Internet]. 2020 Dec 1 [cited 2023 Oct 5];20(1):1–12. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-09885-x

9. Butler T, Gall A, Garvey G, Ngampromwongse K, Hector D, Turnbull S, et al. A Comprehensive Review of Optimal Approaches to Co-Design in Health with First Nations Australians. International Journal of Environmental Research and Public Health 2022, Vol 19, Page 16166 [Internet]. 2022 Dec 2 [cited 2023 Oct 5];19(23):16166. Available from: https://www.mdpi.com/1660-4601/19/23/16166/htm

10. Tembo D, Hickey G, Montenegro C, Chandler D, Nelson E, Porter K, et al. Effective engagement and involvement with community stakeholders in the co-production of global health research. BMJ [Internet]. 2021 Feb 16 [cited 2023 Oct 5];372. Available from: https://www.bmj.com/content/372/bmj.n178

11. Abdalla SM, Koya SF, Jamieson M, Verma M, Haldane V, Jung AS, et al. Investing in trust and community resilience: lessons from the early months of the first digital pandemic. BMJ [Internet]. 2021 Nov 29 [cited 2023 Oct 5];375. Available from: https://www.bmj.com/content/375/BMJ-2021-067487

12. Abdollahi B, Tomita N, Hassanpour S. Data augmentation in training deep learning models for medical image analysis. Intelligent Systems Reference Library [Internet]. 2020 [cited 2023 Oct 5];186:167–80. Available from: https://link.springer.com/chapter/10.1007/978-3-030-42750-4_6

13. Bhuse P, Singh B, Raut P. Effect of Data Augmentation on the Accuracy of Convolutional Neural Networks. Lecture Notes in Networks and Systems [Internet]. 2022

[cited 2023 Oct 5];191:337–48. Available from: https://link.springer.com/chapter/10.1007/978-981-16-0739-4_33

14. He C, Liu J, Zhu Y, Du W. Data Augmentation for Deep Neural Networks Model in EEG Classification Task: A Review. Front Hum Neurosci. 2021 Dec 17;15:765525.

15. Savage N. Synthetic data could be better than real data. Nature. 2023;

16. Delaney AM, Brophy E, Ward TE. Synthesis of Realistic ECG using Generative Adversarial Networks. 2019 Sep 19 [cited 2023 Oct 5]; Available from: https://arxiv.org/abs/1909.09150v1

17. Odaibo SG, D. M, S. M, S. M. Generative Adversarial Networks Synthesize Realistic OCT Images of the Retina. 2019 Feb 18 [cited 2023 Oct 5]; Available from: https://arxiv.org/abs/1902.06676v1

18. Suresh K, Cohen MS, Hartnick CJ, Bartholomew RA, Lee DJ, Crowson MG. Generation of synthetic tympanic membrane images: Development, human validation, and clinical implications of synthetic data. PLOS Digital Health [Internet]. 2023 Feb 24 [cited 2023 Oct 5];2(2):e0000202. Available from: https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000202

19. Krawczyk B. Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence. 2016;5:221–32.

20. Pot M, Kieusseyan N, Prainsack B. Not all biases are bad: equitable and inequitable biases in machine learning and radiology. Insights Imaging [Internet]. 2021 Dec 1 [cited 2023 Oct 5];12(1):1–10. Available from: https://insightsimaging.springeropen.com/articles/10.1186/s13244-020-00955-7

21. Guo Y, Wang X, Xiao P, Xu X. An ensemble learning framework for convolutional neural network based on multiple classifiers. Soft comput [Internet]. 2020 Mar 1 [cited 2023

Oct 5];24(5):3727–35. Available from: https://link.springer.com/article/10.1007/s00500-019-04141-w

22. Yazdizadeh A, Patterson Z. Ensemble Convolutional Neural Networks for Mode Inference in Smartphone Travel Survey.

23. Zeng X, Jiang Z, Luo W, Li H, Li H, Li G, et al. Efficient and accurate identification of ear diseases using an ensemble deep learning model. Sci Rep [Internet]. 2021;11(1):1–10. Available from: https://doi.org/10.1038/s41598-021-90345-w

24. Cha D, Pae C, Seong SB, Choi JY, Park HJ. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. EBioMedicine [Internet]. 2019;45:606–14. Available from: https://doi.org/10.1016/j.ebiom.2019.06.050

25. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak [Internet]. 2020 Dec 1 [cited 2023 Oct 5];20(1):1–9. Available from: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6

26. Suara S, Jha A, Sinha P, Sekh AA. Is Grad-CAM Explainable in Medical Images? 2023 Jul 20 [cited 2023 Oct 5]; Available from: https://arxiv.org/abs/2307.10506v1