IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 December 2023, accepted 5 January 2024, date of publication 9 January 2024,
date of current version 18 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3351668

RESEARCH ARTICLE

OTONet: Deep Neural Network for Precise Otoscopy Image Classification

DIVYA RAO[1], ROHIT SINGH[2], SUDIKSHA KOTTACHERY KAMATH[1], SANJEEV KUSHAL PENDEKANTI[1], DIVYA PAI[3], SUCHETA V. KOLEKAR[1], M. RAVIRAJA HOLLA[1], AND SAMEENA PATHAN[1]

[1]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
[2]Department of Otorhinolaryngology, Kasturba Medical College, Manipal Academy of Higher Education, Manipal 576104, India
[3]Department of Orthodontics and Dentofacial Orthopedics, Manipal College of Dental Sciences, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding author: Sameena Pathan (sameena.bp@manipal.edu)

ABSTRACT Otoscopy is a diagnostic procedure to visualize the external ear canal and eardrum, facilitating the detection of various ear pathologies and conditions. Timely otoscopy image classification offers significant advantages, including early detection, reduced patient anxiety, and personalized treatment plans. This paper introduces a novel OTONet framework specifically tailored for otoscopy image classification. It leverages octave 3D convolution and a combination of feature and region-focus modules to create an accurate and robust classification system capable of distinguishing between various otoscopic conditions. This architecture is designed to efficiently capture and process the spatial and feature information present in otoscopy images. Using a public otoscopy dataset, OTONet has reached a classification accuracy of 99.3% and an F1 score of 99.4% across 11 classes of ear conditions. A comparative analysis demonstrates that OTONet surpasses other established machine learning models, including ResNet50, ResNet50v2, VGG16, Dense-Net169, and ConvNeXtTiny, across various evaluation metrics. The research's contribution to improved diagnostic accuracy reduced human error, expedited diagnostics, and its potential for telemedicine applications.

INDEX TERMS Artificial intelligence, medical image analysis, diagnosis, convolutional neural networks, otology, applied engineering, healthcare.

I. INTRODUCTION

Otoscopy is a vital diagnostic procedure used by medical professionals to visualize the external ear canal and eardrum [1], aiding in the detection of ear pathologies and conditions [2]. This non-invasive examination plays a crucial role in diagnosing various ear conditions and abnormalities. It can identify issues like otitis media (middle ear infections), tympanic membrane perforations [3], excessive earwax accumulation, foreign bodies lodged in the ear canal, ear infections (including otitis externa or swimmer's ear), ear-drum abnormalities, Eustachian tube dysfunction [4], and, in rare cases, tumors or growths within the ear. The otoscope enables medical professionals to assess inflammation, fluid buildup, structural abnormalities, and other indicators that guide accurate diagnosis and appropriate treatment.

Images are captured in an otoscope by integrating specialized cameras or imaging sensors within the device [5]. The otoscope emits light onto the area under examination using LEDs, which are then directed and focused onto the ear canal and eardrum through lenses and mirrors. The integrated camera captures the illuminated image, which is instantly displayed on a screen for real-time visualization by healthcare professionals. This technology, illustrated in Figure 1 enables clear and magnified views of the internal ear structures, aiding in accurate diagnoses and medical evaluations. Timely otoscope image classification brings numerous advantages to medical diagnostics and patient care. It enables early detection and intervention, ensuring swift identification of ear conditions and tailored treatments. This leads to reduced patient anxiety, prevention of complications, and optimized

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

FIGURE 1. Ear exam using otoscope [7].

resource allocation. With precise diagnoses, personalized treatment plans can be promptly implemented, contributing to improved patient outcomes and cost savings.

Applying Convolutional Neural Networks (CNNs), directly to otoscopy image classification [6] faces unique challenges due to the distinctive characteristics of medical imaging data. Otoscopy images often exhibit variations in lighting, orientation, and the presence of artefacts, making it essential to design robust and optimized deep-learning models for accurate classification. AI holds promise in reshaping otoscope image analysis. Its automated image classification capability offers efficient diagnostic avenues, ensuring the timely detection of subtle ear conditions. By extracting features and recognizing patterns, AI enhances diagnostic accuracy, ultimately assisting medical professionals in delivering precise and timely diagnoses. Classifying otoscope images presents challenges due to varying image quality, subtle visual differences in ear conditions, limited training data, class imbalance, and the need for domain adaptation. Anatomical variations, interclass variability, and the lack of annotations further complicate accurate classification. Factors like ethnicity, age, and evolving medical knowledge require adaptable models.

The motivation behind this research lies in the need to develop an efficient and automated approach to optimize deep neural networks for otoscopy image classification [6]. The primary aim of this research is to design an accurate and robust image classification system that can effectively differentiate between these various otoscopic conditions. The proposed model leverages the innovative OTONet framework, which incorporates octave 3D convolution technique with submodules that focus on specific features and regions in the images. The study seeks to enhance classification performance, reduce region redundancy, and increase the receptive field, ultimately providing a valuable tool for clinical diagnosis and decision-making in the field of otoscopy. The research will evaluate the model's performance against these multiple otoscopy categories, contributing to improved healthcare practices and patient outcomes.

The problem at hand is to develop an accurate and efficient deep learning model for the classification of otoscopy images into distinct categories, representing various ear conditions and abnormalities. Otoscopy image classification plays a crucial role in assisting medical professionals in diagnosing ear pathologies, such as otitis media [8], tympanic membrane perforations, and ear infections.

The challenges in otoscopy image classification arise from the variability in otoscopy images, including differences in lighting conditions, image orientations, and the presence of artefacts. The intricate structures of the ear [9] make distinguishing subtle variations challenging, necessitating the need for a robust and optimized classification approach.

The objective of our work is to design a custom architecture capable of accurately and efficiently classifying otoscopy images, allowing for timely and accurate diagnosis. By addressing the complexities of otoscopy image classification through an optimized model, this research aims to improve patient outcomes, facilitate early detection of ear conditions, and enhance the overall efficiency of medical diagnostics [10], [11].

The proposed solution involves leveraging CNNs and exploring optimization strategies, including hyperparameter tuning and transfer learning. Through rigorous experimentation and validation, we aim to identify the most effective model that achieves high classification accuracy while ensuring generalization and robustness across diverse otoscopy image datasets. By effectively solving the problem of otoscopy image classification, this research contributes to advancing computer-aided medical diagnostics, supporting healthcare professionals in making informed decisions. The optimized model can serve as a valuable tool for early diagnosis, personalized treatment plans, and improved patient care in the field of otolaryngology [12].

The contributions of this paper are as follows:

1. This paper presents a novel OTONet framework tailored specifically for otoscopy image classification, featuring octave 3D convolution and feature and region focus modules created from the ground up.

2. The model's ability to accurately classify a diverse range of otoscopic conditions, such as Acute Otitis Media, Otitis Externa, Ventilation Tubes, and more, showcases its practical utility for classification tasks..

3. This paper provides a benchmark for otoscopy image classification by comparing the custom OTONet with well-established models.

II. RELATED WORK

In the field of otoscopy image analysis, a significant body of research has been dedicated to the development of advanced techniques for the detection and classification of ear conditions.

Table 1 provides an overview of notable research studies focused on otoscopy image analysis. These studies aim to enhance the diagnosis and classification of various ear conditions using advanced computational methods, primarily deep

TABLE 1. Review of recent advances in automated otoscopy image analysis.

S. No	**Author, Year**	**Aim**	**Dataset**	**No. of Images/Videos used**	**Metrics**	**GPU**	Novel Model
1	Akriti Singh et al., 2021 [13]	To suggest a CNN model that uses a collection of ear imagery to detect ear infections.	Ear Imagery Dataset, Universidad de Chile	880	Accuracy = 96%	_	CNN with 6 blocks and 2 dense layers
2	Michelle Viscaino et al., 2021[14]	To supply a CNN-LSTM hybrid learning framework for video otoscopy analysis as a component of an ear disease computer-aided diagnosis system.	The study team's medical collaborators gathered the video otoscopy dataset.	365	Accuracy=98.15% Precision= 91.94% Recall= 91.67% Specificity= 98.96% F1-score= 91.51%	_	No
3	L. Hu et al., 2018[15]	To formulate and evaluate an altered spectroscopic otoscope for otitis media optical detection.	Not Mentioned	NA	Accuracy=89.7%, Sensitivity=87.5%, Specificity=91.9%, AUC-ROC=0.94	_	No
4	Yiqing Zheng et al., 2021 [16]	To construct an automated system that uses a two-stage attention-aware convolutional neural network to diagnose otitis media from photographs of the tympanic membrane.	Tympanic membrane images, collected from a hospital in China.	4,000	Accuracy=89.5%, Sensitivity=89.3%, Specificity=89.7%, AUC-ROC=0.95	NVIDIA TITAN Xp GPU	No
5	Michelle Viscaino et al., 2022 [17]	To use color dependence analysis in a CNN-based system to increase the accuracy of computer-aided diagnosis systems for middle and external ear ailments.	Ear imagery database	22,000	Accuracy=98.5% AUC-ROC=0.999	NVIDIA GeForce GTX1080 GPU	Yes

TABLE 1. *(Continued.)* **Review of recent advances in automated otoscopy image analysis.**

6	Yiqing Zheng, 2021 [18]	To create a deep learning model employing otoscope pictures for the automated classification of pediatric otitis media.	Otoscope images of pediatric patients with otitis media, collected from a hospital in China.	10,703	Accuracy =0.936 (Xception model) Accuracy=0.921 (MobileNet-V2)	NA	No
7	Seda Camalan et al., 2021 [19]	Involves merging left and right eardrum otoscopy pictures using the OtoPair technique, to increase the precision of automated image analysis in ear exams.	Eardrum otoscopy images, collected from a hospital in Turkey.	300	NA	NA	No
8	Hamidullah Bino et al., 2022 [20]	To create OtoXNet, a deep learning-based technique for automatically identifying eardrum disorders from otoscope footage.	Dataset used was collected from a tertiary hospital in Taiwan.	394	Accuracy=91.5%, Sensitivity=91.5%, Specificity=91.5%, Precision=91.5%, F1-Score=91.5%.	NA	OtoXNet
9	Dongchul Cha et al., 2019 [21]	To create a machine learning model that uses a sizable library of otoendoscopic pictures to diagnose ear conditions.	Publicly available dataset.	10,544	Accuracy=94.5%	Deep Learning Toolbox in MATLAB	No
10	Kamel K. Mohammed et al. ,2022 [22]	Classifying ear imaging databases utilizing CNN-LSTM architecture and Bayesian optimization in order to distinguish between four conditions affecting the middle and outer ears.	Publicly available dataset.	880	Accuracy=100% (CNN-LSTM) Accuracy=86.3% (CNN classifier)	NA	Yes, novel CNN arch. using BiLTSM

learning and computer vision techniques. The table highlights key aspects of each study, including the primary objective, the dataset used, the number of images or videos analyzed, evaluation metrics, GPU utilization, the introduction of novel models, and the overall conclusions.

III. METHODOLOGY

The methodology section of this research presents a comprehensive process to address the classification of otoscopy images, as illustrated in Figure 2. This section delineates the critical stages that include data acquisition, preprocessing methods, the architectural structure of OTONet, consisting of Octave Convolution and Feature and Region Focus submodules, and the rigorous evaluation using various performance metrics. The methodology is segmented into distinct subsections to provide an in-depth understanding of each phase of the image classification process. These components include dataset acquisition, preprocessing steps, the architecture of the innovative OTONet model, focusing on its crucial components like Octave Convolution and Feature and Region Focus submodules, and the comprehensive evaluation strategy employing diverse performance metrics to assess the model's effectiveness and robustness. Each phase is described in detail in the subsections that follow.

A. DATASET

The otoscopy image dataset used for this study presented herein has been meticulously curated and evaluated to facilitate accurate and reliable classification of various ear conditions. In total, the dataset comprises 956 otoscope samples, categorized into distinct classes that reflect a spectrum of ear conditions: Normal Tympanic Membrane, Acute Otitis Media (AOM), Chronic Suppurative Otitis Media(CSOM), Excessive amount of Earwax, Otitis Externa, Ear Ventilation Tube, Foreign Bodies in the Ear, Pseudo Membranes, Tympanosclerosis. The classes and the number of samples in each class are represented in Figure 3.

The dataset undergoes a series of preprocessing steps. First, the original frame, representing the unprocessed, raw image frame captured during otoscopy, serves as the foundational starting point. In this process, specific attention is given to the isolation of low-quality images resulting from factors like insufficient lighting or inadvertent camera movement. This meticulous organization ensures that the model is exposed to high-quality, pertinent data, effectively preventing noise or irrelevant details from interfering with the learning process. Subsequently, the original color images are converted into grayscale, simplifying each pixel's intensity representation to a single value. This grayscale conversion reduces the data's complexity while preserving essential information, making it more amenable to further processing. The application of a Gaussian filter through convolution follows, serving the purpose of smoothing out the original images and mitigating noise, especially critical in medical imaging where high-frequency noise and fine details can be present. The outcome is a cleaner image representation that emphasizes relevant features. To enhance image contrast and highlight specific details, thresholding is applied by subtracting the smoothed image from the original. Employing a threshold value of 4, pixels with intensity values greater than or equal to 4 are designated as high, while those below four are classified as low. Finally, a crucial step in the segmentation process involves computing a circular Region of Interest (ROI) for each image. This segmentation isolates the area of interest within the image, specifically focusing on the circular frame encapsulating pertinent information. This meticulous preprocessing sequence collectively contributes to refining the dataset and facilitating improved feature extraction during the subsequent model training process.

A notable challenge in otoscopy image classification with this dataset is the inherent class imbalance, where certain ear conditions may be underrepresented in the dataset. To address this issue, oversampling is performed on minority classes by replicating images while undersampling is applied to majority classes by reducing the number of instances. This balances the class distribution, for underrepresented classes such as foreign objects and pseudo membranes, synthetic data points are generated using the Synthetic Minority Over-sampling Technique [23].

To enhance the diversity of the dataset, data augmentation is applied to the images. Data augmentation plays a crucial role in enhancing the robustness of the otoscopy image classification model presented in the paper. By introducing simulated variations in viewpoints through random rotations and employing mirroring effects via horizontal and vertical flips, the model becomes adept at discerning ear conditions from different perspectives encountered during otoscopy procedures. The inclusion of minor changes in zooming values, alterations in brightness and saturation, and the addition of Gaussian noise during augmentation addresses real-world challenges, such as varying lighting conditions and image quality. Collectively, these techniques contribute to the depth of the training data, exposing the model to a diverse set of scenarios and improving its adaptability to unpredictable conditions. By training on augmented data, the model learns to classify ear conditions more effectively under a broader range of circumstances, ultimately resulting in improved classification performance, especially in scenarios not adequately represented in the original dataset.

B. OTONET: OCTAVE CONVOLUTION

In this section, we outline the OTONet framework that constitutes a multiscale octave 3-Dimensional CNN [24] incorporated with both feature and region-focus submodules designed for the classification of otoscope images.

The primary characteristic of the OTONet revolves around the notable observation that both the network's parameters and the memory requisites experience a pronounced escalation on par with the augmentation of convolutional layers. This phenomenon is particularly prominent in the context of 3D CNNs. An approach termed octave convolution [25]

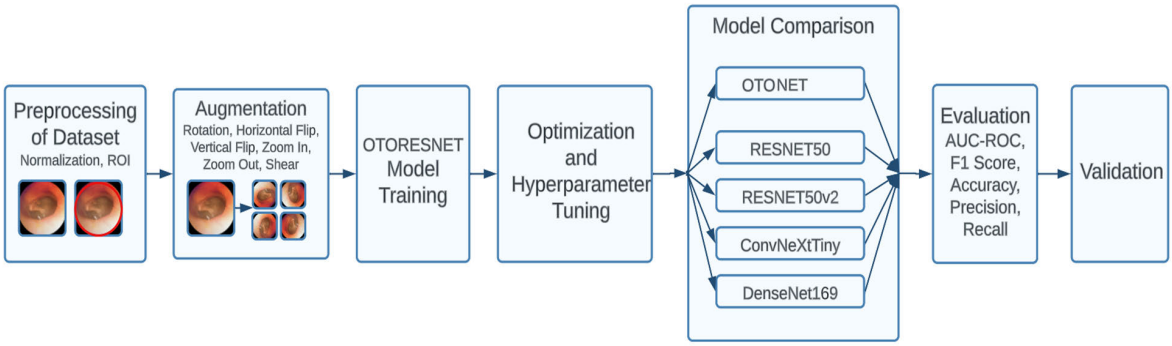

FIGURE 2. Methodology process flowchart followed in this study.

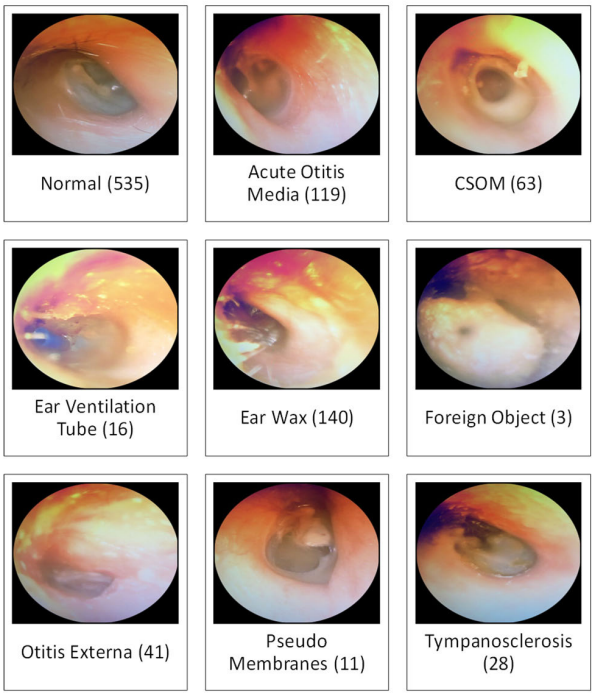

FIGURE 3. Representative otoscope images from 11 classification classes.

employs a decomposition strategy [26] to segregate the map of features generated by the model into discrete regions and space features. These segregated components are subsequently updated individually before being exchanged and eventually fused, leading to a reduction in region redundancy and an efficient expansion of the output space.

Our current study advocates the incorporation of three tiers of octave 3D convolution. The feature maps in the *j*th layer are denoted by

$$\mathrm{A}_j = \left\{ \mathrm{a}_1, \mathrm{a}_2, \cdots, \mathrm{a}_i, \cdots, \mathrm{a}_{\mathrm{n}channels} \mid \mathrm{a}_i \in \mathbb{R}_1^{nbands^{\times} l_1 \times l_2} \right\} \tag{1}$$

where each *ai* is an element of $\mathbb{R}_1^{nbands \times l_1 \times l_2}$. Here, $l_1 \times l_2$ signifies the region dimensions, and *nbands* indicates the number of color channels (R, G,B).

The input feature maps are divided into two groups, the highfrequency group A_1^{High} and the low-frequency group A_1^{Low} long the feature dimension, as the first stage in the octave three-dimensional convolution process. A hyperparameter α, which denotes the fraction of the low-frequency category in relation to the total, controls this division. The feature counts of A_1^{High} and A_1^{Low} are calculated as $(1-\alpha) \times C$ and $\alpha \times c$ respectively, where c signifies the total number of features. The result of the first layer of the three-dimensional octave convolution is expressed as follows, given the input map of attributes A_1:

$$A_1^{High} = 3\,\mathrm{DConv}\,(A_1) \tag{2}$$

$$A_1^{Low} = 3\,\mathrm{DConv}\,(Avg\_pool\,(A_1)) \tag{3}$$

In this context, the operation denoted as 3DConv(.) represents the standard 3D convolution process, while Avg_pool corresponds to the mean of the pooling operation [27]. The high-frequency and low-frequency groups communicate between features and update across features during the intermediate layer. The information that follows is a breakdown of the calculation for the central layer's output:

$$A_2^{High} = 3\,\mathrm{DConv}\left(A_1^{High}\right) + UpSample\left(F\left(A_1^{Low}\right)\right) \tag{4}$$

$$A_2^{Low} = 3\,\mathrm{DConv}\left(A_1^{Low}\right) + F\left(Avg\_pool\left(A_1^{High}\right)\right) \tag{5}$$

Here, *Up*Sample(.) signifies the up-sampling operation. In the context of hyperspectral image classification, during the final layer, adjustments are made to the high-frequency and lowfrequency groups to ensure they share the same shape to mitigate feature redundancy [28]. The output of this concluding layer, denoted as B, is determined by:

$$\mathrm{B} = 3\mathrm{DConv}\left(Avg\_pool\left(A_2^{High}\right)\right) + F\left(A_2^{Low}\right) \tag{6}$$

As a result, within the framework of octave 3D convolution, the region resolution of the low-frequency group is effectively diminished through the exchange of information across adjacent regions. Two key benefits of the octave three-dimensional convolution method are the decrease in region redundancy and the increase in the receptive field. In order to improve the precision of otoscope image classification, we thus suggest incorporating octave three-dimensional convolution instead of standard three-dimensional convolution.

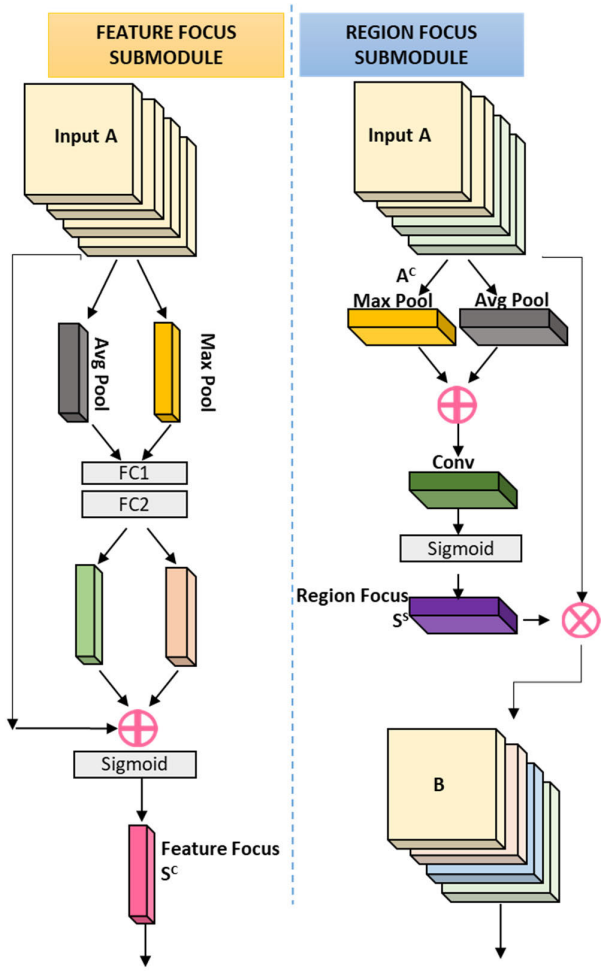

FIGURE 4. Feature and region focus submodules.

C. OTONET: FEATURE AND REGION FOCUS SUBMODULES

To enhance the representational capabilities of the model network, considering the wealth of colour features and region data intrinsic to HSIs, integration of both the feature focus sub-module and region focus submodule is proposed. These subsystems aim to bestow distinct significance upon feature map features and determine the salient portions within a feature map. We accomplish this by utilizing the convolutional block focus module, which was first presented by Woo et al. [29]. It is a flexible and fully trainable supplement to basic CNN architectures. The configurations of the submodules are visually depicted in Figure 4. Feature attention [30] operates by accentuating the reduction of feature redundancy and constructing a feature attention map that captures inter-feature relationships within features. As illustrated in the top region of Figure 4,

Let $A = \{a1, a2, \cdots, ai, \cdots, anchannels \mid ai \in R1nbands \times l\ 1 \times l2\}$ represent intermediate layer feature maps. Both average pooling and max pooling are simultaneously used to compress and combine the features, creating two distinct feature maps: max-pooled features Xmax and average-pooled features Aavg. To facilitate training, a shared network consisting of two dense layers processes these pooled features. The learned weights from this network are applied to both Xmax and Aavg. Consequently, the submodule feature focus submodule, denoted as SC∈Rlnfeatures $\times 1 \times 1 \times 1$, is derived as a tensor of dimensions *nchannels* $\times 1 \times 1 \times 1$. We also introduce a reduction ratio, r, to optimize parameters and set the concealed activation size to $nchannels/r \times 1 \times 1 \times 1$[32].

The calculation of focus in the feature-focus submodule is summarised as:

$$SC = \sigma\left(FC\left(Max\_pool(A)\right) + FC\left(Avg\_pool_{(}(A)\right)\right) \quad (7)$$

where the max pooling operation is denoted by *Max_pool* and σ stands for the sigmoid function.

We complement the focus submodule features with a region mechanism to further explore the focal points within a feature of a feature map. The region focus module comes after the feature-wise focus module, as seen in the bottom part of Figure 4. The region focus module,

The input for the region focus submodule, denoted as AC, consists of feature-refined feature maps, calculated as:

$$A^c = A \otimes S^c \quad (8)$$

where $\otimes$ stands for multiplication of individual elements. Global mean and max pooling operations are used to effectively utilize the feature information, producing 3D feature maps *ACmax* and *ACavg*. The 3D region focus map is then created by merging and convolving these maps through a conventional convolutional layer. The region focus is ascertained as follows:

$$S^S = \sigma(F^{3\times3\times3}([Max\_pool(AC; Avg\_pool()AC)])) \quad (9)$$

In this case, $F^{3\times3\times3}$ represents a standard 3D convolution with a $3 \times 3 \times 3$ kernel size. The output feature map B is obtained by SS to AC through element-wise multiplication:

$$B = A^c \otimes S^s \quad (10)$$

The final B represents the optimal feature map obtained from A by running both attention modules in succession. It is possible that this improved feature map will improve the classification performance.

D. OTONET: CLASSIFICATION

In this section, we outline the design of our OTONet architecture, structured as follows:

Prior to official network training, we first apply Principal Component Analysis [32] (PCA) to reduce the dimensionality of the data. As shown in Figure 5, this step aids in parameter reduction and the preservation of important information. Subsequently, the 3D patch undergoes processing through three network branches. Each branch comprises three consecutive 3D octave convolution layers for the extraction of multi-scale features [33]. Batch Normalization-Rectified Linear Unit [34] (BN-RELU) activation comes after each of these layers. The outputs of these branches are denoted as A1, A2, and A3. It's important to note that the three branches vary

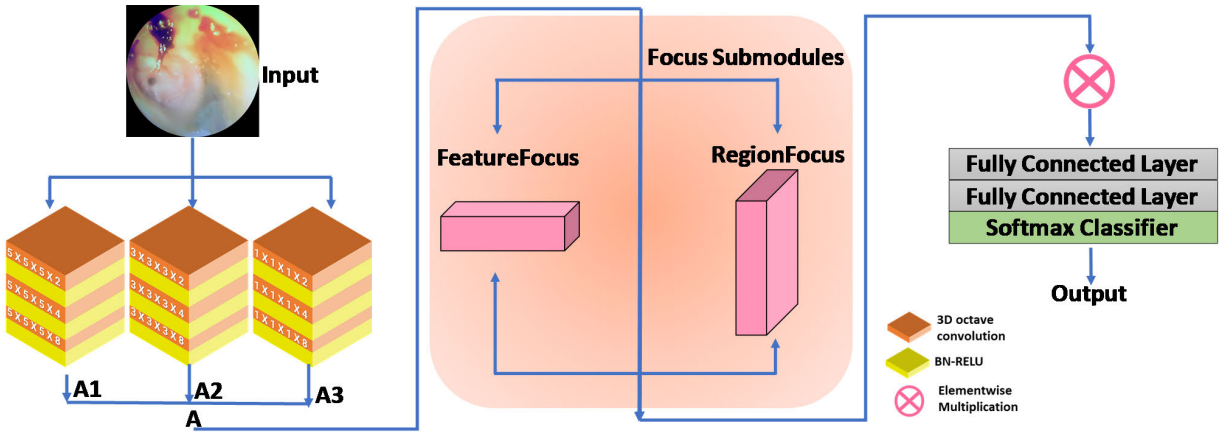

FIGURE 5. OTONet architecture.

in octave 3D convolution kernel sizes—1 × 1 ×1, 3 × 3 ×3, and 5 × 5 ×5. Figure 5 illustrates the design of each branch with distinct numbers of convolution kernels.

For ease of concatenation, we ensure the consistency of the original data size and feature map dimensions. Thus, the outputs of the three branches (A_1, A_2, A_3) are concatenated to form A:

$$A = \text{Concat}(A_1, A_2, A_3) \tag{11}$$

In this case, the concatenation operation is represented by Concat(.). To further abstract the feature map, a standard 3D convolutional layer is then applied.

Next, in order to improve the acquired feature maps' ability to discriminate, a focus module is presented that includes both feature-wise focus and region-wise focus.

Lastly, for classification, our architecture makes use of a softmax classifier and two fully connected layers. The fully connected layer uses 'dropout' to reduce overfitting in a sensible way without adding too many parameters. Categorical cross-entropy is chosen as the loss function as follows:

$$E = -\sum t_o * \log(y_o) \tag{12}$$

where y_o denotes the network output and t_o denotes the correct label. For the ongoing reduction of loss and parameter updates, we adopt the Adam optimization method. OTONet is an innovative multi-scale octave 3D CNN designed specifically for the classification of otoscopy images. It is enhanced with feature-focus and region-focus mechanisms.

OTONet utilizes Octave Convolution, Feature Focus, and Region Focus submodules to boost its classification capabilities. Octave Convolution addresses the challenge of increasing model parameters and memory requirements in 3D CNNs by decomposing feature maps into high and low-frequency components. It updates these components individually and then fuses them to reduce region redundancy and efficiently expand the output space. This three-tiered octave 3D convolution diminishes region redundancy and increases the receptive field. The Feature and Region Focus submodules further enhance OTONet's architecture. The Feature Focus submodule reduces feature redundancy and constructs a feature attention map to capture inter-feature relationships. Simultaneously, the Region Focus submodule explores focal points within feature maps, improving the model's ability to discern salient portions. The seamless integration of these submodules works cohesively to bestow distinct significance upon feature map features. This results in a comprehensive framework excelling in capturing intricate patterns and features crucial for accurate otoscopy image classification.

E. EVALUATION METRICS

The assessment of the proposed OTONet framework's performance in otoscopy image classification is vital for understanding its effectiveness. These metrics encompass accuracy, sensitivity, specificity, and the F1 score, providing a comprehensive view of the model's classification capabilities. Accuracy reflects the overall correctness of predictions, while sensitivity measures the model's ability to correctly identify positive cases. Specificity assesses the model's aptitude for accurately recognizing negative cases, and the F1 score balances precision and recall. The formulae for the computation are given in equations 13-16.

$$F1 - \text{score} = \frac{2.precision.recall}{precision + recall} \tag{13}$$

$$\text{Sensitivity} = \frac{\sum_{i=1}^{c} \frac{TP_i}{TP_i + FN_i}}{c} \tag{14}$$

$$\text{Precision} = \frac{\sum_{i=1}^{c} \frac{TP_i}{TP_i + FP_i}}{c} \tag{15}$$

$$\text{Specificity} = \frac{\sum_{i=1}^{c} \frac{TN_i}{TN_i + FP_i}}{c} \tag{16}$$

where True Positives (TP) signify instances that have been correctly classified into a specific class, True Negatives (TN) denote instances that have been accurately classified as not belonging to a particular class, False Positives (FP) represent instances that have been erroneously classified as belonging to a class, False Negatives (FN) refer to instances that have been incorrectly classified as not belonging to a specific class.

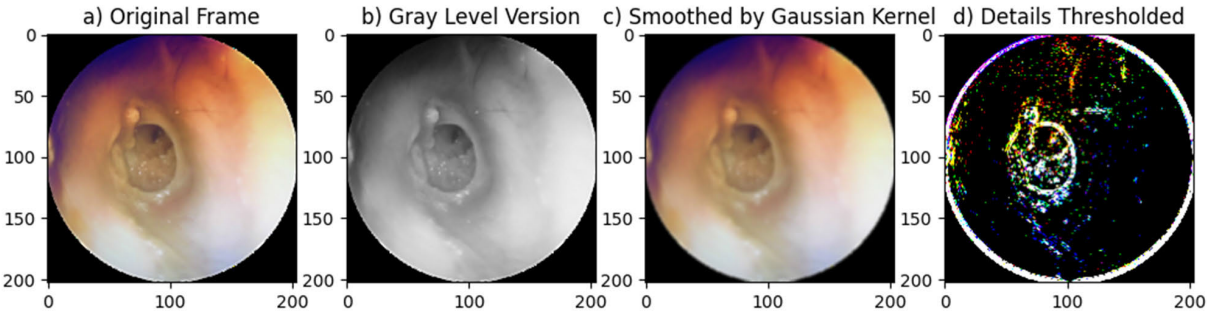

FIGURE 6. Otoscopy image preprocessing steps on an image in the dataset - (a) Original frame, (b) Grayscale version, (c) Smoothing by gaussian kernel, and (d) Thresholded details.

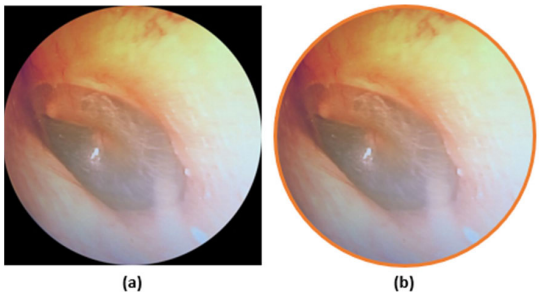

FIGURE 7. Example of image (a) acquired from the dataset, (b) cropped based on ROI.

IV. RESULTS

A. DATA PREPROCESSING

The dataset was organized manually through the segregation of samples. Efforts have been made to ensure data quality by isolating low-quality images resulted from factors such as insufficient lighting or inadvertent camera movement. The preprocessing described in detail in the methodology section were applied to the dataset to prepare it for further processing by the OTONet and is demonstrated in Figure 6. Figure 6a shows the original frame which refers to the unaltered, raw image frame captured during the otoscopy procedure. It serves as the starting point for all subsequent preprocessing steps. The original color image is converted into a grayscale version as shown in 6b, where each pixel's intensity is represented by a single value (gray level) instead of multiple color channels causing image simplification for further processing. In order to smooth out the original and reduce noise, Figure 6c illustrates the process of convolution of the input with a Gaussian filter. It helps to eliminate high-frequency noise and detail, resulting in a cleaner representation. The frame in 6d shows thresholded values when smoothed image is subtracted from the original image. In our work, a threshold value of 4 is applied, meaning that pixels with intensity values greater than or equal to 4 are set to high, and those below 4 are set to low. A circular region of interest is computed for each of the images and segmented from the binary image to isolate the area of interest within the image which is the circular frame that captures relevant information as illustrated in Figure 7.

In Table 2, we demonstrate the data split for the OTONet model. The dataset is categorized into various ear conditions, with images divided into training, testing, and validation sets. The allocation of separate validation and test datasets is crucial for assessing the model's performance, validating its generalization, and ensuring its accuracy in classifying otoscopy images across diverse categories. These datasets serve as independent benchmarks to gauge the model's robustness and reliability in real-world scenarios, which is essential for the development of an effective otoscopy image classification system.

Table 3 provides an overview of the class distribution before and after applying the synthetic minority oversampling technique to address class imbalance in our dataset. The table displays the original number of images in each class and the percentage of over-sampling applied to the

TABLE 2. The number of samples for each class for training, testing and validation.

Image Class	Train	Test	Validation	Total
Normal	374	107	53	535
Acute Otitis Media	83	24	11	119
Chronic Otitis Media	44	13	6	63
Earwax Plug	98	28	14	140
Otitis Externa	28	9	4	41
Foreign Object	2	1	0	3
Ventilation Tube	11	4	1	16
Pseudo Membrane	7	3	1	11
Tympanosclerosis	19	6	2	28

TABLE 3. Addressing the imbalance in classes using oversampling techniques.

Image Class	Original Train Data	Over sampling Ratio	Updated Train Data
Acute Otitis Media	83	1	166
Chronic Otitis Media	44	1	88
Earwax Plug	98	1	196
Otitis Externa	28	3	112
Foreign Object	2	4	12
Ventilation Tube	11	3	55
Pseudo Membrane	7	5	35
Tympanosclerosis	19	3	76

TABLE 4. Techniques used for image augmentation on the fly.

Augmentation Technique	Parameter
Rotation	[0, 350]
Horizontal Flip	[0, 1]
Vertical Flip	[0, 1]
Zoom	[0.0, 0.5]
Brightness	[0.1, 0.1]
Saturation	[0.5, 1.5]
Gaussian Noise	[0, 0.05]

minority classes using SMOTE. This technique helps create a more balanced dataset by oversampling the minority classes effectively increasing the representation of these classes. As seen in the table, the oversampling percentage varies for each class depending on the initial class distribution. The highest oversampling percentage is applied to the "Foreign Object" class, which originally had only two images. The normal image class had 374 images and has not been under or oversamples. This is done ensure that the machine learning model is not biased towards the majority classes. We aim to provide the model with sufficient examples to learn and generalize from, ultimately improving its ability to accurately classify otoscopy images across all classes.

We employed a suite of on-the-fly data augmentation techniques as illustrated in the accompanying Table 4, brought substantial value to our classification task. Random rotations, within the range of 0 to 350 degrees, offered simulated variations in viewpoint, essential for modeling the wide array of orientations seen in otoscopy procedures. Horizontal and vertical flips provided mirroring effects, enabling the model to discern ear conditions from both left and right perspectives. Furthermore, minor changes in the zooming in values, alterations in brightness and saturation, as well as adding Gaussian noise, collectively introduced variations mirroring the real-world challenges encountered during otoscopy. These augmentation techniques added depth to our training data, fostering a more robust and generalized model for accurate ear condition classification.

B. LOSS ANALYSIS

The loss analysis of OTONet, as depicted in Figure 8, offers valuable insights into the training and validation performance over 110 epochs. In this analysis, the blue curve represents the training loss, reflecting how effectively OTONet learned the nuances of the training dataset with each successive epoch. The orange curve illustrates the validation loss, which gauges how well the model generalizes its knowledge to previously unseen data. A small difference between the training and validation loss curves is indicative of the model's ability to minimize overfitting, ensuring robust generalization to real-world otoscopy images. As the loss values for both training and validation data approach zero, it underscores the proficiency of OTONet in capturing the intricate patterns within the dataset. The diminishing loss values as the epochs progress signify

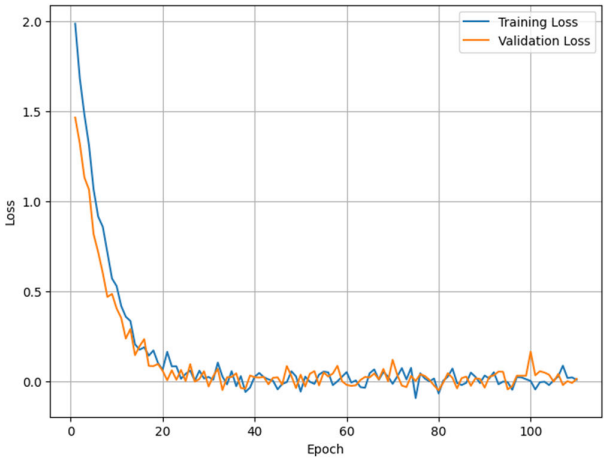

FIGURE 8. OTONet train and validation loss curves for 110 epochs.

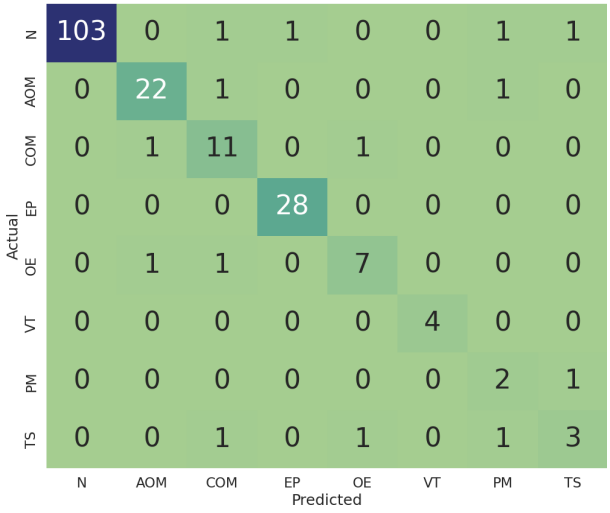

FIGURE 9. OTONet confusion matrix depicting the classification results for different ear conditions. [N (Normal), AOM (Acute Otitis Media), COM (Chronic Otitis Media), EP (Earwax Plug), OE (Otitis Externa), VT (Ventilation Tube), PM (Pseudo Membrane), TS (Tympanosclerosis)].

the model's rapid convergence and its capability to swiftly enhance its performance. These findings collectively endorse the efficacy of OTONet in learning and classifying ear conditions from otoscopy images, demonstrating confidence in its applicability and potential for accurate diagnoses.

C. CONFUSION MATRIX ANALYSIS

The confusion matrix, presented in Figure 9, provides a comprehensive assessment of the classification performance of OTONet across the spectrum of ear condition classes. This matrix is instrumental in understanding the model's strengths and areas for improvement. The diagonal elements represent true positive counts, indicating the number of instances correctly classified. Notably, the "Normal" and "Acute Otitis Media" classes exhibit robust classification, with the majority of instances correctly identified. However, off-diagonal elements signify misclassifications, revealing areas where OTONet may benefit from further refinement. Instances of misclassification are observed in classes such

TABLE 5. Performance comparison of OTONet with state-of-the-art ML image classification architectures.

Method	Accuracy	Sensitivity	Specificity	F1 Score
ResNet50	95.5%	96.8%	92.4%	95.6%
ResNet50v2	96.2%	98.1%	97.8%	96.4%
VGG16	98.0%	98.9%	96.3%	98.0%
DenseNet169	97.8%	96.6%	95.4%	97.9%
ConvNeXtTiny	93.3%	96.4%	95.3%	94.5%
OTONet	**99.3%**	**99.3%**	**98.8%**	**99.4%**

as "Chronic Otitis Media," "Earwax Plug," and "Otitis Externa." The misclassifications observed in certain classes, such as "Chronic Otitis Media," "Earwax Plug," and "Otitis Externa," can be partially attributed to the imbalanced distribution of samples in these classes. Since these classes have fewer instances, the model may have had less exposure to them during training. As a result, possibly struggled to distinguish subtle differences in these less-represented classes, leading to misclassifications even after augmentations and oversampling techniques.

D. COMPARATIVE ANALYSIS

In this comparative analysis section, we evaluate the performance of OTONet against several state-of-the-art machine learning image classification architectures, including ResNet50, ResNet50v2, VGG16, DenseNet169, and ConvNeXtTiny on the dataset.

Table 5 provides an in-depth analysis of the performance of OTONet in comparison to other state-of-the-art ML image classification architectures. The exceptional accuracy of OTONet at 99.3% indicates its remarkable capability to correctly classify images, outperforming competitors like VGG16. The high sensitivity of 99.3% underscores OTONet's proficiency in identifying positive instances, while the specificity of 98.8% highlights its accuracy in recognising negative cases. The F1 score of 99.4% further accentuates OTONet's ability to balance precision and recall, showcasing its robust performance. These results collectively highlight the superior classification capabilities of OTONet when compared to other state-of-the-art machine learning models, emphasizing its potential for advanced and accurate ear condition classification in otoscopy images.

Figure 10 complements these quantitative results by illustrating the ROC graph and AUC comparison. OTONet stands out with the highest AUC of 0.89, reinforcing its exceptional ability to distinguish between different classes. ResNet50, VGG16, DenseNet169, and ConvNeXtTiny, while exhibiting reasonable AUC values, fall short of OTONet's performance.

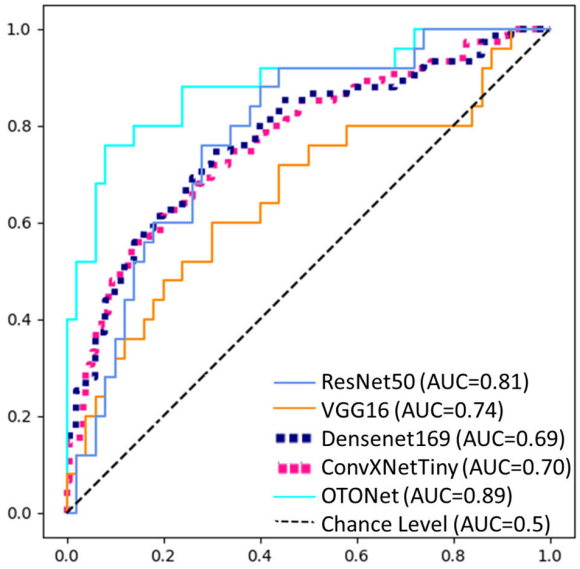

FIGURE 10. Comparative ROC graph.

E. LIMITATIONS OF THE STUDY

In the course of this research, potential limitations and challenges have been identified as follows. The publicly available dataset used in the study may not fully encapsulate the diversity of otoscopy images encountered in authentic clinical scenarios. This limitation could influence the model's generalizability across various patient demographics, otoscope devices, and clinical conditions. Despite efforts to mitigate class imbalance through oversampling techniques, certain classes, notably "Foreign Object" and "Pseudo Membrane," remain underrepresented, posing a challenge to the model's accurate classification of these conditions. The study also acknowledges the need for further clinical validation on a larger and more diverse patient population to assess the model's real-world applicability. Sensitivity to augmentation parameters is identified as another potential limitation, emphasizing the importance of fine-tuning these parameters for optimal model performance. Lastly, the proposed OTONet architecture, with its multiscale octave 3D CNN, may require substantial computational resources, potentially limiting its accessibility in resource-constrained environments. These limitations underscore the necessity for an open and thorough discussion and serve as a valuable guide for future research endeavours in this domain.

V. CONCLUSION

In conclusion, this paper presents an in-depth exploration of the proposed OTONet framework designed for the purpose

of automated classification of ear conditions in otoscopy images. The OTONet architecture, with its unique combination of Octave Convolution, Feature Focus, and Region Focus submodules, has demonstrated exceptional performance in classifying various ear pathologies. Leveraging state-of-the-art deep learning techniques, we have showcased its effectiveness in detecting conditions like Acute Otitis Media, Chronic Otitis Media, Earwax Plug, Otitis Externa, Ventilation Tube, Pseudo Membrane, Tympanosclerosis, and more. Our results underscore the remarkable accuracy and robustness of OTONet when compared to other well-established machine-learning models, such as ResNet50, ResNet50v2, VGG16, DenseNet169, and ConvNeXtTiny. OTONet exhibited superior performance across various metrics, including accuracy, sensitivity, specificity, and F1 score, which are crucial for precise diagnosis and classification of ear conditions. In this work, we have highlighted the critical role of data preprocessing, augmentation, and class imbalance handling techniques in improving the model's performance. Augmentation, in particular, significantly contributed to enhancing OTONet's capability to classify ear conditions accurately. The use of synthetic minority oversampling techniques effectively addressed class imbalances, ensuring a more balanced and reliable model.

The findings of this study are promising for the field of otoscopy image analysis, with the potential to support medical professionals in making accurate and timely diagnoses. OTONet's superior performance, combined with its efficiency and robustness, positions it as a valuable tool in revolutionizing the diagnosis of ear conditions. This work opens doors to future enhancements and wider applications in the field of medical image analysis, providing invaluable support to medical professionals and contributing to the overall well-being of patients worldwide.

FUNDING

This research article was conducted without any external funding. The authors are grateful to Manipal Academy of Higher Education, Manipal, for their essential role in facilitating the publication of this research.

CONFLICT OF INTEREST

The authors confirm that there are no known conflicts of interest associated with this publication and here have been no financial gains for this work that could have influenced its outcome.

REFERENCES

[1] S. I. Pelton, "Otoscopy for the diagnosis of otitis media," *Pediatric Infectious Disease J.*, vol. 17, no. 6, pp. 540–543, Jun. 1998.

[2] S. Anandamurugan, M. S. Kumar, E. G. Prashanth, and K. Nithin, "Ear disease detection using R-CNN," in *Proc. 5th Int. Conf. Comput. Intell. Commun. Technol. (CCICT)*, Sonepat, India, Jul. 2022, pp. 543–549, doi: 10.1109/CCiCT56684.2022.00101.

[3] E. Basaran, Z. Cömert, A. Sengür, Ü. Budak, Y. Çelik, and M. Togaçar, "Chronic tympanic membrane diagnosis based on deep convolutional neural network," in *Proc. 4th Int. Conf. Comput. Sci. Eng. (UBMK)*, Samsun, Turkey, Sep. 2019, pp. 1–4, doi: 10.1109/UBMK.2019.8907070.

[4] S. Hamrang-Yousefi, J. Ng, and C. Andaloro, "Eustachian tube dysfunction," in *StatPearls [Internet]*. Treasure Island, FL, USA: StatPearls, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK555908/

[5] T. C. Cavalcanti, H. M. Lew, K. Lee, S.-Y. Lee, M. K. Park, and J. Y. Hwang, "Intelligent smartphone-based multimode imaging otoscope for the mobile diagnosis of otitis media," *Biomed. Opt. Exp.*, vol. 12, no. 12, pp. 7765–7779, Dec. 2021.

[6] M. A. Khan, S. Kwon, J. Choo, S. M. Hong, S. H. Kang, I.-H. Park, S. K. Kim, and S. J. Hong, "Automatic detection of tympanic membrane and middle ear infection from oto-endoscopic images via convolutional neural networks," *Neural Netw.*, vol. 126, pp. 384–394, Jun. 2020, doi: 10.1016/j.neunet.2020.03.023.

[7] (Oct. 2023). *Columbia Doctors*. [Online]. Available: https://www.columbiadoctors.org/health-library/multimedia/ear-exam-using-otoscope/

[8] A. Danishyar and J. V. Ashurst, "Acute otitis media," in *StatPearls [Internet]*. Treasure Island, FL, USA: StatPearls, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK470332/

[9] P. S. Sundar, C. Chowdhury, and S. Kamarthi, "Evaluation of human ear anatomy and functionality by axiomatic design," *Biomimetics*, vol. 6, no. 2, p. 31, May 2021, doi: 10.3390/biomimetics6020031.

[10] L. T. Braun, J. M. Zottmann, C. Adolf, C. Lottspeich, C. Then, S. Wirth, M. R. Fischer, and R. Schmidmaier, "Representation scaffolds improve diagnostic efficiency in medical students," *Med. Educ.*, vol. 51, no. 11, pp. 1118–1126, Nov. 2017, doi: 10.1111/medu.13355.

[11] M. Juuti, S. Szyller, S. Marchal, and N. Asokan, "PRADA: Protecting against DNN model stealing attacks," in *Proc. IEEE Eur. Symp. Secur. Privacy*, Jun. 2019, pp. 512–527, doi: 10.1109/EUROSP.2019.00044.

[12] D. Goldenberg and B. L. Wenig, "Telemedicine in otolaryngology," *Amer. J. Otolaryngol.*, vol. 23, no. 1, pp. 35–43, Jan. 2002, doi: 10.1053/ajot.2002.28770.

[13] A. Singh and M. K. Dutta, "Diagnosis of ear conditions using deep learning approach," in *Proc. Int. Conf. Commun., Control Inf. Sci. (ICCISc)*, vol. 1, Idukki, India, Jun. 2021, pp. 1–5, doi: 10.1109/ICCISc52257.2021.9484919.

[14] M. Viscaino, J. C. Maass, P. H. Delano, and F. A. Cheein, "Computer-aided ear diagnosis system based on CNN-LSTM hybrid learning framework for video otoscopy examination," *IEEE Access*, vol. 9, pp. 161292–161304, 2021, doi: 10.1109/ACCESS.2021.3132133.

[15] L. Hu, W. Li, H. Lin, Y. Li, K. Svanberg, G. Zhao, H. Zhang, and S. Svanberg, "Optical detection of otitis media using modified spectroscopic otoscope," in *Proc. Asia Commun. Photon. Conf. (ACP)*, Hangzhou, China, Oct. 2018, pp. 1–4, doi: 10.1109/ACP.2018.8595778.

[16] Y. Cai, J.-G. Yu, Y. Chen, C. Liu, L. Xiao, E. M. Grais, F. Zhao, L. Lan, S. Zeng, J. Zeng, M. Wu, Y. Su, Y. Li, and Y. Zheng, "Investigating the use of a two-stage attention-aware convolutional neural network for the automated diagnosis of otitis media from tympanic membrane images: A prediction model development and validation study," *BMJ Open*, vol. 11, no. 1, Jan. 2021, Art. no. e041139, doi: 10.1136/bmjopen-2020-041139.

[17] M. Viscaino, M. Talamilla, J. C. Maass, P. Henríquez, P. H. Délano, C. Auat Cheein, and F. Auat Cheein, "Color dependence analysis in a CNN-based computer-aided diagnosis system for middle and external ear diseases," *Diagnostics*, vol. 12, no. 4, p. 917, Apr. 2022, doi: 10.3390/diagnostics12040917.

[18] Z. Wu, Z. Lin, L. Li, H. Pan, G. Chen, Y. Fu, and Q. Qiu, "Deep learning for classification of pediatric otitis media," *Laryngoscope*, vol. 131, no. 7, pp. 2344–2351, Jul. 2021, doi: 10.1002/lary.29302.

[19] S. Camalan, A. C. Moberly, T. Teknos, G. Essig, C. Elmaraghy, N. Taj-Schaal, and M. N. Gurcan, "OtoPair: Combining right and left eardrum otoscopy images to improve the accuracy of automated image analysis," *Appl. Sci.*, vol. 11, no. 4, p. 1831, Feb. 2021, doi: 10.3390/app11041831.

[20] H. Binol, M. K. K. Niazi, C. Elmaraghy, A. C. Moberly, and M. N. Gurcan, "OtoXNet—Automated identification of eardrum diseases from otoscope videos: A deep learning study for video-representing images," *Neural Comput. Appl.*, vol. 34, no. 14, pp. 12197–12210, Jul. 2022, doi: 10.1007/s00521-022-07107-6.

[21] D. Cha, C. Pae, S.-B. Seong, J. Y. Choi, and H.-J. Park, "Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database," *EBioMedicine*, vol. 45, pp. 606–614, Jul. 2019, doi: 10.1016/j.ebiom.2019.06.050.

[22] K. K. Mohammed, A. E. Hassanien, and H. M. Afify, "Classification of ear imagery database using Bayesian optimization based on CNN-LSTM architecture," *J. Digit. Imag.*, vol. 35, no. 4, pp. 947–961, Aug. 2022, doi: 10.1007/s10278-022-00617-8.

[23] W. Satriaji and R. Kusumaningrum, "Effect of synthetic minority oversampling technique (SMOTE), feature representation, and classification algorithm on imbalanced sentiment analysis," in *Proc. 2nd Int. Conf. Informat. Comput. Sci. (ICICoS)*, Semarang, Indonesia, Oct. 2018, pp. 1–5, doi: 10.1109/ICICOS.2018.8621648.
[24] J. Tan, Y. Gao, Z. Liang, W. Cao, M. J. Pomeroy, Y. Huo, L. Li, M. A. Barish, A. F. Abbasi, and P. J. Pickhardt, "3D-GLCM CNN: A 3-dimensional gray-level co-occurrence matrix-based CNN model for polyp classification via CT colonography," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2013–2024, Jun. 2020, doi: 10.1109/TMI.2019.2963177.
[25] Y. Chen, H. Fan, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, S. Yan, and J. Feng, "Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution," 2019, *arXiv:1904.05049*.
[26] B. Yu, Z. Zhang, X. Shu, Y. Wang, T. Liu, B. Wang, and S. Li, "Joint extraction of entities and relations based on a novel decomposition strategy," 2019, *arXiv:1909.04273*.
[27] J. Yang, F. Xie, H. Fan, Z. Jiang, and J. Liu, "Classification for dermoscopy images using convolutional neural networks based on region average pooling," *IEEE Access*, vol. 6, pp. 65130–65138, 2018, doi: 10.1109/ACCESS.2018.2877587.
[28] D. He, S. Chan, X. Ni, and M. Guizani, "Software-defined-networking-enabled traffic anomaly detection and mitigation," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1890–1898, Dec. 2017, doi: 10.1109/JIOT.2017.2694702.
[29] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Sanghyun_Woo_Convolutional_Block_Attention_ECCV_2018_paper.html
[30] H. Peng, N. Pappas, D. Yogatama, R. Schwartz, N. A. Smith, and L. Kong, "Random feature attention," 2021, *arXiv:2103.02143*.
[31] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of image classification algorithms based on convolutional neural networks," *Remote Sens.*, vol. 13, no. 22, p. 4712, Nov. 2021, doi: 10.3390/rs13224712.
[32] M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," *Nature Rev. Methods Primers*, vol. 2, no. 1, p. 100, Dec. 2022, doi: 10.1038/s43586-022-00184-w.
[33] L. A. Lim and H. Y. Keles, "Learning multi-scale features for foreground segmentation," *Pattern Anal. Appl.*, vol. 23, no. 3, pp. 1369–1380, Aug. 2020, doi: 10.1007/s10044-019-00845-9.
[34] J. Ge, X. Cui, K. Xiao, C. Zou, Y. Chen, and R. Wei, "BNReLU: Combine batch normalization and rectified linear unit to reduce hardware overhead," in *Proc. IEEE 13th Int. Conf. ASIC (ASICON)*, Oct. 2019, pp. 1–4, doi: 10.1109/ASICON47005.2019.8983577.

DIVYA RAO received the Ph.D. degree in artificial intelligence applied to oncology. She is currently an Assistant Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. Her research interests include machine learning, natural language processing, and healthcare informatics.

ROHIT SINGH received the M.B.B.S. and M.S. degrees and the D.N.B. degree in otorhinolaryngology-head and neck surgery from Kasturba Medical College, Manipal, in 2000 and 2005, respectively. He is an Additional Professor in otorhinolaryngology with Kasturba Medical College. He is the Director of alumni relations, fosters alumni connections at the Manipal Academy of Higher Education. He teaching both undergraduate and postgraduate students, he conducts Rhinology research engages in clinical and surgical practices and contributes to departmental administration. He actively involved in continuing medical education programs and conferences, he also serves as the Staff Advisor for the Student's Council's Literary Committee.

SUDIKSHA KOTTACHERY KAMATH is currently pursuing the bachelor's degree in information technology with the Manipal Institute of Technology, Manipal, with a focus on a minor in business management. She has a deep interest in machine learning and its implementations in the healthcare industry and aspires to contribute to the integration of information technology and the healthcare sector.

SANJEEV KUSHAL PENDEKANTI is currently pursuing the bachelor's degree in information technology with the Information and Communication Technology Department, Manipal Institute of Technology. He is deeply interested in the field of data mining and machine learning and their applications in the healthcare domain. He aspires to contribute to impactful research in this domain and use technology to make significant improvements to the existing medical sector.

DIVYA PAI received the bachelor's degree in dental sciences and the M.D.S. degree in orthodontics and dentofacial orthopedics from the Manipal College of Dental Sciences, Manipal, in 2012 and 2017, respectively. She is currently an Assistant Professor with the Department of Orthodontics and she is engaged in teaching undergraduate programs, including the Dental Mechanics and DORA Programs.

SUCHETA V. KOLEKAR received the Ph.D. degree in adaptive e-learning from MIT. She is currently an Associate Professor with the Department of Information and Communication Technology, MIT, Manipal Academy of Higher Education, Karnataka, India. She is extensively engaged in research areas, such as e-learning, web usage mining, human–computer interaction, serious game development, and cloud computing. She received the E-Learning Excellence Award by the Academic Conferences International, in 2017, for her pioneering work in adaptive e-learning.

M. RAVIRAJA HOLLA received the B.E. (C.S.E.) degree from Bangalore University, India, the M.Tech. (C.S.E.) degree from KSOU, Mysuru, India, and the Ph.D. degree from the Department of Computer Engineering, National Institute of Technology Karnataka (NITK), Surathkal. He is currently an Assistant Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. His research interests include information security, high-performance computing, and the semantic web.

SAMEENA PATHAN is currently an Assistant Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. Her research interests include pattern recognition, medical image analysis, artificial intelligence, and machine learning.