A Multimodal Machine Learning Framework for Diagnosis of Otitis Media with Effusion using 3D Wideband Acoustic Immittance

1st Tariq Rahim*
Department of Computer Science
Kingston University London
United Kingdom
T.Rahim@kingston.ac.uk

2nd Fei Zhao
Cardiff School of Sport and Health Sciences)
Cardiff Metropolitan University
Cardiff, United Kingdom
fzhao@cardiffmet.ac.uk

***Abstract*—Wideband acoustic immittance (WAI) technology has been known for over a decade, delivering an enhanced diagnosis of middle ear (ME) diseases across a wider frequency range than standard tympanometry. Nevertheless, its clinical usage confronts the limitations of restricted interpretation and insufficient explanation of the WAI outcomes. This paper proposes a multimodal machine learning (MML) approach for classifying ME diseases into normal ear and ear with abnormalalty i.e., otitis media with effusion. The proposed MML model is grounded on the integration of a 3 layered convolutional neural network and a multi-layer perception network. The outcomes exhibited that the proposed MML model surpasses the available methods by achieving 98.27% accuracy for classifying ME diseases using the WAI measurements.**

***Index Terms*—Accuracy, convolutional neural network, machine learning, multi-layer perception , Wideband acoustic immittance.**

I. Introduction

Otitis media (OM) is a middle ear (ME) inflammation that is split clinically into two diagnostic classes: otitis media with effusion (OME) and acute otitis media (AOM). AOM is defined as a critical disease with a prompt onset, whereas OME is defined as a liquid in the ME. Both of these classes are prevalent among kids but OM is a considerably common reason for consultations in the case of kids at primary care doctors [1]. The lack of apparent contagious symptoms i.e., fever and earache, and inadequate investigative precision can lead to delayed investigation. Therefore, it can cause language and speech growth delays with potential behavioral and academic issues [2]. At the 2012 Eriksholm workshop, the Wideband Acoustic Immittance (WAI) term was selected to define a class of middle-ear measurements based on acoustic power and immittance, including energy absorbance (EA) and energy reflectance. Interacoustics (Denmark) introduced the Titan system, a commercially available instrument developed for assessing middle-ear conditions. The system estimates the acoustic transfer function, represented as EA, across a broad frequency span (226 Hz-8000 Hz). The system's (Titan) results are displayed via both 2D and 3D plots of EA, spanning various pressures and frequencies [3]. Contemporary investigations have also demonstrated the significant benefits of WAI in delivering added information on ME function employing a vast frequency range as a function of pressure at ambient pressure and peak pressure, plotted in 2D and 3D graphs. These graphs from WAI allow the clinician to sufficiently comprehend the dynamic features of the ME by identifying specific tympanometric patterns related to the ME pathologic change [4]. The primary hurdle in maximizing the WAI application lies in the extensive amount of data, which contains thousands of absorbance values across the pressure-frequency plane. While this data is useful for comprehending ME conditions, it poses problems for physicians in investigating, understanding, and classifying ears as abnormal or normal.

II. Related Work

Artificial Intelligence (AI) including its subclasses such as machine learning (ML), and deep learning (DL) are employed for feature engineering, prediction, and classification in different applications including medical imaging [5]. To handle the current problem of analyzing the ME conditions, AI, ML, and DL techniques have been used for the processing of complex WAI data [6]. A CNN-based model using an augmentation technique is presented to enhance the accuracy of the classification of OM using WAI measurements. The 1014 measurements (WAI) were collected from patients ages ranging from two months to twelve years and were divided into 3 classes of diagnosis such as AOM, OME, and normal ear. The outcomes demonstrated an accuracy of 92.6% in specifying OM. Nevertheless, the accuracy for classification was low to 79% in distinguishing the OME and AOM [7]. Moreover, for tackling the noise within ear images, a DL-based approach integrated with Bayesian optimization is implemented where the input images were labeled either infected or not infected. The results demonstrated effective outcomes with the other benchmarked models [8].

Furthermore, the ML approach is employed to specify the WAI absorbance features across different frequency-pressure regions in the ears with OME and normal ME to allow the diagnosis of ME conditions automatically [9]. Employing

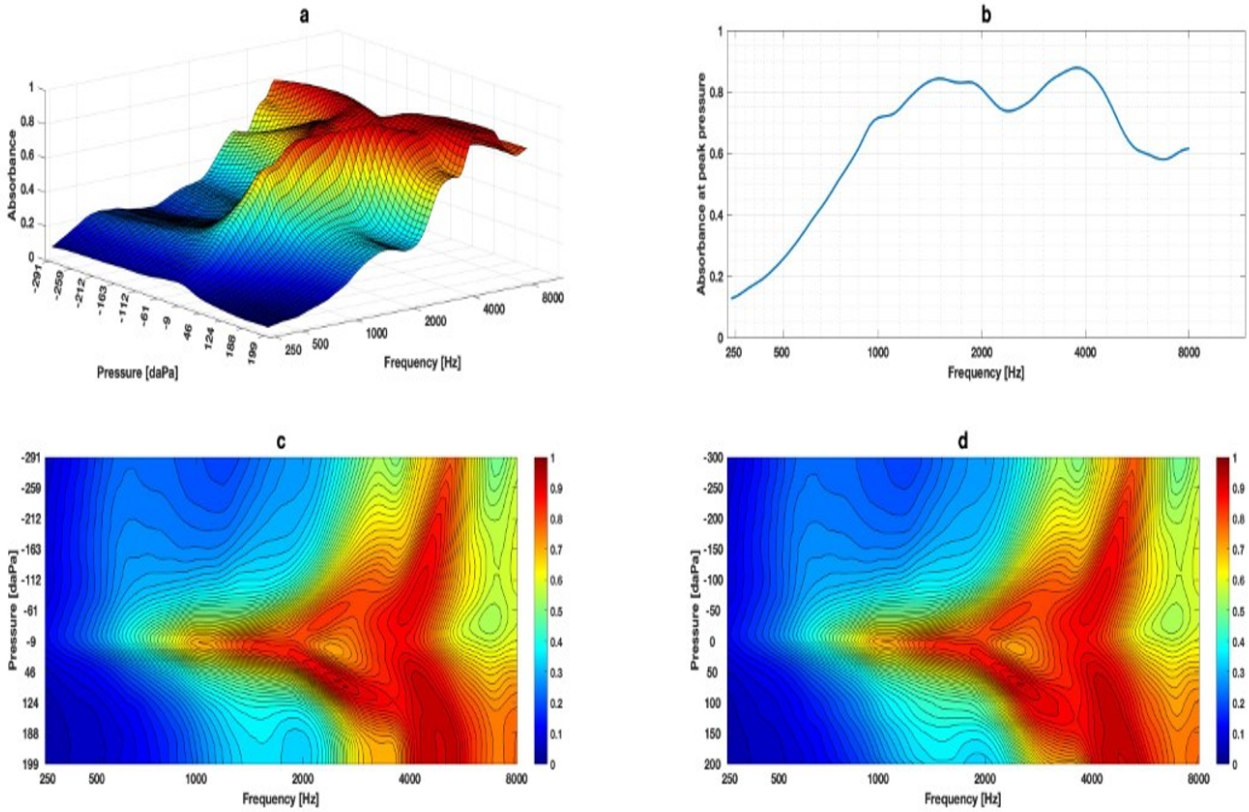

Fig. 1: A depiction of 3D WAI imaging and data pre-processing: (a) 3D WAI data with normal ME function from an adult participant; (b) 2D frequency-absorbance plot at peak pressure from the same participant; (c) 2D frequency-pressure image transformed from (a); (d) 2D frequency-pressure image from (a) after Y-axis pressure interpolation.

class activation mapping, a two-stage DL-based classification approach is implemented for the automatic diagnosis of OM employing tympanic membrane images. The proposed approach attained an overall accuracy of 93.4% employing ResNet50 as the backbone network, while the F1 Score of classification for OME was 96.8% and 94.3% for the normal images. Our recent work presents a cascaded DL approach that uses a convolutional neural network (CNN) followed by a self-attention approach. The first stage classifies OME into specific age groups followed by a self-attention approach that classifies the data's discriminative parts. The two-stage ML technique achieved classification accuracy of 96.6%, 94.1%, and 90.7% for the three age groups, respectively [3].

Multimodal machine learning (MML) is a class of DL operating on the combination of images, videos, data, audio, and signals such as (heart rate). In distinction, unimodal models can process exclusively one data type, such as images or text (commonly defined as feature vectors). MML is distinct from fusing unimodal models trained independently. It integrates information from other modalities to produce better predictions [10]. Although the MML has shown considerable success in diagnostics and medical imaging, its application to WAI data for the classification of ME conditions remains unexplored. WAI, which supplies a vast spectrum of acoustic responses, has displayed the potential to differentiate between normal and pathological ears. Yet, existing models often fail to effectively fuse WAI data with other modalities that could improve diagnostic precision. The lack of a robust multimodal framework integrating temporal, acoustic, and visual data shows a critical research gap, making it challenging for physicians to leverage all available data for precise middle-ear disease prediction. To address this gap, this initial work proposes an MML framework for this first using WAI data for the classification of ME conditions into normal and abnormal as OME.

III. Proposed MML Approach for the Classification

This section details the materials and methods adopted for the classification of ME diseases into normal and abnormal as OME.

A. Dataset Specification

The WAI dataset comprises 1177 sets of WAI measurements and was collected from volunteers and clients in 7 hospitals in China i.e., Chongqing, Shanghai, Xuzhou, Beijing, and Guangzhou. Of the total data set, 551 WAI measurements are normal while 626 WAI measurements are OME ears.In

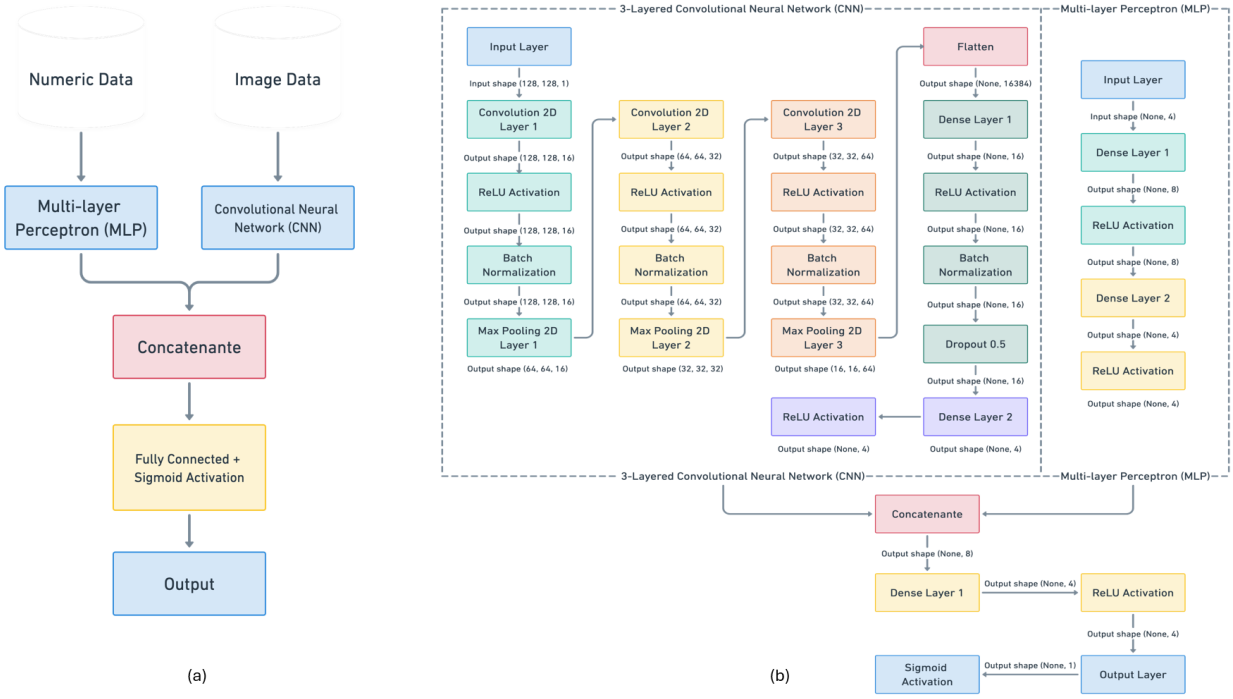

Fig. 2: Illustration of the proposed MML model where (a) display the general flow of the MML while (b) display the proposed MML model.

addition to the inadequate quality of the measurements of WAI i.e., having missing pressure weights exclusion, the data from infants and neonates younger than 1-year-old were also removed [3]. The development of an ML-based model using the WAI plots of pressure- -frequency has been detailed in our previous work i.e., Interacoustic from Denmark provided Titan IMP440. It is operated to estimate EA for a broad spectrum of frequencies (Hz) and pressure, spanning from +200 daPa to -300 daPa. Figure. 1a depicts an instance of a plot for WAI in 3D for a normal ME function having an age of 25 years of participants. The first dimension represents frequency, varying from 226 Hz to 8000 Hz in intervals of 1/24-octave; the second dimension represents pressure, running from -300 to +200 daPa; and the third dimension represents EA, ranging from 0 to 1. Fig. 1b depicts an EA curve at peak pressure across a broad frequency range while Fig. 1c illustrates a 2D image employing preprocessing pressure and the domains of frequency that correspond to the WAI data as in Fig. 1a. More detailed explanation can be found in [3].

B. Implementation of the proposed MML Model

The general flow of the proposed MML for the classification of WAI data into normal ear and ear with OME is shown in Fig. 2(a) where it can be seen that the input is fed both in the form of numeric structured and image (grayscale) data. Before implementation of the proposed MML model, extensive preprocessing is done to remove outliers in the WAI data. The relationship of pressure, frequency, and absorbance aided

TABLE I: Simulation parameters for the proposed MML model

Network parameter	Configuration values
Input size	128 x 128
Learning rate ((η))	0.001
Batch size	64
Optimizer	Adam
Activation function	ReLU
Decay	$1e-3/200$

in this process. It was found that some of the columns in the WAI measurements were missing due to the handling of the Titan IMP440 device and patient comfort. The data was divided into 70. 20, and 10 percent for training, testing, and validation. A multi-layer perception (MLP) and CNN model is employed that is concatenated after operating on the WAI measurements. The detailed architecture of the proposed MML is depicted in Fig. 2(b). For the CNN model, it employs a 3-layered convolutional layer of two dimensions while the MLP has one layer followed by two dense layers. For both CNN and MLP, the rectifier linear unit (ReLU) is used as an activation function. The features extracted from both CNN and MLP are then concatenated generating a feature vector having a size of 8 i.e., 4 from each CNN and MLP. These features are further processed by a dense layer and ReLU to integrate the information both from the grayscale image and structure data. Finally, a dense layer having a sigmoid activation is used for binary classification of the WAI measurements into normal

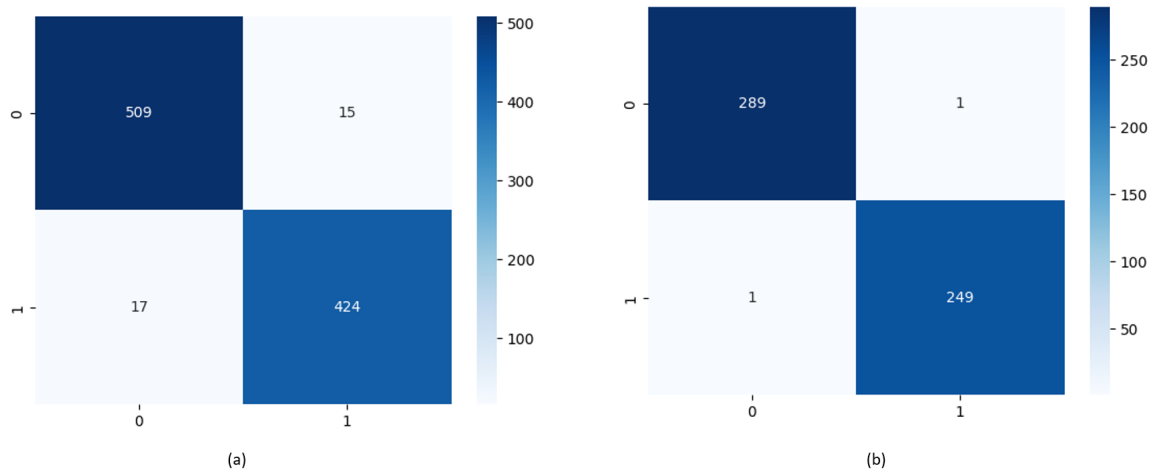

Fig. 3: Confusion matrix where (a) represents confusion matrix for all data while (b) represents confusion matrix for training data.

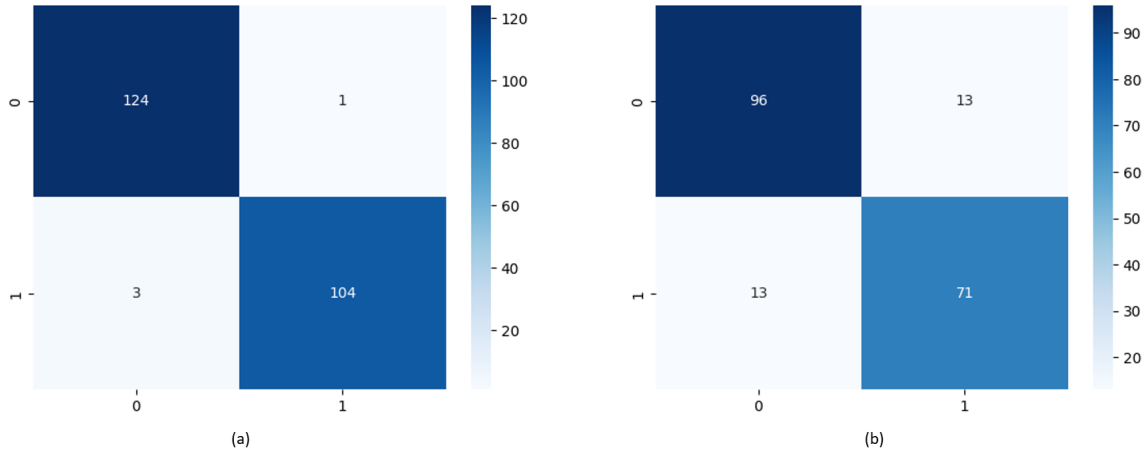

Fig. 4: Confusion matrix where (a) represents confusion matrix for validation while (b) represents confusion matrix for testing data.

ears and ears with OME. The loss function is the binary-entropy loss and is given as:

$$\mathcal{L}(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{y}_i\right) + (1 - y_i)\log\left(1 - \hat{y}_i\right)\right] \quad (1)$$

where y_i is the true label i.e., either 0 or 1, $\hat{y}_i$ is for the predicted probability i.e., output from the sigmoid function, and N is the number of samples in the batch.

$$z = \text{concat}\left(z_{\text{CNN}}, z_{\text{MLP}}\right) \quad (2)$$

where z_{CNN} and z_{MLP} are the resultant vectors from the respective CNN and MLP branches. The simulation parameters used for the proposed MML model are given in Table 1.

IV. Results and Discussion

Accuracy, precision, sensitivity or recall, specificity, and F1-score are employed for the model evaluation. The following parameters are defined as follows:

- **True Positive (TP):** The TP stands for the case/cases when the model correctly predicts the positive class. In other words, if there is a normal ear or ear with OME, the proposed model accurately classifies it.
- **False Positive (FP):** The FP stands for the case/cases when the model incorrectly predicts the positive class. In other words, if there is no ear with OME or normal ear, the proposed model accurately classifies.
- **True Negative (TN):** The TN stands for the case/cases when the model correctly predicts the negative class.
- **False Negative (FN):** The FN stands for the case/cases when the model incorrectly predicts the negative class.

Employing these parameters, the subsequent metrics of performance can be computed to assess the performance of the presented MML mode.

TABLE II: Simulation parameters for the proposed MML model

Model	Performance Metrics				
	Accuracy	Precision	Recall	Specificity	F1-score
CNN [9]	82.00	83.00	89.00	X	86.00
2D network [7]	92.60	X	92.20	92.90	92.60
FNN	81.00	83.00	89.00	X	86.00
2-stage attention DL [3]	96.80	X	97.70	95.40	96.80
Proposed MML model	98.27	99.00	97.20	99.20	98.10

Precision: This metric of performance assess how precisely the model classify normal ear or ear with OME.

$$\text{Precision (Pre)} = \frac{TP}{TP + FP} \times 100 \tag{3}$$

- **Sensitivity**: This metric of performance is also called recall and assesses the proportion of the actual normal or ear with OME classified correctly.

$$\text{Sensitivity or recall (recall)} = \frac{TP}{TP + FN} \times 100 \tag{4}$$

- **F1-score**: This performance metric is the harmonic average among precision and recall and spans between 0 to 1. F1-score is calculated as:

$$\text{F1-score} = \frac{2 \times \text{Sen} \times \text{Pre}}{\text{recall} + \text{Pre}} \times 100 \tag{5}$$

Figure 3 depicts the confusion matrix after implementing the proposed MML model for the classification of ME conditions into normal and ear, where Fig. 3(a) shows the confusion matrix for overall data while Fig. 3 shows the confusion matrix for the training data. Furthermore, Fig. 4 illustrates the confusion matrix for validation (Fig.4(a)) while Fig. 4 (b) illustrates the confusion matrix for the testing phase. The performance of the model is reflected in the form of high accuracy and the model achieved accuracy of 99.6%, 98.27%, and 86.52% for training, validation, and testing, respectively.

Furthermore, performance metrics such as accuracy, precision, sensitivity (recall), specificity, and F1-score are selected to evaluate the performance of the proposed MML model. As shown in Table 2, the presented MML is extensively benchmarked with the available methods employed for the classification of ME disease. Unlike the previous approaches, this paper uses specificity and precision to evaluate the model performance.

V. Conclusion

This paper is an initial attempt in the development of a multi-modal framework by proposing a DL-based approach i.e., MMML for the classification of ME disease into normal ear and ear with OME employing WAI measurements. The proposed MML model takes input both in the form of numeric and images using WAI measurements and performing the decision process. The initial results demonstrated outperformed results in terms of the performance metrics such as accuracy, precision, recall, specificity, and F1-score. These results are benchmarked with the current available studies and models. For future work, age categorization is an issue where the specific dataset comprises three age categories and the future work will focus on generating a higher accurate prediction, especially for the age category less than 3 years.

Acknowledgment

This work is supported by Kingston University London, United Kingdom.

References

[1] J. V. Sundgaard, M. R. Hannemose, S. Laugesen, P. Bray, J. Harte, Y. Kamide, C. Tanaka, R. R. Paulsen, and A. N. Christensen, "Multi-modal deep learning for joint prediction of otitis media and diagnostic difficulty," *Laryngoscope Investigative Otolaryngology*, vol. 9, no. 1, p. e1199, 2024.

[2] A. Espeso, D. Owens, and G. Williams, "The diagnosis of hearing loss in children: Common presentations and investigations," *Current Paediatrics*, vol. 16, no. 7, pp. 484–488, 2006.

[3] E. M. Grais, L. Nie, B. Zou, X. Wang, T. Rahim, J. Sun, S. Li, J. Wang, W. Jiang, Y. Cai *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*, vol. 87, p. 105525, 2024.

[4] Y. Cai, J.-G. Yu, Y. Chen, C. Liu, L. Xiao, E. M. Grais, F. Zhao, L. Lan, S. Zeng, J. Zeng *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*, vol. 11, no. 1, p. e041139, 2021.

[5] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[6] L. Nie, C. Li, F. Marzani, H. Wang, F. Thibouw, and A. B. Grayeli, "Classification of wideband tympanometry by deep transfer learning with data augmentation for automatic diagnosis of otosclerosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 2, pp. 888–897, 2021.

[7] J. V. Sundgaard, P. Bray, S. Laugesen, J. Harte, Y. Kamide, C. Tanaka, A. N. Christensen, and R. R. Paulsen, "A deep learning approach for detecting otitis media from wideband tympanometry measurements," *IEEE journal of biomedical and health informatics*, vol. 26, no. 7, pp. 2974–2982, 2022.

[8] I. M. Mehedi, M. S. Hanif, M. Bilal, M. T. Vellingiri, and T. Palaniswamy, "Artificial intelligence with deep learning based automated ear infection detection," *IEEE Access*, 2024.

[9] E. M. Grais, X. Wang, J. Wang, F. Zhao, W. Jiang, Y. Cai, L. Zhang, Q. Lin, and H. Yang, "Analysing wideband absorbance immittance in normal and ears with otitis media with effusion using machine learning," *Scientific Reports*, vol. 11, no. 1, p. 10643, 2021.

[10] N. Srivastava and R. R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," *Advances in neural information processing systems*, vol. 25, 2012.