



Toward automated severe pharyngitis detection with smartphone camera using deep learning networks

Tae Keun Yoo^{a,*}, Joon Yul Choi^{b,1}, Younil Jang^c, Ein Oh^d, Ik Hee Ryu^{e,f}

^a Department of Ophthalmology, Aerospace Medical Center, Republic of Korea Air Force, Cheongju, South Korea

^b Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, OH, USA

^c Department of Otorhinolaryngology-Head & Neck Surgery, 10th Fighter Wing, Republic of Korea Air Force, Suwon, South Korea

^d Department of Anesthesiology and Pain Medicine, Seoul Women's Hospital, Bucheon, South Korea

^e B&VIIT Eye Center, Seoul, South Korea

^f VISUWORKS, Seoul, South Korea



ARTICLE INFO

Keywords:
Pharyngitis
Tonsillitis
Deep learning
Smartphone
Automated diagnosis
Telemedicine

ABSTRACT

Purpose: Severe pharyngitis is frequently associated with inflammations caused by streptococcal pharyngitis, which can cause immune-mediated and post-infectious complications. The recent global pandemic of coronavirus disease (COVID-19) encourages the use of telemedicine for patients with respiratory symptoms. This study, therefore, purposed automated detection of severe pharyngitis using a deep learning framework with self-taken throat images.

Methods: A dataset composed of two classes of 131 throat images with pharyngitis and 208 normal throat images was collected. Before the training classifier, we constructed a cycle consistency generative adversarial network (CycleGAN) to augment the training dataset. The ResNet50, Inception-v3, and MobileNet-v2 architectures were trained with transfer learning and validated using a randomly selected test dataset. The performance of the models was evaluated based on the accuracy and area under the receiver operating characteristic curve (ROC-AUC).

Results: The CycleGAN-based synthetic images reflected the pragmatic characteristic features of pharyngitis. Using the synthetic throat images, the deep learning model demonstrated a significant improvement in the accuracy of the pharyngitis diagnosis. ResNet50 with GAN-based augmentation showed the best ROC-AUC of 0.988 for pharyngitis detection in the test dataset. In the 4-fold cross-validation using the ResNet50, the highest detection accuracy and ROC-AUC achieved were 95.3% and 0.992, respectively.

Conclusion: The deep learning model for smartphone-based pharyngitis screening allows fast identification of severe pharyngitis with a potential of the timely diagnosis of pharyngitis. In the recent pandemic of COVID-19, this framework will help patients with upper respiratory symptoms to improve convenience in diagnosis and reduce transmission.

1. Introduction

Diagnostic support in remote healthcare services has shown the ability to minimize the exposure of ill patients to healthcare providers and other patients [1]. The recent global pandemic of coronavirus disease (COVID-19) has encouraged the use of telemedicine for patients with upper respiratory symptoms [2]. Because smartphones have

become ubiquitous, many researchers are interested in utilizing them in telemedicine. Deep learning technology can assist with patient examination using a smartphone when clinicians deal with limited information in a remote patient monitoring setting. In particular, a smartphone is useful to take a picture of the throat [3]. Therefore, home monitoring using a smartphone will help in the diagnosis and treatment of patients with upper respiratory symptoms to improve convenience and to reduce

* Corresponding author. Department of Ophthalmology, Aerospace Medical Center, Republic of Korea Air Force, 635 Danjae-ro, Namil-myeon, Cheongwon-gun, Chungcheongbuk-do, 363-849, Cheongju, South Korea.

** Corresponding author. Epilepsy Center, Neurological Institute, Cleveland Clinic, 9500 Euclid Ave, Cleveland, Ohio, USA.

E-mail addresses: eyetaekeunyoo@gmail.com (T.K. Yoo), jychoi717@gmail.com (J.Y. Choi).

¹ Tae Keun Yoo and Joon Yul Choi contributed equally to this work.

transmission. There have been several approaches adopting deep learning for automated diagnosis of several diseases using images captured by smartphones [4,5].

Pharyngitis, which is diagnosed in more than 11 million patients in the United States annually, is a common condition associated with acute upper respiratory tract infection [6]. Pharyngitis is an inflammation of the back of the throat and tonsils. Sore throat caused by pharyngitis is one of the main causes of medical visits for young patients [7]. The most common cause of acute pharyngitis is a self-limiting viral infection. However, *Streptococcus pyogenes* is the major bacterial infectious cause of pharyngitis and is responsible for an estimated 20–30% of cases of sore throat [8]. Frequently, severe pharyngitis with fever and exudative tonsillitis is associated with streptococcal pharyngitis, which can cause immune-mediated and post-infectious complications, such as acute rheumatic fever [9]. Therefore, timely diagnosis of pharyngitis for treatment is important to reduce symptoms, fever, and complications [10]. However, many patients with upper respiratory infection ignore their symptoms in the early stage and medical visits do not routinely take place. Moreover, in recent days, many patients hesitate to visit clinics because of the COVID-19 outbreak.

The importance of a mobile-based monitoring system for patients with acute upper respiratory infections has been raised because of its applicability and effectiveness [3]. A previous study endeavored to collect throat images using additional equipment in conjunction with the smartphone and used the k-nearest neighbor algorithm in color distribution space to classify images with streptococcal tonsil [11]. However, the need for additional equipment limited their effectiveness of this method in a real-world setting. Moreover, color distribution was unable to represent the characteristic features of pharyngitis. Throat images exhibit variation in the size, illumination, and shape of the oral cavity.

Here, we present a deep learning model with smartphone-based throat images facilitating the detection of severe pharyngitis using self-taken throat images (Fig. 1). We performed automated detection of severe pharyngitis using a convolutional neural network (CNN) framework.

2. Methods

2.1. Data collection

The basic concept of our framework is throat examination using a self-taken smartphone image with computer-aided diagnosis system, which is similar to the previous dermatology study [12]. This study was performed using publicly accessible self-taken throat images on the web. We collected throat images from web-based open social Q&A systems including Naver Korea (<https://kin.naver.com>), and Yahoo Japan (<https://chiebukuro.yahoo.co.jp>). The additional throat image datasets were extracted using the Google image search engine. Most throat

images were posted by users asking for medical advice via the open social Q&A. The search strategy was based on the key terms “sore throat,” “pharyngitis,” “tonsillitis,” “exudative tonsillitis,” “tonsillopharyngitis,” “throat image,” and “smartphone” in Korean, Japanese, and English languages. The most recent image from the database search was achieved on April 30, 2020. Images that were not taken with smartphones were manually picked and excluded for this study. Images with the characteristics of either pharyngitis or normal throat were manually classified by two clinicians, and the ambiguous images were excluded to clarify the image domains. Finally, we collected the initial dataset with two classes including 131 throat images with pharyngitis and 208 normal throat images. The dataset was randomly separated into training (50%, N = 169), validation (25%, N = 85), and test sets (25%, N = 85) to apply deep learning to an independent dataset. Detailed data distribution and augmentation are described in Table 1. Only the throat and tonsils images were used for the input data without further manipulation to reduce the intra-class variance. Original images were extracted from the database in the PNG (Portable Network Graphics) format. The images were resampled to a pixel resolution of 256 × 256 × 3 in the PNG format for CycleGAN and deep learning models.

All procedures were performed in accordance with the ethical standards of the institutional and national research committee, and the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study did not require ethics committee approval; instead, the researchers used open web-based and de-identified data. All datasets for the development of the deep learning model are available at Mendeley Data repositories (<https://doi.org/10.17632/8ynyhnj2kz>).

2.2. Data augmentation using GAN

Because of the limited number of datasets and their imbalanced distribution, data augmentation is required for deep learning training. Basic data augmentation techniques such as flip, translation, rotation, and brightness change have been applied to train deep learning models. Several previous studies have attempted to train deep learning models using generative adversarial network (GAN)-based synthetic images to increase the classification performance [13]. Inspired by previous works using a generative adversarial network, we adopted the CycleGAN-based data augmentation to increase the accuracy of diagnosis. The cycle consistency in Fig. 2 allows CycleGAN to capture the characteristics of two image domains and automatically learn how these characteristics should be translated to transfer to domains without any paired datasets [14]. CycleGAN was developed to overcome the limitations of paired data when two generators and two discriminators are used. It is considered to be a powerful technique that performs image domain transfer and face transfer [15]. Previous studies have demonstrated that CycleGAN can improve deep learning models by generating training situations to learn better decision boundaries between classes.

We built the CycleGAN augmentation model to increase the

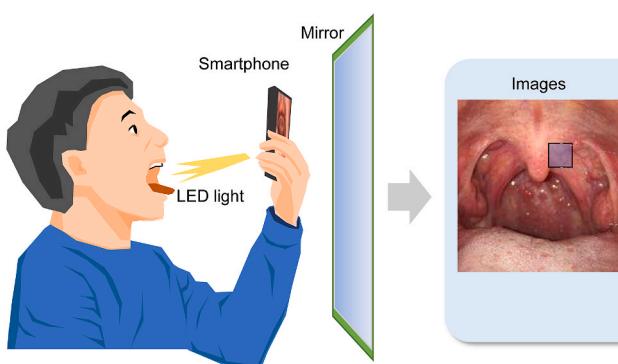


Fig. 1. Workflow of building a deep learning model for pharyngitis diagnosis using a smartphone.

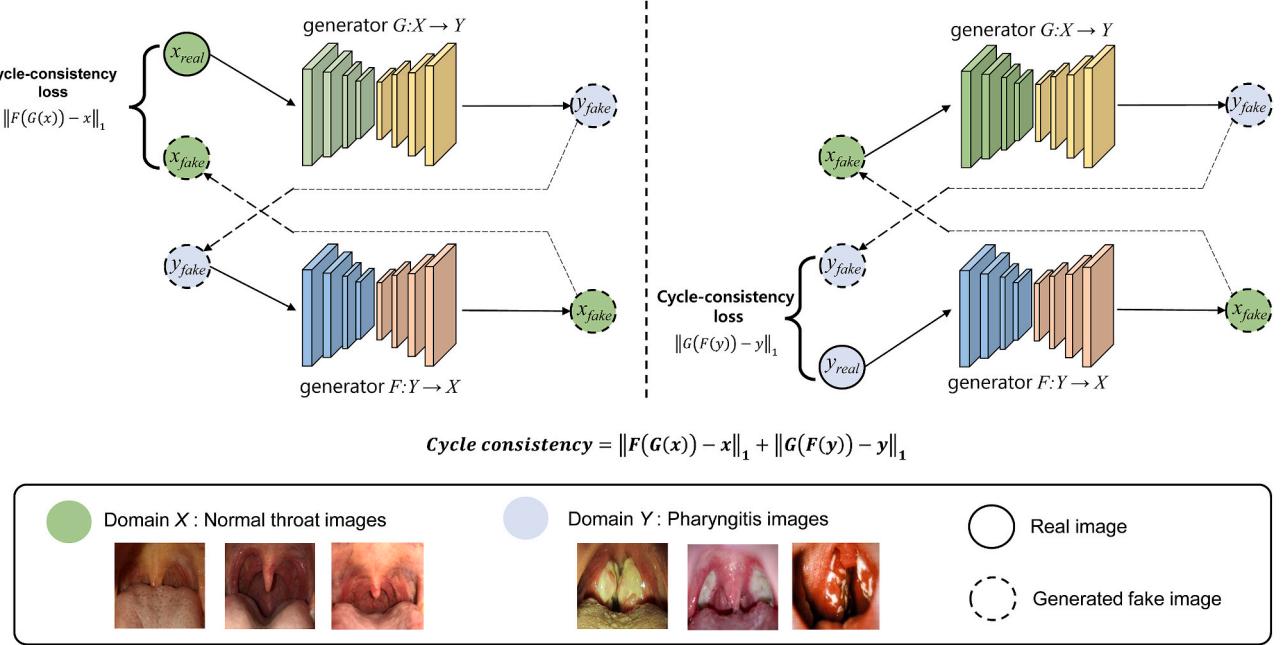


Fig. 2. Schematics of the CycleGAN model generating new normal images and pathologic throat images with pharyngitis.

generalizability of the dataset and to improve the classification performance in the imbalanced dataset. Before training the CycleGAN, the throat images were augmented using linear transformation including left and right flip, width and height translation from 5% to +5%, random rotation from 10° to 10° , zooming from 0% to 20%, and random brightness change from 10% to 10%. We defined these transformations as the basic augmentation step. In this step, 1000 normal throat and 600 pharyngitis images were randomly sampled for the training set, and 500 normal throat and 300 pharyngitis images were randomly sampled for the validation set. Using the data with basic augmentation, we trained the CycleGAN models to transform both normal to pharyngitis throat images and pharyngitis to normal images. The trained CycleGAN model augmented the training set (1000 normal throat and 600 pharyngitis images before augmentation), and a total of 3000 throat images including 1500 normal and 1500 pharyngitis images were prepared to train the diagnostic classifier model after CycleGAN-based augmentation. It should be noted that supervised GAN techniques including conditional GAN and Pix2px were unable to be applied in this study because of a lack of paired normal and pharyngitis images.

To use a verified and pre-designed image generator, all the input images required resizing to a pixel resolution of $256 \times 256 \times 3$, which is the basic setup of a CycleGAN. Therefore, we used the default parameter settings, that is, the ADAM optimizer with a batch size of 1, to optimize the GAN networks.

2.3. Development of CNN model

We trained conventional deep learning models after data augmentation. Because of a small image dataset in this study, developing a custom deep learning method is challenging due to difficulty and time-consuming (e.g. a small training dataset can easily result in an overfitted model and low performance) [16]. To overcome the problem of the small dataset, transfer learning was widely used to train deep learning models using pre-trained architectures [17,18]. This study also applied pre-trained learning models to the classification task of throat images. The last fully connected layer of the CNNs was only trained, and the study used the pre-trained conventional model as a feature extractor [17].

The CNN models including ResNet50, Inception-v3, and MobileNet-

v2 were adopted to build binary classifiers [19]. These CNN models have been used successfully in many studies, demonstrating state-of-the-art performance with the saliency map [20,21]. The models were trained using the training set, and the validation set was used to estimate how well the model had been trained. We downloaded the CNN models, which were pre-trained on the ImageNet database, and performed fine-tuning of the weights of the pre-trained networks. This process generally maintains the weights of some bottom layers to avoid over-fitting and performs delicate modification of the high-level features. To use the images generated by CycleGAN, the size of the input images for the deep learning models was set to a pixel resolution of $256 \times 256 \times 3$ and the images were resized for each pretrained model. Most conventional deep learning models adopted a pixel resolution of $256 \times 256 \times 3$ or $224 \times 224 \times 3$ [22]. One deep learning research showed that the best performance was achieved at an image resolution of between 256×256 and 448×448 pixels for binary classification [23]. Therefore, the resolution of our study was appropriate to detect pharyngitis in a binary decision. The model was trained with 250 epochs and a batch size of 20. The ADAM optimizer with a learning rate of 0.0001 was also used with a cross-entropy loss for all CNN models. The cross-entropy loss function is defined as:

$$L_{\text{cross entropy}} = - \sum_i^N p_i \log(q_i), \quad (1)$$

Where p_i represents the ground truth value, and q_i represents predicted probability value from a classifier for the i th image. The optimizer updated the network parameters to minimize the loss function. In our experiments, it tuned a fully connected layer of the CNN models. For example, the first 49 layers of ResNet50 were left frozen and we trained the last fully connected layer using the training dataset, which is described in Table 1, the ADAM optimizer. We chose the final classifier model which maximized the accuracy in the validation dataset.

To visualize the clusters to see if the classes are separable by a considerable margin, the t-distributed stochastic neighbor embedding (*t*-SNE) algorithm was executed using sampled instances. The feature vectors from the last layer of the pre-trained Inception-v3 model were extracted to train the *t*-SNE [24]. As there is a growing demand for explainable artificial intelligence methods in medicine [25], we adopted the Grad-CAM technique (<https://github.com/jacobgil/pytorchGradCam>)

Table 1
Throat image dataset and augmentation used in this study.

	Class	Number of training set (%)	Number of validation set (%)	Number of test set (%)
Raw data	Normal	104 (50.0)	52 (25.0)	52 (25.0)
	Pharyngitis	65 (49.6)	33 (25.2)	33 (25.2)
Basic augmentation	Normal	1000 (64.4)	500 (32.2)	52 (3.4)
	Pharyngitis	600 (64.3)	300 (32.2)	33 (3.5)
GAN-based image synthesis augmentation	Normal	1500 (73.1)	500 (24.4)	52 (2.5)
	Pharyngitis	1500 (81.8)	300 (16.4)	33 (1.8)

GAN, generative adversarial network.

-grad-cam) to generate the attention map [26]. The Grad-CAM visualizes the decisional areas of the CNN model using the gradients of any target flowing into the final convolutional network. The heatmap has a low resolution and it was up-sampled via bicubic interpolation. Finally, it produces heatmaps that highlight the area of interest and interprets the decision of the deep learning models.

The performance of the CNN models was evaluated based on the accuracy and area under the curve (AUC) of the receiver operating characteristic curve (ROC) and the precision-recall curve (PRC). The Youden index, which is an estimate of the optimal diagnostic threshold, was adopted in this study [27]. After obtaining the sensitivity and specificity, the Youden index was calculated at each cut-off point. We selected the optimal value which maximized the Youden index. These performance indexes are expressed as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2)$$

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{TP + FN}, \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (5)$$

$$\text{Youden index} = \text{Sensitivity} + \text{Specificity} - 1, \quad (6)$$

Where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives, respectively. We also performed a 4-fold cross-validation using the entire dataset to evaluate a generalized performance. Google Colab Pro, a cloud service for disseminating the deep learning research, was adopted to implement the CycleGAN and CNN models [28]. Google Colab Pro provides a development environment using the Tensorflow-based deep learning libraries and a robust graphic processing unit (GPU). This enables the rapid processing of a heavy deep learning network without the need for a personal GPU [15]. The available hardware for each virtual machine varied by session, but typically included top products of NVIDIA GPUs (K80, T4, or P100), around 8–12 GB of RAM, and 50–70 GB of free space on the virtual machine hard drive [29]. We used the Colab CycleGAN tutorial page to develop and validate the CycleGAN model, and all of the code is available on the Tensorflow webpage (<https://www.tensorflow.org/tutorials/generative/cyclegan>). We only modified the input pipeline to import our dataset. The code of the CycleGAN and CNN models is presented in Supplementary Material.

3. Results

We developed a deep learning model using GAN-based augmentation in the challenging context of pharyngitis detection. To build a balanced training dataset, the CycleGAN models generated normal and pathologic throat images using the initial training dataset. The color intensity

distributions of the pharyngitis and normal throat images were not significantly different, although most throat images had exudate regions (Fig. 3 A). The t-SNE algorithm demonstrates that both groups are clustered and they are separable by a considerable margin (Fig. 3 B). The final CycleGAN model for the pharyngitis throat image was trained for 20 epochs, which required 10 h. After training, normal throat images were translated into pathologic images, and throat images with pharyngitis were translated into normal images. Finally, we constructed a balanced augmented training dataset including 1500 normal and 1500 pharyngitis images using CycleGAN. The CycleGAN-based synthetic images were realistic and reflected the characteristic features of pharyngitis (Fig. 4 A). The CycleGAN model synthesized white or gray patches and increased the redness on the throat wall and tonsils from normal images (Fig. 4 B). Generated throat images were reviewed by three clinicians including an otolaryngologist and an anesthesiologist. All pharyngitis images generated were deemed by the three clinicians to show “more pathologic and inflammatory” results when compared to the original images. This feature generation based on the CycleGAN model can be effective for generating a new sample to increase the intra-class variation and generalizability.

The CNN models were trained using the final augmented dataset via the transfer learning technique. Fig. 5 demonstrates the training process of the ResNet50 CNN model using the training and validation sets. The training process for CNN took approximately 12 h for 1000 epochs with fine-tuning for each training step. After the 200th epoch, the validation accuracy was not improved and the cross-entropy of the validation result increased. Therefore, we considered that training for 200 epochs was optimal in the ResNet50 model and selected the trained model at the 200th epoch. Fig. 6 A shows the training, validation, and test datasets of CNN models with and without CycleGAN-based augmentation. The ROC-AUCs of ResNet50, Inception-v3, and MobileNet-v2 without GAN-based augmentation were 0.943, 0.932, and 0.934, respectively (Fig. 6 B). We obtained the best ROC-AUC of pharyngitis detection using ResNet50 with GAN-based augmentation corresponding to 0.988 (Fig. 6 C). At the optimal diagnostic cut-off value, the ResNet50 model predicted pharyngitis with an accuracy of 95.3%, sensitivity of 97.0%, and specificity of 94.2%. The ROC-AUC of Inception-v3 and MobileNet-v2 corresponded to 0.981 and 0.970, respectively. The models with GAN-based augmentation demonstrated superior performance in comparison with the models with only basic augmentation. We also evaluated the performance of the models using PRCs (Fig. 7). The PRC-AUC also demonstrated that deep learning models with GAN-based augmentation had better performance than those without GAN-based augmentation. ResNet50 with GAN-based augmentation predicted pharyngitis with the highest PRC-AUC of 0.986. When a custom deep learning network without the transfer learning technique was trained, the validation accuracy was lower than 80% and was not improved during the training epochs (Supplementary materials).

Table 2 shows the performance of CNN models via 4-fold cross-validation in the entire dataset. A similar tendency is observed for the average accuracy and AUC values in the cross-validation. The result shows that the highest detection accuracy and ROC-AUC achieved were 95.3% and 0.992, respectively, by using the ResNet50. Other CNN models showed lower performance than the ResNet50 with GAN-based augmentation, but there were no significant differences between CNN models in the 4-fold cross-validation. A saliency map using the Grad-CAM technique was generated to visualize the characteristic pathologic features of pharyngitis for the predicted evidence (Fig. 8 A). In normal throat images, however, some regions were highlighted (Fig. 8 B). When synthesized images were tested using the trained model, synthesized exudates were highlighted correctly in the images as shown in Fig. 8 C.

Furthermore, external clinical cases from previous literature were analyzed to investigate to show the capability of detecting pharyngitis of the framework developed in this study [30,31]. The representative cases with severe pharyngitis are shown in Fig. 9. One case presents an

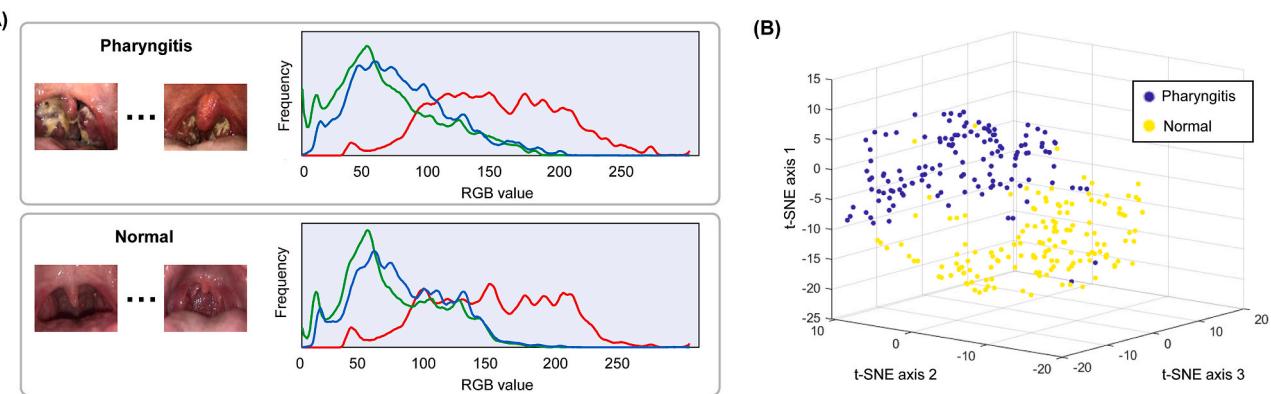


Fig. 3. Visual data exploration of pharyngitis and normal throat images. (A) The mean red-green-blue (RGB) distributions. (B) The feature space visualized using the t-SNE technique. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

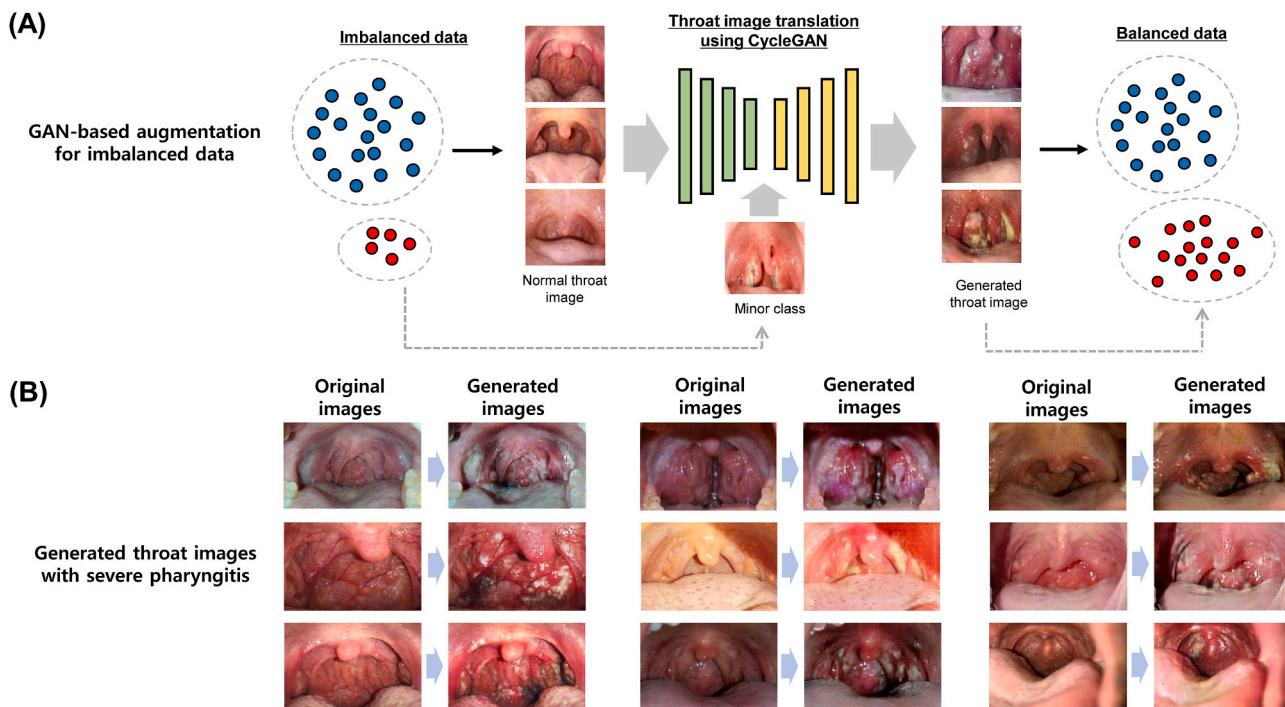


Fig. 4. Data augmentation using CycleGAN to improve the diagnostic performance of pharyngitis. (A) CycleGAN-based augmentation for imbalanced data. (B) Examples of pathologic throat images with pharyngitis generated by the CycleGAN.

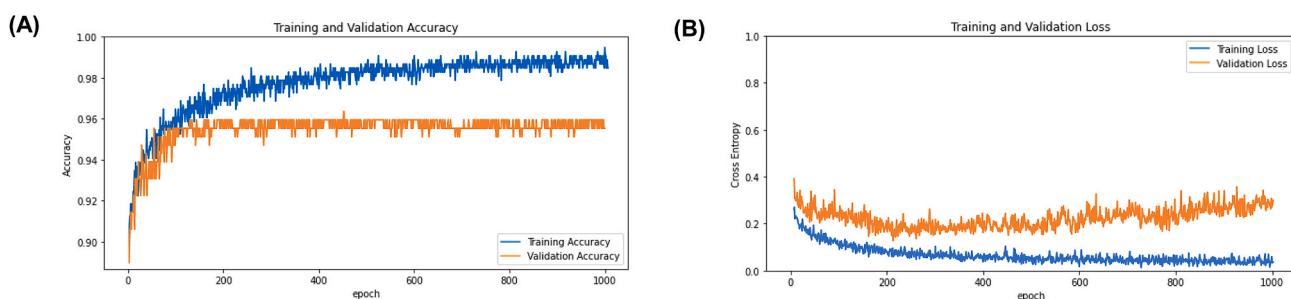


Fig. 5. Training process of the ResNet50 CNN model for pharyngitis detection. (A) Accuracy learning curves of the training and validation sets. (B) Loss learning curves of the training and validation sets.

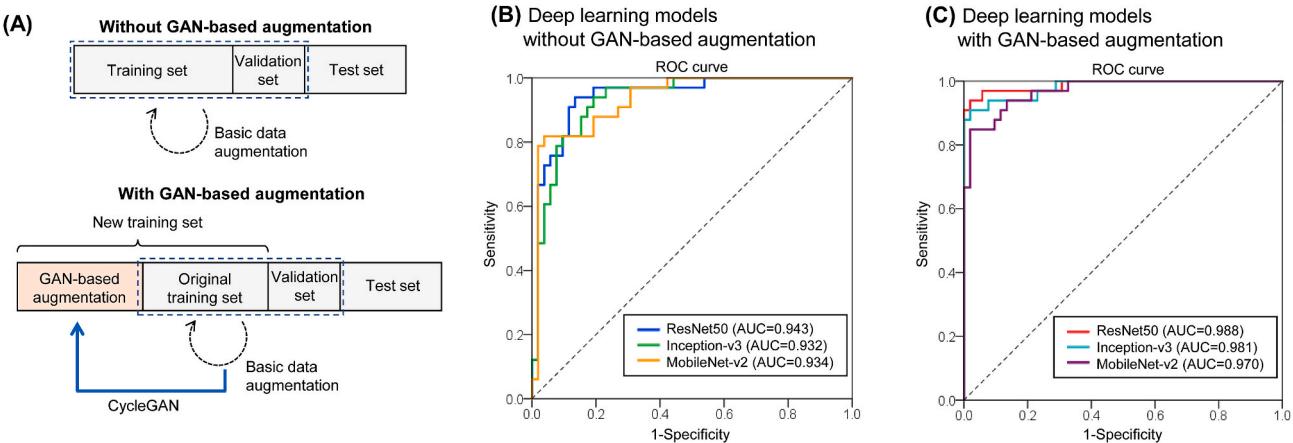


Fig. 6. Performance of pharyngitis detection deep learning models. (A) Schematics of basic augmentation and GAN-based augmentation. (B) The receiver operating characteristic curves of deep learning models using basic augmentation. (C) The receiver operating characteristic curves of deep learning models using GAN-based augmentation.

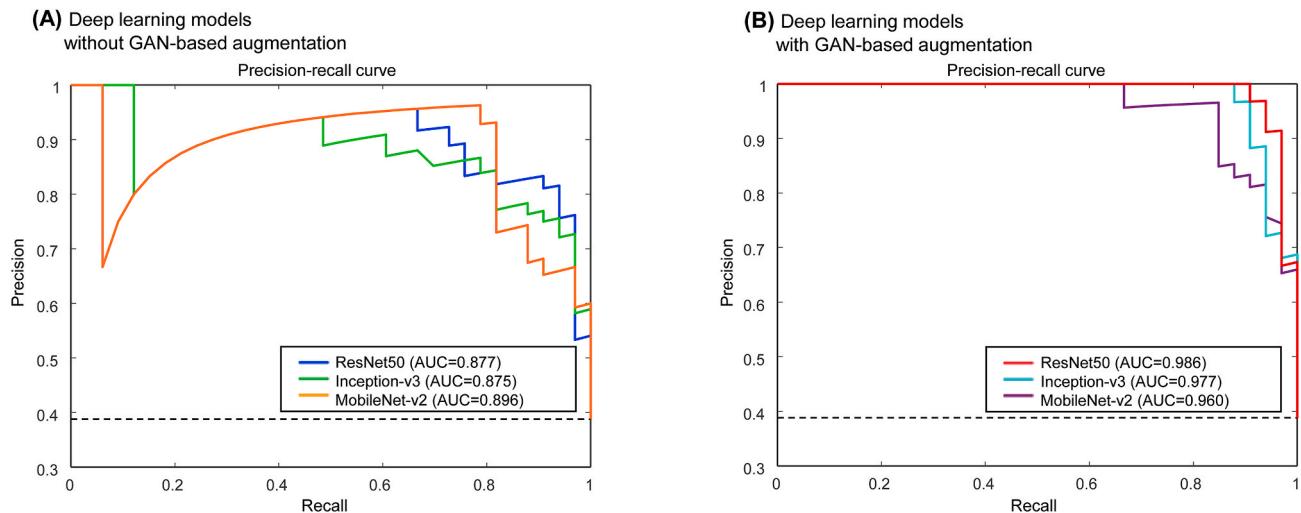


Fig. 7. Precision-recall curves of pharyngitis detection deep learning models. (A) The deep learning models using basic augmentation. (B) The deep learning models using GAN-based augmentation.

Table 2

Classification performance for severe pharyngitis detection in the 4-fold cross-validation using the developmental set.

	AUC (95% CI)	Accuracy (%) (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	P-value*
Basic augmentation					
ResNet50	0.968 (0.948–0.988)	92.5 (88.6–95.4)	89.8 (82.0–95.0)	94.2 (89.3–97.3)	0.207
Inception-v3	0.966 (0.947–0.986)	91.7 (87.6–94.8)	88.8 (80.8–94.3)	93.6 (88.5–96.9)	0.182
MobileNet-v2	0.963 (0.943–0.984)	91.3 (87.2–94.5)	91.8 (84.5–96.4)	91.0 (85.4–95.0)	0.150
GAN-based image synthesis augmentation					
ResNet50	0.992 (0.985–0.999)	95.3 (91.9–97.5)	92.9 (85.8–97.1)	96.8 (92.7–98.9)	Reference
Inception-v3	0.987 (0.978–0.996)	94.9 (91.4–97.3)	92.9 (85.8–97.1)	96.2 (91.8–98.6)	0.718
MobileNet-v2	0.981 (0.969–0.994)	93.7 (90.0–96.4)	90.8 (83.3–95.7)	95.5 (91.0–98.2)	0.483

AUC, area under the receiver operating characteristic curve; CI, confidence interval; GAN, generative adversarial network.

*Comparison of receiver operating characteristics curves with the single best technique (ResNet50 with GAN-based image synthesis augmentation) according to the Delong test.

81-year-old man with odynophagia (Fig. 9 A) and the other case presents a 41-year-old diabetic man with throat pain (Fig. 9 B). The ResNet50 model was able to detect pharyngitis in both cases. The Grad-CAM technique highlighted white patches on the throat wall as markers of severe pharyngitis in the deep learning model.

4. Discussion

The current proof-of-concept study investigated the possibility of deep learning using a smartphone for detecting pharyngitis. A recent study demonstrated the ability of self-taken throat images to detect pharyngitis based on the k-nearest neighbor algorithm in a color space [11]. That study used additional equipment as well as a

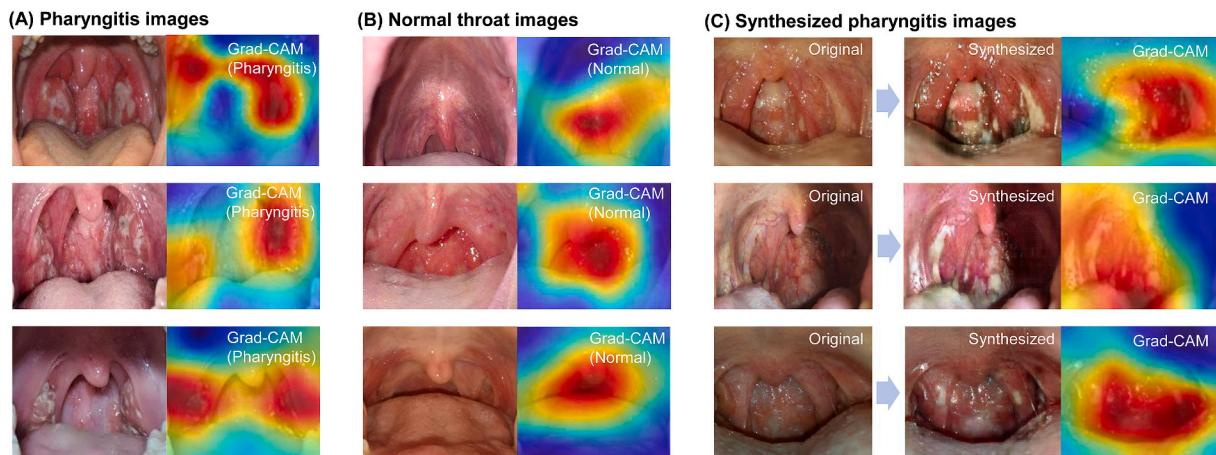


Fig. 8. Example of deep learning classification results with a saliency map using the Grad-CAM technique (A) Pharyngitis throat images. (B) Normal throat images. (C) Synthesized images.

smartphone to obtain throat images. We utilized self-taken throat images without additional equipment using deep learning techniques for the detection of pharyngitis. We showed that the trained GAN models were able to generate new realistic synthetic throat images which can improve the diagnostic accuracy. The final deep learning model achieves high accuracy for automated diagnosis of pharyngitis using smartphone images. Therefore, this framework could be targeted toward patients with a sore throat who need screening for severe pharyngitis. We believe that our study could be extended to computer-aided diagnosis using images from an endoscope system in otolaryngology clinics, similar to what has been done with colonoscopy images using a deep learning model [32].

To the best of our knowledge, no study has been performed to detect pharyngitis based on deep learning using smartphone images. However,

it should be noted that this study is considered as only a preliminary and proof-of-concept study because of its technical limitations due to using only Google Colab. The current study framework is similar to that of Chamier, which showed a deep learning framework using Google Drive and training and prediction on Google Colab [33]. Our proposed conceptual workflow is shown in Fig. 10. It needs the Flask servers and interfaces implemented using HTML and JavaScript to be applied in a real clinical setting. This work could be part of a larger project to enable smartphone-based pharyngitis detection via a cloud-based app-level mobile data analysis. We believe that a future app-based model can provide a robust solution for the cost-effective and convenient screening of pharyngitis in a telemedicine setting.

We have highlighted the feasibility of a smartphone-based approach with deep learning to detect pharyngitis. Our approach does not require

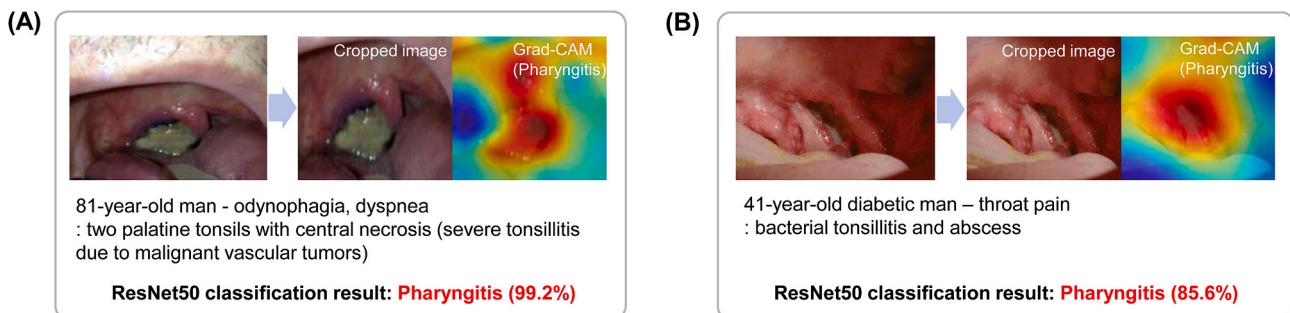


Fig. 9. Classification results from the deep learning model applied to clinical cases with severe pharyngitis. (A) An 81-year-old man with odynophagia [30]. (B) A 41-year-old diabetic man with throat pain [31].

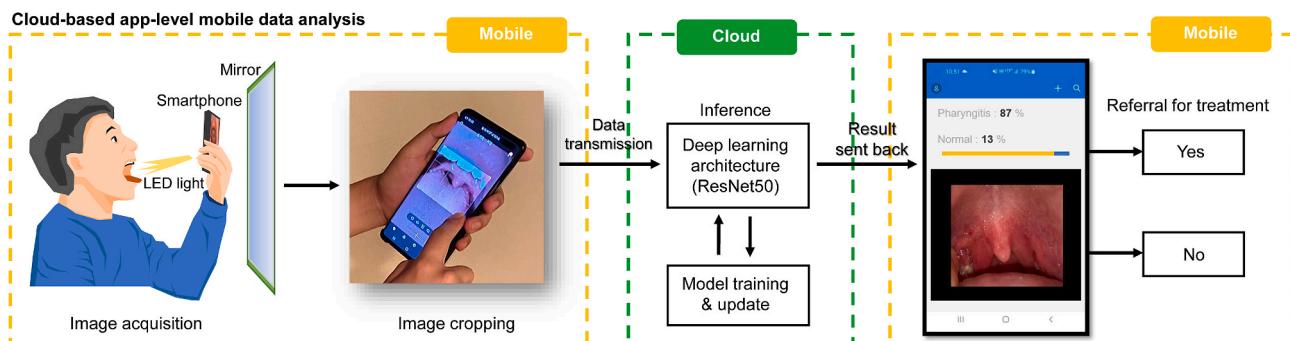


Fig. 10. Example of a proposed smartphone-based system for pharyngitis detection.

additional equipment to collect throat images, and we collected self-taken images using a smartphone. The use of web-based image capture of users from various devices and GAN-based image augmentation may allow robust image processing. While in the current study the implementation was performed on a desktop computer and the Colab's remote server, a smartphone application will also be possible to perform this identification for detecting pharyngitis. The light deep learning models such as MobileNet can be executed on a smartphone without transmitting the images to a server [34]. Our result demonstrates that MobileNet-v2 also has a high diagnostic performance similar to ResNet50 and Inception-v3. In future research, an increased amount of training data is needed to improve the detection accuracy using light models. The utility of deep learning and smartphones may facilitate the widespread adoption of telemedicine and physical examination platforms via our approach. Furthermore, the presented framework enabled using a smartphone camera and deep learning techniques will help the patients in self-screening for severe pharyngitis and may accelerate diagnostic support in remote healthcare services. Because of the recent pandemic of COVID-19, patients with respiratory symptoms need self-monitoring to evaluate their pathologic status [35]. Advances in technology will require clinicians to embrace remote healthcare services. The framework presented in this paper was designed for the timely diagnosis of pharyngitis for treatment. Similar smartphone-image-based diagnostic approaches have been introduced in several other medical image domains including skin diseases [36], hematologic diseases [5], oral diseases [37], and eye diseases [38].

The main concern of deep learning models in this study is the inability to generalize decision boundaries from a small number of datasets. Recently, deep learning using few-shot learning techniques enables the model to learn a new task with limited information from a few instances by incorporating prior knowledge [39]. As most few-shot learning methods are based on pre-trained CNN networks, our approach is also that of transfer learning based on pretrained CNN. GAN could improve deep learning models with small datasets by generating new training instances in a few-shot learning setting [40]. Our dataset is relatively small and imbalanced because it is difficult to collect self-taken pathologic throat images with pharyngitis. This study adopted a GAN technique to increase the accuracy of deep learning in detecting pharyngitis using a smartphone. Several previous studies in other medical domains focused on building GAN models for data augmentation [41,42]. Recently, progressive growing GAN (PGGAN) [41], Pix2Pix, and CycleGAN have been widely used to generate synthetic data in medical fields. CycleGAN performed well using a small dataset and has been successfully applied to breast mass classification [43], CT and MRI image augmentation [44], and face transformation for surgical outcome prediction [45]. Moreover, the CycleGAN has the advantages of very intuitive image domain transfer without matching paired images. Consistent with the previous studies, our work showed improved diagnostic accuracy using the CycleGAN model in the throat image.

This study has several limitations. First, we did not consider the Mallampati score which is a simple airway classification system to evaluate the risk for difficult tracheal intubation [46]. Some patients with a higher Mallampati score are unable to expose their throat without the use of a tongue depressor. To improve the utility of our framework, the Mallampati score needs to be automatically evaluated to confirm the visibility of the pharynx. In addition, throat images with tongue depressors should also be included in the training dataset for patients with a higher Mallampati score. Second, our deep learning model analyzed images with a low resolution of 256×256 pixels. The image size was regularized by the CycleGAN augmentation step. The resolution may affect the classification of mild pharyngitis cases. Future studies should improve the resolution of the throat image to provide a more accurate diagnosis. Third, the datasets for training and validation included a limited number of throat images. Although we attempted to collect throat images from web-based sources, we did not include real throat images from actual clinical settings. One previous study on skin diseases

explored clinical datasets and web-based datasets and showed that the labeling of web-based images is often uncertain [47]. Therefore, further clinical datasets with validation will be required to confirm the effectiveness of our framework. Fourth, because we collected only throat images via image search engines, there was no meta-data including gender, race, season, or age. According to a previous epidemiologic study, these factors could affect the pharyngitis detection performance [48]. Fifth, although the datasets were reviewed by authors, images could be potentially duplicated in the dataset. The duplicated images would affect the independence of the validation dataset.

We have shown the feasibility of deep learning for the detection of tonsil swelling and exudates in throat images. The limitations of our study should be overcome by the availability of a sufficient number of throat images taken by a smartphone with a well-organized study protocol. To validate the effectiveness of our framework, a prospective study with many patients should be performed in a clinic once the app-based framework is developed. This will solve the problem of possible duplicated images in the dataset and the absence of meta-data.

5. Conclusion

In this study, we proposed a transfer learning-based deep learning model to detect severe pharyngitis with self-taken throat images using a smartphone. The CycleGAN model successfully generated new throat images to improve the classification performance of the deep learning model. Our model performed binary classification with an accuracy and ROC-AUC of 95.3% and 0.988, respectively. The proposed method of smartphone-based screening for pharyngitis allows the fast identification of severe pharyngitis and will improve the timely diagnosis of pharyngitis for appropriate treatment to prevent complications. In addition, this framework will help patients with upper respiratory symptoms to obtain a diagnosis and treatment more conveniently and will protect them from transmission in the recent COVID-19 pandemic.

Declaration of competing interest

Ik Hee Ryu is an executive of VISUWORKS, Inc. The remaining authors declare no competing interests.

We wish to confirm that there has been no significant financial support for this work that could have influenced its outcome.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2020.103980>.

References

- [1] N. Pappot, G.A. Taarnhøj, H. Pappot, Telemedicine and e-health solutions for COVID-19: patients' perspective, *Telemed. J. E-Health Off. J. Am. Telemed. Assoc.* (2020), <https://doi.org/10.1089/tmj.2020.0099>.
- [2] S.E. Maurrasse, J.C. Rastatter, S.R. Hoff, K.R. Billings, T.S. Valika, Telemedicine during the COVID-19 Pandemic: A Pediatric Otolaryngology Perspective, *Otolaryngol.-Head Neck Surg.* 2020, <https://doi.org/10.1177/0194599820931827>.
- [3] D.C. Ertugrul, A.H. Ulusoy, Development of a knowledge-based medical expert system to infer supportive treatment suggestions for pediatric patients, *ETRI J.* 41 (2019) 515–527, <https://doi.org/10.4218/etrij.2018-0428>.
- [4] M.A. Zulkifley, S.R. Abdani, N.H. Zulkifley, Pterygium-Net: a deep learning approach to pterygium detection and localization, *Multimed. Tools. Appl.* 78 (2019) 34563–34584, <https://doi.org/10.1007/s11042-019-08130-x>.
- [5] K. de Haan, H. Ceylan Koydemir, Y. Rivenson, D. Tseng, E. Van Dyne, L. Bakic, D. Karinca, K. Liang, M. Ilango, E. Gumustekin, A. Ozcan, Automated screening of sickle cells using a smartphone-based microscope and deep learning, *Npj Digit. Med.* 3 (2020) 1–9, <https://doi.org/10.1038/s41746-020-0282-y>.
- [6] T.T. Van, K. Mata, J. Dien Bard, Automated detection of *Streptococcus pyogenes* pharyngitis by use of colorex strep A CHROMagar and WASP Lab artificial intelligence chromogenic detection module software, *J. Clin. Microbiol.* 57 (2019), <https://doi.org/10.1128/JCM.00811-19>.

- [7] N. Shaikh, E. Leonard, J.M. Martin, Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis, *Pediatr.* 126 (2010) 557–564, <https://doi.org/10.1542/peds.2009-2648>.
- [8] A. Rao, B. Berg, T. Quezada, R. Fader, K. Walker, S. Tang, U. Cowen, D. Duncan, J. Sickler, Diagnosis and antibiotic treatment of group a streptococcal pharyngitis in children in a primary care setting: impact of point-of-care polymerase chain reaction, *BMC Pediatr.* 19 (2019) 24, <https://doi.org/10.1186/s12887-019-1393-y>.
- [9] M.N. Vazquez, J.E. Sanders, Diagnosis and management of group A streptococcal pharyngitis and associated complications, *Pediatr. Emerg. Med. Pract.* 14 (2017) 1–20.
- [10] L.E. Norton, B.R. Lee, L. Harte, K. Mann, J.G. Newland, R.A. Grimes, A.L. Myers, Improving guideline-based streptococcal pharyngitis testing: a quality improvement initiative, *Pediatrics* 142 (2018), <https://doi.org/10.1542/peds.2017-2033>.
- [11] B. Askarian, S.-C. Yoo, J.W. Chong, Novel image processing method for detecting strep throat (streptococcal pharyngitis) using smartphone, *Sensors* 19 (2019), <https://doi.org/10.3390/s19153307>.
- [12] M. Janda, C. Horsham, D. Vagenas, L.J. Loescher, N. Gillespie, U. Koh, C. Curiel-Lewandrowski, R. Hofmann-Wellenhof, A. Halpern, D.C. Whiteman, J.A. Whitty, B. M. Smithers, H.P. Soyer, Accuracy of mobile digital teledermoscopy for skin self-examinations in adults at high risk of skin cancer: an open-label, randomised controlled trial, *Lancet Digit. Health* 2 (2020) e129–e137, [https://doi.org/10.1016/S2589-7500\(20\)30001-7](https://doi.org/10.1016/S2589-7500(20)30001-7).
- [13] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification, *Neurocomputing* 321 (2018) 321–331, <https://doi.org/10.1016/j.neucom.2018.09.013>.
- [14] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proc. IEEE Int. Conf. Comput. Vis.* (2017) 2223–2232.
- [15] T.K. Yoo, J.Y. Choi, H.K. Kim, CycleGAN-based deep learning technique for artifact reduction in fundus photography, *Graefes Arch. Clin. Exp. Ophthalmol.* (2020), <https://doi.org/10.1007/s00417-020-04709-5>.
- [16] R. Nakasi, E. Mwebaze, A. Zawedde, J. Tusubira, B. Akera, G. Maiga, A new approach for microscopic diagnosis of malaria parasites in thick blood smears using pre-trained deep learning models, *SN Appl. Sci.* 2 (2020) 1255, <https://doi.org/10.1007/s42452-020-3000-0>.
- [17] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G. Jamalipour Soufi, Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning, *Med. Image Anal.* (2020) 101794, <https://doi.org/10.1016/j.media.2020.101794>.
- [18] J.Y. Choi, T.K. Yoo, J.G. Seo, J. Kwak, T.T. Um, T.H. Rim, Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database, *PLoS One* 12 (2017), e0187336, <https://doi.org/10.1371/journal.pone.0187336>.
- [19] K. El Asnaoui, Y. Chawki, Using X-ray images and deep learning for automated detection of coronavirus disease, *J. Biomol. Struct. Dyn.* (2020) 1–12, <https://doi.org/10.1080/07391102.2020.1767212>.
- [20] D.S. Kermany, M. Goldbaum, W. Cai, C.C.S. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M.K. Prasadha, J. Pei, M.Y.L. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V.A.N. Huu, C. Wen, E.D. Zhang, C.L. Zhang, O. Li, X. Wang, M. A. Singer, X. Sun, J. Xu, A. Afreshi, M.A. Lewis, H. Xia, K. Zhang, Identifying medical diagnoses and treatable diseases by image-based deep learning, *Cell* 172 (2018) 1122–1131, <https://doi.org/10.1016/j.cell.2018.02.010>, e9.
- [21] T.K. Yoo, J.Y. Choi, J.G. Seo, B. Ramasubramanian, S. Selvaperumal, D.W. Kim, The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: a preliminary experiment, *Med. Biol. Eng. Comput.* 57 (2019) 677–687, <https://doi.org/10.1007/s11517-018-1915-z>.
- [22] L. Lu, B.J.D. Jr, Prognostic analysis of histopathological images using pre-trained convolutional neural networks: application to hepatocellular carcinoma, *PeerJ* 8 (2020), e8668, <https://doi.org/10.7717/peerj.8668>.
- [23] C.F. Sabottke, B.M. Spieler, The effect of image resolution on deep learning in radiography, *Radiol. Artif. Intell.* 2 (2020), e190015, <https://doi.org/10.1148/rhai.2019190015>.
- [24] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (2008) 2579–2605.
- [25] T.K. Yoo, I.H. Ryu, H. Choi, J.K. Kim, I.S. Lee, J.S. Kim, G. Lee, T.H. Rim, Explainable machine learning approach as a tool to understand factors used to select the refractive surgery technique on the expert level, *Transl. Vis. Sci. Technol.* 9 (2020), 8–8.
- [26] T. He, J. Guo, N. Chen, X. Xu, Z. Wang, K. Fu, L. Liu, Z. Yi, MediMLP: using grad-CAM to extract crucial variables for lung cancer postoperative complication prediction, *IEEE J. Biomed. Health Inform.* 24 (2020) 1762–1771, <https://doi.org/10.1109/JBHI.2019.2949601>.
- [27] R. Fluss, D. Faraggi, B. Reiser, Estimation of the Youden Index and its associated cutoff point, *Biom. J.* 47 (2005) 458–472.
- [28] T. Carneiro, R.V. Medeiros Da Nóbrega, T. Nepomuceno, G.-B. Bian, V.H.C. De Albuquerque, P.P.R. Filho, Performance analysis of Google colabatory as a tool for accelerating deep learning applications, *IEEE Access* 6 (2018) 61677–61685, <https://doi.org/10.1109/ACCESS.2018.2874767>.
- [29] M.J. Nelson, A.K. Hoover, Notes on using Google colabatory in AI education, in: *Proc. 2020 ACM Conf. Innov. Technol. Comput. Sci. Educ.*, 2020, pp. 533–534.
- [30] F. Chamberland, T. Maurina, S. Degano-Valmary, T. Spicarolen, L. Chaigneau, Angiosarcoma: a case report of gingival disease with both palatine tonsils localization, *Rare Tumors* 8 (2016) 113–117.
- [31] B. Zilberman, R. de Cleva, R.S. Testa, U. Sene, R. Eshkenazy, J.J. Gama-Rodrigues, Cervical necrotizing fasciitis due to bacterial tonsillitis, *Clinics* 60 (2005) 177–182.
- [32] E. Ribeiro, A. Uhl, G. Wimmer, M. Häfner, Exploring deep learning and transfer learning for colonic polyp classification, *Comput. Math. Methods Med.* (2016) 6584725, <https://doi.org/10.1155/2016/6584725>, 2016.
- [33] L. von Chamier, R.F. Laine, J. Jukkala, C. Spahn, M. Lerche, S. Hernández-Pérez, P. K. Mattila, E. Karinou, S. Holden, A.C. Solak, A. Krull, T.-O. Buchholz, F. Jug, L. A. Royer, M. Heilemann, G. Jacquemet, R. Henriques, ZeroCostDL4Mic: an open platform to simplify access and use of deep-learning in microscopy, *bioRxiv* (2020) 2020, <https://doi.org/10.1101/2020.03.20.000133>, 03.20.000133.
- [34] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, H. Omata, Road damage detection and classification using deep neural networks with smartphone images, *Comput. Aided Civ. Infrastruct. Eng.* 33 (2018) 1127–1141, <https://doi.org/10.1111/mice.12387>.
- [35] T. Greenhalgh, G.C.H. Koh, J. Car, Covid-19: a remote assessment in primary care, *BMJ* 368 (2020), <https://doi.org/10.1136/bmj.m1182>.
- [36] S.S. Han, I. Park, S.E. Chang, W. Lim, M.S. Kim, G.H. Park, J.B. Chae, C.H. Huh, J.-I. Na, Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders, *J. Invest. Dermatol.* (2020), <https://doi.org/10.1016/j.jid.2020.01.019>, 0.
- [37] M. Petrucci, M. De Benedittis, WhatsApp: a telemedicine platform for facilitating remote oral medicine consultation and improving clinical examinations, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 121 (2016) 248–254, <https://doi.org/10.1016/j.oooo.2015.11.005>.
- [38] A. Pujari, R. Mukhija, A.B. Singh, R. Chawla, N. Sharma, A. Kumar, Smartphone-based high definition anterior segment photography, *Indian J. Ophthalmol.* 66 (2018) 1375–1376, https://doi.org/10.4103/ijo.IJO_544_18.
- [39] S. Feng, M.F. Duarte, Few-shot learning-based human activity recognition, *Expert Syst. Appl.* 138 (2019) 112782, <https://doi.org/10.1016/j.eswa.2019.06.070>.
- [40] R. Zhang, T. Che, Z. Ghahramani, Y. Bengio, Y. Song, Metagan: an adversarial approach to few-shot learning, *Adv. Neural Inf. Process. Syst.*, 2018, pp. 2365–2374.
- [41] C. Han, L. Rundo, R. Araki, Y. Furukawa, G. Mauri, H. Nakayama, H. Hayashi, Infinite brain MR images: PGGAN-based data augmentation for tumor detection, in: A. Esposito, M. Faundez-Zanuy, F.C. Morabito, E. Pasero (Eds.), *Neural Approaches Dyn. Signal Exch.*, Springer, Singapore, 2020, pp. 291–303, https://doi.org/10.1007/978-981-13-8950-4_27.
- [42] Y. Onishi, A. Teramoto, M. Tsujimoto, T. Tsukamoto, K. Saito, H. Toyama, K. Imaizumi, H. Fujita, Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks, *Int. J. Comput. Assist. Radiol. Surg.* 15 (2020) 173–178, <https://doi.org/10.1007/s11548-019-02092-z>.
- [43] C. Muramatsu, M. Nishio, T. Goto, M. Oiwa, T. Morita, M. Yakami, T. Kubo, K. Togashi, H. Fujita, Improving breast mass classification by shared data with domain transformation using a generative adversarial network, *Comput. Biol. Med.* 119 (2020) 103698, <https://doi.org/10.1016/j.combiomed.2020.103698>.
- [44] D. Lee, W.-J. Moon, J.C. Ye, Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks, *Nat. Mach. Intell.* 2 (2020) 34–42, <https://doi.org/10.1038/s42256-019-0137-x>.
- [45] T.K. Yoo, J.Y. Choi, H.K. Kim, A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease, *Comput. Biol. Med.* (2020) 103628, <https://doi.org/10.1016/j.combiomed.2020.103628>.
- [46] T.J. Nuckton, D.V. Glidden, W.S. Browner, D.M. Claman, Physical examination: Mallampati score as an independent predictor of obstructive sleep apnea, *Sleep* 29 (2006) 903–908, <https://doi.org/10.1093/sleep/29.7.903>.
- [47] S.S. Han, G.H. Park, W. Lim, M.S. Kim, J.I. Na, I. Park, S.E. Chang, Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network, *PLoS One* 13 (2018), <https://doi.org/10.1371/journal.pone.0191493>.
- [48] M.-H. Lin, P.-F. Chang, W.-K. Fong, C.-W. Yen, K.-L. Hung, S.-J. Lin, Epidemiological and clinical features of group A Streptococcus pharyngitis in children, *Acta Paediatr. Taiwanica Taiwana Er Ke Yi Xue Hui Za Zhi.* 44 (2003) 274–278.