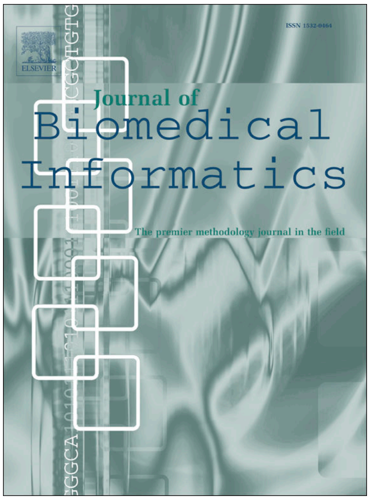


Original Research

A multimodal machine learning algorithm improved diagnostic accuracy for otitis media in a school aged aboriginal population

Jacqueline H. Stephens, Phong Phu Nguyen, Amanda Machell, Linett Sanchez, Eng H. Ooi, A. Simon Carney, Trent Lewis

PII: S1532-0464(25)00030-9
DOI: https://doi.org/10.1016/j.jbi.2025.104801
Reference: YJBIN 104801

To appear in: *Journal of Biomedical Informatics*

Received Date: 18 September 2024
Revised Date: 18 December 2024
Accepted Date: 24 January 2025

Please cite this article as: Stephens, J.H., Nguyen, P.P., Machell, A., Sanchez, L., Ooi, E.H., Carney, A.S., Lewis, T., A multimodal machine learning algorithm improved diagnostic accuracy for otitis media in a school aged aboriginal population, *Journal of Biomedical Informatics* (2025), doi: https://doi.org/10.1016/j.jbi.2025.104801

A MULTIMODAL MACHINE LEARNING ALGORITHM IMPROVED DIAGNOSTIC ACCURACY FOR OTITIS MEDIA IN A SCHOOL AGED ABORIGINAL POPULATION

Jacqueline H Stephens, MPH PhD [1] [ORCID 0000-0002-7278-1374]

Phong Phu Nguyen, MSS [2]

Amanda Machell, B.Ed, MPH, PhD [1] [ORCID 0000-0002-9058-3435]

Linett Sanchez, BA (Hons), B.Sp.Thy, MPH [3] [ORCID 0000-0001-8736-5501]

Eng H Ooi, FRACS PhD [1,4] [ORCID 0000-0002-7201-815X]

A Simon Carney, FRCS, FRACS, DM [1] [ORCID 0000-0001-7507-7983]

Trent Lewis, BSc(Hon) PhD [2] [ORCID 0000-0002-4025-4019]

1) Flinders Health and Medical Research Institute, College of Medicine and Public Health, Flinders University, Adelaide, Australia
2) College of Science and Engineering, Flinders University, Adelaide, Australia.
3) Audiology, College of Nursing & Health Sciences, Flinders University, Adelaide, Australia.
4) Otolaryngology Head and Neck Surgery Unit, Flinders Medical Centre, Adelaide, Australia.

Corresponding Author:

Associate Professor Jacqueline Stephens

Jacqueline.stephens@flinders.edu.au

Flinders Health and Medical Research Institute

College of Medicine and Public Health

Flinders University

GPO Box 2100, Adelaide, South Australia 5001

Word Count:

3633

ABSTRACT

Objective

Otitis Media (OM) - ear infection - can lead to hearing loss and associated developmental delay. There are several subgroups of OM which can be difficult to diagnose accurately, even for experienced clinicians. AI and machine learning algorithms for OM diagnosis are evolving but typically only focus on one defined diagnostic feature of OM. This study aimed to establish if combining otoscopic and tympanometry data improves the diagnostic accuracy of a ML algorithm for diagnosing OM and its various subgroups.

Methods

We used an existing dataset containing data from 813 school-aged children (aged five to eight years) from 10 Aboriginal communities in remote South Australia. Data were collected between 2009 and 2011. All children underwent video otoscopy and tympanometry assessment of both ears and diagnosis of OM was made by otorhinolaryngology (ENT) surgeons. After data augmentation and preprocessing, the database contained 15,057 samples with matched video otoscopy and tympanometry data (normal: n=8,239; abnormal: n=6,746). Support Vector Machine models were used to build the ML system.

Results

By combining tympanometry data with the probability prediction of the single otoscopy model, the accuracy of the system increased from 78% (otoscopy data) to 82% (otoscopy and tympanometry data).

Conclusion

Compared to diagnosis based solely on otoscopy data, combining otoscopy and tympanometry data increased the diagnostic accuracy of the ML algorithm. This approach could be used to support the accurate diagnosis of OM in children and can facilitate timely and appropriate treatment, especially in rural and remote areas.

Keywords

Otitis media [MeSH], Machine Learning [MeSH], Diagnosis [MeSH], Australian Aboriginal and Torres Strait Islander Peoples [MeSH]

INTRODUCTION

Otitis media (OM) is one of the most common childhood infections, with most children experiencing at least one episode of OM by their third birthday.[1] OM, or inflammation of the middle ear, is an umbrella term encompassing acute otitis media (AOM), otitis media with effusion (OME) and chronic otitis media (COM). While AOM is characterized by signs of an acute infection (e.g. pain and fever), and can resolve spontaneously without complications, OME is often asymptomatic and can go undiagnosed.[2] OM causes hearing loss and, without timely diagnosis and treatment, may lead to developmental delays including negative effects on auditory processing, language and speech development, school readiness, social competence, psychosocial wellbeing, and sleep.[3] Importantly, these various types of OM have differing degrees of associated hearing loss and educational prognoses, so distinguishing these is clinically important and, accordingly, early detection is crucial for ensuring timely and appropriate treatment.

Diagnostic tools for OM include otoscopy and tympanometry.[2] However, the use of those tools requires extensive training, and an accurate diagnosis is dependent on the technical skills and clinical experience of the practitioners.[4] The initial diagnosis of OM largely falls on general practitioners (GPs) but the rate of correct diagnosis is lower than that of specialist otolaryngologists due to less clinical experience and formal training.[5] For example, the average correct diagnosis of OM by GPs was approximately 50%,[6-8] compared to around 74% by otolaryngologists.[7, 9] However, access to such specialists is limited, especially in rural and remote areas.[10] Poor diagnostic accuracy leads to misdiagnosis and subsequent delay in treatment, with the main preventable complication being significant and persisting hearing loss. Thus, a system to support timely and accurate diagnosis of OM, especially in rural and remote areas, is needed to support less skilled and experienced practitioners in accurately diagnosing OM.

Machine learning (ML) is the process of developing computer algorithms to learn and make decisions from data,[11] and has been proposed as a method to provide automated diagnosis of OM using tympanic membrane (ear drum) images.[12, 13] ML can be divided into three main subtypes: supervised learning, unsupervised learning, and reinforcement learning.[11] Supervised learning is commonly used in the OM literature and classifies data into discrete labelled categories.[11] Supervised learning uses several different algorithms, including Support Vector Machine (SVM) which is used in this study.

Despite being in its infancy, ML diagnosis has been shown to be at least as accurate or even more accurate than diagnosis performed by medical practitioners. For example, utilizing 89 otoscopic images of various ear pathologies, Livingstone et al. demonstrated ML diagnosis can achieve 86% accuracy, compared to 59% for ENT specialists.[10] Results from recent systematic reviews,[11, 12, 14] and meta-analyses,[13, 15] also demonstrate the potential for ML to accurately diagnose OM. Ding and colleagues[14] reviewed 79 studies of artificial intelligence-based approaches to diagnose (n=52 studies), treat (n=21 studies), or manage (n=6 studies) OM patients, with results showing ML had a diagnostic accuracy of between 74% and 100% for both binary and multi-label classification problems. Similarly, Ngombu et al[11] reviewed 25 articles describing applications of AI technologies specific to diagnosing OM, including machine learning and natural language processing (n = 23) and prototype approaches (n = 2), with diagnostic accuracy greater than 80% in the majority of studies (19/23 studies). Meta-analytic results from Cao et al[15] (n=16 studies) and Habib et al[13] (n=39 studies) also support the accuracy of ML for diagnosing OM. Results demonstrated that ML can accurately diagnose OM in between 76% and 98% of cases for binary classification of normal vs. abnormal for middle ear disorders.

While it is clear ML has the potential to significantly improve the timely identification and treatment of OM, those reviews highlight several limitations of the current literature. First, few studies included data from pediatric populations.[11, 15] This is likely due to ethical considerations and data collection challenges inherent in using pediatric datasets.[16] Children can be uncooperative, making it difficult to obtain a clear image and this may affect the accuracy of the ML diagnosis. However, it is important to consider pediatric populations separately from adult populations due to the structural differences in the ear.[17, 18] Furthermore, some children, particularly Australian First Nations children, have been shown to have earlier onset, higher frequency, and greater severity of otitis media than non-Indigenous children.[19, 20] More complex OM can result in difficulty to clearly visualize the tympanic membrane or obtain clear otoscopic imagery. As a result, OM may on occasion be difficult to accurately diagnose using only otoscopy images.[21] Pragmatically, in clinical practice clinicians use multiple measures in their clinical decision-making. Despite this, few studies have explored the potential to combine data from multiple measures of OM.[14] Combining multiple measures can enhance the accuracy of the ML diagnosis.[14] By combining clinical data with both tympanometry and digital otoscopy data, Binol et al[22] provides support for this notion by achieving a classification accuracy of 84.9%, compared to 76.7% for tympanometry data alone and 74.0% for the otoscopy data alone. Our current study, as reported herein, fills these identified gaps by developing a ML-based diagnosis support system for OM-related illness utilising both otoscopy and tympanometry data for improved diagnostic accuracy in a school-aged population. Therefore, the objective of this study was to establish if combining two types of clinical data, in this case otoscopic images with tympanometry data, could be achievable and if this could improve the diagnostic accuracy of a ML algorithm for diagnosing OM and its various subgroups.

Problem
Ear disease is common in childhood and is diagnosed using specialized clinical tools, such as otoscopy and tympanometry. However, an accurate diagnosis is dependent on the technical skills of practitioners and in many resource poor locations, where ear disease is common, the practitioners do not have the skills to interpret clinical audiometry data.
What is Already Known
Machine learning algorithms using otoscopy data alone have shown potential to support clinical decision making by practitioners who do not have the skills to interpret clinical audiometry data.
What this Paper Adds
We demonstrate that compared to diagnosis based solely on otoscopy data, combining otoscopy and tympanometry data increased the diagnostic accuracy of a machine learning algorithm. A multimodal diagnostic algorithm could support the accurate diagnosis of otitis media in resource

poor settings, such as for First Nations children living in rural and remote areas.

MATERIALS AND METHODS

This manuscript was prepared in accordance with the Standards for Reporting Diagnostic Accuracy (STARD) Guidelines.[23] This study was approved by the Flinders University Human Research Ethics Committee (HREC no. 4600).

Dataset

Non-identifiable data were extracted from an existing cohort study dataset.[24] Original data collection took place twice yearly in schools between 2009 and 2011. The dataset contains 4,454 distinct video otoscopy files (stored in Windows Media Video (WMV) format at 30 frames per second) with corresponding tympanometry data for both left and right ears, as well as corresponding metadata (labelling) in an MS Excel spreadsheet.

Images were mapped to the corresponding record in the MS Excel file and classified as either: normal, OME, AOM, healed COM, inactive mucosal COM, active mucosal COM, inactive squamous COM, or active squamous COM. Data were later collapsed into two categories, with normal and healed COM classified as "Normal" (OM negative) and all other categories classified as "Abnormal" (OM positive). The original tympanometry categories of Pass, Fail, or Indeterminate were harmonised with the image data to a binary Normal or Abnormal classification.

Participants

Participants included 813 school-aged children, aged 5 to 18-years, from 10 Aboriginal communities in remote South Australia (n=4 communities with swimming pools; n=6 communities without swimming pools). All children who were at school on the day of testing completed ear health and hearing assessments. The mean age of children in the swimming pool communities (10.7 years) and non-pool (10.6 years) communities was similar (p=0.668).

Software

All data augmentation, processing, and machine learning model training was performed using Python software. The code is available at GitHub (https://github.com/phunp/otitis-media).

Reference standard (ground truth labels)

OM can be classified in several different ways with no agreed reference standard.[15] For our study, ground truth labels were classified by otolaryngologists using the criteria according to Browning in the following order of priority: normal, OME, AOM, healed COM, inactive mucosal COM, active mucosal COM, inactive squamous COM, or active squamous COM.[25]

Data augmentation and preprocessing

After removing duplicate video files and invalid records (records with missing video mapping, invalid video file names, or invalid results), 3,552 of the original 4,454 videos remained. However, still images are needed to train the ML model, so video frames were extracted from the videos. For this, each video was loaded using OpenCV library (version 3.4.2) and the first 18% and the last 12% of each video was removed as this typically captured the device insertion and removal from the ear. For the remaining video duration, an image was captured and saved at every 10% interval for the total video duration. This process resulted in approximately six to seven images extracted per video and a total of 22,458 still images. However, not all images could be used due to poor quality, e.g. blurry, too bright, or the tympanic membrane was obscured by cerumen (Figure 1).

To overcome this problem, a subset of 3368 images (15% of the total images) were selected for manual filtering. From these, 1,301 images were removed due to poor quality, leaving 2,607 images. A simple Convolutional Neural Network (CNN) model was built to automate this process. The model contained two pairs of convolutional layers, following by a max pooling layer. The data were flattened out and connected to three fully connected layers. The accuracy of the filtering model was 84%, with an f1-score of 0.83. The filtering model was used to classify the whole dataset of 22,458 extracted images with 7,227 images removed in the filtering process due to poor quality, leaving 15,231 high quality otoscopic images. Of these, 15,057 records were available with matching otoscopy and tympanometry data (Table 1). Results are available from authors on request.

Filter extraction

Binary thresholding was applied in the grey scale version of the original image to extract the region of interest (ROI) as only a portion of each otoscopy image contained the eardrum and the quality of image varied. However, in the final model the original images were used as the accuracy of the ML algorithm was higher with the original, compared to cropped images.

Machine learning model training

The researchers tested a range of machine learning approaches, including multilayer perceptron (MLP), convolutional neural networks (CNN), and decision tree algorithms, however, the support vector machine (SVM) outperformed these other ML approaches. As such, the SVM model was chosen to train and test the classification system on 70% and 30% of the data, respectively. Stratified sampling on the output was performed to ensure the training set and test set have the same percentage of each group samples in comparison with the original dataset. Given the large size of the original dataset, a single train/test split created sets with representative variety and complexity of diagnostic outcomes, therefore a more complex splitting strategy was unnecessary. GridSearch utility in *sklearn* package was used to select the best hyperparameters for the SVM pipeline. The parameters were tuned with several values, as presented in Table 2.

The GridSearch accuracy score was determined using a 5-fold cross-validation of each of the hyperparameter combinations on the training set. The mean of the five scores was used to select the best model. Stratified technique was also applied as the number of samples in each category was different. After the best group of hyperparameters was found, the model with those parameters was trained again on the whole training set and then validated against the test set.

To combine different types of data in a single pipeline SVM was chosen. First, otoscopy images were fed to the SVM model, where the SVM model predicted a probability number for the two classes. That is, how definite the model is in predicting the sample belongs to a given class. The probability number for the Normal class was extracted and combined with the tympanometry data (Figure 2). Next, another model was introduced and trained on the combination of SVM probability prediction and the tympanometry data. The combined data, with two features as the input: 1) SVM probability prediction and 2) tympanometry data; was fed into a simple multilayer perceptron (MLP) model with 10 neurons in a single layer. The subsequent model gave prediction on the test set. The performance of the validation set after training with the combined data was compared with the performance of the SVM model on the otoscopy data alone.

Evaluation metrics

A confusion matrix indicates the number of instances which correspond to the predicted and the actual classes; and is often used in binary classification.[26] The performance of the ML model was measured from computing a confusion matrix. From the confusion matrix scores accuracy, precision, recall, F1-score and Matthews Correlation Coefficient (MCC) were computed.

Accuracy represents the ratio between the correctly classified instances and all the instances in the dataset. As accuracy does not consider false positives and false negatives, precision and recall metrics were also calculated.

Precision determines the proportion of positive predictions that were actually correct, while recall determines the proportion of actual positives that were predicted correctly. Precision and recall values range from 0 (worst) to 1 (best). As a good model needs to strike the right balance between precision and recall, F1-score was also calculated. F1-score is the harmonic mean of precision and recall. But, as the F1-score ignores the ability of the model to accurately classify negative cases and can consequently give overinflated results, MCC was also calculated. MCC will only produce a high score if the prediction obtained good results in all four confusion matrix categories (true positives, false negatives, true negatives, and false positives).[27] Values generally range from 0 to 1, where 0 equates to random guessing and 1 is perfect classification.

Role of the funding source

The funding body had no input into the study design, collection, analysis, nor interpretation of the data. Nor did they have any input into the writing of the report or the decision to submit the paper for publication.

RESULTS

The optimal SVM parameters from the GridSearch were for the PCA to retain 100 components, a RBF Kernel, and a C value of 10. This SVM model, using only otoscopy images achieved an overall accuracy of 78%, but still made errors by classifying 445/2488 (18%) normal samples as abnormal, and 571/2030 (28%) abnormal samples as normal (Figure 3). The average precision, recall and F1-score were all 0.77. The MCC value was 0.54.

To boost the accuracy of single SVM model, another model was introduced and trained on the combination of SVM prediction and the tympanometry data. The simple MLP, trained with only 50 iterations, boosted the accuracy of the pipeline from 78% for SVM alone to 82% for SVM+MLP. The precision (0.83), recall (0.84) and F1-score (1.84) for the SVM+MLP model was higher than for SVM model alone (Precision=0.78, Recall=0.82, f1-score=0.80). The MCC value for the combined model was 0.63 (the value for SVM alone is 0.55). The performance metrics for both models are shown in Figure 3.

DISCUSSION

We developed a ML system to support the accurate diagnosis of OM using multiple sources of data with the aim of improving timely and appropriate diagnosis and treatment. Results indicate performance of the ML model was highest when otoscopy and tympanometry data were combined. The results from this study support our hypothesis that the addition of multimodal data into machine learning models could increase the performance of a ML for OM diagnosis. We demonstrate the combination of multiple data sources improved accuracy of ML to diagnose OM in a multimodal system containing two ML models: SVM and MLP. The SVM model gave predictions based on otoscopy images with the resulting probability combined with the tympanometry data and fed to the MLP model. This process increased the diagnostic accuracy from 78% (for otoscopy data alone) to 82% (for combined otoscopy and tympanometry data) for a set of otoscopy images. This approach of using multiple diagnostic techniques reflects real world processes, where the clinicians often perform more than one assessment before making a diagnosis.

Accuracy of ML diagnoses of ear disease have been shown to be higher than when performed by medical practitioners alone.[13] The demonstrated superiority of an ML diagnosis of OM may be particularly relevant in remote settings where access to otolaryngologists and audiologists is limited and long wait times exist. Frontline health workers for the initial identification of OM and related hearing loss include remote area nurses, Aboriginal health workers and visiting audiologists. Post-diagnosis treatment is typically prescribed by remote area GPs and/or by otolaryngologists. However, lack of targeted training, the very wide scope of practice demanded of remote area nurses and Aboriginal health workers, and high levels of staff turnover often result in under-identification of OM in children in remote settings. The availability of ML diagnosis at the remote community level would support more confident and timely referrals to otolaryngologists and audiologists for treatment and intervention. ML diagnoses provided at regular intervals to otolaryngologists by telehealth could enhance the monitoring of a child’s ear health over time.

The potential for telehealth to change the delivery of audiology and ENT services for people living in remote settings has been acknowledged for many years.[28] Furthermore, ENT telehealth services have been shown to reduce costs for the health care system and provide access to ENT specialist care within clinically recommended timeframes.[29] However, it was not until the COVID-19 pandemic impacted healthcare delivery that clinicians became more positive about – and started more widely implementing – the use of ear and hearing telehealth services.[30] Evidence shows ENT specialists can use electronically stored otoscopic images collected during telehealth appointments to make accurate diagnoses.[31] However, clinicians can have differing opinions when diagnosing OM using still otoscopy images alone, with greater agreeance shown in the diagnosis of AOM and less when diagnosing OME.[32] The addition of a wider range of clinical data, including audiometry, tympanometry, and observations by nurses, significantly improved clinician's diagnostic accuracy and agreeance.[31, 32] Advances in technology, including the use of video otoscopy rather than still otoscopy,[33] has been shown to aid clinicians in the delivery of ear and hearing healthcare in remote areas. Given ML algorithms can accurately classify ear disease from otoscopic images,[34] and have been shown to more accurately classify otoscopy images than clinicians,[13] there is no doubt the integration of validated algorithms within existing telemedicine initiatives will support effective triage of ear disease and hearing loss. However, potential barriers to the successful delivery of telehealth services – including the integration of ML algorithms into telehealth – for ear and hearing healthcare must be overcome through collaboration between device manufacturers, clinicians, and professional bodies to ensure there is appropriate guidance, training, funding, and technology to support clinicians to deliver effective and appropriate care.[35-38]

Strengths and limitations

To the authors' knowledge, this is one of the first studies to incorporate multiple types of diagnosis data into a multimodal ML-based OM diagnosis system. Moreover, while there have been some recently published studies,[39, 40] this is the first to use Australian data and the first to use data from a population with a high incidence otitis media and which is often more complex than among other populations. Data were collected in a real-world context during a school-based ear health and hearing research study and screening program implemented over three years in a remote Australian location. The children participating in the study presented with OM disease spanning the wide spectrum of disease, however, often First Nations children have more complex otitis media presentations than non-Indigenous children. As such, our findings provide evidence for the feasibility of ML-based OM diagnosis among children experiencing complex OM and who live in remote locations.

Several limitations should be noted. The data were filtered by the auto-filter model, which was built based on a manual image selection by a non-ENT expert with no prior training. Although the data analyst worked with clinicians to understand the dataset, the selection could be problematic, especially for different ear conditions. This might directly affect the performance of the target model, in comparison with the other models in the literature, which have been built mostly by still-image of the ears that are well-selected and processed specifically by an otolaryngologist. The model could be more accurate if it was built using a dataset of still images captured from otoscopy directly. Furthermore, while understanding the clinical features driving the model performance was outside of scope this project, this warrants future investigation in future. An additional limitation was our stratification of data at class level, rather than at patient level, which may mean our model is affected by leakage bias, resulting in an overoptimistic prediction. Finally, this approach utilised a very simple mechanism of stacking multiple models in a single pipeline. However, more advanced ensemble techniques could be applied to get better performance. In the future, more advanced techniques, like feature fusion and decision fusion,[41] could be applied to build a more robust system to utilise multiple data sources and effectively combine results of different models to give better prediction on different types of OM.

CONCLUSION

This study combined multiple data types to boost the performance of an OM diagnosis system based on ML. The overall accuracy increased from 78% to 82% by combining tympanometry data with otoscopy data. In the future, multiple enhancements could be made to the system in the image processing, hyperparameter tuning, CNN architecture exploration, and other fusion approaches. More accurate systems with an effective deployment model could expose the technology capability in supporting the OM diagnosing procedure. This, therefore, has potential utility to improve diagnosis of children experiencing OM, especially for children living in remote locations.

LIST OF ABBREVIATIONS

AOM – acute otitis media

CNN - convolutional neural networks

COM – chronic otitis media

ENT – ear, nose and throat

GP – general practitioner

MCC - Matthews Correlation Coefficient

ML – machine learning

MLP – multilayer perceptron

OM – otitis media

OME – otitis media with effusion

ROI - region of interest

SVM - support vector machine

WMV - Windows Media Video

DECLARATIONS

Acknowledgements

The authors would like to acknowledge Dr Brianna Poirier and Dr Charmaine Woods for administrative support in the preparation of the ethics application and extraction of the research dataset from the original dataset.

The authors are non-Indigenous with differing levels of experience working in Australian First Nations Health. We thank the First Nations people who participated in the original data collection and whose data has been used in this study.

The authors acknowledge the people of the Anangu, Pitjantjatjara, and Yankunytjatjara Lands, of the Yalata Community, and of the Maralinga Tjarutja Lands as the Traditional Owners and Custodians of the places where the original data collection took place. The authors acknowledge the Kaurna people as the Traditional Owners and Custodians of the lands on which the analysis was conducted.

Conflicts Of Interests

No authors have any conflicts of interest to declare.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This project received a Flinders University Cross College grant.

Author Contributions

JHS and TL conceptualised the study. LS and ASC performed the original data collection. EO, ASC and LS provided clinical guidance. PPN and TL performed the data analysis. AM and JHS drafted the manuscript. All authors had input into, and approved, the final manuscript.

Data Sharing

Due to the nature of the ethical restrictions, access to the data is not available.

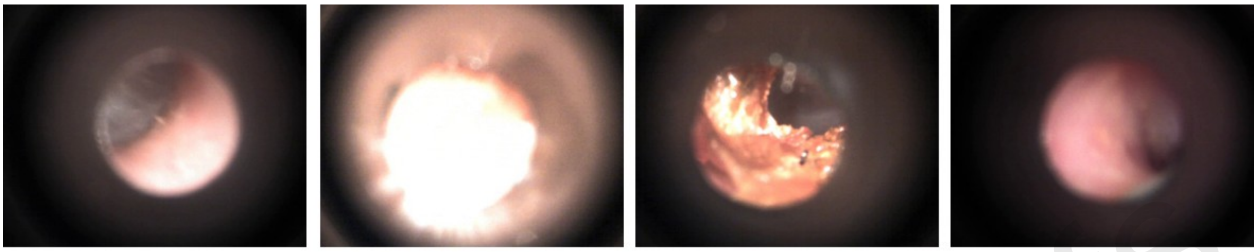
Table 1: Otoscopy data samples with matching tympanometry data (n=15,057).

Label	Tympanometry and otoscopy data samples (n)
Normal (OM negative)	8,293
Abnormal (OM positive)	6,764
Total	*15,057*

Table 2: Support Vector Machine (SVM) pipeline parameters and their tuning values.

Parameter	Values
Number of components for PCA to retain	10, 50 and 100
Kernel coefficient	rbf, linear and sigmoid
C value	0.1, 1 and 10

Figure 1: Extracted frames demonstrating poor quality otoscopy imaging.



(a) Outer ear (b) Too bright (c) Obscured by cerumen (d) Blurry

Figure 2: Schematic of the multimodal machine learning model. SVM: support vector machine, MLP: Multilayer perceptron.

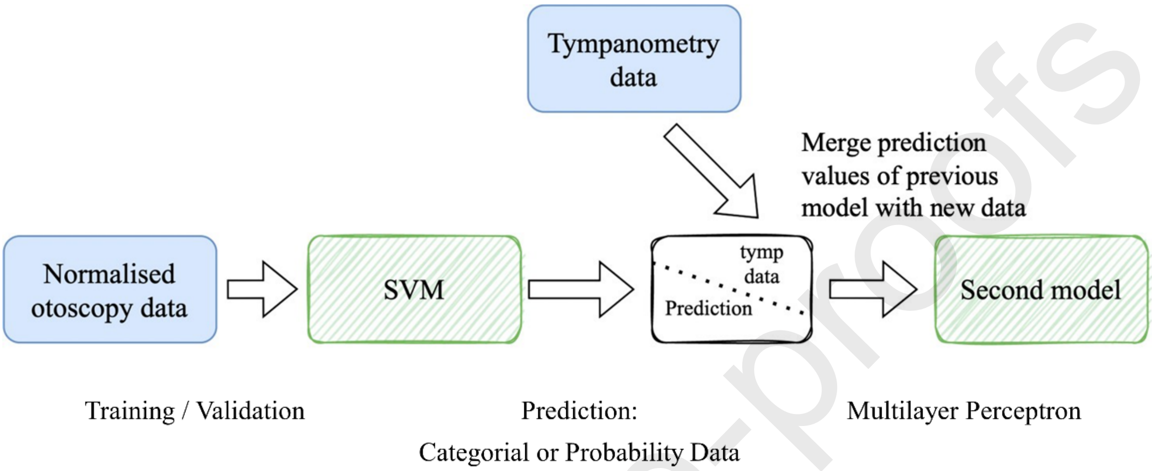
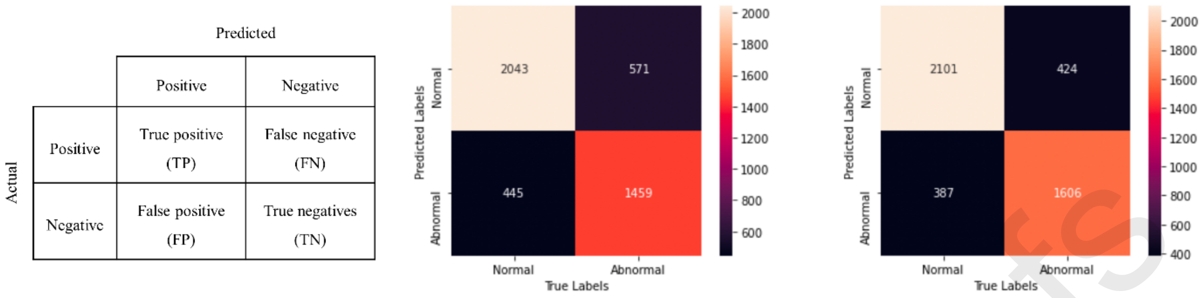


Figure 3: Confusion matrices for pipelines with: a) single SVM model and b) SVM + MLP models. Note: SVM = support vector machine, MLP = multilayer perceptron.



		Predicted	
		Positive	Negative
Actual	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negatives (TN)

a) Standard

b) SVM

Accuracy: 78%
Precision: 0.78
Sensitivity: 0.82
Specificity: 0.72
Recall: 0.77
F1-score: 0.80
MCC: 0.54

c) SVM +MLP

Accuracy: 82%
Precision: 0.83
Sensitivity: 0.84
Specificity: 0.79
Recall: 0.82
F1-score: 0.84
MCC: 0.64

REFERENCES

1. Monasta L, Ronfani L, Marchetti F, et al. Burden of disease caused by otitis media: systematic review and global estimates. *PLoS One*. 2012;7(4):e36226.

2. Schilder AGM, Chonmaitree T, Cripps AW, et al. Otitis media. *Nature Reviews Disease Primers*. 2016/09/08 2016;2(1):16063.

3. Homøe P, Heidemann CH, Damoiseaux RA, et al. Panel 5: Impact of otitis media on quality of life and development. *Int J Pediatr Otorhinolaryngol*. Mar 2020;130 Suppl 1(Suppl 1):109837.

4. Habib AR, Wong E, Sacks R, Singh N. Artificial intelligence to detect tympanic membrane perforations. *J Laryngol Otol*. Apr 2020;134(4):311-315.

5. Abbott P, Rosenkranz S, Hu W, Gunasekera H, Reath J. The effect and acceptability of tympanometry and pneumatic otoscopy in general practitioner diagnosis and management of childhood ear disease. *BMC Family Practice*. 2014/12/12 2014;15(1):181.

6. Buchanan CM, Pothier DD. Recognition of paediatric otopathology by General Practitioners. *Int J Pediatr Otorhinolaryngol*. May 2008;72(5):669-73.

7. Pichichero ME, Poole MD. Comparison of performance by otolaryngologists, pediatricians, and general practioners on an otoendoscopic diagnostic video examination. *Int J Pediatr Otorhinolaryngol*. Mar 2005;69(3):361-6.

8. Asher E, Leibovitz E, Press J, Greenberg D, Bilenko N, Reuveni H. Accuracy of acute otitis media diagnosis in community and hospital settings. *Acta Paediatr*. Apr 2005;94(4):423-8.

9. Pichichero ME, Poole MD. Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. *Arch Pediatr Adolesc Med*. Oct 2001;155(10):1137-42.

10. Livingstone D, Talai AS, Chau J, Forkert ND. Building an Otoscopic screening prototype tool using deep learning. *J Otolaryngol Head Neck Surg*. Nov 26 2019;48(1):66.

11. Ngombu S, Binol H, Gurcan MN, Moberly AC. Advances in Artificial Intelligence to Diagnose Otitis Media: State of the Art Review. *Otolaryngol Head Neck Surg*. Apr 2023;168(4):635-642.

12. Wu Q, Wang X, Liang G, et al. Advances in Image-Based Artificial Intelligence in Otorhinolaryngology-Head and Neck Surgery: A Systematic Review. *Otolaryngol Head Neck Surg*. Nov 2023;169(5):1132-1142.

13. Habib AR, Kajbafzadeh M, Hasan Z, et al. Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. *Clin Otolaryngol*. May 2022;47(3):401-413.

14. Ding X, Huang Y, Tian X, Zhao Y, Feng G, Gao Z. Diagnosis, Treatment, and Management of Otitis Media with Artificial Intelligence. *Diagnostics*. 2023;13(13). doi:10.3390/diagnostics13132309 https://mdpi-res.com/d_attachment/diagnostics/diagnostics-13-02309/article_deploy/diagnostics-13-02309-v2.pdf?version=1688981775

15. Cao Z, Chen F, Grais EM, et al. Machine Learning in Diagnosing Middle Ear Disorders Using Tympanic Membrane Images: A Meta-Analysis. *Laryngoscope*. Apr 2023;133(4):732-741.

16. Vesoulis ZA, Husain AN, Cole FS. Improving child health through Big Data and data science. *Pediatr Res*. Jan 2023;93(2):342-349.

17. Kanka N, Murakoshi M, Hamanishi S, et al. Longitudinal changes in dynamic characteristics of

neonatal external and middle ears. *Int J Pediatr Otorhinolaryngol*. Jul 2020;134:110061.

18. Mishra SK, Dinger Z, Renken L. Maturation of middle ear transmission in children. *Hear Res*. Feb 2017;344:62-67.

19. Jervis-Bardy J, Sanchez L, Carney AS. Otitis media in Indigenous Australian children: review of epidemiology and risk factors. *J Laryngol Otol*. Jan 2014;128 Suppl 1:S16-27.

20. Kong K, Lannigan FJ, Morris PS, Leach AJ, O'Leary SJ. Ear, nose and throat surgery: All you need to know about the surgical approach to the management of middle-ear effusions in Australian Indigenous and non-Indigenous children. *J Paediatr Child Health*. Nov 2017;53(11):1060-1064.

21. Moberly AC, Zhang M, Yu L, et al. Digital otoscopy versus microscopy: How correct and confident are ear experts in their diagnoses? *J Telemed Telecare*. Aug 2018;24(7):453-459.

22. Binol H, Moberly A, Niazi MKK, et al. *Decision Fusion on Image Analysis and Tympanometry to Detect Eardrum Abnormalities*. 2020.

23. Patrick MB, Johannes BR, David EB, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ : British Medical Journal*. 2015;351:h5527.

24. Sanchez L, Carney AS, Esterman A, Sparrow K, Turner D. Does access to saltwater swimming pools reduce ear pathology and hearing loss in school children of remote arid zone aboriginal communities? A prospective 3-year cohort study. *Clin Otolaryngol*. Sep 2019;44(5):736-742.

25. Browning G, Weir J, Kelly G, Swan IRC. Chronic Otitis Media. In: Watkinson JC, Clake RW, eds. *Scott-Brown's Otorhinolaryngology Head & Neck Surgery*. 8th Edition ed. CRC Press; 2019:977-1019.

26. Almeida H, Meurs MJ, Kosseim L, Butler G, Tsang A. Machine learning for biomedical literature triage. *PLoS One*. 2014;9(12):e115892.

27. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020/01/02 2020;21(1):6.

28. Swanepoel de W, Hall JW, 3rd. A systematic review of telehealth applications in audiology. *Telemed J E Health*. Mar 2010;16(2):181-200.

29. Altamimi AA, Brennan-Jones CG, Robinson M, et al. A hospital-based asynchronous ENT telehealth service for children with otitis media: Cost-minimisation and improved access. *J Telemed Telecare*. Jan 31 2024:1357633X231223994.

30. Eikelboom RH, Bennett RJ, Manchaiah V, et al. International survey of audiologists during the COVID-19 pandemic: use of and attitudes to telehealth. *Int J Audiol*. Apr 2022;61(4):283-292.

31. Habib AR, Perry C, Crossland G, 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*. May 2023;168:111494.

32. Sundgaard JV, Värendh M, Nordström F, et al. Inter-rater reliability of the diagnosis of otitis media based on otoscopic images and wideband tympanometry measurements. *Int J Pediatr Otorhinolaryngol*. Feb 2022;153:111034.

33. Quick ME, Brennan-Jones CG, Kuthubutheen J. Remote paediatric ear examination comparing video-otoscopy and still otoscopy clinician rated outcomes. *Int J Pediatr Otorhinolaryngol*. Feb 2024;177:111871.

34. Habib AR, Crossland G, Patel H, et al. An Artificial Intelligence Computer-vision Algorithm to Triage Otoscopic Images From Australian Aboriginal and Torres Strait Islander Children. *Otol Neurotol*. Apr 1 2022;43(4):481-488.

35. Evans T, Nejman T, Stewart E, Windmill I. Increasing Pediatric Audiology Services via Telehealth. *Semin Hear*. May 2021;42(2):136-151.

36. Gonzalez JN, Axiotakis LG, Jr., Yu VX, Gudis DA, Overdevest JB. Practice of Telehealth in Otolaryngology: A Scoping Review in the Era of COVID-19. *Otolaryngol Head Neck Surg*. Mar 2022;166(3):417-424.

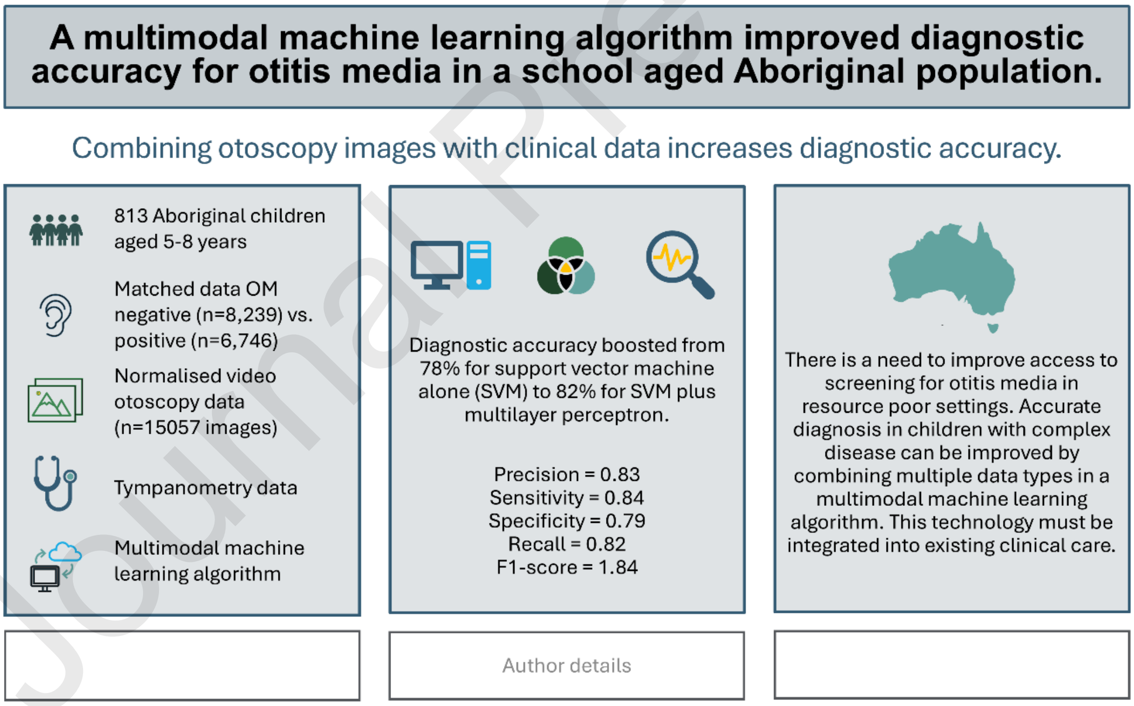
37. Parmar B, Beukes E, Rajasingam S. The impact of COVID-19 on provision of UK audiology services & on attitudes towards delivery of telehealth services. *Int J Audiol*. Mar 2022;61(3):228-238.

38. Swanepoel de W, Clark JL, Koekemoer D, et al. Telehealth in audiology: the need and potential to reach underserved communities. *Int J Audiol*. Mar 2010;49(3):195-202.

39. Grais EM, Nie L, Zou B, et al. An advanced machine learning approach for high accuracy automated diagnosis of otitis media with effusion in different age groups using 3D wideband acoustic immittance. *Biomedical Signal Processing and Control*. 2024;87

40. Sundgaard JV, Hannemose MR, Laugesen S, et al. Multi-modal deep learning for joint prediction of otitis media and diagnostic difficulty. *Laryngoscope Investig Otolaryngol*. Feb 2024;9(1):e1199.

41. Boulahia SY, Amamra A, Madi MR, Daikh S. Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition. *Machine Vision and Applications*. 2021;32(6)



A multimodal machine learning algorithm improved diagnostic accuracy for otitis media in a school aged Aboriginal population.

Combining otoscopy images with clinical data increases diagnostic accuracy.

813 Aboriginal children aged 5-8 years

Matched data OM negative (n=8,239) vs. positive (n=6,746)

Normalised video otoscopy data (n=15057 images)

Tympanometry data

Multimodal machine learning algorithm

Diagnostic accuracy boosted from 78% for support vector machine alone (SVM) to 82% for SVM plus multilayer perceptron.

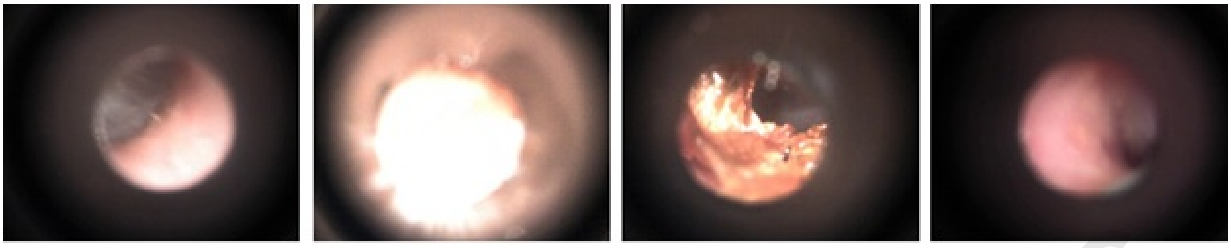
Precision = 0.83
Sensitivity = 0.84
Specificity = 0.79
Recall = 0.82
F1-score = 1.84

There is a need to improve access to screening for otitis media in resource poor settings. Accurate diagnosis in children with complex disease can be improved by combining multiple data types in a multimodal machine learning algorithm. This technology must be integrated into existing clinical care.

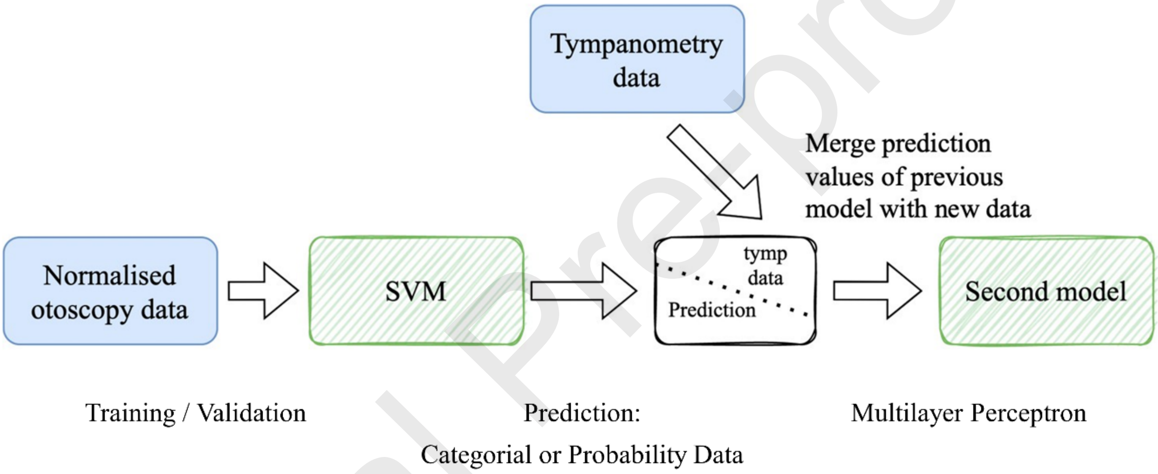
Author details

SIGNIFICANCE

Most otitis media diagnostic algorithms only use image data, whereas we demonstrate using multiple data types in a diagnostic algorithm is possible. Compared to diagnosis based solely on otoscopy data, combining otoscopy and tympanometry data increased the diagnostic accuracy of the ML algorithm. A multimodal diagnostic algorithm could support the accurate diagnosis of otitis media in resource poor settings, such as for First Nations children living in rural and remote areas.

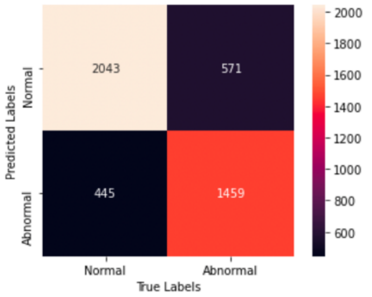


(a) Outer ear (b) Too bright (c) Obscured by cerumen (d) Blurry



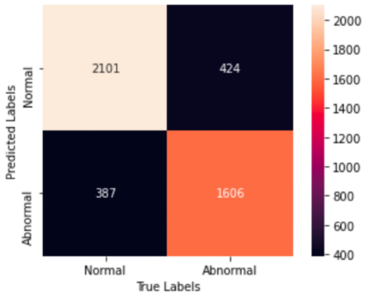
		Predicted	
		Positive	Negative
Actual	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negatives (TN)

a) Standard



b) SVM

Accuracy: 78%
Precision: 0.78
Sensitivity: 0.82
Specificity: 0.72
Recall: 0.77
F1-score: 0.80
MCC: 0.54



c) SVM +MLP

Accuracy: 82%
Precision: 0.83
Sensitivity: 0.84
Specificity: 0.79
Recall: 0.82
F1-score: 0.84
MCC: 0.64