Journal of Otology 20 (2025) 26 – 32

Journal of Otology

journal homepage: www.sciopen.com/journal/1672-2930

Developing a diagnostic support system for audiogram interpretation using deep learning-based object detection

Titipat Achakulvisut[1], Suchanon Phanthong[2], Thanawut Timpitak[1], Kanpat Vesessook[3], Sirinan Junthong[2], Withita Utainrat[2] and Kanokrat Bunnag[2] ✉

[1] *Department of Biomedical Engineering, Faculty of Engineering, Mahidol University*

[2] *Department of Otolaryngology, Faculty of Medicine Vajira Hospital, Navamindradhiraj University*

[3] *International Community School, Bangkok, Thailand*

ABSTRACT

Objective: To develop and evaluate an automated system for digitizing audiograms, classifying hearing loss levels, and comparing their performance with traditional methods and otolaryngologists' interpretations.Designed and Methods: We conducted a retrospective diagnostic study using 1,959 audiogram images from patients aged 7 years and older at the Faculty of Medicine, Vajira Hospital, Navamindradhiraj University. We employed an object detection approach to digitize audiograms and developed multiple machine learning models to classify six hearing loss levels. The dataset was split into 70% training (1,407 images) and 30% testing (352 images) sets. We compared our model's performance with classifications based on manually extracted audiogram values and otolaryngologists' interpretations.Result: Our object detection-based model achieved an F1-score of 94.72% in classifying hearing loss levels, comparable to the 96.43% F1-score obtained using manually extracted values. The Light Gradient Boosting Machine (LGBM) model is used as the classifier for the manually extracted data, which achieved top performance with 94.72% accuracy, 94.72% f1-score, 94.72 recall, and 94.72 precision. In object detection based model, The Random Forest Classifier (RFC) model showed the highest 96.43% accuracy in predicting hearing loss level, with a F1-score of 96.43%, recall of 96.43%, and precision of 96.45%.Conclusion: Our proposed automated approach for audiogram digitization and hearing loss classification performs comparably to traditional methods and otolaryngologists' interpretations. This system can potentially assist otolaryngologists in providing more timely and effective treatment by quickly and accurately classifying hearing loss.

KEYWORDS

Audiogram, Deep machine learning, Training set, Validation set, Testing set, Automatic Machine Learning (AutoML), Random Forest Classifier (RFC), Support Vector Machine (SVM), XGBoost.

1 Introduction

Hearing loss is a significant global health concern that adversely affects individuals' quality of life and imposes a substantial burden on healthcare systems(Committee on et al. 2016; Olusanya, Neumann, and Saunders 2014). Currently, approximately 6% of the global population is affected by hearing impairment, and this prevalence is projected to rise to 10% by year 2050. Various factors contribute to this increasing trend, including urban noise pollution, ototoxic medications, and inner ear infections(Chadha 2018). The diagnosis of hearing loss employs multiple methods, with standard audiometry being the most widely used and accepted technique. This method categorizes hearing loss into conductive, sensorineural, or mixed types and assesses the severity of the condition(Davies 2016).

Moreover, the availability of otolaryngologists and trained audiologists, who are essential for diagnosing and interpreting hearing conditions, is limited in some resource-constrained hospitals. This scarcity poses a significant problem for individuals lacking access to appropriate treatment. In certain regions, particularly in Southeast Asia and Africa, there are reports of fewer than one audiologist per million people, exacerbating the challenges in addressing hearing health effectively(Organization 2013).

Accessibility to diagnosis and treatment for hearing loss remains a challenge, particularly in remote or underdeveloped regions with limited healthcare resources(Organization 2013). To address this issue, an automated hearing test via smartphone has been developed and is widely used as a screening tool in many countries(Wasmann et al. 2021), including Thailand(Bunnag et al. 2023). However, interpreting these tests remains challenging due to the shortage of audiologists and otolaryngologists, which further complicates effective diagnosis and treatment.

Peer review under responsibility of PLA General Hospital Department of Otolaryngology Head and Neck Surgery.

* Corresponding author Kanokrat Bunnag; Kanokrat.b@nmu.ac.th

https://doi.org/10.26599/JOTO.2025.9540005

Received 4 August 2024; Accepted 16 January 2025

The audiogram is a commonly used clinical evaluation tool for diagnosis hearing loss(Committee on et al. 2016). It is typically created using an audiometer that generates pure tones at various frequencies ranging from 250 Hz to 8000 Hz to gauge the patient's hearing capabilities (Olusanya, Neumann, and Saunders 2014). The results are generally used to diagnose hearing loss with the Hughson-Westlake (HW) procedure (Tsilivigkos et al. 2023).

Deep learning has been increasingly applied within the field of otolaryngology(Crowson, Ranisau, et al. 2020; Tsilivigkos et al. 2023; You et al. 2020) and biomedical engineer(Charih and Green 2022; Elkhouly et al. 2023; Kassjański, Kulawiak, and Przewoźny 2022; Yang et al. 2024). Crowson et al.(Crowson, Lee, et al. 2020) utilized ResNet-101 to classify types of hearing loss with an accuracy rate of 97.5%; however, the severity of hearing loss was not classified. Other approaches, such as unsupervised spectral clustering to classify audiograms based on detected shapes and Gaussian process regression to find the boundary between undetected and detected tones, are also employed in this context(Charih and Green 2022; Elkhouly et al. 2023; Kassjański, Kulawiak, and Przewoźny 2022).

In some instances, using image classification may not be suitable when otolaryngologists need access to the extracted data from the audiogram. Extracting numerical values from audiograms for classification is easier to interpret as it aligns with how otolaryngologists classify hearing loss. Yang et al.(Yang et al. 2024) employed YOLOv5 and OCR technologies to digitize audiograms, achieving a 97.68% accuracy rate for PTA symbols, yet the analysis did not include categorization of hearing loss type and severity.

Our study aimed to employ deep learning techniques to analyze both the severity and type of hearing loss, thereby rendering the hearing test results more comprehensive and applicable than prior research. By combining an object detection model with a classifier, we can significantly improve access to preliminary diagnoses and prompt treatments, potentially reducing the workload on healthcare professionals, including otolaryngologists and audiologists. This approach can bridge the gap in healthcare delivery, enabling early detection and intervention, thus improving outcomes for individuals with hearing impairment in underserved areas.

2 Materials and Methods

A retrospective diagnostic study was conducted after obtaining approval from the Research Ethics Review Committee for Research Involving Human Subjects at the Faculty of Medicine, Vajira Hospital (IRB No.: 123/65E). The study recruited 1959 participants, aged over 7 years, who underwent an audiogram at Vajira Hospital, Navamindradhiraj University, between July 1, 2021, and June 30, 2022. The exclusion criteria included audiograms aged under 7 years old, no audiologic record, or audiograms from another hospital.

The sample size(Daniel 1999; Rutterford, Copas, and Eldridge 2015) was determined by using the formula for estimating an infinite population proportion. Crowson MG et al.(Crowson, Lee, et al. 2020) using deep learning for automatic audiogram interpretation showed least accuracy (96.0%) with Resnet-32. (As below)

$$n = \frac{Z^2_{\alpha/2} P(1-P)}{d^2}$$

a = 0.05 (two-sides test), $Z_{0.025} = 1.96$, d = 0.01
P = accuracy was Resnet-32 at 96%

$$\text{Samplesize} = 1476$$

Calculated sample size was divided into training set (70%), validation set (15%) and test set (15%).

A total of 1,959 image files in JPEG format with corresponding hearing test data were collected (Figure 1). The audiograms contain 8 classes of audiological symbols, including air right unmasked (o), air left unmasked (x), air right masked (△), air left masked (□), bone right unmasked (<), bone left unmasked (>), bone right masked ([), and bone left masked (]). These symbols represent air and bone conduction at 6 frequencies, ranging from 250 to 8,000 Hz in both the left and right ears, and are marked with different symbols (Figure 2). We allocated 200 images for training and evaluating the object detection performance for graph and table detection, as well as audiological symbol detection. This system allows the accurate extraction of varying data from the audiogram images, enhancing the overall diagnostic process (Figure 3).

In this study, 1959 images were divided into 2 tasks: 200 images for digitalization and the remaining 1,759 images for further utilization. In the image digitization task, we aimed to train a series of object detection models to extract graphs, tables, and audiological symbols from audiograms. The object detection models consist of two main components: 1) graph and table detection, and 2) audiological symbol detection (Figure 4). In the first step, we trained detection models for graph and table detection. Subsequently, we trained a model for audiological symbol detection from the extracted graph using YOLO v5 and YOLO v8. The numerical data from the tables, including the PB, SRT, SL values, are extracted using EasyOCR framework(Github September 2023 - Version 1.7.1) to recognize and extract the handwriting numbers.

Combining object detection and the OCR model, we can extract numerical features from the audiogram. Each symbol's position is mapped to the appropriate frequency and decibel level. Numerical values are then extracted from the table using OCR. The extracted data are used for the hearing level classification models. The linear interpolation method is used during inference to replace any missing values that could have come from the object detection or OCR model.

In the audiogram severity classification task, the images were randomly assigned to the Training set and the Testing set, with proportions of 70% (1759 images) and 30% (352 images), respectively. Additionally, 30% of the training set was used for validation using a random stratified 10-fold method, all images were inputted into an Excel spreadsheet by the researchers for comparison with the data extracted by the machine learning model from the images.

In this task, we develop two models using different datasets: one labeled by researcher (group truth dataset) and the other automatically extracted from the digitalization task. We considered 16 different features found on the audiogram, including the hearing threshold at 250, 500, 1000, 2000, 4000, and 8000 Hz. Each model was trained separately, with one using manual labeling by otolaryngologists and the other utilizing automatic feature extraction from object detection model and OCR. We use AutoML library, PyCaret (Ali April 2020) to train various models including a Random Forest Classifier, Light Gradient Boosting Machine, Gradient Boosting Classifier, etc. The models were evaluated using different metrics such as accuracy, recall, precision, F1-score, and ROC-AUC score.

The study categorizes results into three parts: (1) Identifying the best model for alphanumeric data extraction from images, compared with researcher-inputted numerical data; (2) Determining the optimal model for severity level interpretation of hearing loss, comparing with otolaryngologist interpretation; (3) Assessing the best model for hearing loss type interpretation compared with

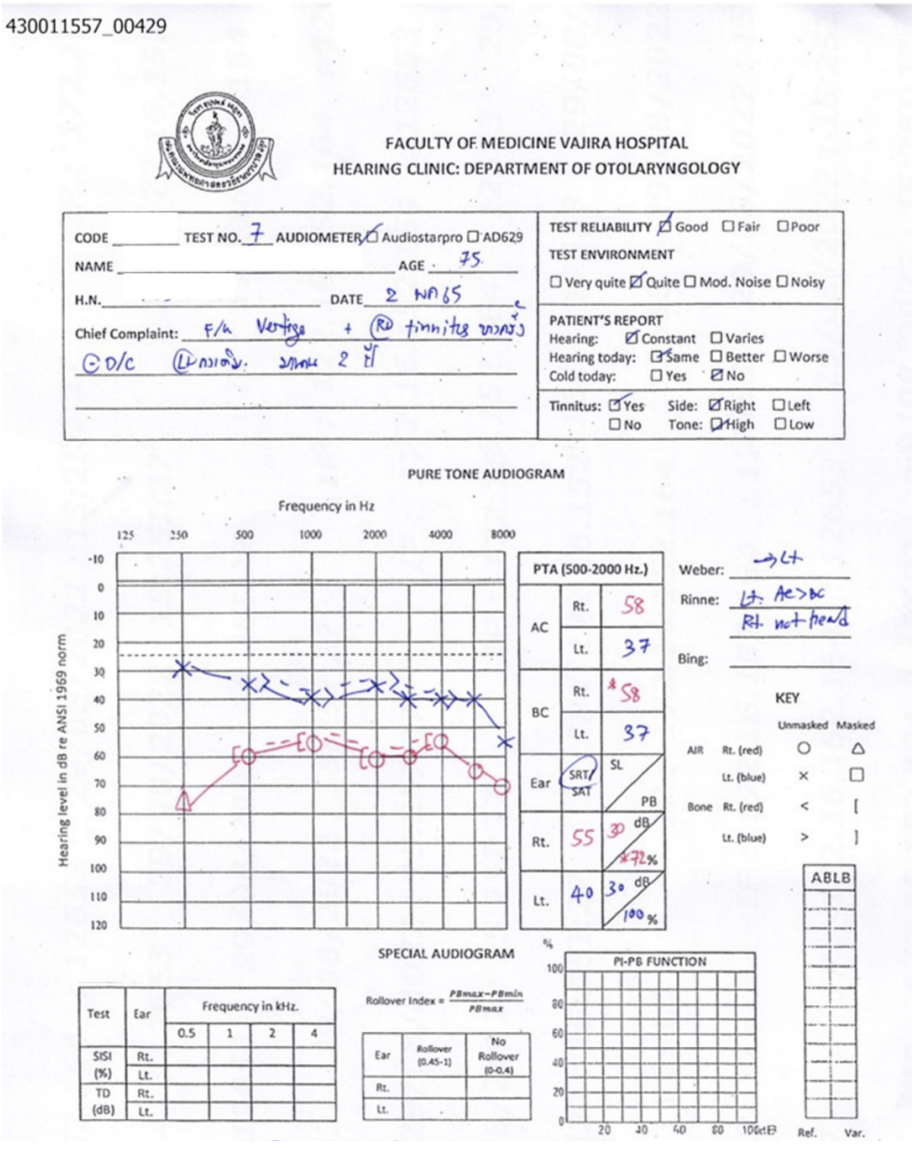

Figure 1 Sample of audiogram image used for data extraction

		Unmasked	Masked
Air	Right	○	△
	Left	×	□
Bone	Right	〈	[
	Left	〉	]

Figure 2 Symbol for data extraction air right (red color) unmasked (○), air left (blue color) unmasked (x), air right (red color) masked (△), air left (blue color) masked (□), bone right (red color) unmasked (<), bone left (blue color) unmasked (>), bone right (red color) masked ([), and bone left (blue color) masked (])

otolaryngologist interpretation.

The results were assessed using various metrics including accuracy, recall, precision, F1 score, and AUC score. Additionally, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were determined using confusion matrix tables. These comprehensive evaluations provide a detailed understanding of model performance across multiple dimensions in the study.

3 Result

In object detection tasks, 200 images from a total of 1,959 were used. Thirty percent of these 200 images were utilized to evaluate the performance of object detection models. For graph and table detection models, the results indicate similar performance between YOLOv5 and YOLOv8, with YOLOv8 demonstrating slightly higher performance across most metrics.

We used the same data split method for audiological symbol detection as for the graph and table models. The mean Average Precision (mAP) score and Intersection over Union (IoU) of validation on both YOLOv5 and YOLOv8 were relatively low compared to graph and table detection. However, the mAP50 score was comparable to other models, with 0.886 and 0.898 for YOLOv5 and YOLOv8, respectively. The right bone masked ([) symbol had the lowest mAP score for each class on both YOLOv5 and YOLOv8, with scores of 0.826 and 0.817, respectively. The extracted audiogram features were used to compute the MAE (mean absolute error) and MAPE (mean absolute percentage error) from the original data. The results showed MAE and MAPE values of 25.872 and 2.4896, respectively. (Table1)

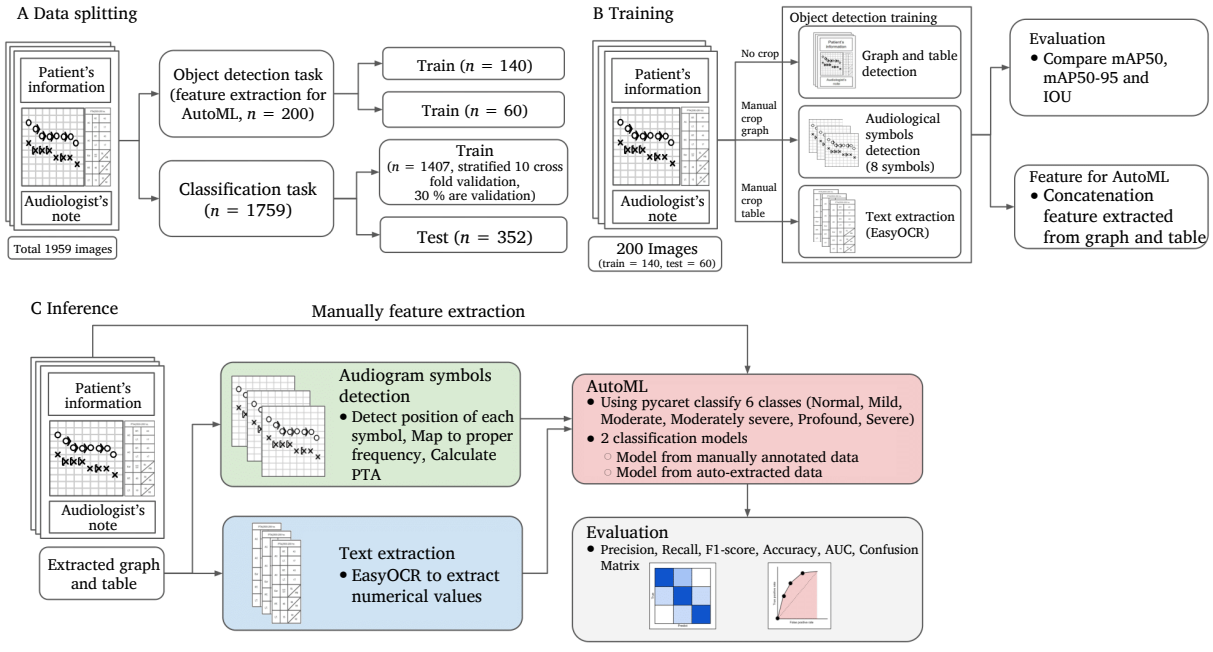

Figure 3 Schematic of audiological symbol detection and hearing loss classification. A. Data Splitting. Audiograms are splitted for object detection and hearing loss classification. B. Training. The training process contains the data splitting to training object detection models for automatic feature extraction and hearing loss classification. C. Model Inference. Object detection models are used for automatic feature extraction. We compare the performance of hearing loss classification from manual feature extraction and automatic feature extraction with AutoML pipeline.

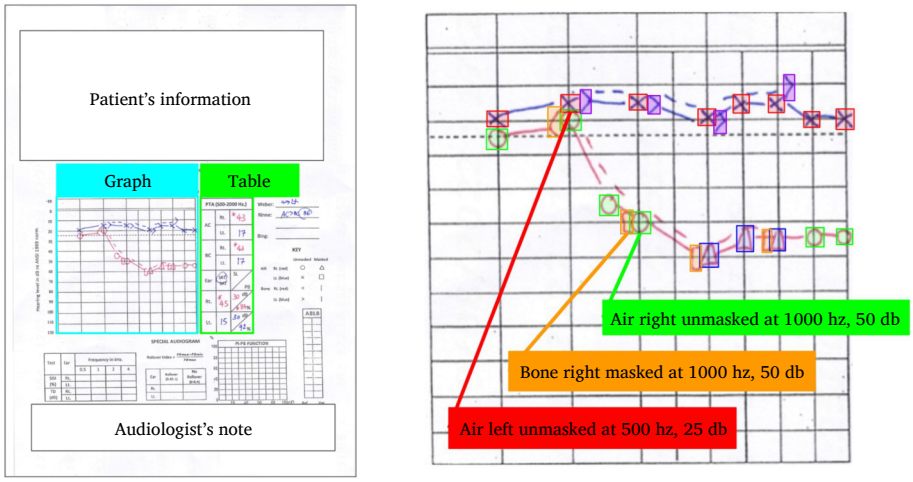

Figure 4 Graph, table, and audiological symbols annotation (A) example of graph and table annotation that use to train graph and table detection model (B)example of audiological symbol annotation that used to train audiological detection model (C) The table shows the symbolic labels that were used to train audiological symbol detection model.

In the hearing loss classification task, the 1,759 individuals out of 1,959 undergoing hearing assessment, divided into a training set (1,407 individuals) and a test set (352 individuals) for train and evaluate the hearing loss classification model. In the training set, the average age was 57.99 years, with 872 females and 580 males, and 631 individuals reported loud noise exposure. The test set included 231 females, 121 males, and 156 individuals with noise exposure. Hearing ability was assessed using Pure Tone Average (PTA) of Air Conduction (AC) and Bone Conduction (BC) at various frequencies, with severity categorized as mild to profound (Clark 1981). Hearing loss types included Normal Hearing, Sensorineural (SNHL), Conductive (CHL), and Mixed. These classifications were applied to both ears. (Table 2) (Table 3) (Table 4) (Figure 5)

We have developed 2 classification models, which include manual extraction data from the physician and automatic extraction data from the object detection model. for training

Regarding the model train with automatic extraction, the top five models exhibit comparable outcomes. The light gradient boosting model has the highest score in four metrics across the validation and test sets: accuracy (0.9551 on validation and 0.9472 on test), recall (0.9557 on validation and 0.9472 on test), precision (0.9566 on validation and 0.9472 on test), and F1 score (0.9553 on validation and 0.9472 on test). Gradient boosting and light gradient boosting achieved the highest AUC scores of 0.9903 on the test set and 0.9883 on the validation set, respectively.

Compared with automatic extracted models, the model trained using manual extracted data performs marginally better. The

Table 1 Object detection performance and mAP50 of each audiological symbol. (Top) We measure the performance of models on graph, table, and audiological symbols detection. (Bottom) The mAP score comparison of each class between YOLOv5 and YOLOv8 are shown in the table.

Graph						
Model	Train			Test		
	mAP	mAP50	IoU	mAP	mAP50	IoU
YOLOv5	0.995	0.995	0.934	0.995	0.995	0.981
YOLOv8	0.995	0.995	0.937	0.995	0.995	0.981
Table						
	Train					
YOLOv5	0.990	0.995	0.914	0.988	0.995	0.971
YOLOv8	0.988	0.995	0.911	0.993	0.995	0.980
Audiological Symbols						
	Train			Test		
YOLOv5	0.889	0.995	0.910	0.411	0.886	0.62
YOLOv8	0.935	0.995	0.931	0.411	0.898	0.65

mAP50 of each audiological symbol								
Symbol	O	X	<	>	△	□	[	]
YOLOv5	0.957	0.925	0.848	0.841	0.833	0.986	0.826	0.875
YOLOv8	0.945	0.945	0.857	0.876	0.915	0.85	0.817	0.944

Table 2 Characteristic of patients

Characteristic	Training set (n = 1407)	Test group (n = 352)
Age (year)	57.99 ± 17.29	58.63 ± 17.91
Sex		
Female	827	231
Male	580	121
Hearing today		
Same	1032	279
Better	89	18
Worse	286	55
Tinnitus		
Present with tinnitus	631	156
No tinnitus	776	196

Table 3 Audiogram data from right ear

Characteristic (Audiogram interpretation on right side)	Training set (n = 1407)	Test group (n = 352)
Right PTA AC (dB ± SD)	37.17 ± 23.87	37.63 ± 23.46
Right PTA BC (dB ± SD)	31.49 ± 17.94	32.55 ± 17.51
Right AC 250 Hz (dB ± SD)	35.87 ± 22.43	35.82 ± 21.94
Right AC 500 Hz (dB ± SD)	34.75 ± 23.65	34.62 ± 23.13
Right AC 1000 Hz (dB ± SD)	37.28 ± 24.88	37.64 ± 24.53
Right AC 2000 Hz (dB ± SD)	39.74 ± 25.34	40.33 ± 25.00
Right AC 4000 Hz (dB ± SD)	46.14 ± 28.39	47.54 ± 39.95
Right AC 6000 Hz (dB ± SD)	47.23 ± 28.49	47.97 ± 28.81
Right AC 8000 Hz (dB ± SD)	49.50 ± 30.40	50.45 ± 31.33
Right BC 500 Hz (dB ± SD)	28.44 ± 16.76	28.80 ± 16.47
Right BC 1000 Hz (dB ± SD)	30.71 ± 19.56	31.72 ± 18.56
Right BC 2000 Hz (dB ± SD)	35.16 ± 19.90	36.22 ± 19.39
Right BC 4000 Hz (dB ± SD)	39.41 ± 21.37	39.50 ± 21.09
Degree of hearing loss right ear		
- Mild	292	77
- Moderate	204	49
- Moderately severe	170	37
- Severe	77	22
- Profound	46	11
Type of hearing loss right ear		
- Normal hearing	371	98
- Sensorineural hearing loss (SNHL)	833	1210
- Conductive hearing loss (CHL)	23	3
- Mixed hearing loss	180	41

Table 4 Audiogram data from left ear

Characteristic (Audiogram interpretation on left side)	Training set (n = 1407)	Test group (n = 352)
Left PTA AC (dB ± SD)	37.94 ± 24.08	36.92 ± 23.16
Left PTA BC (dB ± SD)	31.82 ± 18.33	31.56 ± 17.61
Left AC 250 Hz (dB ± SD)	36.38 ± 22.81	34.19 ± 21.45
Left AC 500 Hz (dB ± SD)	35.71 ± 23.83	33.68 ± 22.13
Left AC 1000 Hz (dB ± SD)	38.25 ± 25.48	36.96 ± 24.05
Left AC 2000 Hz (dB ± SD)	40.17 ± 25.64	39.26 ± 24.53
Left AC 4000 Hz (dB ± SD)	47.06 ± 27.96	45.31 ± 26.62
Left AC 6000 Hz (dB ± SD)	48.53 ± 28.04	49.13 ± 27.66
Left AC 8000 Hz (dB ± SD)	45.85 ± 30.19	50.34 ± 29.67
Left BC 500 Hz (dB ± SD)	28.86 ± 16.65	28.02 ± 16.17
Left BC 1000 Hz (dB ± SD)	31.05 ± 19.70	31.11 ± 19.06
Left BC 2000 Hz (dB ± SD)	35.50 ± 19.88	35.45 ± 19.99
Left BC 4000 Hz (dB ± SD)	39.74 ± 20.87	38.94 ± 20.70
Degree of hearing loss left ear		
- Mild	272	76
- Moderate	215	45
- Moderately severe	171	41
- Severe	96	26
- Profound	59	8
Type of hearing loss left ear		
- Normal hearing	335	85
- Sensorineural hearing loss (SNHL)	845	215
- Conductive hearing loss (CHL)	33	6
- Mixed hearing loss	194	46

random forest, on the other hand, outperforms all four metrics in this model, including accuracy (0.9674 on validation and 0.9643 on test), recall (0.9674 on validation and 0.9643 on test), precision

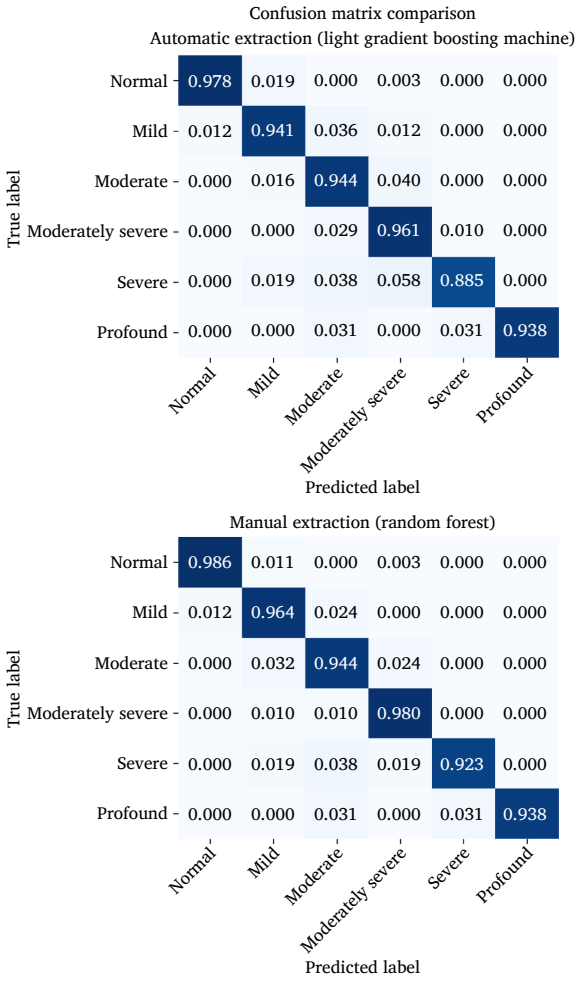

Figure 5 Confusion matrices comparison. Confusion matrix comparison between hearing loss classification using automatic extracted features and manually extracted features.

(0.9685 on test and 0.9645 on test), and F1 score (0.9673 on validation and 0.9643 on test). While gradient boosting has the best AUC score on the test set by 0.9892 and light gradient boosting has the best AUC score on validation by 0.9887. (Table 5) (Figure 6)

4 Discussion

Currently, there are several methods for diagnosing hearing loss, The standard method is known as standard audiometry. It can be used to identify the causes of a patient's hearing loss, which is typically classified into three types: Conductive hearing loss, Sensorineural hearing loss, and Mixed hearing loss. Additionally, the severity of hearing loss can be classified into five levels: mild, moderate, moderately severe, severe, and profound.

After collecting data and providing it to all 14 models, including Naive Bayes, Ridge Classifier, Logistic Regression, Dummy Classifier, SVM-Linear Kernel, Linear Discriminant Analysis, AdaBoost Classifier, Quadratic Discriminant Analysis, Random Forest Classifier, Decision Tree Classifier, K Neighbors Classifier, Extra Trees Classifier, Gradient Boosting Classifier, and Light Gradient Boosting Machine, it was found that the machine learning interpretation of the standard audiogram yielded results close to those of otolaryngologist interpretations. The result is consistent with the research by Matthew G. Crowson, which found that Resnet-101 could interpret standard audiograms with an accuracy of up to 97.5%. Biomedical engineers have noted that if the F1-score exceeds 90% in testing, the data can be considered highly reliable.

Regardless, both the digitization process and the hearing classification process can be improved in our future research. For digitization of audiograms using object detection, we find some missing or misclassified symbols in the graph due to symbol overlaps or small-size symbols. One approach is to increase the number of annotated audiological symbol images. Although a more robust object detection model or specialized models for small objects [site] may help improve the digitization of audiograms, For tabular extraction, handwritten OCR models may give better recognition results compared to our current use of EasyOCR.

Moreover, training a noise-robust hearing classification model may also improve the classification performance, as automatic feature extraction and OCR may provide inaccurate features for the model. In addition, the program may not work as expected in cases of different forms and symbols of annotations for audiologic symbols or different scales on either the x or y axis.

This research holds significant potential for application across

Table 5 Hearing loss classification performance of top-5 models using AutoML with automatic extracted data (top) and manually extracted data (bottom).

Model	Accuracy		AUC		Recall		Precision		F1 Score	
	validation	test	validation	test	validation	test	validation	test	validation	test
Gradient Boosting Classifier	0.9527	0.9401	0.9901	0.9856	0.9527	0.9401	0.9539	0.9419	0.9524	0.9388
Light Gradient Boosting Machine	0.9557	0.9472	0.9909	0.9842	0.9557	0.9472	0.9566	0.9472	0.9553	0.9472
Random Forest Classifier	0.9461	0.9330	0.9907	0.9806	0.9461	0.9330	0.9472	0.9352	0.9452	0.9317
Decision Tree Classifier	0.9217	0.9344	0.9496	0.9522	0.9217	0.9344	0.9243	0.9346	0.9217	0.9335
Extra Trees Classifier	0.9125	0.9044	0.9891	0.9815	0.9125	0.9044	0.9134	0.9043	0.9108	0.9033
Model	Accuracy		AUC		Recall		Precision		F1 Score	
	validation	test	validation	test	validation	test	validation	test	validation	test
Gradient Boosting Classifier	0.9654	0.9629	0.9879	0.9892	0.9654	0.9629	0.9664	0.9631	0.9653	0.9629
Light Gradient Boosting Machine	0.9654	0.9629	0.9887	0.9861	0.9654	0.9629	0.9663	0.9630	0.9653	0.9629
Random Forest Classifier	0.9674	0.9643	0.9886	0.9865	0.9674	0.9643	0.9685	0.9645	0.9673	0.9643
Decision Tree Classifier	0.9298	0.9272	0.9538	0.9528	0.9298	0.9272	0.9319	0.9286	0.9298	0.9275
Extra Trees Classifier	0.9527	0.9501	0.9884	0.9834	0.9527	0.9501	0.9549	0.9506	0.9523	0.9500

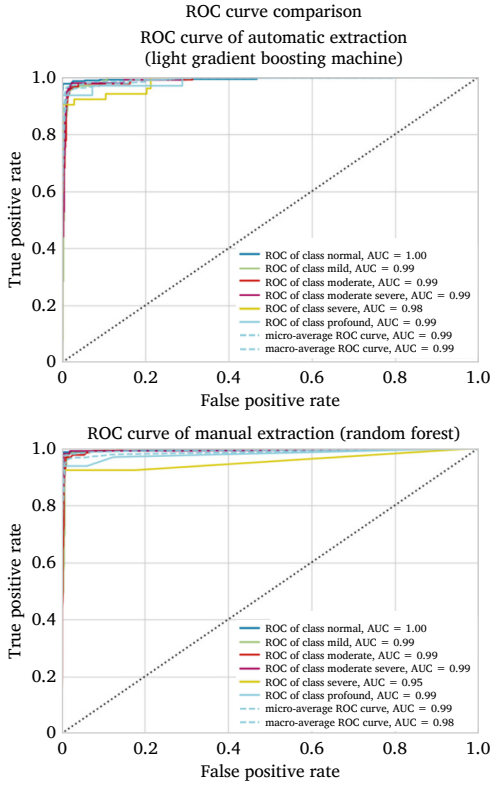

Figure 6 ROC curve comparison. ROC curve comparison between hearing loss classification using automatic extracted features and manually extracted features.

various areas of Thailand, addressing the shortage of healthcare personnel such as otolaryngologists and audiologists, particularly in smaller healthcare facilities. By leveraging deep learning technology, this system can alleviate workload burdens, enhance efficiency in medical practices, and expedite the interpretation of audiograms by physicians. Consequently, it enables better resource allocation, reduces waiting times for diagnosis and treatment, and ultimately enhances the quality of life for patients and their families.

This research has several limitations. Firstly, data was collected solely from the Department of Otolaryngology, Faculty of Medicine, Vajira Hospital, using the GSI Audiostarpro Audiogram device. Results may vary when utilizing data from different institutions or employing different tools. Secondly, hand-recorded data may introduce human error or inconsistencies. Lastly, there might be other deep learning models with similar effectiveness that were not tested in this research, suggesting the need for further exploration and validation.

5 Conclusion

Interpreting audiogram results through an object detection based machine learning system yields outcomes closely resembling those of interpretations by otolaryngologists. This suggests the potential of machine learning in accurately analyzing and diagnosing hearing loss, offering promising applications in clinical settings.

Acknowledgements

The authors would like to express their gratitude and appreciation to Navamindradhiraj University and Mahidol University for their support and facilities throughout the research.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors have no funding, financial relationships, or conflicts of interest to disclose.

References

Ali, M., 2020. PyCaret: an open source, low-code machine learning library in Python.

Bunnag, K., Kaewsalubsri, W., Junthong, S., et al., 2023. A study on the IOS application "uHear" as a screening tool for hearing loss in Bangkok. *Laryngoscope Investig. Otolaryngol.* 8(1), 253–261.

Chadha, D.S., 2018. Global action for hearing loss. *Vestn Otorinolaringol.* 83(4), 5–8.

Charih, F., Green, J.R., 2022. Audiogram digitization tool for audiological reports. *IEEE Access* 10, 110761–110769.

Clark, J.G., 1981. Uses and abuses of hearing loss classification. *ASHA* 23(7), 493–500.

Committee on, Accessible, Adults Affordable Hearing Health Care for, Policy Board on Health Sciences, Health, Division Medicine, Engineering National Academies of Sciences, and Medicine, 2016. The national academies collection: reports funded by national institutes of health. In: Hearing Health Care for Adults: Priorities for Improving Access and Affordability. Blazer, D.G., Domnitz, S., Liverman, C.T., Eds. Washington DC: National Academies Press. Copyright, 2016. By the National Academy of Sciences. All rights reserved: Washington (DC).

Crowson, M.G., Lee, J.W., Hamour, A., et al., 2020a. AutoAudio: deep learning for automatic audiogram interpretation. *J. Med. Syst.,* 44(9), 163.

Crowson, M.G., Ranisau, J., Eskander, A., et al., 2020b. A contemporary review of machine learning in otolaryngology–head and neck surgery. *Laryngoscope* 130(1), 45–51.

Daniel, W.W., 1999. Biostatistics: A Foundation for Analysis in the Health Sciences. 7th ed. Hoboken: Wiley.

Davies, R.A., 2016. Audiometry and other hearing tests. *Handb. Clin. Neurol.* 137, 157–176.

Elkhouly, A., Andrew, A.M., Rahim, H.A., et al., 2023. Data-driven audiogram classifier using data normalization and multi-stage feature selection. *Sci. Rep.* 13(1), 1854.

Github. September 2023-version 1.7.1. EasyOCR.

Kassjański, M., Kulawiak, M., Przewoźny, T., 2022. Development of an AI-based audiogram classification method for patient referral. In: Proceedings of the 17th Conference on Computer Science and Intelligence Systems, Sofia, Bulgaria, pp. 163–168.

Olusanya, B.O., Neumann, K.J., Saunders, J.E., 2014. The global burden of disabling hearing impairment: a call to action. *Bull. World Health Organ.* 92(5), 367–373.

Rutterford, C., Copas, A., Eldridge, S., 2015. Methods for sample size determination in cluster randomized trials. *Int. J. Epidemiol.* 44(3), 1051–1067.

Tsilivigkos, C., Athanasopoulos, M., Micco, R.D., et al., 2023. Deep learning techniques and imaging in otorhinolaryngology-a state-of-the-art review. *J. Clin. Med.* 12(22), 6973.

Wasmann, J.W., Pragt, L., Eikelboom, R., et al., 2022. Digital approaches to automated and machine learning assessments of hearing: scoping review. *J. Med. Internet Res.* 24(2), e32581.

World Health Organization, 2013. Multi-country assessment of national capacity to provide hearing care. Available at https://www.who.int/publications/i/item/9789241506571. Accessed 11 Mar.

Yang, T.W., Yang, C.Y., Lee, Y.H., et al., 2024. A novel method for audiogram digitization in audiological reports. *IEEE Access* 12, 37862–37872.

You, E., Lin, V., Mijovic, T., et al., 2020. Artificial intelligence applications in otology: a state of the art review. *Otolaryngol.-Head Neck Surg.* 163(6), 1123–1133.