EarMonitor: Non-clinical Assessment of Ear Health Conditions Using a Low-cost Endoscope Camera on Smartphones

XIAOFU JIN, The Hong Kong University of Science and Technology, China
MINGMING FAN, The Hong Kong University of Science and Technology (Guangzhou), China and The Hong Kong University of Science and Technology, China

Hearing loss affects 20% of the global population, a rate that is increasing dramatically as the world's population ages. Early prevention and identification of ear diseases can significantly reduce the risk of becoming disabled with hearing impairment. We propose *EarMonitor*, an interactive, vision-based ear health monitoring system that enables users to examine their ear conditions with a low-cost hand-held endoscope. *EarMonitor* can detect six ear health conditions suitable for self-assessment. It can particularly recognize complications from ear diseases, helping users better understand the results. In the wild, our computer vision algorithm achieves a detection sensitivity of 0.949 for earwax buildup and blockage in 100 external auditory canal photos; our deep learning model achieves an average detection sensitivity of 0.861 for the other five conditions considering complications in 350 tympanic membrane photos. We validated *EarMonitor*'s effectiveness through a user study involving 17 participants and two experts, leading to valuable insights regarding the design and interpretation of non-clinical assessment devices.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; *User studies*; • **Applied computing → Health care information systems**.

Additional Key Words and Phrases: mobile health, ear health, computer vision

ACM Reference Format:
Xiaofu Jin and Mingming Fan. 2024. EarMonitor: Non-clinical Assessment of Ear Health Conditions Using a Low-cost Endoscope Camera on Smartphones. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 254 (September 2024), 20 pages. https://doi.org/10.1145/3676499

1 Introduction

Approximately 1.57 billion individuals worldwide are affected by hearing loss due to age-related issues, infections, and chronic exposure to loud noises [8, 23, 25, 26]. The early detection and treatment of hearing loss are crucial; failure to do so can result in severe impairment, significantly impacting various aspects of an individual's life, such as socioeconomic status, mental and physical health, and opportunities in education and employment [7]. However, identifying potential ear issues at an early stage remains challenging. One key reason is that the initial symptoms of ear conditions, such as earwax buildup, otitis media with effusion, and early-stage cholesteatoma, are not easily noticeable to many individuals [1]. Additionally, the scarcity of medical resources and the financial burden, particularly in developing regions, hinder access to timely medical assessment

Authors' Contact Information: Xiaofu Jin, xjinao@connect.ust.hk, The Hong Kong University of Science and Technology, Hong Kong SAR, China; Mingming Fan, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China and The Hong Kong University of Science and Technology, Hong Kong SAR, China, mingmingfan@ust.hk.

ACM 2573-0142/2024/9-ART254
https://doi.org/10.1145/3676499

and care. In areas with limited resources, the availability of specialized equipment like otoscopes or tympanometry devices is scarce, and a shortage of otolaryngology specialists leads to long waiting periods for patients. In some remote areas, only general practitioners with limited ENT training are available [6]. Therefore, there is a pressing need for accessible and affordable support for the general public, especially those in rural or underdeveloped areas, to facilitate a better understanding of their ear conditions before resorting to medical assistance, which is often considered as a last resort due to financial constraints.

To address this challenge, the Human-Computer Interaction (HCI) community has ventured into the realm of do-it-yourself (DIY) early assessment tools, leveraging the advancements in mobile and AI technologies to deepen the understanding of user health conditions [2, 18, 41, 42]. In a notable effort to facilitate at-home detection and early identification of ear diseases, Jin et al. investigated the use of acoustic reflectometry, aiming to discern three significant ear health conditions within daily environments [10]. While promising, it's important to note that acoustic reflectometry's principles inherently limit its detection capabilities to ear diseases with specific symptoms. However, the majority of existing image-based diagnostic methodologies are restricted to laboratory settings, dependent on high-resolution images obtained from sophisticated devices, which may not align with the practicalities of self-management [4, 11]. Moreover, prevailing research often simplifies the diagnosis of ear diseases to a mere multi-classification issue, neglecting the fact that ear infections frequently manifest with complications, such as otitis media accompanied by tympanic membrane perforation [1]. The identification of such complications is critical in the prevention and treatment of ear diseases, since each complication may necessitate a distinct therapeutic approach. Equally important is the value of user input in reporting symptoms, which is essential for spotting potential complications, understanding the progression of the disease, and deciding on the appropriate treatment measures. For instance, patients experiencing otitis media with effusion (OME) for less than three months might be advised to adopt a 'watch and wait' approach whereas those enduring it for longer periods may need to seek medical intervention [1, 32]. Additionally, the insight gained from user-reported symptoms regarding the cause of the disease is invaluable in identifying its origin and guiding subsequent health management to prevent the recurrence or worsening of the condition [1].

In this study, we adopted a user-centered design approach to develop an ear health self-assessment system that integrates user inputs to evaluate common ear conditions and offer relevant management advice autonomously. Initially, we engaged in discussions with three otology experts, identifying six prevalent ear conditions amenable to home monitoring: earwax buildup and blockage (EB), otitis media with effusion (OME), suppurative otitis media (SOM), cholesteatoma (CH), tympanic membrane perforation (TMP), and healthy ear condition (HE). These consultations informed the system's design principles, emphasizing self-administration safety, adherence to a clinical consultation framework by aligning symptoms with assessment outcomes, and the maintenance of a consistent assessment log. While experts express caution regarding the direct application of AI-generated outcomes in treatment decisions, they advocate for its use in monitoring health conditions, enhancing the precision and clarity of clinical consultations when further examination is necessary.

We introduce *EarMonitor*, a pioneering interactive, vision-based system for ear health monitoring that enables ear assessment using an affordable, handheld endoscope connected to smartphones. Users can employ an external endoscope to capture images of the external auditory canal and tympanic membrane, as depicted in Figure 1. Leveraging computer vision and deep learning techniques, the system autonomously assesses the ear condition, amalgamating user inputs to deliver a thorough evaluation and actionable recommendations. For sustained self-care, EarMonitor maintains a log of routine assessments, enabling users to track changes over time and furnishing healthcare professionals with detailed information for enhanced diagnostic accuracy. We conducted three evaluations to validate *EarMonitor*: (i) A technical evaluation of model performance shows our computer vision algorithm achieves a detection sensitivity of 0.949 for earwax buildup and blockage in 100 external auditory canal photos; our deep learning model average detection sensitivity of 0.904 for the other five conditions in 350 tympanic membrane photos collected in the wild and is compatible or even better than average general practitioners and pediatricians [17, 30, 31]. Notably, general practitioners and pediatricians diagnose acute otitis media (AOM) and otitis media with effusion (OME) with varying degrees of accuracy: general practitioners accurately diagnose AOM without any doubt in only 54% of cases [17], pediatricians' average correct diagnosis ranges from 25% to 73% in AOM and OME [30, 31].(ii) A user evaluation with 17 participants indicates that our system enables users to conduct self-examination and present effective assessments and suggestions. (iii) An expert evaluation with two otologists suggests that our system is valid and useful, and can be further improved by providing remote communication between otologists and users.

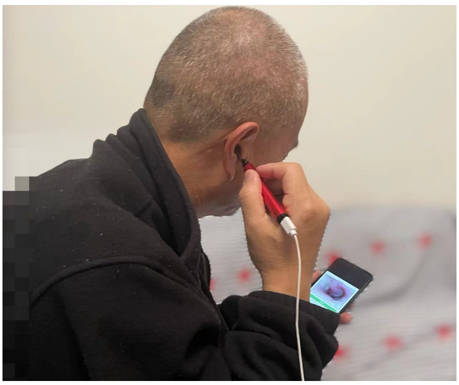

Fig. 1. Figure a shows an end-user using *EarMonitor*.

Our main research contributions include:

- EarMonitor - We offer a low-cost, user-friendly solution designed to assess six common ear health conditions that are prevalent in daily life. This innovative tool allows users to conveniently input their symptoms, observe their ear conditions in real time, and track assessing history. It features voice interaction capabilities for triggering automatic assessments, providing users with comprehensive interpretations of their ear health conditions along with tailored suggestions based on their input and AI-driven evaluations.
- Algorithm - We designed a two-stage detection process following the otologist's examining procedure, enabled and improved the deep learning model to detect ear diseases with complications, which matters in the early prevention strategy.
- Evaluation - Our comprehensive assessment includes a technical evaluation of images captured outside the controlled laboratory environment, a user study, and expert analysis, collectively demonstrating the feasibility of this tool for empowering end-users to assess their ear health. Furthermore, we garnered valuable insights into the development of non-clinical self-assessment tools, aiming to enhance end-users' understanding of AI-generated results and encourage regular monitoring of their ear health.

2 Background and Related Work

2.1 Background of Ear Diseases

Hearing impairment is one of the leading contributors to years lived with a disability, with over 5 percent of the world's population (360 million people) currently living with a disabling hearing loss [27]. It not only heavily impacts many aspects of a person's life [7] but also represents a substantial financial drain on society as it is estimated to cost 750 billion dollars internationally per year [7, 27]. Hearing impairment, which can result from an injury or ear disease, may become permanent [7, 39]. In this section, we provide some background information about ear diseases and brief clinical procedures for an initial assessment. The ear is divided into three basic sections– external, middle, and inner ears as shown in the left figure of Fig 2. The initial assessment of ear disease often happens with the external and middle ear. The otologist uses a hand-held endoscope to stretch into patients' ears following the external auditory canal. Combining the results of the interrogation with conditions observed in the external ear and tympanic membrane (TM), the otologist makes an initial assessment. This includes diagnosing diseases in the external and middle ear or identifying conditions that require further examination (peculiar middle ear disease and potential diseases in the inner ear). As mentioned before, this work explores approaches to detect common ear diseases at home for initial assessment, including earwax buildup and blockage (EB), TM perforation (TMP), otitis media with effusion (OME), suppurative otitis media (SOM), cholesteatoma (CH), which are shown in the right figure of Fig 2. Below is a brief description of these ear diseases.

- **Earwax Buildup and Blockage.** Earwax protects the auditory canal by providing a barrier against water penetration and suppressing the growth of bacteria and fungi. However, in some persons, the earwax does not extrude but accumulates (Fig 2 b) and becomes impacted in the canal preventing the normal transmission of sound [1]. It can commonly bring out temporary hearing loss or tinnitus [39].
- **Otitis Media with Effusion (OME).** OME, also known as secretory otitis media, means the otitis media with the fluid accumulating behind an intact tympanic membrane (TM) Fig 2 d), and thus the most commonly reported symptom is hearing loss [1, 32]. The hearing loss in OME is often transient as the middle ear effusion frequently resolves spontaneously, for this reason, a "watch and wait" period should be adopted and treatment only offered to those in whom an effusion is persistent [32].
- **Suppurative Otitis Media (SOM).** Patients with SOM often experience persistent otorrhea, but this symptom is not obligatory; they can also experience hearing loss, tinnitus, otalgia, and pressure sensation [33]. If SOM develops into a chronic one, it often happens with permanent perforation, which means that treatment is usually multifaceted, requiring antimicrobial agents and surgery (Fig 2 c) [32].
- **Cholesteatoma.** Cholesteatoma can be classified as either congenital or acquired, which occurs when keratinizing squamous epithelium (skin) is present in the middle ear [1, 32]. It typically presents with chronic smelly ear discharge (Fig 2 f) and can be diagnosed when squamous epithelium and keratin are seen in the middle ear. It can cause symptoms such as hearing loss and otorrhea, and the only curative treatment is surgical [1, 32].
- **TM Perforation (ruptured eardrum).** The tympanic membrane (TM) serves as a key component of the tympano-ossicular system for sound transmission. Perforation of the TM, which also refers to a ruptured eardrum, often means a hole or tear in the eardrum (Fig 2 c). It is common in an otologic practice and can be caused by various reasons such as trauma, CSOM, and cholesteatoma [1, 21]. If TM perforation does not heal and causes a long-term infection, people may get temporary or even permanent hearing loss [21].

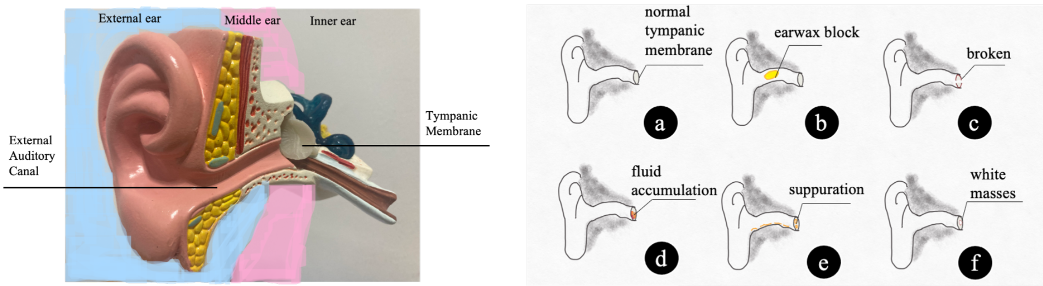

Fig. 2. Background information of the ear. The left figure shows the anatomy of the ear. The right figure demonstrates different ear health conditions.

2.2 Hearing Condition Monitoring

Researchers are exploring artificial intelligence approaches to recognize ear diseases recently [4, 11, 38]. However, they are limited in detecting specific ear diseases which can not cover common diseases that happen in both the external auditory canal and middle ear, thus can not cater to users' need to monitor at home and lack in exploring beyond the laboratory. Moreover, their depending clinical devices are usually expensive and sophisticated to use. To facilitate users to detect hearing health conditions at home and help them recognize ear diseases at an earlier stage for timely intervention and treatment, Chan et al. explored solutions using speakers and microphones within existing smartphones to detect middle ear fluid by assessing eardrum mobility [5]. Jin et al. also explored the acoustic reflectometry approach to detect three hearing health conditions [10]. Although it can detect more hearing health conditions rather than one, it also limits the detection of diseases that have specific symptoms. Moreover, almost all of the previous research focuses on the detection of independent ear diseases [5, 10, 11, 38], but in fact, ear infections usually have complications [1]. Therefore, we aim to build a non-clinical self-assessment system, providing assessment results and corresponding suggestions.

3 Formative Study

To understand the standard procedure of assessment and system requirements of *EarMonitor*, we conducted semi-structured interviews with three otologists, two of whom had over 10 years of experience and one had over 20 years. We first introduced the initial idea and purpose of the system. Then we asked questions including 1) ear conditions they often assess, 2) the outpatient standard procedure, and 3) possible benefits/challenges of this self-examination system. One researcher stayed with a doctor in the clinic, observing the entire assessment process and taking notes on their interactions. The audio recordings of interviews were transcribed through an online transcribing tool and then manually reviewed by the authors to correct for errors. The transcripts were coded using a thematic analysis methodology [3]. The observation notes were added to each theme as supplementary details.

We have summarized six ear conditions, the most common ones experts reported in the clinical. They are earwax buildup and blockage (EB), otitis media with effusion (OME), suppurative otitis media (SOM), cholesteatoma (CH), and tympanic membrane (TM) perforation and those can be reflected on the external and middle ear. The outpatient procedure in a clinical setting begins with an interrogation, covering disease history, symptoms experienced by the patient, duration, and other follow-up questions. Then the otologist uses an otoscopy to check the external auditory canal status and the TM status. After combining the results of the interrogation with observations, the otologist decides whether to conduct further examinations, adopt the 'watch and wait' method, or make a precise assessment and arrange the corresponding treatment. Experts highlighted the scarcity of qualified otologists in many parts of the world, which often leads to misdiagnoses in remote regions where diseases can significantly advance before a correct diagnosis is made. They noted that a self-examination system could lessen the clinical assessment burden, enable earlier intervention, and improve patient understanding of their conditions, thus enhancing communication. All experts agreed on the importance of adhering to a standardized clinical consultation process to ensure a consistent medical record for potential future clinical transfers. Additionally, we have outlined the principal design requirements for *EarMonitor*, presented below.

3.1 Design Requirements

Drawing on our interview findings, previous research, and observations, we have synthesized design requirements and solution integrating with domain knowledge of *EarMonitor*.

- **D1. Ensure safety during self-operation.** Otologist experts expressed the safety concerns of self-operation. The human ear canal, extending from the pinna to the tympanic membrane (TM), is approximately 2.5 cm (1 inch) long and curved. For the general examination procedure, the ear canal should be as straight as possible to make sure that it does not damage the TM. The device should keep within a safe length as well to avoid hurting the TM.
- **D2. Include the detection of complications.** Ear diseases may not happen independently. For example, tympanic membrane perforation (TMP) often happens as a complication of suppurative otitis media (SOM) and cholesteatoma. The assessment of the complications may better reveal the reason or nature of the TMP, thus helping adopt a more accurate treatment.
- **D3. Interpret assessment results targeting users' major concerns.** Expert participants reported that some patients searched online and found similar images for self-comparison. However, presenting reference images for users to compare their images and assess potential diseases is inappropriate because users lack the training needed to identify key disease features. Even general practitioners achieve a low accuracy of assessment [30] in such identifications. Providing predictive results is more instructive and inspiring for users. Due to the variety of specific ear diseases, automatic detection algorithms may not always perform perfectly. It is crucial that the level of confidence in the results is clearly communicated to users to ensure they fully understand their ear conditions. Additionally, the assessment results should be interpreted alongside their symptoms. If users are advised to consult a doctor, this preliminary assessment can expedite the clinical interview process by reducing the need for patients to recount their symptoms, thereby saving time for both doctors and patients. Furthermore, this approach can enhance the accuracy of self-reported information, as patients have more time to recall the onset of symptoms and organize their descriptions, potentially reducing the rate of errors.
- **D4. Keep track of continuous assessment records and enable comparison.** The result of the assessment should be saved and a long-term regular assessment record because people have variations of symptoms. We implemented the basic function to track the records, and further advanced functions will include automatic recognition of abnormal conditions based on previous records when we conduct follow-up long-term studies.

3.2 System Design

Following the design considerations, we built *EarMonitor*. Instead of exploiting expensive pneumatic otoscopy used in hospitals, we only need a pen-like hand-held endoscope, which only costs $10 and is light and convenient to hold in hand. The length of the endoscope inside the ear is set to 2 cm for

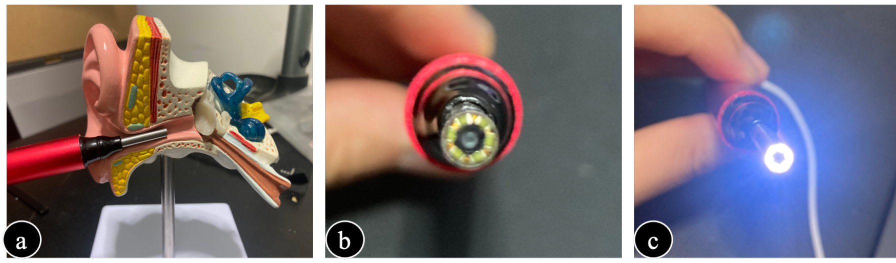

Fig. 3. The description of the hand-held endoscope. Fig a shows that the length of the endoscope is safe. Fig b shows there is a circle of lights around the endoscope's camera and Fig c shows the effect when the lights are on.

safety purposes with a diameter of only 3.9 mm, and a focal length of 15 mm as Figure 3 a shows. There is a circle of lights around the endoscope camera to provide sufficient light as Figure 3 b, c shows.

When the user first enters the system, they should complete a consultation questionnaire simulating the process of the doctor interrogating as Figure 4 a shows. After submitting the questionnaire, the user will be redirected to the detection interface, where the user can see the real-time video that the endoscope captures as Figure 4 b shows. The user can try adapting to the camera of the endoscope freely. After they feel that they can control the endoscope easily, they can click the submit button. The system first detects the condition of the external auditory canal. The user can trigger the detection of the external auditory canal by saying sentences including 'external auditory canal', for example, 'now I want to detect external auditory canal. We use the open-source speech recognition SDK of iFLYTEK to detect the keywords, then the system will capture the current of the external auditory canal photo and call the model of detecting earwax buildup and blockage and show the result. If the user has an earwax buildup and blockage, they will be prompted that they need to clean the earwax first before further examination. If the condition of the external auditory canal is normal, the user can move the endoscope a little deeper into the ear to capture the image of the tympanic membrane. The user can then trigger the detection of the tympanic membrane by saying sentences including 'tympanic membrane'. The system will capture the current TM photo from the real-time video and show the result of the multi-label classification detection with the level of confidence to ensure a comprehensive understanding of their ear conditions as Figure 4 c shows. Users can then select different detected diseases for further exploration. Besides showing the detection result and summary from the questionnaire, the system presents a sample TM image of the disease for reference, listing the typical symptoms and highlighting those the user mentioned as Figure 4 d shows. All comprehensive suggestions follow the clinical diagnostic logic of these three experienced otologists. Every assessment will be documented to facilitate future tracking and comparison as Figure 4 e shows.

4 Implementation

In this section, we introduce the implementation details. We have deployed the system using a web-based server-client architecture to enhance flexibility for mobile devices. Following the procedure of the otologist's examining ears, we design a two-stage detection process as shown in Fig 4.

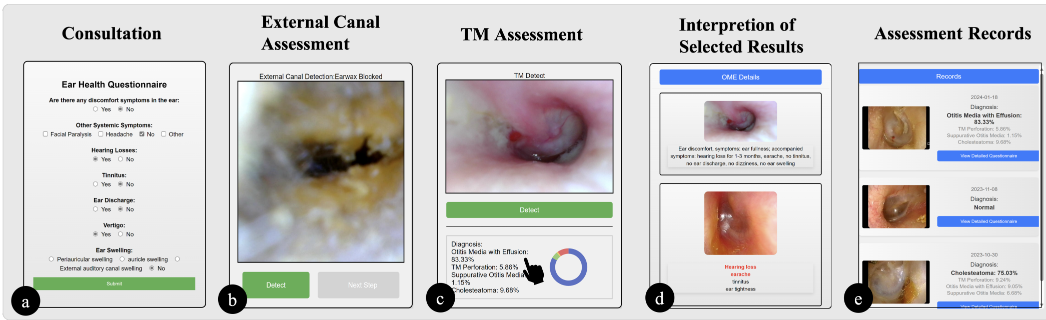

Fig. 4. The system overview of *EarMonitor*.

4.1 Data Collection

We collaborated with the local hospital and acquired 250 external auditory canal images and 1567 TM images from 500 people of the local hospital without including any sensitive data such as name and ID number. These images were captured by professional devices in hospitals and have been saved in the database. To enhance the accuracy of the in-the-wild data, we collected 320 external auditory canal images from 120 people (220 for training, 100 for testing), and 883 TM images (533 for training, 350 for testing) using our endoscope with the approval of the local hospital. All the images have been assessed and tagged, including complications, by otologists. The distribution shows as Table 1.

We built dataset A which consists of 470 external auditory canal images (250 from the hospital data and 220 from in-the-wild collection) and dataset B which consists of 2100 TM images (1567 from the hospital data and 533 from in-the-wild collection) for training. Then we use all the left in-the-wild data for testing, that is 100 external auditory canal images and 350 TM images for testing. We partitioned the dataset into training and testing sets based on individual subjects, ensuring that images in the training set and testing set are from different individuals.

Table 1. Data Distribution of hospital data and in-the-wild data on different ear conditions for healthy ear (HE), earwax buildup and blockage (EB) (1 means getting the disease, 0 means healthy external canal), TM perforation (TMP), otitis media with effusion (OME), suppurative otitis media (SOM), cholesteatoma (CH). We have incorporated complicating conditions that will be included in the calculations for each respective condition.

Condition	Hospital	Wild
EB-1	175	224
EB-0	75	96
HE	557	364
TMP	296	207
OME	208	197
SOM	391	70
CH	297	45

4.2 Hardware Formation

We adopted the cheapest medical camera with 1280 * 720 resolution to provide sufficient information on the ear images. Compared to the ordinary endoscope with the same price, it provides sharper images and depth of field with a higher resolution. Six LED lights around the camera can provide sufficient light when capturing the images inside the ear. We adopted the DVP interface which supports real-time overview. We adopted a high compression rate video chip for decoding the signals. Our final cost of the device is only around $10.

4.3 Examination from the External Auditory Canal Photo

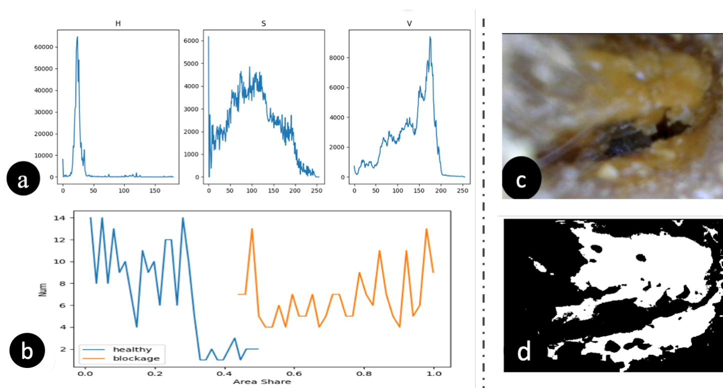

Fig. 5. Figure a shows the HSV signal distribution of an earwax buildup and blockage photo. Figure b shows the earwax area share comparison between normal photos and earwax buildup and blockage photos.

The main objective of the external auditory canal is to detect the presence of earwax buildup and blockage. Through the analysis of the images of earwax build-up and blockage, we found that the color feature is obvious for distinguishing if the external auditory canal is blocked by earwax. To mitigating lighting issues, we map all features from RGB space to HSV (THue-saturation-value) to abstract color (hue) by separating it from saturation and pseudo-illumination [34]. Figure 5 a shows an example HSV value distribution of one earwax buildup and blockage image. We then performed HSV color statistical analysis on these images and calculated the peak values across three channels for each image. After getting the range of peak values for all earwax buildup and blockage images, we removed outliers and decided on the threshold intervals Rh, Rs, and Rv. We completed the dichotomous classification of the images based on the following formula as Figure 5 c,d shows:

$$M(i,j) = \begin{cases} 1 & H(i,j) \in Rh \ and \ S(i,j) \in Rs \ and \ V(i,j) \in Rv \\ 0 & \text{else} \end{cases} \tag{1}$$

Subsequently, we calculated the number of pixels whose value equals 1 in the figure, that is the occupied area of the earwax region Sd. Then we calculated the earwax area share of the whole figure Sd / S, noted as AS. We plotted the AS' distribution of the external auditory canal training set as shown in Figure 5 b, line blue shows the area share of earwax in normal conditions, and line orange shows that in earwax blockage conditions. To distinguish the two, we set the threshold T=0.4 based on the distribution. When the earwax is occupied greater than T, the user will be labeled as having earwax buildup and blockage. We adopted this approach because the task is straightforward to solve and does not require the use of a neural network, which would be more time-consuming. Additionally, the intuitive method provides a visual representation that allows users to better interpret the results.

4.4 Examination from the Tympanic Membrane Photo

After examining the condition of the external auditory canal, if there is no earwax blockage, the user can proceed to take a photo of the tympanic membrane (TM) to detect possible diseases. Our model takes the TM image as input with pre-processing tools including center crop, resize, and color normalization to improve model stability [15]. Due to the significant advances in deep learning in image classification in recent years, we selected the DenseNet model [9], a well-performing convolutional neural network (CNN), which was also confirmed to have better results on ear disease classification problems [11]. We use one-hot to transform the label of images into a vector of length N. Unlike single-label classification, there may be multiple positions of 1 in the label vector, which correspond to the labels of the ground truth. The model is forward propagated to obtain features at various levels, which are finally used for multi-label classification of the images. Moreover, in order to enable the model to better learn the possible dependencies and associations between different diseases, we added correction terms based on BCE LOSS (Binary Cross Entropy), as shown below:

$$L = BCE(P, G) + \frac{\alpha}{2}(Pi - Pj)^2 \tag{2}$$

where L refers to loss, P refers to predictive probability, G refers to ground truth, Pi, Pj refers to the diseases that typically have correlations with each other. In addition, to prevent the model from being optimized to an absolute correlation, we also add a coefficient α to control the balance between BCE loss and the regular term.

To enhance the model's robustness and reduce overfitting issues, we employed data augmentation techniques such as horizontal flipping, random rotation, color enhancement, and random cropping. We used the Adam optimizer[14], training the model for 50 epochs with a batch size of 16. The model outputs the predicted probability of each ear health condition being present after analyzing the predicted result and the ground truth label, we decide to let the system present the ear health condition whose predicted possibility exceeds 0.7. If none of the predicting possibilities exceeds 0.7, the system outputs "unlabelled disease".

5 Evaluation

5.1 Technical Evaluation

We evaluated our model using a laptop, with an Intel(R) i7-3632QM 2.2 GHz CPU, 7 GB GPU, and 64-bit Windows 7 operating system as 1 b shows. We choose commonly-used metric accuracy, positive predictive value (PPV, often called precision), sensitivity, and specificity to evaluate our approach [16]. Accuracy represents the overall effectiveness of a classifier. PPV answers the question: "How likely is it that this user has the disease given that the test result is positive?" [16, 22]. Sensitivity refers to the ability of the test to correctly identify those patients with the disease. [16, 22, 29]. The specificity of an assessment model refers to the ability of the test to correctly identify those patients without disease [16, 22, 29].

We initially evaluated our model separately for each ear condition as the Table 2 lists. Among them, the earwax buildup and blockage (EB) were evaluated using the algorithm described in Sec. 4.3, and other conditions including healthy ear (HE), earwax buildup and blockage (EB), TM perforation (TMP), otitis media with effusion(OME), suppurative otitis media (SOM), cholesteatoma (CH) were evaluated using the deep-learning model described in Sec. 4.4. Overall, our model achieves high accuracy, precision, and specificity on these six ear health conditions, with accuracies

Table 2. Performance on different ear conditions for healthy ear (HE), earwax buildup and blockage (EB), TM perforation (TMP), otitis media with effusion (OME), suppurative otitis media (SOM), cholesteatoma (CH).

Condition	Accuracy	PPV (Precision)	Sensitivity(Recall)	Specificity
EB	0.940	0.930	0.949	0.931
HE	0.972	0.946	0.963	0.975
TMP	0.949	0.953	0.803	0.989
OME	0.949	0.963	0.765	0.993
SOM	0.954	0.926	0.906	0.973
CH	0.960	1.000	0.794	1.000

over 90%. It indicates that the model has a strong ability and high accuracy in confirming diseases it detects as positive and identifying those patients without the disease. EB, HE, and SOM, also have a high sensitivity, which suggests that they can more correctly identify those patients with the disease. Especially for HE, its high sensitivity avoids the situation that people in a healthy state are misdiagnosed as getting ear diseases and can make notification that if the model did not output a healthy state and other conditions, the user may probably get ear diseases which are not labeled in the common cases. TMP, OME, and CH have relatively lower sensitivity, which means that they may miss out some certain cases which getting these diseases, while the high precision indicates that once the system identifies you have a certain disease, you may have a great probability of really getting one. Although our conditions generally exhibit high accuracy, we have identified several reasons for failure cases, including lighting, angle, and atypical presentations, and discuss the effect of data distribution: 1) Lighting: In some in-the-wild data collections, reflections caused by clinic ceiling lighting were mistakenly identified as suppurative otitis media (SOM). This misrecognition highlights the sensitivity of our model to variations in lighting conditions. That shows users should use it without extra strong light. 2) Angle: Despite having a variety of angles in our training dataset, certain cases with abnormal angles were incorrectly detected. This issue may stem from the device's round shape, which allows it to rotate freely, affecting model performance. To address this, we are considering modifications to restrict the device to a fixed angle, potentially simplifying condition detection. 3) Atypical Cases: For tympanic membrane perforation (TMP), some cases involve small holes that may be overlooked by our model. Cholesteatoma (CH) presents variably, and due to limited data, our model struggles to recognize these variations, contributing to lower recall rates. 4) Data Distribution: Our dataset included a higher proportion of healthy ear (HE) conditions to teach the model about the general diversity of healthy ears, which indeed resulted in high performance for HE detection. However, this imbalance may have led to relatively low recall for TMP, OME, and CH. Moreover, the presence of conditions like OME+SOM complicates detection, potentially confusing the model when differentiating between OME and SOM. This could be another factor in reduced recall. For future improvements, we need to reconsider the most effective methods for including complication detection. Our model demonstrates high specificity, indicating a conservative approach: it generally does not falsely diagnose a disease when it is absent. However, this conservatism may inadvertently miss some disease cases, which is a limitation that needs addressing.

5.2 User and expert evaluation

We validated the effectiveness of *EarMonitor* through a user study with 17 participants and two experts, which provided valuable insights into the design and interpretation of non-clinical assessment devices.

5.2.1 User Evaluation. **Participants.** We conducted the user studies with 17 participants (aged 26 to 82; 6 male and 11 female; M = 52.9; SD = 17.3) as Table 3 shows. Among 17 participants, six reported not having any symptoms, while 11 reported having one.

Procedure. We first introduced the background of our research project and clarified that the study was completely voluntary and no personally identifiable information would be collected. After obtaining their informed consent, we demonstrated how to use *EarMonitor* and let them freely try the sensing camera to adapt the visual video to their operations. When they felt that they were able to use it independently, they were asked to experience the whole process of self-assessment. Afterward, participants rated the usability experience using the 7-Likert scale. We clarified to participants that usability means how easy they feel to use the system and if there are any difficulties during the process. The easier it is to use, the higher the score is. Finally, we interviewed them about their experience and attitude toward using *EarMonitor*, challenges, and expectations for the tool.

Table 3. Participants' Information, Self-assessment Result, and Expert Result

Id	Age	Sex	Having Symptom	Self-assessment Result	Expert Result
P1	63	M	no	healthy	healthy
P2	27	F	earache	swelling in the external auditory canal	otitis external
P3	65	F	intermittent tinnitus	healthy	healthy
P4	61	M	no	healthy	healthy
P5	64	M	hearing loss	otitis media with effusion	otitis media with effusion
P6	43	F	stuffy ears	otitis media with effusion	otitis media with effusion
P7	35	F	tinnitus	healthy	healthy
P8	68	F	hearing loss	healthy	healthy
P9	60	F	hearing loss	earwax buildup and blockage	earwax buildup and blockage
P10	82	F	hearing loss	cholesteatom	cholesteatom
P11	27	M	no	healthy	healthy
P12	65	F	no	healthy	otitis external
P13	47	F	earache	TM perforation and suppurative otitis media	TM perforation and suppurative otitis media
P14	26	M	no	healthy	healthy
P15	72	F	hearing loss	healthy	healthy
P16	40	M	hearing loss	TM perforation	TM perforation
P17	55	F	no	healthy	healthy

User experience and usability. We asked our participants to rate the user experience and usability of the four stages of using *EarMonitor* (from 1-strongly low to 7-strongly high), which are filling in the questionnaire, detecting the external auditory canal, detecting TM, and interpreting the comprehensive assessment. Figure 6 shows the result. Overall, participants reported easy to use *EarMonitor*. P4 reported that it is easy and quick to use, and easily carried. It recalled his memory, that "*one time, when I traveled to a small town, after getting off the airplane, I felt earache badly. I went to the hospital in the small town immediately and was told that my eardrum ruptured. I was so terrified and hurried to the nearby city for further assessment, but was told I was fine. This accident ruined my trip, if I had this tool with me at that time, I could at least notify the doctor in that small town and ask him to test it again to avoid the unpleasant experience.*" P5 highly appraised *EarMonitor* for detecting his potential ear disease, "*I got slightly hearing loss recently, but I thought it was the normal procedure as I am aging gradually. If not you, I could not find that I have already got the ear disease.*" Moreover, P13 appreciated the usefulness of using *EarMonitor* to detect and monitor the ear disease she had before. P13 said, "*I had got the chronic suppurative otitis media before, and suspect that my earache this time also raised by this. Your tool confirms my suspicion and I will take the same medicine I used to take as an instant solution and go to the hospital soon. I expect that I can also use it to monitor the following recovering process.*" Participants further praised the comprehensive assessment, which not only considers the model's output but also considers their input. P7 reported that she sometimes got tinnitus, "*I think that I may get a certain type of ear disease, but the system shows that I am healthy. For the result, I felt doubt at first, while the system prompted that according to my self-reported questionnaire, I may have just gotten too tired recently. I like the comprehensive interpreting result, I feel more relieved and have decided to have more rest these days.*" P16 reported that he liked that the system would list his reported symptoms and the symptoms that he was assessed with, "*I can see that these symptoms are consistent, which I am more sure about the result.*"

While some participants raised concerns and difficulties. P3 felt a little afraid to put it into their ear deeper although she was told about our safety design. "*Since I was a child I have been clumsy and not good at crafts. I am afraid that I can't control it well.*"- P3. P2 reported that the color of the ear through the camera looks distorted, and may influence his general justification based on his life experience. Several older adult participants complained that the questionnaire and assessment part consisted of too many words, which made them "hurting eyes". As P10 suggested, she preferred having a conversation rather than filling out the questionnaire during the consultation. Similarly, P17 also suggested that "*at least let the system read these questions in the meantime.*" Moreover, P3 complained that some choices in questionnaires make her confused, "*I feel hard to recognize which type of tinnitus I got, buzzing or roaring sounds.*" While P11 preferred the questionnaire format, he said that it made him feel more comfortable than talking to real doctors because he gets nervous when seeing doctors. Moreover, several participants raised doubt about the result of the analysis, "*Is it accurate enough?*" After we explained that the evaluation result was compatible or even better than general practitioners and pediatricians, they generally showed their surprise and expressed approval of using in daily lives. P15 further explained, "*I would like to use it for monitoring my health since I am old and easily get diseases. For me, I can accept that if I am in a healthy state, I but be misjudged as unhealthy. The situation I am afraid of is that it won't recognize it. So if I have symptoms and can get an appointment, I'll still go to the hospital.*" As for the endoscope sensing device, several participants complained that with the use time increasing, the endoscope gets a little hot, although it impacts not much, they expect the device could have a low temperature. Overall, our system was found usable by the majority of the participants, although improvements can be made to further address the aforementioned complaints.

5.2.2 Expert Evaluation. **Participants.** To validate *EarMonitor*, we invited two otologists (E1: 16 years of practice, E2: 23 years of practice) to confirm our previous participants' disease assessment and attend an interview study to evaluate on *EarMonitor*.

Procedure. We asked two experts to review the images and video captured in the previous study. They were consistent in their assessment of the disease as shown in Table 3. Notably, although P12

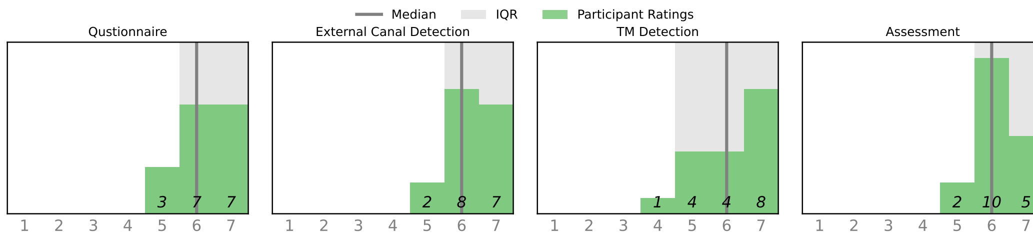

Fig. 6. Participants' ratings on four steps: Questionnaire, External Auditory Canal Detection, TM detection, and Assessment Result. 1 means strongly low, and 7 means strongly high.

did not get those diseases our system covers, she was found to get a slightly otitis external. We further asked their opinions on the validation of operating *EarMonitor* at home.

Results. Both experts agreed that *EarMonitor* can give a reasonable assessment. E1 said, "*There are only a few common clinical ear diseases, and the system seems to give a relatively accurate assessment, especially for complications, it is great that it can be pointed out at the same time.*" E2 further mentioned, "*The assessment combines the result of the consultation questionnaire, which is more reasonable and follows the same pipeline in the clinic.*" E2 also expressed his compliment on the instant solution given by the system, he considered that it was beneficial for the patients who can not have a doctor's appointment in time because all of the recent outpatient dates had been booked. While E1 raised questions about the possibility of making mistakes in the system. After explaining the principle and evaluation result of our model, E1 confirmed that the ability to detect whether the user is in healthy ear condition is important, even if the result of a specific disease is not perfectly accurate, at least the user knows that they may need to go to the hospital. E1 suggested that the system should add more notifications on certain severe symptoms like severe earache, and the user should go to the hospital for double checking to keep safe. Furthermore, for the external auditory canal, E2 suggested adding more diseases that may happen in the external auditory canal such as otitis external.

The overall procedure participants operated by themselves are considered valid after being introduced to the safety design of *EarMonitor*. E1 touched the camera, then tried it by himself and suggested that "*It can be better with softer materials, because the human ear canal is zigzag, even after pulling the ear as straight as possible by hand, the hard material may not be good for some people and cause discomfort.*"

Moreover, they confirmed that the images captured by users themselves are clear enough for home usage to recognize the common ear diseases as Fig 7 shows. Both of them considered that video can contain more information than images because video can capture different angles of the ear canal. They also mentioned that the light is a little distorted compared with the real eyes although it may not affect the system's detection. Interestingly, E2 raised an expectation that she hoped the system could also be online for access for doctors so that doctors can take a further check for those patients who get alerted by the system and have trouble coming to hospitals in a short time.

6 Discussion

In this section, we outline current limitations and lessons from the user and expert studies and discuss possible solutions for future work and implications for the self-assessment system.

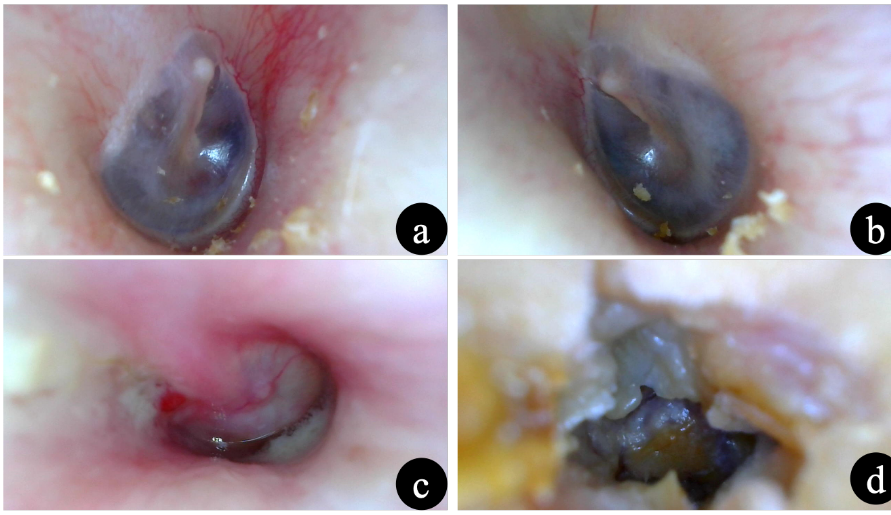

Fig. 7. Images that were taken by our participants. Generally, they are well captured, figure c and figure d appear to have a little focus problem but do not impact the model's detection result.

6.1 Implications of Self-assessment Tools

Device requirement. Participants reflected that the tools for self-assessment need to be safe, easy to use, and comfortable. Our current hand-held endoscope is designed as a straight tube-shaped with a circle of small light bubbles. The light bubbles are used for lighting the view inside the ear, however, they become hot as usage time increases, as participants have complained. Similar to Liang's study on the oral cavity, sufficient lighting is also essential for ear health examination [18]. These lights utilize low-cost RTD light sources, suggesting that the brighter the lighting required, the hotter the device becomes. Therefore, we may explore cold light sources to keep the temperature cool while keeping sufficient lighting without increasing too much cost. The current housing of this endoscope is made of stainless steel, which is hard and may cause discomfort. Indeed, P3 reported that she felt it was not comfortable when it touched the ear. Based on E1's suggestion, we could use a softer material, such as rubber, to cover the endoscope and enhance the user experience. Furthermore, our error case analysis revealed that the round shape of the device can lead to varying angles in captured images, which complicates the model's ability to deliver accurate results. For future developments in self-assessment image-based tools, it would be beneficial to incorporate design features that help users maintain a fixed angle. For instance, adding a holder could make it easier for users to hold the device at a consistent angle.

Effective support for self-report with multi-modal description and interactive conversation. In the self-assessment setting, users feel less burdened and are likely willing to report more symptoms [24, 28]. However, some participants feel the experience is not as effective as visiting doctors due to a lack of contextual information and interaction. Especially, when they would like to know more context information such as the types of tinnitus (e.g., buzzing or roaring sounds), they ask for sample sounds for more accurate expression. A multi-modal description providing more context information will enhance their experience. Moreover, chatbots have been used and evaluated the feasibility of conducting interviews in one-on-one text-based conversations [12, 40] and many patients preferred the chatbot questionnaire over the regular questionnaire [35]. The design and strategy of chatbots in self-assessment settings are worth exploring in the future.

Potentials in managing chronic diseases and recovery process. From the user study, we identify an extra need for tracking personal ear health beyond the general monitoring of ear health conditions in daily life. P13 suggested that she expects that *EarMonitor* can keep tracking her ear health condition because she once got the SOM, and may recur. In addition, she expressed her expectation that during the treatment period, she also would like to see how much she progresses to be healthy. Tracking one person's ear health conditions can be a challenging topic because it needs the model to be very sensitive to the small changes that users get. We may further investigate how to design a system of long-term tracking of a single person's ear health for future work.

6.2 Limitations and Insights on Improving the Credibility of EarMonitor.

Our current approach has an average 95.7% classification accuracy, which is compatible or even better than average general practitioners and pediatricians [17, 30, 31]. Therefore, it shows a great possibility that *EarMonitor* can make the automated assessment of common ear diseases in medically underserved populations. Although *EarMonitor* shows its promising prospects, it has several limitations as we pointed out in Sec. 5.1. We then discuss these limitations and raise the following improvement points that can further enhance the credibility of the system.

Collect more diverse data and enhance the data robustness. One limitation is the data component. Due to collection difficulties, not all data is gathered from our devices — about half comes from hospitals using professional devices. While this hospital data helps the model learn features more effectively, incorporating more in-the-wild data could enhance model robustness. Additionally, hospital data often comes from patients with serious illnesses, which might skew the model toward recognizing more severe disease symptoms. This could lead to neglect of milder symptoms, which are crucial for broader self-assessment applications. Therefore, further data collection efforts should aim to include a more diverse range of symptom severity to improve the model's robustness.

Influence on false negative and credibility improvement. Our current model demonstrates high specificity, which indicates a conservative approach but may inadvertently miss some disease cases. The occurrence of false negatives, where serious conditions are mistakenly identified as healthy conditions, presents potential risks. Such errors may delay necessary medical treatment, exacerbate underlying conditions, and lead to severe health outcomes, including complications such as hearing loss or systemic infections. For patients, the reliability of *EarMonitor* is crucial for both immediate health decisions and long-term health management. Ensuring accurate health assessments is paramount for maintaining trust in the device. For medical professionals and experts, while the model reassures those without ear diseases, minimizing unnecessary clinical visits, it also prompts early intervention for those recognized as having diseases. Despite the occurrence of false negatives, using the system is still preferable to not using it at all. Experts recommend that symptomatic individuals seek further clinical evaluation. However, for those in areas with inadequate medical resources, immediate access to a doctor may not be feasible. To mitigate the risks associated with false negatives and to bolster the credibility of EarMonitor, we propose expert involvement and human-AI collaboration strategies to enhance the device's performance. We are committed to engaging with medical experts throughout the development and evaluation processes to ensure that EarMonitor adheres to clinical standards and builds professional trust. For instance, we could adopt a crowd-sourcing approach, inviting experts to periodically review and evaluate a selection of the system's results during their available time. To improve the system's reliability, we are considering the implementation of an online interface that allows doctors to interact with patients remotely. This would enable healthcare professionals to verify the system's results and provide detailed medical advice, particularly benefiting patients who face difficulties in accessing hospital care promptly. *EarMonitor* is designed to complement, not replace, traditional medical diagnostics by facilitating preliminary assessments, especially valuable in regions with limited medical resources. E2 suggested enhancing the system to provide doctors with online access for further evaluations of patients flagged by the system, which could expedite care and refine treatment plans. Additionally, feedback from these consultations could be used to refine the training dataset, provided users consent to share their data, thereby enhancing the accuracy of *EarMonitor*. We aim to explore the best ways to integrate AI-based assessments with professional healthcare to offer more precise and affordable ear health monitoring.

Consider more relationships beyond dependency between the diseases. Multi-class classification means a classification task with more than two classes and assumes that the classes are mutually exclusive. It assumes that each sample is assigned to one and only one label. Previous research has adopted a multi-class classification method to detect ear diseases [10]. While, ear diseases are often having complications, which means when you get disease A, you may probably get disease B at the same time [1]. *EarMonitor* adopts multi-label classification and can identify more than one ear health condition at one time, which is consistent with the reality that ear diseases are not independent of each other [37]. We also consider the correlation between the diseases and add the feature to enhance our model. However, according to the analysis result of our bad cases, we found that certain images that should only be categorized to one label were predicted to be two labels. We suggest that further dependency between the diseases needs to be considered as an important improvement for self-examination, including the probability of having other diseases at the same time, and the existence of mutually exclusive diseases. By addressing these concerns through robust technological improvements and transparent communication, the potential for false negatives can be reduced, thereby safeguarding patient health and strengthening the credibility and reliability of EarMonitor in medical and consumer markets.

Develop specific models for different age groups. Aging is one of the main factors in increasing the risk of getting hearing loss [8, 23, 25, 26]. P15 also expressed her conservative attitude toward using *EarMonitor* with high expectations of recognizing more potential ear diseases. In our current system, we did not train specific models for different age groups, which may not meet all their needs. We may conduct the following study to investigate whether there are different expectations of different groups and whether training specific models for them has better performance for future work.

Include more external auditory canal disease detection. In the current system, for the external auditory canal, we only involve the detection of earwax buildup and blockage as much previous research was done in the laboratory [19**?** , 20]. However, there are still kinds of external ear diseases such as external otitis, whose feature varies, and is more difficult to achieve a high accuracy rate by the technical means of image recognition at present. Notably, during our user evaluation, P12 was found to have otitis external, which indicates that we should include more conditions in the external ear. Beyond the machine learning and deep learning approach, We will further explore the solution considering involving useful traditional computer vision technology and detailed information of consultation results as an extra feature in the subsequent work.

Combine with other detection methods. Through the analysis of bad cases, we found out that there exist certain diseases that are ignored by the current system. Especially for OME, features of certain images are not obvious due to the special case or lighting (e.g., reflective), which leads the model to predict a probability lower than the threshold. The acoustic reflectometry approach can detect the middle ear fluid, which can complement detecting OME very well [13, 36]. Extended approach *EarHealth* can detect three ear health conditions through the acoustic reflectometry approach [10]. We may explore combining the result of acoustic reflectometry technology as a complementing method to improve the credibility of the system.

7 Conclusion

With the increasing population getting hearing loss, prevention, identification, and management of ear diseases are imperative. While, medical resources and qualified otologists are limited, especially in low-resourced countries. In this study, we have presented *EarMonitor*, a low-cost, feasible, portable system that enables users to self-examine their ear health condition easily at home. *EarMonitor* utilizes a hand-held endoscope camera to capture ear auditory canal photos and tympanic membrane photos with the input from users' self-report questionnaires for accurate detection of six common ear health conditions. Our result shows that *EarMonitor* can achieve a compatible performance with general practitioners and pediatricians, which makes progress toward making affordable ear health assessment feasible. We also identified several challenges that need to be addressed in the future, along with insights to enhance the development of self-assessment tools and bolster their reliability.

Acknowledgements

We would like to express our sincere gratitude to Yanyan He for providing professional support on ear health. We also extend our thanks to Xian Wang and Zhengyang Ma for their assistance in contacting hospital resources. Special thanks to the ophthalmologists from our collaborating hospitals for their invaluable contributions to this research.

References

[1] John Jacob Ballenger and James Byron Snow. 2003. *Ballenger's otorhinolaryngology: head and neck surgery.* Pmph-usa.

[2] Colin Barry, Jessica de Souza, Yinan Xuan, Jason Holden, Eric Granholm, and Edward Jay Wang. 2022. At-Home Pupillometry using Smartphone Facial Identification Cameras. (4 2022), 1–12. https://doi.org/10.1145/3491102.3502493

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[4] Therese Canares, Weiyao Wang, Mathias Unberath, and James Clark. 2021. Artificial intelligence to diagnose ear disease using otoscopic image analysis: a review. *Journal of Investigative Medicine* 70 (09 2021), jim–2021. https://doi.org/10.1136/jim-2021-001870

[5] Justin Chan, Sharat Raju, Rajalakshmi Nandakumar, Randall Bly, and Shyamnath Gollakota. 2019. Detecting middle ear fluid using smartphones. *Science translational medicine* 11, 492 (2019), eaav1102.

[6] L Dimitrov, S Unadkat, A Khanna, C Rennie, and H Saleh. 2020. ENT training amongst general practitioners: results from a questionnaire. *The Journal of Laryngology & Otology* 134, 2 (2020), 109–115. https://doi.org/10.1017/S0022215120000201

[7] K Graydon, C Waterworth, H Miller, and H Gunasekera. 2019. Global burden of hearing impairment and ear disease. *The Journal of Laryngology & Otology* 133, 1 (2019), 18–25. https://doi.org/10.1017/S0022215118001275

[8] Lydia M Haile, Kaloyan Kamenov, Paul Svitil Briant, Aislyn U Orji, Jaimie D Steinmetz, Amir Abdoli, Mohammad Abdollahi, Eman Abu-Gharbieh, Ashkan Afshin, Haroon Ahmed, et al. 2021. Hearing loss prevalence and years lived with disability, 1990–2019: findings from the Global Burden of Disease Study 2019. *The Lancet* 397, 10278 (2021), 996–1009.

[9] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 4700–4708.

[10] Yincheng Jin, Yang Gao, Xiaotao Guo, Jun Wen, Zhengxiong Li, Zhanpeng Jin, and Zhan-Peng Jin. 2022. EarHealth: An Earphone-based Acoustic Otoscope for Detection of Multiple Ear Diseases in Daily Life; EarHealth: An Earphone-based Acoustic Otoscope for Detection of Multiple Ear Diseases in Daily Life. *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services* 1 (2022). https://doi.org/10.1145/3498361

[11] Mohammad Azam Khan, Soonwook Kwon, Jaegul Choo, Seok Min Hong, Sung Hun Kang, Il-Ho Park, Sung Kyun Kim, and Seok Jin Hong. 2020. Automatic detection of tympanic membrane and middle ear infection from oto-endoscopic images via convolutional neural networks. *Neural Networks* 126 (2020), 384–394.

[12] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems.* 1–12.

[13] Sandra Kimball. 1998. Acoustic reflectometry: Spectral gradient analysis for improved detection of middle ear effusion in children. *The Pediatric infectious disease journal* 17, 6 (1998), 552–555.

[14] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2017. Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 6 (2017), 84–90.

[16] Abdul Ghaaliq Lalkhen and Anthony McCluskey. 2008. Clinical tests: sensitivity and specificity. *Continuing education in anaesthesia critical care & pain* 8, 6 (2008), 221–223.

[17] J. M. Legros, H. Hitoto, F. Garnier, C. Dagorne, E. Parot-Schinkel, and S. Fanello. 2008. Clinical qualitative evaluation of the diagnosis of acute otitis media in general practice. *Int J Pediatr Otorhinolaryngol* 72, 1 (2008), 23–30. https://doi.org/10.1016/j.ijporl.2007.09.010 Epub 2007 Oct 31.

[18] Yuan Liang, Hsuan Wei Fan, Zhujun Fang, Leiying Miao, Wen Li, Xuan Zhang, Weibin Sun, Kun Wang, Lei He, and Xiang Anthony Chen. 2020. OralCam: Enabling Self-Examination and Awareness of Oral Health Using a Smartphone Camera. *Conference on Human Factors in Computing Systems - Proceedings* 20 (4 2020). https://doi.org/10.1145/3313831.3376238

[19] Devon Livingstone and Justin Chau. 2020. Otoscopic diagnosis using computer vision: An automated machine learning approach. *The Laryngoscope* 130, 6 (2020), 1408–1413.

[20] Devon Livingstone, Aron S Talai, Justin Chau, and Nils D Forkert. 2019. Building an Otoscopic screening prototype tool using deep learning. *Journal of Otolaryngology-Head & Neck Surgery* 48, 1 (2019), 1–5.

[21] Ritvik P Mehta, John J Rosowski, Susan E Voss, Ellen O'Neil, and Saumil N Merchant. 2006. Determinants of hearing loss in perforations of the tympanic membrane. *Otology & neurotology: official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology* 27, 2 (2006), 136.

[22] Thomas F Monaghan, Syed N Rahman, Christina W Agudelo, Alan J Wein, Jason M Lazar, Karel Everaert, and Roger R Dmochowski. 2021. Foundational statistical principles in medical research: sensitivity, specificity, positive predictive value, and negative predictive value. *Medicina* 57, 5 (2021), 503.

[23] Lorenzo Monasta, Luca Ronfani, Federico Marchetti, Marcella Montico, Liza Vecchi Brumatti, Alessandro Bavcar, Domenico Grasso, Chiara Barbiero, and Giorgio Tamburlini. 2012. Burden of disease caused by otitis media: systematic review and global estimates. *PloS one* 7, 4 (2012), e36226.

[24] Tom Nadarzynski, Oliver Miles, Aimee Cowie, and Damien Ridge. 2019. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital health* 5 (2019), 2055207619871808.

[25] World Health Organization. 2022. Deafness and Hearing Loss. https://www.who.int/health-topics/hearing-loss#tab=tab_1

[26] World Health Organization et al. 2015. Hearing loss due to recreational exposure to loud sounds: a review. (2015).

[27] World Health Organization et al. 2017. *Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report, 2017.* World Health Organization.

[28] Adam Palanica, Peter Flaschner, Anirudh Thommandram, Michael Li, Yan Fossat, et al. 2019. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *Journal of medical Internet research* 21, 4 (2019), e12887.

[29] Rajul Parikh, Annie Mathai, Shefali Parikh, G Chandra Sekhar, and Ravi Thomas. 2008. Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology* 56, 1 (2008), 45.

[30] Michael E Pichichero and Michael D Poole. 2001. Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. *Archives of pediatrics & adolescent medicine* 155, 10 (2001), 1137–1142.

[31] Michael E Pichichero and Michael D Poole. 2005. Comparison of performance by otolaryngologists, pediatricians, and general practioners on an otoendoscopic diagnostic video examination. *International journal of pediatric otorhinolaryngology* 69, 3 (2005), 361–366.

[32] Ali Qureishi, Yan Lee, Katherine Belfield, John P Birchall, and Matija Daniel. 2014. Update on otitis media–prevention and treatment. *Infection and drug resistance* 7 (2014), 15.

[33] Michael Reiss and Gilfe Reiss. 2010. Suppurative chronic otitis media: etiology, diagnosis and therapy. *Medizinische Monatsschrift fur Pharmazeuten* 33, 1 (2010), 11–6.

[34] Alvy Ray Smith. 1978. Color Gamut Transform Pairs. In *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '78)*. Association for Computing Machinery, New York, NY, USA, 12–19. https://doi.org/10.1145/800248.807361

[35] Mariska E Te Pas, Werner GMM Rutten, R Arthur Bouwman, and Marc P Buise. 2020. User experience of a chatbot questionnaire versus a regular computer questionnaire: prospective comparative study. *JMIR Medical Informatics* 8, 12 (2020), e21982.

[36] David W Teele and John Teele. 1984. Detection of middle ear effusion by acoustic reflectometry. *The Journal of pediatrics* 104, 6 (1984), 832–838.

[37] Grigorios Tsoumakas and Ioannis Katakis. 2007. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)* 3, 3 (2007), 1–13.

[38] Michelle Viscaino, Juan C Maass, Paul H Delano, Mariela Torrente, Carlos Stott, and Fernando Auat Cheein. 2020. Computer-aided diagnosis of external and middle ear conditions: A machine learning approach. *PLoS One* 15, 3 (2020), e0229226.
[39] Tony Wright. 2015. Ear wax. *BMJ* 351 (2015).
[40] Ziang Xiao, Michelle X Zhou, Q Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 3 (2020), 1–37.
[41] Xuhai Xu, Ebrahim Nemati, Korosh Vatanparvar, Viswam Nathan, Tousif Ahmed, Md Mahbubur Rahman, Daniel McCaffrey, Jilong Kuang, and Jun Alex Gao. 2021. Listen2Cough: Leveraging End-to-End Deep Learning Cough Detection Model to Enhance Lung Health Assessment Using Passively Sensed Audio. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 42 (mar 2021), 22 pages. https://doi.org/10.1145/3448124
[42] Hanbin Zhang, Li Zhu, Viswam Nathan, Jilong Kuang, Jacob Kim, Jun Alex Gao, and Jeffrey Olgin. 2021. Towards Early Detection and Burden Estimation of Atrial Fibrillation in an Ambulatory Free-Living Environment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 86 (jun 2021), 19 pages. https://doi.org/10.1145/3463503

Received February 2024; revised May 2024; accepted June 2024