Intelligent smartphone-based multimode imaging otoscope for the mobile diagnosis of otitis media

THIAGO C. CAVALCANTI,[1] [iD] HAH MIN LEW,[1] [iD] KYUNGSU LEE,[1] SANG-YEON LEE,[2] MOO KYUN PARK,[2,3,4] AND JAE YOUN HWANG[1,3,5]

[1]*Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, Republic of Korea*
[2]*Department of Otorhinolaryngology-Head and Neck Surgery, Seoul National University Hospital, Seoul, Republic of Korea*
[3]*co-corresponding authors*
[4]*aseptic@snu.ac.kr*
[5]*jyhwang@dgist.ac.kr*

Abstract: Otitis media (OM) is one of the most common ear diseases in children and a common reason for outpatient visits to medical doctors in primary care practices. Adhesive OM (AdOM) is recognized as a sequela of OM with effusion (OME) and often requires surgical intervention. OME and AdOM exhibit similar symptoms, and it is difficult to distinguish between them using a conventional otoscope in a primary care unit. The accuracy of the diagnosis is highly dependent on the experience of the examiner. The development of an advanced otoscope with less variation in diagnostic accuracy by the examiner is crucial for a more accurate diagnosis. Thus, we developed an intelligent smartphone-based multimode imaging otoscope for better diagnosis of OM, even in mobile environments. The system offers spectral and autofluorescence imaging of the tympanic membrane using a smartphone attached to the developed multimode imaging module. Moreover, it is capable of intelligent analysis for distinguishing between normal, OME, and AdOM ears using a machine learning algorithm. Using the developed system, we examined the ears of 69 patients to assess their performance for distinguishing between normal, OME, and AdOM ears. In the classification of ear diseases, the multimode system based on machine learning analysis performed better in terms of accuracy and F1 scores than single RGB image analysis, RGB/fluorescence image analysis, and the analysis of spectral image cubes only, respectively. These results demonstrate that the intelligent multimode diagnostic capability of an otoscope would be beneficial for better diagnosis and management of OM.

1.   Introduction

Otitis media (OM) is a spectrum of ear disorders that is fairly common, especially in children [1]. OM with effusion (OME) is defined as the presence of fluid in the middle ear without acute ear infection [2]. OME is one of the main types of OM and is responsible for 10%–15% of visits to the clinic in childhood [3]. OME is the most common cause of hearing loss in children in developed countries [4]. OME is related to poor language learning and school performance during childhood [5,6]. OME and adhesive OM (AdOM) are two common middle ear inflammations. Both are characterized by the presence of effusion in the middle ear chamber and often appear because of a previous acute bacterial infection in that area, which is distinguished by the formation of tympanic membrane adhesions with other middle ear structures in the case of AdOM [7,8]. They are nearly asymptomatic disorders and tend to evolve into chronic cases. If not diagnosed and treated promptly, they may cause ossicular chain damage and progress to cholesteatoma

development [8]. Prompt and accurate diagnosis and management of ear disorders are critical to avoid these issues.

The detection of middle ear fluid is a requirement for a specific diagnosis of OM [9]. Commonly, the presence of effusion in the middle ear was confirmed using a pneumatic otoscope. However, the need to adequately seal the ear canal poses a challenge for the use of such tools. Alternatively, acoustic reflectometry may be applied for this, but this approach also presents a limitation because it only allows the acquisition of one-dimensional data. Many efforts have been made to include various functionalities in traditional otoscopy, such as spectroscopy, short-wave infrared imaging, and optical coherence tomography [10–13].

The clinical environment requires nurses and physicians to be in constant motion, meeting many patients a day in different locations. In that milieu, portable tools are essential to increase the efficiency of health care professionals [14]. Consequently, smartphones have been increasingly incorporated in the medical practice and thus allow aiding physicians in the clinical decision-making, monitoring patients with chronic diseases, and increasing their productivity and efficiency [15–20]. Additionally, smartphone-based otoscopy has also been reported as a diagnostic and monitoring tool for ear diseases [21–23].

Fluorescence was used to diagnose ear diseases in previous reports. Using chinchilla models, Spector et al. studied the potential of fluorescence spectroscopy to assess bacteria causing acute OM infections [24]. Using fluorescence spectroscopy, four bacteria causing acute OM infection were successfully distinguished in animal models. Levy et al. and Yim et al. tested the potential of fluorescence imaging for the diagnosis of cholesteatoma and OM, respectively [25,26]. Fluorescence imaging is useful for diagnosis; however, exogenous fluorophores are required [25,26]. In situ autofluorescence imaging has been used as a useful methodology for assessing biological tissues in medicine and biology. Cells and tissues naturally emit fluorescence, which is also known as autofluorescence, when excited by light at suitable wavelengths (ultraviolet, visible, or near-infrared). Autofluorescence from cells and tissues is closely related to their morphological and metabolic conditions. Therefore, comprehensive diagnostic information on diseases can be obtained via a noninvasive analysis of tissue autofluorescence without contrast agents [27]. Valdez et al. explored autofluorescence imaging of the middle ear in the detection of ear diseases [28]. They described a study of the potential of a multiwavelength autofluorescence imaging otoscope for the diagnosis of cholesteatoma.

In our previous work, we developed a smartphone-based spectral imaging otoscope, capable of obtaining a spectral image cube that contains spectral information for the diagnosis of middle ear diseases [29]. The system allowed differentiation between the normal and abnormal tympanic membranes. Therefore, it showed a high capability for the quantitative detection of chronic OM with high contrast, implying that the smartphone-based spectral imaging otoscope may have the potential for mobile diagnosis of various middle ear diseases. Previous studies demonstrated that the incorporation of multiple imaging modalities, acquiring complementary information from the sample, provides an important performance improvement compared to single imaging modality diagnostic tools [30–32]. Therefore, an advanced smartphone-based otoscope with multimode imaging and analysis capability must be developed for better versatility in the diagnosis of various ear diseases with high sensitivity and specificity.

We, therefore, developed an intelligent smartphone-based multimode imaging otoscope for better mobile diagnosis of ear diseases and demonstrated the potential of machine learning-based multimode image analysis in distinguishing between normal, OME, and AdOM ears. To acquire different but complementary information about the tympanic membrane of ears, RGB, autofluorescence, and spectral imaging modalities are integrated with a smartphone-based otoscope, and further machine-learning-based multimode image analysis is implemented on the smartphone-based multimode otoscope. The intelligent smartphone-based multimode imaging otoscope was used to examine the ears of 69 patients with and without middle ear diseases such as

OME or AdOM to evaluate it. To differentiate normal ears, OME, and AdOM, various machine learning techniques such as multilayer perceptron (MP), random forest (RF), logistic regression (LR), decision trees (DTs), and Naïve Bayes (NB) are trained and tested with a multimode dataset, acquired using our developed system. The machine learning techniques were further compared to conventional spectral classification algorithms, such as the spectral angle mapper (SAM) and Euclidean distance (ED), demonstrating the potential of the intelligent smartphone-based multimode imaging otoscope for mobile ear disease diagnosis.

2. Materials and methods

2.1. Intelligent smartphone-based multimode otoscope

We developed an intelligent multimode smartphone-based otoscope. The devised otoscope comprises an interface module, an imaging module, an illumination module, a smartphone (Galaxy S8+, Samsung), and a custom-designed Android application. Figure 1(a) shows the schematics of the overall system components. The rear imaging module of the smartphone is an RGB camera with a resolution of 12 MP, a sensor size of 1/2.55 inches, and a pixel size of 1.4 μm. The aperture ratio and focal length of the lens are f/1.7 and 4 mm, respectively. The interface module includes an interface circuit and a 3.7 V Li-ion battery. The imaging module includes a set of optical lenses for collecting light from ear regions of interest, a high-pass filter for rejecting the excitation light from the collected light, and a smartphone for recording the light. The illumination module incorporates visible range light-emitting diodes (LEDs), high-power ultraviolet (UV) LEDs, band-pass filters to guarantee that excitation light at a selected wavelength from the UV LEDs are delivered onto a sample, and coupling lenses for delivering the light from the UV LEDs into optical fibers. The light from the optical fibers was directed to the sample regions of interest. A photographic image of the assembled apparatus is shown in Fig. 1(b).

The interface circuit mainly comprises of a microcontroller unit (MCU) (Atmega 128A, Atmel), a Bluetooth low energy (BLE) module (RN4871, Microchip), and two LED drivers, one of which is for driving the visible range LEDs (TLC5926, Texas Instruments), and the other is for driving the high-power LEDs (STP04CM05, ST microelectronics). The BLE module enables wireless connection with a smartphone and controls the system via our custom-designed Android application. Three voltage regulators are also included in the interface circuit to supply power to the components on the board. While a 3.7V-to-3.3V regulator (TPS73033, Texas Instruments) provides power to the BLE module, a 3.7V-to-5 V regulator (NCP1402, On semiconductors) supplies power to the MCU, the visible range LEDs, and LED driver, and an additional 3.7V-to-5 V regulator (MCP73831, Microchip) capable of high-current output (~400 mA) is used for high-power LEDs. The interface circuit also enables the recharging of a lithium-ion battery via a micro-USB female connector. Figure 1(c) shows a block diagram of the interface board.

Twelve LEDs were used for white light, spectral, and UV autofluorescence imaging. Among the ten LEDs within the visible range, eight LEDs (USHIO LEDs) with peaks at 429.8 nm (FWHM: 20.98 nm), 448.14 nm (FWHM: 26.07 nm), 478.55 nm (FWHM: 28.77 nm), 523.26 nm (FWHM: 38.16 nm), 596.21 nm (FWHM: 20.23 nm), 613.19 nm (FWHM: 19.4 nm), 643.69 nm (FWHM: 20.58 nm), and 664.18 nm (FWHM: 22.5 nm) were incorporated into the system for spectral illumination. One extra light source at 555.83 nm (FWHM: 16.55 nm) was achieved by using a white LED (SST-20-WCS-A120-L4600, Luminus Devices) and an optical filter (65-098, Edmund Optics). Another white LED was used for white light imaging. Figure 2 shows the emission spectra of the LEDs used in the system. The emission spectra of the eight narrow-band color sources are shown in Fig. 2(a). Figure 2(b) exhibits the spectra of the white LED and the filtered white LED with peak emission at 555.83 nm. All visible-range LEDs were placed inside an LED multiplexer, as described in our previous work [29]. The output of the LED multiplexer was attached to one end of a bundle of visible range optical fibers with a diameter of
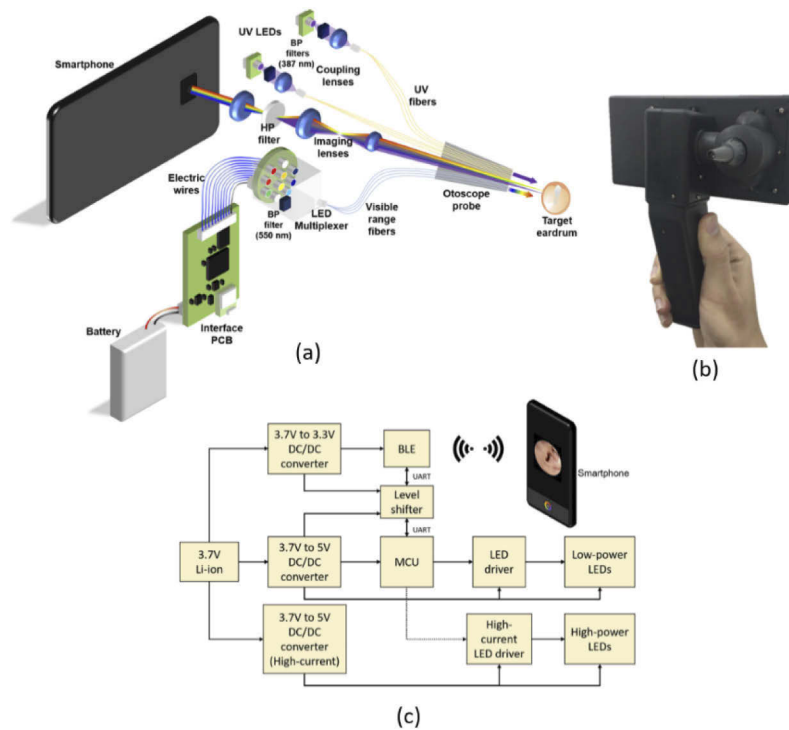
Fig. 1. Smartphone-based multimode imaging otoscope. (a) Schematics of the system's internal components. (b) Photographic image of a smartphone-based multimode imaging otoscope. (c) A block diagram of the components of the interface circuit board.

250 µm (02-531, Edmond Optics), while the other end of the bundle was attached to the tip of an otoscope speculum. The remaining two LEDs (SST-10-B130-G385-00, Luminus Devices) emit high-power UV light at 385 nm for excitation of ear regions of interest. To remove light at unwanted wavelengths from the excitation light, band-pass filters (84-078, Edmund Optics) were placed after the UV LEDs. Figure 2(b) also displays the spectra of the UV LED before and after passing through an optical band-pass filter. After the band-pass filter, a coupling lens (43-480, Edmund Optics) focused an excitation beam onto the bundle of fibers with a diameter of 200 µm (57-068, Edmund Optics), fit to the transmission of UV light. The measured irradiance of the UV light emitted by the optical fibers was 69.75µW/cm^2.

Light from the fibers interacts with target regions of the ear, and the light reflected or emitted from the target regions is then collected by the lens system, which includes a high-pass filter with a cut-off frequency of 400 nm (62-974, Edmund Optics) to ensure the removal of the excitation light for UV autofluorescence imaging. Images were then recorded using the smartphone camera. Finally, the acquired images were transferred to a server via either LTE or wi-fi where the images were analyzed, and the results were then returned to the smartphone.

2.2. Processing of a multimode image cube using machine learning algorithms

2.2.1. Preprocessing of multimode image data

For spectral image analysis, key preprocessing steps must be performed to ensure reliable spectral data classification. As the first preprocessing step, except for the white light and fluorescence images, all RGB images were converted to grayscale, followed by a flat-field correction. Image registration for all images was performed to compensate for the misalignments of images owing
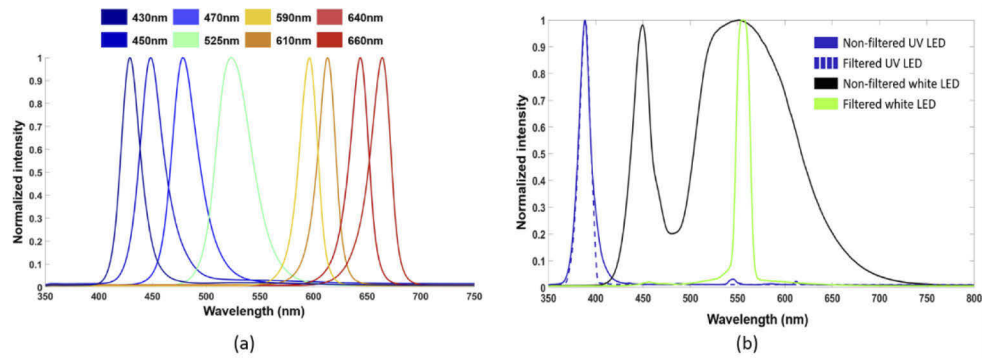
Fig. 2. Emission spectra of the LEDs included in the system; (a) Emission spectra of narrow-band LEDs in the visible range (430 nm to 660 nm). (b) Emission spectra of white, and UV LEDs before and after filtering. The blue lines show the spectra of the UV LED before (solid line) and after (dashed line) optical filtering. An extra narrow band source centered at 555.83 nm was achieved through filtering a white LED. The green line shows the emission spectrum of the white LED after filtering. The LED spectra of the narrow-band sources in the visible range are distributed in average steps of 29.3 nm, with an average bandwidth of 23.69 nm.

to hand movements during the intervals between the image acquisitions. In the preprocessing of an autofluorescence image, the CLAHE algorithm was applied to the image for contrast enhancement [33], and image registration of the contrast-enhanced image was performed. Finally, the preprocessed images were stacked in a multimode image cube that contained 12 images, which are the nine grayscale spectral images and the red (R), green (G), and blue (B) channels of the autofluorescence image. In addition, we compare the performance of the analysis of the multimode image cubes with various other data combinations to find the one that would provide the best data for the classification of OM. The data flow, from preprocessing to the generation of a segmented image, is shown in Fig. 3 for all the data combinations.
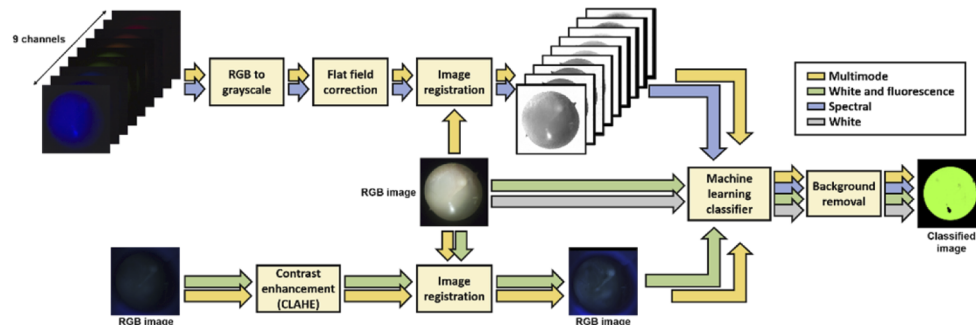

Fig. 3. Block diagram of the image processing and classification data path for all the tested data combinations, that is, multimode (yellow arrows), spectral (blue arrows), the combination of white and fluorescence (green arrows), and only white light images (gray arrows). The data follow different processing paths depending on the imaging modality through which it was acquired. Spectral images undergo flat field correction and image registration, whereas autofluorescence images are contrast-enhanced and then registered. The registration is held using the white-light image as a reference. Now the data is ready for classification using a machine learning algorithm. After the labeled image is generated, background removal is executed.

2.2.2. Spectral classification of the multimode ear images

We compared various machine learning and conventional spectral classification algorithms to find the algorithm that offers the highest classification accuracy using our developed system. The MP, RF, LR, DTs, NB, SAM, and ED algorithms were applied to our developed system for the classification of ear diseases of interest. These classification algorithms have been extensively used to analyze spectral imagery data [17,29,34–41].

For training and testing of machine learning algorithms for classifying normal (Class 1), OM (Class 2), and AdOM (Class 3) ears, ground truths were constructed with the assistance of medical doctors to select the specific regions for each class in all samples. Next, a binary mask was created for regions of interest and applied to the image cube [x, y, and λ (nine wavelengths + R, G, and B channels of an autofluorescence image)], followed by the removal of all null spectral signatures resulting from the masking process. After selecting only pixels from the regions of interest, spectral signatures from spectral images and intensity profiles of R, G, and B channels from the autofluorescence image were extracted at every selected pixel for each class. Additionally, before the classification, data standardization was applied to the extracted data because the maximum intensities in certain channels of autofluorescence images, especially the R channel, were significantly lower than those of other channels. The dataset was then divided in an 80/20 split for the training and test sets, respectively.

MP has been used for the classification of various types of images, including medical, satellite, and spectral images [17,22,35,42,43]. In this study, MP was applied to analyze multimode images. It consists of an input layer, a hidden layer, and an output layer. For the input layer, 12 nodes were used to accept pixels from the nine spectral images at different wavelengths, and the R, G, and B channels of one autofluorescence image. Here, the optimal number of nodes and a regularization parameter were experimentally determined by training and testing cycles while maintaining the other parameters. MP exhibited the best performance at ten nodes in the hidden layer. For the output layer, four nodes were used to output four classes: normal, OME, AdOM, and an extra specular reflection class. The NN model used to classify data composed entirely of the spectral image cubes had nine input nodes, 17 hidden nodes, and four output nodes, and a regularization parameter of 0.03562. We also tested the case when the input data were the R, G, and B channels of both autofluorescence and color eardrum images. Here, the optimal number of nodes was six in the hidden layer. Finally, when the input data were the R, G, and B channels of a white light image of the eardrums, the best number of nodes in the hidden layer was four. A regularization parameter of 0.00028 was determined for the analysis of multimode images and the combination of autofluorescence and white light images, whereas it was determined to be 0.00045 for the analysis of a white light image of eardrums.

We further tested other machine-learning classifiers and conventional spectral classification algorithms. For the analysis of multimode data, an RF classifier was designed with 58 estimators, a maximum tree depth of ten, a minimum-samples-split of 60, and a minimum-samples-leaf of 66, while a DTs classifier had a maximum depth of nine, a minimum-samples-split of 130, and a minimum-samples-leaf of 142. An RF classifier with a maximum depth of seven, 58 estimators, minimum-sample-leaf of 110, and a minimum-sample-split of 174 was found to provide the best classification results when analyzing spectral image cubes only, whereas a DT with a maximum depth of ten, minimum-sample-leaf equal to 146 and minimum-sample-split of 114 yielded the best results with the same data. To analyze the combination of autofluorescence and white light images, the RF classifier had 24 estimators, a maximum tree depth of 8, a minimum-samples-split of 198, a minimum-samples-leaf of 78, while a DT classifier was defined with a max depth of 8, a minimum-samples-split of 182, and a minimum-samples-leaf of 194. Finally, the classification of white light images using an RF classifier was performed with the parameters such as 29 estimators, trees with a maximum depth of 6, a minimum-samples-split of 58, and a minimum-samples-leaf of 78, while the designed DTs classifier had a maximum tree

depth of 9, a minimum-samples-split of 58 and a minimum-samples-leaf of 126. The selected criterion for the split quality measure was entropy. The LR classifier had a penalty of l2 and a C value of 1.8874×10^{-7}, 3.2903×10^{-8}, 1.4873×10^{-5}, and 7.8805×10^{-8} for the multimode, the autofluorescence and white light, the spectral image cube, and white light data types, respectively, while for the NB classifier, var-smoothing parameters defined as 0.2395, 0.0574, 1, and 1×10^{-9} were used to classify for the multimode, the autofluorescence and white light, the spectral image cube, and the white light data types respectively.

2.3. Clinical trials

A clinical trial was performed at Seoul National University, Seoul, South Korea, with approval from the Institutional Review Board of the hospital. One otologist collected data from male and female patients with ages ranging from 2 to 80 years. The medical diagnosis was made based on microscopic images and an audiology test. For patients who had abnormalities in both ears, each ear was counted as a separate sample. This study was conducted in accordance with the Declaration of Helsinki. A total of 69 patients participated in this study. From the 69 patients, 30 normal samples, 30 OME samples, and 29 AdOM samples were acquired

3. Results

3.1. Analysis of a multimode image cube for the classification of normal and OM ears

After multimode images of normal, AdOM, and OME ears were acquired using our developed otoscope, the multimode images were analyzed using MP for their classification. Average spectral signatures of each class are shown in Fig. 4. These signatures were obtained by averaging the pixel values within a 50×50 window in the dataset. The areas for extraction of the signatures were indicated by an otologist. The vertical lines show the standard deviations for each wavelength. The graph in Fig. 4 shows that the most important features to distinguish between different classes are observed at 525 nm and 550 nm. In other wavelengths, a strong overlap of the standard deviation is noticed.

Figure 5 shows white light (column a), autofluorescence (column b), and classified images (column c) of the tympanic membranes of normal, AdOM, and OME ears. Normal tympanic membranes exhibit low autofluorescence with a faint autofluorescence originating at the malleus and bony promontory, as reported in a previous study [28]. Additionally, vascular regions appear darker than normal tympanic membrane regions because of the strong absorption of UV light in the blood. In Fig. 5(b), the AdOM eardrum shows adhesion in the mesotympanum and attic areas along with effusion in the middle ear, resulting in a strong presence of autofluorescence. In the image, the regions of adhesion exhibit strong light-blue autofluorescence, whereas the effusion regions at the center of the eardrum exhibit greenish autofluorescence. However, in the OME autofluorescence image, no strong autofluorescence was observed. This could be because of the type of effusion, which is likely to be mucous in the case of AdOM and serous in the case of OME [44]. As previously suggested [28], the faint autofluorescence from the bony promontory cannot be distinguished in Fig. 5(b) (OME), probably because of the presence of effusion. Interestingly, strong autofluorescence from earwax was observed in the images. These results show that autofluorescence imaging could be used as a powerful additional aid for a specialist during OM diagnosis, allowing the visualization of features that cannot be observed using a regular otoscope.

Moreover, we performed spectral classification of the multimode images after pre-processing using MP. Figure 5(c) shows the classified images of the tympanic membranes of normal, AdOM, and OME ears. Green indicates pixels classified as normal, blue indicates pixels classified as OME, red indicates pixels labeled as adhesion, and black indicates pixels classified as specular reflection. A normal tympanic membrane was successfully classified (Fig. 5(c) (normal)). In the
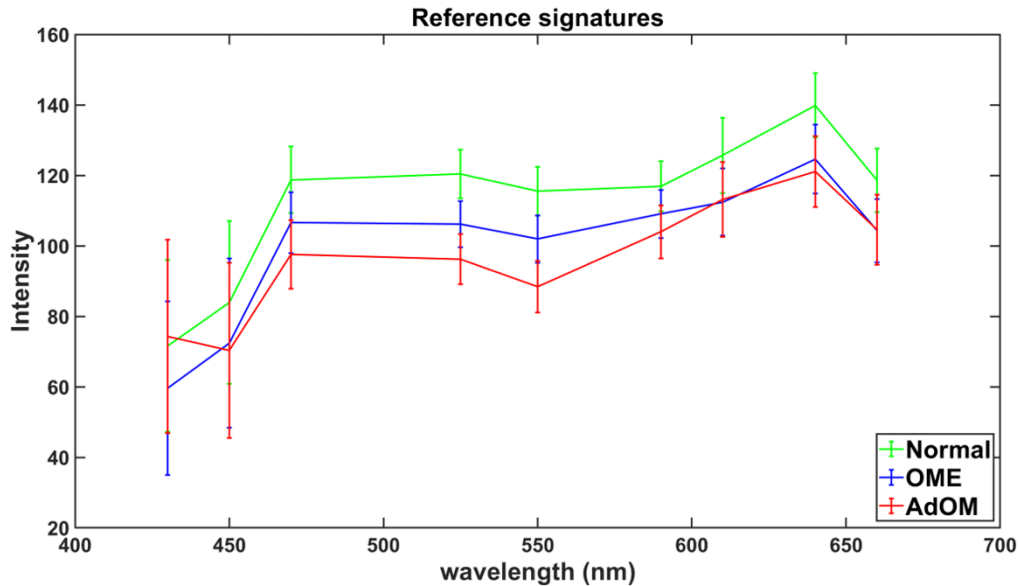
Fig. 4. Reference signatures for classification. The green signature represents the normal class, the blue OME, and the red AdOM. From the signatures we note that the most important spectral features are found at 525 nm and 550 nm.

classified image for OME (Fig. 5(c) (OME)), most pixels in the tympanic membrane region were classified as OME (blue). Here, the pixels in the areas corresponding to earwax were classified as normal regions, however, since we did not train the algorithms to classify earwax, this can be considered misclassification. Additionally, in the classified image for AdOM, the areas of adhesions with the attic were distinguished from the areas where effusion was observed behind the tympanic membrane, except for adhesion regions at the center of the image, where a strong greenish autofluorescence is noted.

3.2. Comparison of various machine learning and conventional techniques on the assessment of multimode otoscopic imagery

To evaluate the benefits of multimode image analysis in distinguishing between normal, AdOM, and OME ears using our developed otoscope, the result of multimode image analysis with a spectral image cube and an autofluorescence image was compared to the results obtained by the analysis of the combination of white light and autofluorescence images, spectral image cube only, and the analysis of white light images only as shown in Table 1. Table 1 is organized in order of accuracy, with the highest accuracy located on the top rows and the lowest on the bottom rows. The best metrics are highlighted in bold. The metrics shown in Table 1 were obtained through a 10-fold cross-validation process. MP provided the best outcome in the metrics considered. In particular, the multimode image analysis resulted in the highest mean F1-score of 0.7320, area under the curve (AUC) of 0.9186, and accuracy of 0.7963. The RF closely followed the performance of the MP, but with a much lower F1 score in the AdOM class. A similar trend was observed for the three top classifiers. As expected, the traditional classification algorithms, such as SAM and ED, exhibit poorer performance than the machine learning algorithms because they are the most vulnerable to inter-and intra-class spectral signature variations. All the algorithms had lower outcomes in the AdOM class than in the other classes, implying that spectral and fluorescence imagery data may not be ideal for the diagnosis of AdOM. Table 2 shows the
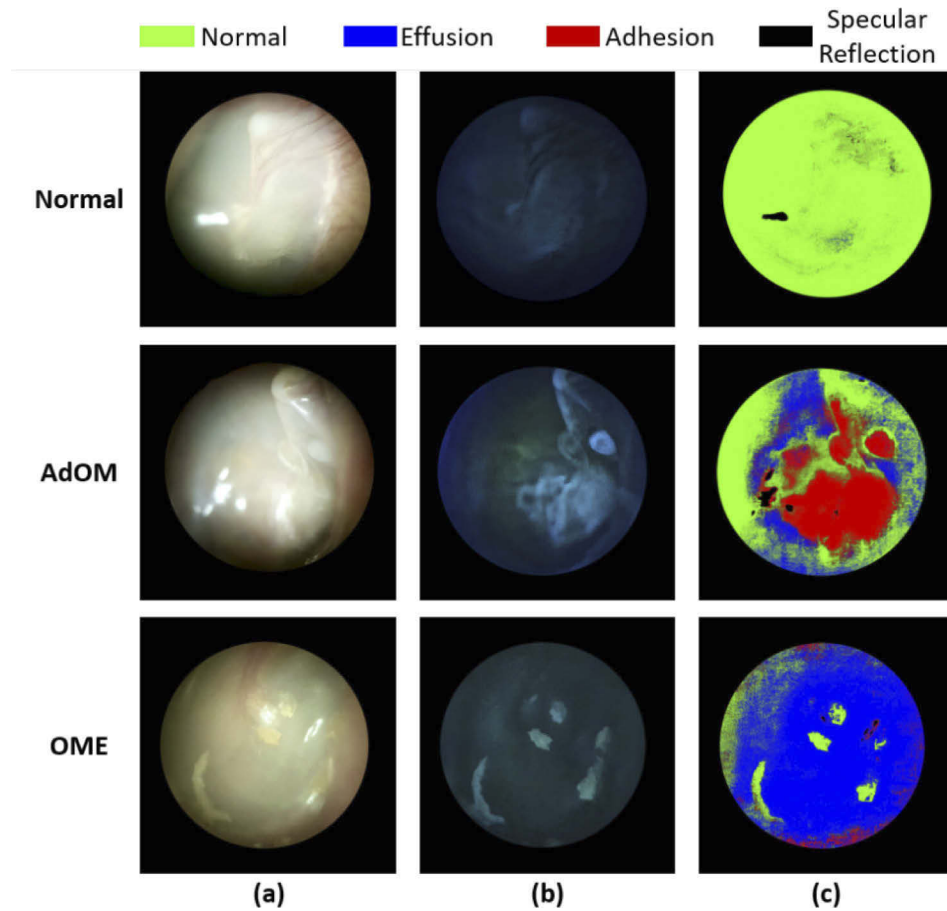
Fig. 5. (column a) white light, (column b) autofluorescence images of normal, and OM eardrums, and (column c) classification maps generated by the neural networks. (a, Normal): a white light image of a normal eardrum; (b, Normal): autofluorescence image of a normal eardrum. See the contrast between the blood vessels and other structures resulting from low autofluorescence in the blood vessel areas. See that the eardrum was correctly labeled in green representing a healthy eardrum (c, Normal); (a, AdOM): a white image of a case of OM; (b, AdOM): fluorescence image of a case of OM. Note the various sources of autofluorescence indicating abnormality in this ear, especially in the regions of adhesion with the attic and mesotimpanum. We can also see a greenish color that may be emitted from fluorophores in the middle ear effusion. (c, AdOM): see that pixels of regions of adhesion with the attic were labeled in red as expected and the pixels corresponding to regions containing effusion were labeled in blue. However, the areas at the center of the image, where strong greenish autofluorescence is seen, were misclassified as adhesion; (a, OME) The white light image of another case of OM; (b, OME): autofluorescence image. Note that here, in contrast to (b, AdOM), less autofluorescence is seen, nevertheless, compared with Fig. 3(b, Normal), autofluorescence coming from the bony promontory cannot be noticed. (c, OME) shows a classification map with most of the eardrum labeled in blue, agreeing with the OME diagnostic given to this ear.

confusion matrix of the MP classifier, which exhibits the best performance in the classification of Normal, OME, and AdOM classes with the multimode data. The confusion matrix demonstrates that MP correctly classified 87.6871% and 82.7153% the Normal and OM classes but it showed

the lowest performance in the classification of the AdOM class. In particular, the AdOM class was often misclassified as OME class. The confusion matrices for all the classifiers at different data combinations that provided the best results are found in the Supplement 1 (Table S1). These results show that the machine learning-based multimode image analysis with a spectral image cube and an autofluorescence image not only provides additional qualitative information that can be visually examined by a specialist, but also enables precise classification of ear diseases compared to image analysis with a single image and dual-mode images.

Table 1. Comparison of the various machine learning classification algorithms on the examination of multimode otoscope imagery. The highest value is highlighted in bold.

Algorithm	Data type[a]	Normal		OME		AdOM		Accuracy
		F1	AUC	F1	AUC	F1	AUC	
MP	MM	**0.8622**	**0.9444**	**0.7867**	**0.9166**	**0.5472**	**0.8948**	**0.7963**
	W/F	0.8345	0.9216	0.7655	0.9024	0.4208	0.8435	0.7668
	S	0.7501	0.8497	0.6525	0.8169	0.4264	0.8391	0.6801
	W	0.7128	0.7951	0.6238	0.8033	0.1567	0.7301	0.6378
RF	MM	**0.8352**	**0.9275**	**0.7684**	**0.9034**	**0.4451**	**0.8858**	**0.7739**
	W/F	0.8060	0.8987	0.7439	0.8835	0.2518	0.7997	0.7395
	S	0.7160	0.8095	0.6245	0.7784	0.2501	0.7897	0.6435
	W	0.6856	0.7731	0.6142	0.7764	0.0094	0.6913	0.6139
DTs	MM	**0.8100**	**0.9048**	**0.7372**	**0.8740**	**0.4617**	**0.8318**	**0.7453**
	W/F	0.7921	0.8856	0.7172	0.8607	0.2498	0.7718	0.7182
	S	0.7342	0.8261	0.6289	0.7894	0.3827	0.8045	0.6603
	W	0.7220	0.8039	0.6426	0.8093	0.2397	0.7490	0.6537
LR	MM	**0.7738**	**0.8370**	0.6853	0.8143	0.1390	0.7287	**0.6868**
	W/F	0.7502	0.8362	**0.6947**	**0.8497**	**0.1899**	**0.7522**	0.6787
	S	0.6552	0.7219	0.4813	0.6639	0.0312	0.6909	0.5474
	W	0.6993	0.7726	0.6150	0.7957	0.1069	0.7268	0.6223
NB	MM	**0.6973**	**0.7767**	0.4226	**0.7458**	**0.2502**	**0.6634**	**0.5709**
	W/F	0.6555	0.6983	**0.5227**	0.6970	0.1825	0.6204	0.5639
	S	0.6540	0.7087	0.1751	0.6764	0.1703	0.6195	0.5045
	W	0.6177	0.6338	0.4888	0.6438	0.0036	0.5045	0.5292
ED	MM	0.5366	0.6401	0.4989	0.6657	**0.2556**	**0.5611**	**0.4519**
	W/F	0.3702	0.4485	**0.5297**	**0.6744**	0.1673	0.4710	0.3756
	S	**0.5434**	**0.6421**	0.3868	0.5990	0.2205	0.5325	0.4118
	W	0.4389	0.5194	0.5165	0.6101	0.1003	0.4794	0.4051
SAM	MM	0.4685	0.6229	0.5357	0.6790	0.2481	0.5747	0.4140
	W/F	0.2591	0.4396	0.4666	**0.7691**	0.2220	0.5427	0.2748
	S	0.5221	0.6279	0.4803	0.6222	0.2260	0.6194	0.4232
	W	**0.5461**	**0.6789**	**0.6140**	0.7524	**0.2502**	**0.6302**	**0.5062**

[a]MM = Multimode; W/F = White light and autofluorescence; S = Spectral; W = White light

Table 2. Confusion matrix for the MP classifier analyzing multimode data

	Normal	OME	AdOM
Normal	**87.6871%**	10.3856%	1.9273%
OME	12.9704%	**82.7153%**	4.3143%
AdOM	19.2094%	35.5074%	**45.2832%**

4. Discussion

We built an intelligent smartphone-based multimode otoscope to collect multiple and complementary data that contain useful diagnostic information and then evaluated its performance for the diagnosis of ear diseases. We hypothesized that the addition of another imaging modality and intelligent multimode analysis capability into our smartphone-based spectral imaging otoscope would increase its potential for the diagnosis of ear diseases. Thus, we conducted clinical trials at a tertiary education hospital in Seoul, where one otology specialist collected data from normal, OME, and AdOM patients. The collected data were then analyzed using various machine learning and conventional spectral classification algorithms. In our previous work, we demonstrated the potential of a multispectral smartphone-based otoscope [29]. The capability of the system as a diagnostic tool was demonstrated using a small number of samples. However, in this work, we conducted extensive clinical trials to collect a large amount of data from patients using an advanced smartphone-based otoscope with additional imaging and analysis capabilities. Therefore, we could ensure that the intelligent multimode otoscope is more accurate than a conventional otoscope in diagnosing ear diseases.

Few studies employed spectroscopy to diagnose ear diseases [10,11]. However, to the best of our knowledge, there is no study on multimode imaging including spectral imaging in the diagnosis of OM. Therefore, we developed the intelligent smartphone-based multimode otoscope including the spectral imaging capability. For the spectral imaging, we tried to minimize the overlap of the spectra of adjacent LEDs selecting the ones with the least FWHMs under considerations of availability in the market, size, and emitted light intensity. For autofluorescence imaging, light from two high-power UV LEDs was filtered and transmitted through optical fibers to the tympanic membrane of a patient. The measured irradiance of UV light was $69.75 \mu W/cm^2$, 14 times lower than the limit value of $1 mW/cm^2$ guided by the American Conference of Governmental Industrial Hygienists [45]. The device was designed to be attached to a Samsung Galaxy S8 + . However, to serve as a telemedicine tool in remote areas with low resources, the cost may appear as an important factor. Thus, a spectral imaging module, which can be attached to more affordable smartphone models, needs to be further optimized. It can be simply realized by constructing a mechanical structure capable of adjusting the position of a key optical system module. In the system, the diameter of the otoscope probe tip, including a removable and sterilizable cap, is 5.8 mm. Therefore, when the system is applied to a patient with a narrow eardrum, the probe may not be placed at the working distance from the tympanic membrane, thus resulting in the acquisition of a defocused image. This may affect the performance of the system. However, to some extent, a defocused image was also shown to hold useful diagnosis information [46]. As mentioned, spectroscopy, which is based on point detection, has been used for the diagnosis of ear diseases [10,11]. Since the areas of interest in the eardrum correspond to multiple pixels in an image, spectral imaging and analysis can be used for the diagnosis of ear diseases, as spectroscopy, even though slightly defocused images are acquired. In those cases, the classified images tend to show errors near the class boundaries. To increase pixel accuracy in spectral classification, defocused image acquisition should be avoided. It can be realized by fast image acquisition with appropriate illumination and an optical lens system with a large depth of focus.

One of the key advancements of our proposed system over the system described in [29] was the ability to acquire fluorescent images of the tympanic membrane of a patient in addition to its spectral imaging capability. This was realized via the inclusion of high-current LED sink drivers, high-power UV LEDs, optical filters, and specialized optical fibers into the otoscope. To date, a fluorescence imaging otoscope has been developed [28]. In this study, autofluorescence images provided better contrast for identifying the margins of the affected regions. As shown in Fig. 5, autofluorescence images provided additional information for the visualization of ear conditions that cannot be seen solely via the examination of images obtained under white light illumination. However, due to the low intensity of the emitted fluorescence from the middle ear structures, contrast enhancement was needed for a better visual analysis of the images. Furthermore, we noticed that the machine learning classifiers performed better on the data that had been through the contrast enhancement process. In contrast to the device described in [28], our system allows us to obtain a multispectral image cube at nine wavelengths and an autofluorescence image in mobile environments. This permits our system to provide more relevant spectral and autofluorescence information on the ear regions of interest, thus allowing a more precise distinction between ear diseases.

The machine learning-based analysis of the various combinations of data emphasized the importance of a multimode imaging device for the diagnosis of ear diseases. As shown in Table 1, MP was superior to other machine learning algorithms when trained with multimode data. Among the designated classes, the lowest metric values were for the classification of AdOM. This might be because of the similar spectral characteristic features of the AdOM and OME ears. Even the analysis of the multimode data was not sufficient to achieve the classification of AdOM class with an accuracy as high as the Normal and OME cases. One of the distinct features of AdOM from other types of OM is the morphological difference caused by the adhesion of the tympanic membrane and the middle ear structure [7,8]. However, the multimode otoscope developed here is not suitable for acquiring 3D morphological information on ears in detail. Therefore, incorporating a 3D imaging modality that allows for quantitative assessment of eardrum morphology into the system may increase the accuracy of the system in classifying normal, OM, and AdOM ears.

In the clinical trials, patients with a broad age range were examined. Interestingly, age-related variations were found in the spectral signatures for each class. However, the significance of these findings needs to be further evaluated because of an unbalanced distribution of patients per age group in the dataset. Furthermore, the level of the severity of the diseases also affects the spectral signatures, but here the level of severity was not a parameter considered when building the datasets. The associated works remain a future study.

5. Conclusions

We developed an intelligent multimode otoscope capable of obtaining different but complementary information on the eardrum. Specifically, the developed otoscope allows obtaining a white light image, which is required for real-time visualization, a spectral image cube containing nine channels in the visible range, and an autofluorescence image of the eardrum. Therefore, the system described here enables the acquisition of more quantitative and qualitative data using a handheld, fully portable, and ubiquitously connectable device, suitable for primary care environments. Using data collected from clinical trials, we showed that an autofluorescence image of the eardrum allows better visualization of features that cannot be distinguished uniquely by the analysis of images obtained using white light illumination. Machine learning-based analysis with multimode images yields better performance than single-mode or dual-mode images. It was found that MP could distinguish between normal, OME, and AdOM ears with a mean F1-score of 0.7320, AUC of 0.9186, and accuracy of 0.7963. However, the F1 score for AdOM was lower than the F1 scores for the normal and OME ears. The addition of 3D morphological

imaging capabilities to the system could improve this. Currently, more advanced algorithms, such as deep learning networks, are available. The application of more advanced algorithms to the system could also improve the overall performance of the system. The associated study remains a future work. Overall, our findings suggest that the intelligent multimode otoscope can be a useful mobile diagnostic tool for the diagnosis of various ear diseases through the acquisition of additional qualitative and quantitative features in the middle ear.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. O.-P. Alho, H. Oja, M. Koivu, and M. Sorri, "Chronic otitis media with effusion in infancy: how frequent is it? How does it develop?" Arch. Otolaryngol., Head Neck Surg. **121**(4), 432–436 (1995).
2. R. M. Rosenfeld, J. J. Shin, S. R. Schwartz, R. Coggins, L. Gagnon, J. M. Hackell, D. Hoelting, L. L. Hunter, A. W. Kummer, and S. C. Payne, "Clinical practice guideline: otitis media with effusion (update)," Otolaryngol.–Head Neck Surg. **154**(1_suppl), S1–S41 (2016).
3. S. Mansour, J. Magnan, K. Nicolas, and H. Haidar, *Middle Ear Diseases: Advances in Diagnosis and Management* (Springer, 2018).
4. A. Qureishi, Y. Lee, K. Belfield, J. P. Birchall, and M. Daniel, "Update on otitis media–prevention and treatment," Infect. Drug Resist. **7**, 15–24 (2014).
5. J. E. Roberts, M. R. Burchinal, and S. A. Zeisel, "Otitis media in early childhood in relation to children's school-age language and academic skills," Pediatrics **110**(4), 696–706 (2002).
6. J. E. Roberts, M. R. Burchinal, A. M. Collier, C. T. Ramey, M. A. Koch, and F. W. Henderson, "Otitis media in early childhood and cognitive, academic, and classroom performance of the school-aged child," Pediatrics **83**(4), 477–485 (1989).
7. S. Hashimoto, "A guinea pig model of adhesive otitis media and the effect of tympanostomy," Auris, Nasus, Larynx **27**(1), 39–43 (2000).
8. A. Larem, H. Haidar, A. Alsaadi, H. Abdulkarim, M. Abdulraheem, S. Sheta, S. Ganesan, A. Elhakeem, and A. Alqahtani, "Tympanoplasty in adhesive otitis media: a descriptive study," Laryngoscope **126**(12), 2804–2810 (2016).
9. J. L. Paradise, "On classifying otitis media as suppurative or nonsuppurative, with a suggested clinical schema," J. Pediatr. **111**(6), 948–951 (1987).
10. L. Hu, W. Li, H. Lin, Y. Li, H. Zhang, K. Svanberg, and S. Svanberg, "Towards an optical diagnostic system for otitis media using a combination of otoscopy and spectroscopy," J. Biophotonics **12**(6), e201800305 (2019).
11. Z. Schmilovitch, V. Alchanatis, M. Shachar, and Y. Holdstein, "Spectrophotometric otoscope: A new tool in the diagnosis of otitis media," J. Near Infrared Spectrosc. **15**(4), 209–215 (2007).
12. T. A. Valdez, J. A. Carr, K. R. Kavanagh, M. Schwartz, D. Blake, O. Bruns, and M. Bawendi, "Initial findings of shortwave infrared otoscopy in a pediatric population," Int. J. Pediatr. Otorhinolaryngol. **114**, 15–19 (2018).
13. D. Preciado, R. M. Nolan, R. Joshi, G. M. Krakovsky, A. Zhang, N. A. Pudik, N. K. Kumar, R. L. Shelton, S. A. Boppart, and N. M. Bauman, "Otitis media middle ear effusion identification and characterization using an optical coherence tomography otoscope," Otolaryngol.–Head Neck Surg. **162**(3), 367–374 (2020).
14. J. E. Bardram and C. Bossen, "Mobility work: the spatial dimension of collaboration at a hospital," Comput. Support. Coop. Work **14**(2), 131–160 (2005).
15. C. L. Ventola, "Mobile devices and apps for health care professionals: uses and benefits," Pharm. Ther **39**(5), 356 (2014).
16. S. Kim, D. Cho, J. Kim, M. Kim, S. Youn, J. E. Jang, M. Je, D. H. Lee, B. Lee, and D. L. Farkas, "Smartphone-based multispectral imaging: system development and potential for mobile skin diagnosis," Biomed. Opt. Express **7**(12), 5294–5307 (2016).
17. S. Kim, J. Kim, M. Hwang, M. Kim, S. J. Jo, M. Je, J. E. Jang, D. H. Lee, and J. Y. Hwang, "Smartphone-based multispectral imaging and machine-learning based analysis for discrimination between seborrheic dermatitis and psoriasis on the scalp," Biomed. Opt. Express **10**(2), 879–891 (2019).
18. T. N. Kim, F. Myers, C. Reber, P. Loury, P. Loumou, D. Webster, C. Echanique, P. Li, J. R. Davila, and R. N. Maamari, "A smartphone-based tool for rapid, portable, and automated wide-field retinal imaging," Transl. Vis. Sci. Technol. **7**(5), 21 (2018).

19. J. K. Bae, H.-J. Roh, J. S. You, K. Kim, Y. Ahn, S. Askaruly, K. Park, H. Yang, G.-J. Jang, and K. H. Moon, "Quantitative screening of cervical cancers for low-resource settings: pilot study of smartphone-based endoscopic visual inspection after acetic acid using machine learning techniques," JMIR mHealth and uHealth **8**(3), e16467 (2020).

20. R. D. Uthoff, B. Song, S. Sunny, S. Patrick, A. Suresh, T. Kolur, K. Gurushanth, K. Wooten, V. Gupta, and M. E. Platek, "Small form factor, flexible, dual-modality handheld probe for smartphone-based, point-of-care oral and oropharyngeal cancer screening," J. Biomed. Opt. **24**(10), 1 (2019).

21. S. Mousseau, A. Lapointe, and J. Gravel, "Diagnosing acute otitis media using a smartphone otoscope; a randomized controlled trial," Am. J. Emerg. Med. **36**(10), 1796–1801 (2018).

22. H. C. Myburgh, S. Jose, D. W. Swanepoel, and C. Laurent, "Towards low cost automated smartphone-and cloud-based otitis media diagnosis," Biomed. Signal Process. Control **39**, 34–52 (2018).

23. M. N. Demant, R. G. Jensen, M. F. Bhutta, G. H. Laier, J. Lous, and P. Homøe, "Smartphone otoscopy by non-specialist health workers in rural Greenland: a cross-sectional study," Int. J. Pediatr. Otorhinolaryngol. **126**, 109628 (2019).

24. B. C. Spector, L. Reinisch, D. Smith, and J. A. Werkhaven, "Noninvasive fluorescent identification of bacteria causing acute otitis media in a chinchilla model," Laryngoscope **110**(7), 1119–1123 (2000).

25. L. L. Levy, N. Jiang, E. Smouha, R. Richards-Kortum, and A. G. Sikora, "Optical imaging with a high-resolution microendoscope to identify cholesteatoma of the middle ear," Laryngoscope **123**(4), 1016–1020 (2013).

26. J. J. Yim, S. P. Singh, A. Xia, R. Kashfi-Sadabad, M. Tholen, D. M. Huland, D. Zarabanda, Z. Cao, P. Solis-Pazmino, and M. Bogyo, "Short-Wave Infrared Fluorescence Chemical Sensor for Detection of Otitis Media," ACS Sens. **5**(11), 3411–3419 (2020).

27. A. C. Croce and G. Bottiroli, "Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis," Eur. J. Histochem. **58**(4), 2461 (2014).

28. T. A. Valdez, R. Pandey, N. Spegazzini, K. Longo, C. Roehm, R. R. Dasari, and I. Barman, "Multiwavelength fluorescence otoscope for video-rate chemical imaging of middle ear pathology," Anal. Chem. **86**(20), 10454–10460 (2014).

29. T. C. Cavalcanti, S. Kim, K. Lee, S. Y. Lee, M. K. Park, and J. Y. Hwang, "Smartphone-based spectral imaging otoscope: System development and preliminary study for evaluation of its potential as a mobile diagnostic tool," J. Biophotonics **13**(6), e2452 (2020).

30. J. Y. Hwang, S. Wachsmann-Hogiu, V. K. Ramanujan, J. Ljubimova, Z. Gross, H. B. Gray, L. K. Medina-Kauwe, and D. L. Farkas, "A multimode optical imaging system for preclinical applications in vivo: technology development, multiscale imaging, and chemotherapy assessment," Mol. Imaging. Biol. **14**(4), 431–442 (2012).

31. J. Kim, A. Seo, J.-Y. Kim, S. H. Choi, H.-J. Yoon, E. Kim, and J. Y. Hwang, "A multimodal biomicroscopic system based on high-frequency acoustic radiation force impulse and multispectral imaging techniques for tumor characterization ex vivo," Sci. Rep. **7**(1), 1–12 (2017).

32. J. Kim, H. Al Faruque, S. Kim, E. Kim, and J. Y. Hwang, "Multimodal endoscopic system based on multispectral and photometric stereo imaging and analysis," Biomed. Opt. Express **10**(5), 2289–2302 (2019).

33. S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," Comput. Gr. Image Process. **39**(3), 355–368 (1987).

34. S. E. Sesnie, P. E. Gessler, B. Finegan, and S. Thessler, "Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments," Remote Sens. Environ. **112**(5), 2145–2159 (2008).

35. P. K. Goel, S. O. Prasher, R. M. Patel, J.-A. Landry, R. Bonnell, and A. A. Viau, "Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn," Comput. Electron. Agric. **39**(2), 67–93 (2003).

36. J. da Rocha Miranda, M. de Carvalho Alves, E. A. Pozza, and H. S. Neto, "Detection of coffee berry necrosis by digital image processing of landsat 8 oli satellite imagery," ITC J. **85**, 101983 (2020).

37. A. Sitthi, M. Nagai, M. Dailey, and S. Ninsawat, "Exploring land use and land cover of geotagged social-sensing images using naive bayes classifier," Sustainability **8**(9), 921 (2016).

38. J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," IEEE Trans. Geosci. Remote Sensing **48**(11), 4085–4098 (2010).

39. M. Khodadadzadeh, J. Li, A. Plaza, and J. M. Bioucas-Dias, "A subspace-based multinomial logistic regression for hyperspectral image classification," IEEE Geosci. Remote Sens. Lett. **11**(12), 2105–2109 (2014).

40. K. Tan, H. Wang, L. Chen, Q. Du, P. Du, and C. Pan, "Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest," J. Hazard. Mater. **382**, 120987 (2020).

41. J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," IEEE Trans. Geosci. Remote Sensing **43**(3), 492–501 (2005).

42. J. D. Paola and R. A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," IEEE Trans. Geosci. Remote Sensing **33**(4), 981–996 (1995).

43. L. Hassan-Esfahani, A. Torres-Rua, A. Jensen, and M. McKee, "Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks," Remote Sens. **7**(3), 2627–2646 (2015).

44. M. H. Chung, J. Y. Choi, W. S. Lee, H. N. Kim, and J. H. Yoon, "Compositional difference in middle ear effusion: mucous versus serous," Laryngoscope **112**(1), 152–155 (2002).

45. A. Richards, "Threshold limit values for ultraviolet radiation measured for sources used in research equipment and some cases of overexposure to UV radiation," in *8. International Congress of the International Radiation Protection Association (IRPA8)* (1992).

46. N. Erkkola-Anttinen, H. Irjala, M. K. Laine, P. A. Tähtinen, E. Löyttyniemi, and A. Ruohola, "Smartphone otoscopy performed by parents," Telemed. e-Health **25**(6), 477–484 (2019).