ORIGINAL ARTICLE

Feasibility study to improve deep learning in OCT diagnosis of rare retinal diseases with few-shot classification

Tae Keun Yoo[1] · Joon Yul Choi[2] · Hong Kyu Kim[3]

Abstract

Deep learning (DL) has been successfully applied to the diagnosis of ophthalmic diseases. However, rare diseases are commonly neglected due to insufficient data. Here, we demonstrate that few-shot learning (FSL) using a generative adversarial network (GAN) can improve the applicability of DL in the optical coherence tomography (OCT) diagnosis of rare diseases. Four major classes with a large number of datasets and five rare disease classes with a few-shot dataset are included in this study. Before training the classifier, we constructed GAN models to generate pathological OCT images of each rare disease from normal OCT images. The Inception-v3 architecture was trained using an augmented training dataset, and the final model was validated using an independent test dataset. The synthetic images helped in the extraction of the characteristic features of each rare disease. The proposed DL model demonstrated a significant improvement in the accuracy of the OCT diagnosis of rare retinal diseases and outperformed the traditional DL models, Siamese network, and prototypical network. By increasing the accuracy of diagnosing rare retinal diseases through FSL, clinicians can avoid neglecting rare diseases with DL assistance, thereby reducing diagnosis delay and patient burden.

Keywords Rare diseases · Optical coherence tomography · Few-shot learning · Deep learning · Generative adversarial network

1 Introduction

In the USA, a rare disease is generally defined as a condition with a prevalence of no more than one in 1250 individuals; however, the exact prevalence rate for most of these diseases is currently not available [1]. In primary care, a lack of awareness and cognitive factors are considered to be the main reasons for frequent misdiagnosis because clinicians cannot focus on all rare diseases at the same time [2]. Rare retinal diseases affect a limited number of patients; however, they impose a significant burden on society. Most patients with such retinal diseases often encounter diagnostic delays during the screening stage. However, recent artificial intelligence-based diagnostic or screening tools have targeted diseases that have a high prevalence, including diabetic retinopathy and age-related macular degeneration [3]. Because of the lack of sufficient clinical data, it is necessary to improve the accuracy of diagnosing rare retinal diseases [4]. Optical coherence tomography (OCT) is the most important diagnostic tool for screening rare retinal and optic nerve diseases, and it uses a light wave-based mechanism to provide three-dimensional retinal structural information [5]. Since the introduction of the deep learning (DL) algorithm, automated diagnosis for detecting multiple diseases from OCT imaging has attracted considerable attention [6]. However, previous studies using OCT images have been unable to detect rare diseases.

Machine learning techniques have successfully improved clinical decision support in the field of ophthalmology [7, 8]. In particular, the recent availability of large volumes of retinal image data has enabled DL techniques to make significant contributions to diagnostic tasks [9]. However, conventional DL models are still unable to accurately extract disease characteristics from the insufficient clinical data that is available. The use of limited datasets for conventional deep learning training brings an over-fitting problem and may cause critically low classification performance in the validation set [10, 11].

✉ Tae Keun Yoo
  eyetaekeunyoo@gmail.com

1 Department of Ophthalmology, Medical Research Center, Aerospace Medical Center, Republic of Korea Air Force, 635 Danjae-ro, Sangdang-gu, Cheongju, South Korea

2 Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, OH, USA

3 Department of Ophthalmology, Dankook University Hospital, Dankook University College of Medicine, Cheonan, South Korea

As large quantities of data are labeled by clinicians, the current approaches have been limited to the few retinal diseases that have a high prevalence. These DL models may disregard the rare diseases for which they are not trained due to the lack of sufficient labeled data [12]. However, humans can learn new disease categories using a few characteristic images that are available. To accurately detect rare diseases using an automated system, this gap between humans and DL needs to be bridged. Recently, few-shot learning (FSL), which is a new research area in the field of machine learning, has been receiving increasing attention because it requires a limited amount of data for pattern extraction similar to human experts [13]. After the introduction of generative adversarial network (GAN) for data augmentation, the performance of FSL was significantly improved due to the generation of synthetic images [14]. This GAN-based FSL technique provides an intuitive solution for utilizing conventional DL methods that have generally been used for large databases.

Recently, few-shot learning techniques have been adopted to diagnose rare diseases. Parbhu et al. showed that a prototypical network, which is a metric learning technique, is effective for dermatological disease diagnosis using real-world imbalanced datasets [15]. Quellec et al. used a similar metric learning technique using the K-nearest neighbor to classify fundus photographs with rare diseases [12]. Few-shot metric learning using Siamese networks has been used to detect plant diseases with very small datasets [16]. A gradient-based meta-learning approach has been used to improve diagnostic performance with a few-shot skin disease dataset [17]. Burlina et al. demonstrated the feasibility of using low-shot learning based on automated data augmentation to classify fundus photographs with rare conditions [18]. Several researchers have utilized generative models to enlarge training datasets in order to improve the detection accuracy of diseases using very small datasets [19, 20]. Few-shot learning based on data augmentation has also been used to detect pathological chest images of patients with COVID-19 [21]. These previous studies demonstrated that few-shot learning techniques could achieve reliable performance and outperform classical machine learning models when using small training datasets.

To the best of our knowledge, no study has been conducted on detecting rare diseases using the concept of FSL with OCT. Therefore, the purpose of this study is to build a convolutional neural network (CNN) model to detect rare diseases using OCT images. Because limited training data is available on rare retinal diseases, our approach was based on FSL using GAN-based data augmentation. In particular, the cycle-consistent GAN (CycleGAN) was adopted to generate images without matching paired images. CycleGAN is a type of unsupervised machine learning model used for mapping different image domains, and it has demonstrated reliable performance in various academic fields. We conducted experiments to evaluate the qualitative effectiveness of our method and to validate this

technique. We also compared the proposed method with other well-known few-shot learning techniques.

2 Methods

This study was conducted using a publicly accessible OCT image database obtained from a previous study by Kermany [6] and additional anonymized OCT images of rare retinal diseases collected by the authors. Figure 1 illustrates the FSL methods used in our study. Our proposed method (Fig. 1(b)) involves transfer learning with GAN-based augmentation, which comprises two stages: (1) development of CycleGAN models for each rare disease for few-shot OCT image augmentation and (2) fine-tuning training and validation of the DL classification model. The backbone DL models for transfer learning were pretrained using the ImageNet database.

2.1 Data collection

Figure 2 shows the data distribution and typical OCT images of the major and rare diseases considered in this study. The large database obtained from Kermany's previous study (https://data.mendeley.com/datasets/rscbjbr9sj/2) consists of OCT images showing the characteristics of a normal retina as well as that of major retinal diseases [6], including diabetic macular edema [22], drusen [23], and choroidal neovascularization [23], which are considered to be highly prevalent diseases. This database was collected from various eye hospitals and includes labeling data confirmed by expert ophthalmologists. The detailed diagnosis procedure is described in Kermany's original work [6]. Additional retinal image datasets were extracted from Google Images and Google search engine by searching for keywords such as central serous chorioretinopathy, macular telangiectasia, macular hole, Stargardt disease, and retinitis pigmentosa. These rare diseases were selected based on a previous review on OCT diagnosis [24]. According to the Orphan database, central serous chorioretinopathy [25], Stargardt disease [26], and retinitis pigmentosa [27] are considered as rare retinal diseases [28]. Because macular telangiectasia [29] and macular hole [30] also have very low prevalence, it is reasonable to consider them as relatively rare diseases. The images showing the characteristics of these rare diseases were manually classified by two board-certified ophthalmologists with prior knowledge about data sources and related documents, and the ambiguous images were isolated to clarify the image domains. Since the OCT images fitted perfectly with the typical characteristics of each disease, OCT examination was sufficient to diagnose rare diseases in the present study. There was no disagreement between the two ophthalmologists. The OCT images with rare retinal diseases collected by our team are available at the Mendeley Data repository
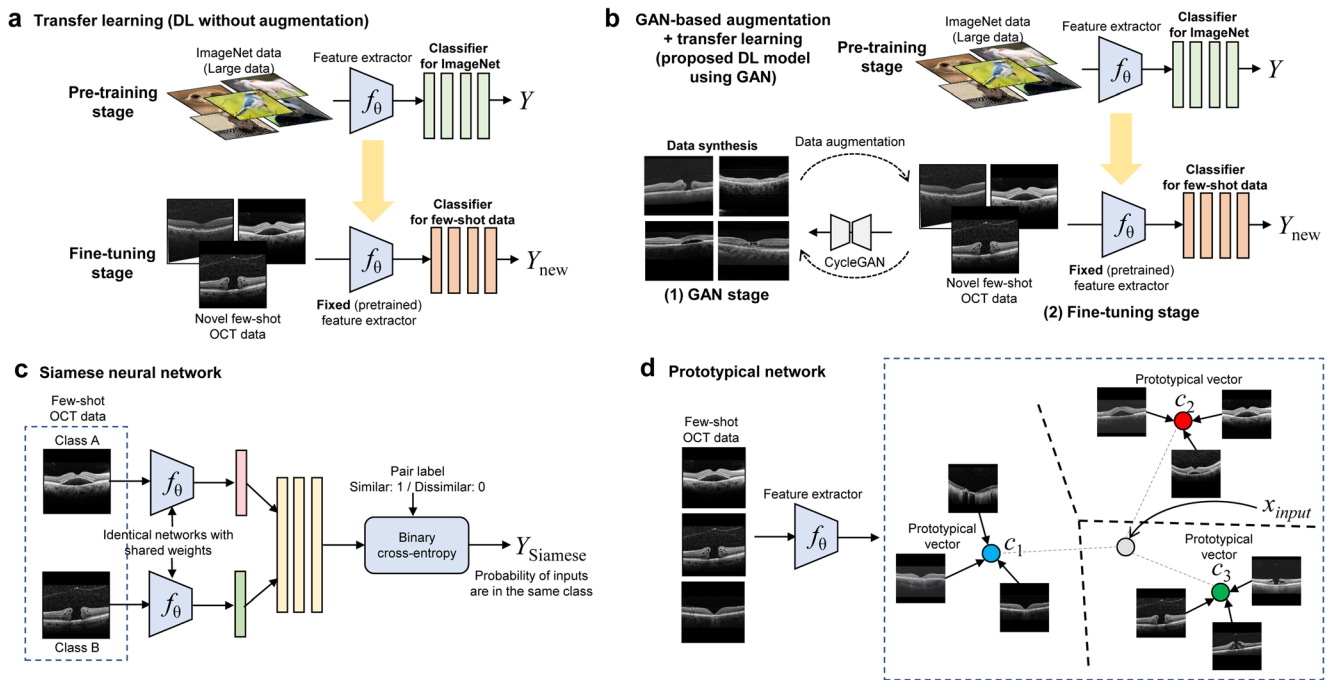
Fig. 1 Few-shot learning techniques for rare disease OCT diagnosis in the present study. **a** Deep learning model using transfer learning without augmentation. **b** Transfer learning model with data augmentation based on generative adversarial networks (GANs). **c** Metric learning model using a Siamese neural network. **d** Metric learning model using a prototypical network

(https://data.mendeley.com/datasets/btv6yrdbmv). The detailed links of the collected OCT image sources are listed in Supplementary Materials.

2.2 Characteristics of the datasets

Table 1 shows the OCT characteristics and epidemiologic data of retinal diseases. The initial training dataset contained a total of nine classes, including 26,860 normal retinas, 11,348 diabetic macular edema, 8616 drusen, 37,205 choroidal neovascularization, 25 central serous chorioretinopathy, 20 macular telangiectasia, 25 macular hole, 15 Stargardt disease, and 12 retinitis pigmentosa images. The aim of extracting these extremely imbalanced datasets was to diagnose rare retinal

diseases using the FSL framework. For the test dataset, we collected 250 normal retinas (sampled from the original test dataset to balance the major classes), 250 diabetic macular edema, 250 drusen, 250 choroidal neovascularization, 5 central serous chorioretinopathy, 4 macular telangiectasia, 5 macular hole, 4 Stargardt disease, and 4 retinitis pigmentosa datasets. The training and test datasets were split randomly, and they exhibited no overlap.

2.3 Few-shot image translation using CycleGAN

FSL learns new patterns from a limited number of training datasets. There are mainly three popular categories of FSL, namely meta-learning, metric learning, and augment-based
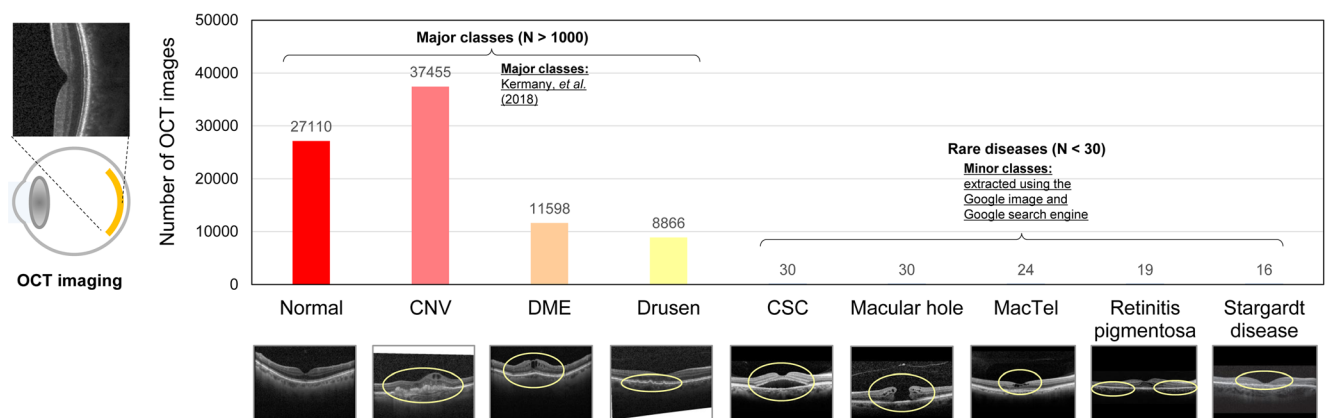

Fig. 2 Datasets pertaining to the optical coherence tomography (OCT) images of major and rare diseases considered in the present study

Table 1 Optical coherence tomography characteristics and prevalence of retinal diseases in the present study

Disease	OCT Characteristic	Prevalence (per 10,000 individuals)	Rare disease category
CNV	Disruption of retinal pigment epithelium, subretinal fluid, cystoid macular edema, exudation, subretinal hemorrhages or scar	2.3% among ≥70 years [23]	Major ocular disease
DME	Thickening of the fovea throughout the whole layer of the retina, cystoid changes	~3.8% among diabetes [22]	Major ocular disease
Drusen	Retinal pigment epithelium layer deformation or thickening with irregularities and undulations	4.8% among ≥50 years [23]	Major ocular disease
CSC	Subretinal fluid under the central macula, foveal distortion	0.061% [25]	Definitely rare disease*
Macular hole	Retinal layer break and tissue defect involving the fovea	0.11% [30]	Relatively rare disease
MacTel	Temporal foveal cystic pit enlargement secondary to loss of outer nuclear layer and ellipsoid zone	~0.022% [29]	Relatively rare disease
Retinitis Pigmentosa	mild inner retinal layer thinning and severe outer retinal layer thinning, cystic macular lesions	0.17% [27]	Definitely rare disease*
Stargardt disease	disruption or complete loss of both outer retinal layers at fovea, thinning of whole retinal layers	0.01% [26]	Definitely rare disease*

CNV, choroidal neovascularization; *CSC*, central serous chorioretinopathy; *DME*, diabetic macular edema; *MacTel*, macular telangiectasia; *OCT*, optical coherence tomography

* Included in the Orphanet rare disease database [28]

techniques [31]. Inspired by previous works using GAN for FSL [32, 33], we adopted CycleGAN-based augmentation for rare retinal diseases to increase the accuracy of diagnosis. CycleGAN was developed to overcome the limitation of paired data when two generators and two discriminators are used. Figure 3 shows the detailed framework of CycleGAN, which is considered to be a powerful DL technique that performs image domain transfer and face transfer. Because there is no database that includes both pathological OCT images and matched normal OCT images, supervised GAN techniques, such as conditional GAN and Pix2Pix, are not applicable in this study. CycleGAN is a type of unsupervised machine learning technique used for mapping different domains, and several researchers have already used it for few-shot and small data domain transfer [32–34]. The detailed mathematical implementation of CycleGAN is described in Supplementary Materials.

We developed CycleGAN augmentation models for each rare retinal disease (central serous chorioretinopathy, macular telangiectasia, macular hole, Stargardt disease, and retinitis pigmentosa). The major classes did not require data augmentation because they had sufficient OCT images to train conventional DL models. Each CycleGAN model was trained based on two domains, including normal retina and one specific rare disease. The few-shot OCT images with rare diseases were augmented using both linear and elastic transformations. Linear transformation included left and right flip, width and height translation from −5 to +5%, random rotation from −30° to +30°, zooming from 0 to 20%, and random brightness change from −10 to +10%. Elastic transformation was achieved using a Gaussian kernel [35]. We defined this transformation as "the basic augmentation step." In our experience, 40% of the original images with basic augmentation

should be retained for training the classifier. In this training step, 2000 normal retinal OCT images were randomly sampled from Kermany's study, and 2000 pathological images were generated by basic augmentation with few-shot samples. The five trained CycleGAN models translated normal OCT images to match the pathological images with each rare disease. Expert ophthalmologists reviewed the generated images and removed images possessing severe artifacts. A total of 5000 pathological OCT images, including 3000 CycleGAN-based and 2000 basic augmented images, were prepared for each rare disease to train the diagnostic classifier model.

To use a verified and pre-designed image generator, all the input images needed to be resized 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. To visualize the effect of CycleGAN-based augmentation, 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.

2.4 Development of CNN model

After data augmentation for rare retinal diseases, we trained the deep CNN using the Inception-v3 model, which is the most popular DL network developed by Google, to build a multi-class diagnosis model. The Inception-v3 model has been used successfully in many previous studies, demonstrating state-of-the-art performance with a saliency map [6, 9]. Figure 4 shows the training and validation processes. The first validation scheme involved fivefold cross-validation using the entire dataset including training and test datasets (Fig. 4(a)). In
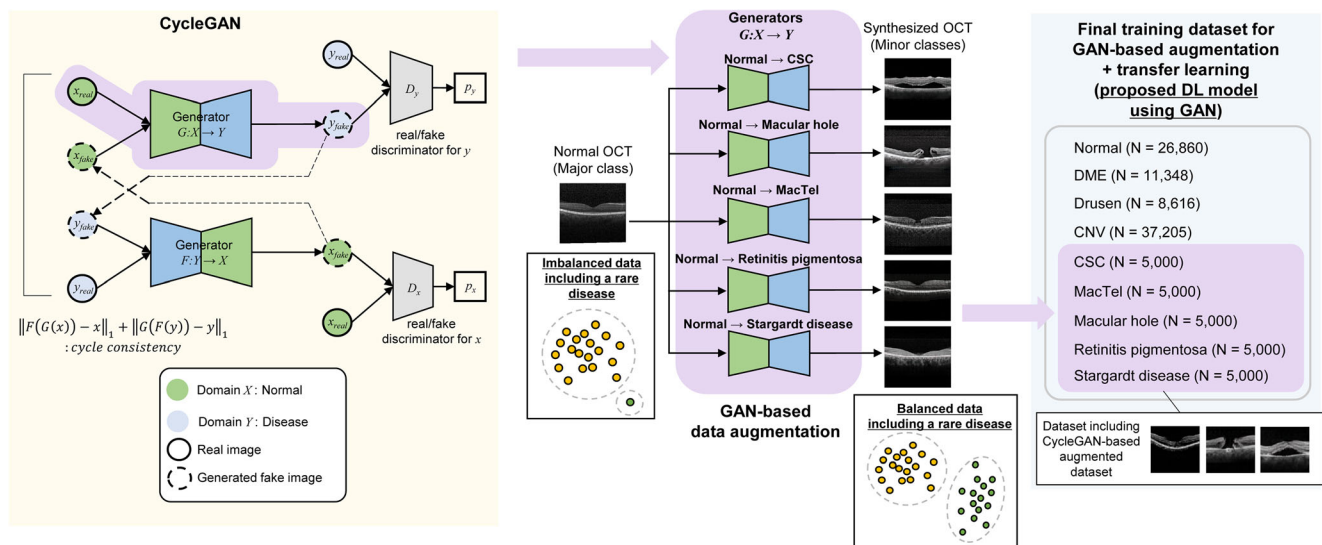
Fig. 3 CycleGAN-based augmentation for rare diseases and image classification processes. The CycleGAN model was trained using the few-shot rare disease OCT image, generating new pathological OCT images with rare diseases

this scheme, even during GAN training, the verification datasets were thoroughly separated from the training sets so that the GAN models could maintain full independence of the verification sets. Because the independent test dataset for the major classes was selected from Kermany's previous work [6], the second scheme involved training the CNN model using the training set and validating it with the test dataset (Fig. 4(b)). The final training dataset for the diagnostic FLS models contained a total of nine classes (Fig. 2), including 26,860 normal retina, 11,348 diabetic macular edema, 8616 drusen, 37,205 choroidal neovascularization, 5000 central serous chorioretinopathy, 5000 macular telangiectasia, 5000 macular hole, 5000 Stargardt disease, and 5000 retinitis pigmentosa images (Fig. 3). A tenth of the training dataset was used as the validation set to estimate how well the model had been trained. We downloaded the Inception-v3 model,

which was pre-trained on the ImageNet database, and performed fine-tuning of the weights of the pre-trained networks (Fig. 1(b)). This process generally keeps 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 for the CNN model, the size of the input images for the Inception-v3 model was resized to a pixel resolution of $299 \times 299 \times 3$. The model was trained with an epoch of 250 and a batch size of 10. The ADAM optimizer was also used with a categorical cross-entropy loss. In our experiments based on transfer learning, it tuned a fully connected layer of the CNNs. The backbone convolutional layers of Inception-v3 were left frozen, and the last fully connected layer was trained using the ADAM optimizer.

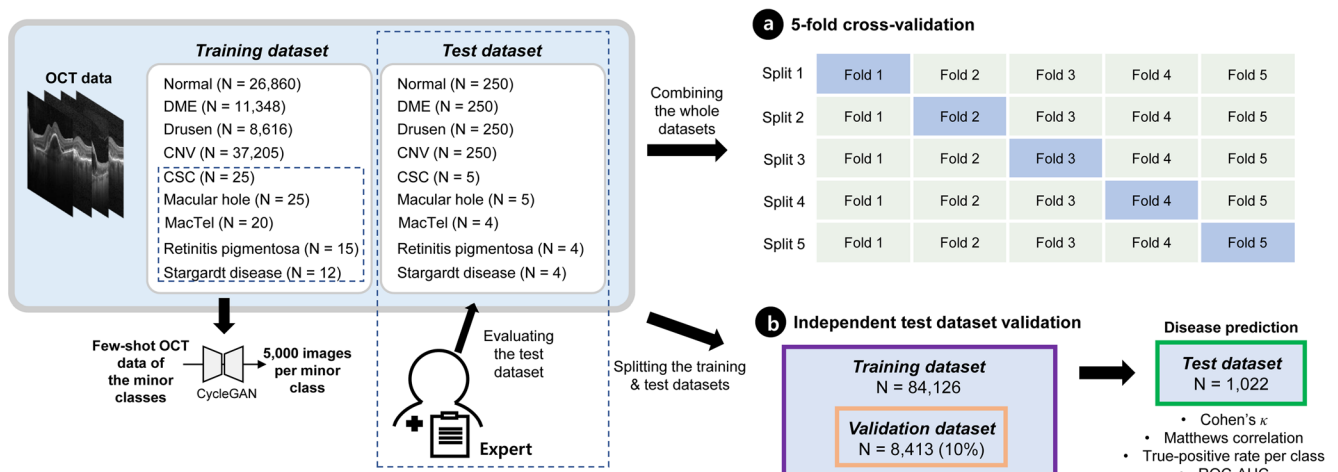Because there is a growing demand for explainable artificial intelligence methods [36], we adopted the Grad-CAM

Fig. 4 Two training and validation schemes for the deep learning model for major and rare disease classification. **a** Five-fold cross-validation using the whole data set. **b** Independent test dataset validation

technique to generate the saliency map. Grad-CAM visualizes the decisional areas of the CNN model using the gradients of any target flowing into the final convolutional network. Finally, it produces heat-maps that highlight the important area of interest and interprets the decision of the Inception-v3 model.

Google CoLab Pro, which is a cloud service for disseminating the DL research, was adopted to implement the CycleGAN and Inception-v3 models. Google CoLab Pro provides a development environment with Tensorflow-based DL libraries and a robust graphic processing unit (GPU). This enables rapid processing of a heavy DL network without the need for a personal GPU.

2.5 Other types of few-shot learning

For comparison, FSL techniques based on metric-learning were also implemented. A convolutional Siamese neural network was developed to find the relationship between two comparable classes [37]. Recently, researchers have reported that Siamese networks perform well in complicated FSL tasks with shared weights of the backbone CNN model [16]. We used Inception-v3 as identical subnetworks for the classes, and the Siamese network was designed as described in the MATLAB 2020b (MathWorks Inc., Natick, MA, USA) example (Fig. 2(c)). In this study, both the prototypical network and K-nearest neighbor learn an embedding based on the Euclidean distance to classify a new instance. To reduce the feature space dimension, we used the Inception-v3 model trained without data augmentation as a backbone CNN model for both prototypical network and K-nearest neighbor techniques. The prototypical network learns a metric space by computing the distance to the prototype representations of each class (Fig. 2(d)) [15]. We set the K value as 3 for the K-nearest neighbor model according to Quellec's work [12].

2.6 Segmentation model using a few-shot dataset

To verify that the segmentation of pathological lesions with few-shot rare disease data is possible, we built an additional segmentation CycleGAN model. The training process was based on a total of 72 ground-truth images, including the images of sampled major diseases and few-shot rare diseases. In these images, the sub-retinal fluid, intra-retinal cyst, and pigmented epithelial detachment were manually labeled by two board-certified ophthalmologists. We performed basic augmentation of these ground-truth images into 1000 images. Finally, 1000 augmented ground-truth segmentation images and 1000 randomly sampled pathological OCT images were used to train the segmentation CycleGAN model.

2.7 Statistical analysis

The main focus of this study was the accuracy of the classification model. The performance of the Inception-v3 model was evaluated based on the accuracies of the whole classes and sub-group of rare diseases. The assessment of diagnostic performance for each class was based on the area under the receiver operating characteristic curve (AUC). To establish the performance of the imbalanced classification, we calculated the unweighted Cohen's κ values, relative classifier information (RCI), and Matthews correlation coefficient from all the classes [38, 39]. To evaluate our FSL from a clinical perspective, all the OCT images in the test dataset were reviewed by an independent expert ophthalmologist who did not have any prior information about the disease names, distribution, and sources.

The basic augmentation step before training the GAN and Inception-v3 models was performed using the imageDataAugmenter and imgaussfilt functions with a Gaussian kernel (with $\sigma = 10$ and $\alpha = 2$) in MATLAB 2020b. We used CoLab's CycleGAN tutorial page to develop and validate the CycleGAN model. All these codes are available on the Tensorflow webpage (https://www.tensorflow.org/tutorials/generative/cyclegan). We modified the data input pipeline of the CycleGAN and Inception-v3 codes to import our dataset.

3 Results

3.1 CycleGAN-based augmentation for rare retinal disease

We developed our DL model using CycleGAN-based augmentation in the challenging context of few-shot OCT images for rare diseases. First, the CycleGAN models generated OCT images with rare diseases, including central serous chorioretinopathy, macular telangiectasia, macular hole, Stargardt disease, and retinitis pigmentosa, using the initial training dataset. The final CycleGAN model for each rare disease was trained for 100 epochs, which required approximately 20 h in the CoLab Pro environment. After training, randomly sampled normal OCT images were translated into pathological images for augmentation while maintaining the structures of the choroid and peripheral retina.

In the initial exploratory experiment, the number of CycleGAN-based augmented data was increased, and it yielded the highest performance at 5000 OCT images per rare disease class (2000 original images with basic augmentation and 3000 CycleGAN-based augmented images) as shown in Fig. 5. Additionally, Fig. 6 shows the acceptance rate for the synthetic OCT images used to train the deep learning model after review by the ophthalmologist. Stargardt disease and
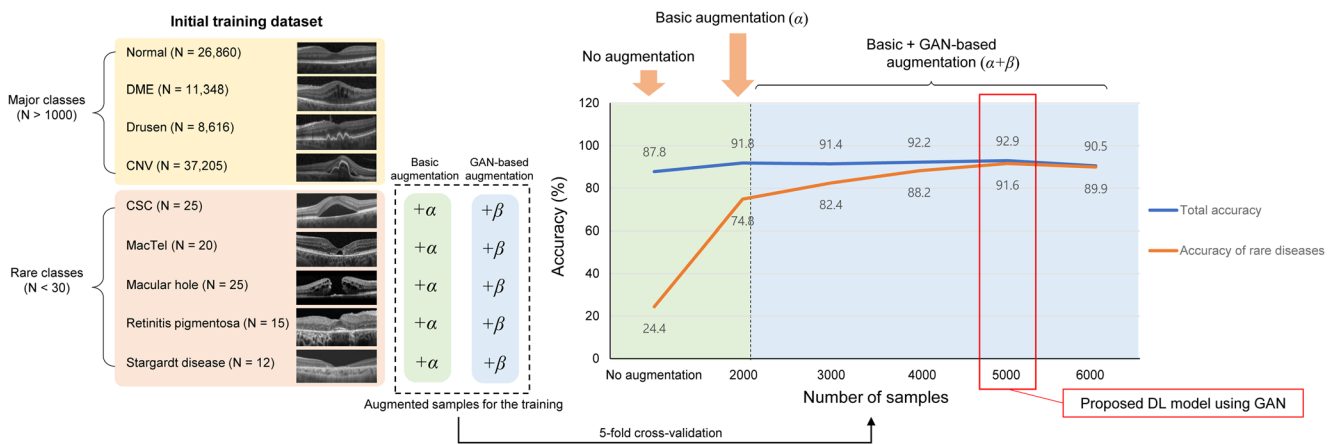
Fig. 5 Exploratory experiment for optimal data augmentation. Inception-v3 with augmentation using 5000 additional images for each rare disease (2000 original images with basic augmentation and 3000 CycleGAN-based augmented images) yielded the highest performance

retinitis pigmentosa showed higher rejection rates than the other rare diseases. The main reasons for the rejection of the synthetic images were the overlapped feature, low quality, and mode collapse. The results of the t-SNE algorithm shows that the initial data without augmentation fails to visualize the minor groups with rare diseases (Fig. 7(a)). After the CycleGAN-based augmentation for rare diseases, the minor groups were easily clustered with improved generalizability (Fig. 7(b)).

The use of CycleGAN-based synthetic images helped in the accurate extraction of the characteristic features of each rare disease, such as the sub-retinal fluid of central serous chorioretinopathy and cavitation of the inner retina in macular telangiectasia. During the image generation process, each case requires approximately 0.2 s for execution. Figure 8 shows examples of the pathological OCT images with rare diseases generated using the CycleGAN model. This feature generation based on normal OCT images can be effective for

generating new samples to increase the intra-class variation of the rare disease classes.

3.2 Performance of CNN diagnostic model

The overall classification performance of the deep learning models for the first validation scheme of the five-fold cross validation using the whole dataset is shown in Table 2, and the best performance was observed in the transfer learning with GAN-based data augmentation (proposed DL model). The multiclass metrics of overall accuracy, Cohen's κ, RCI, and Matthews correlation coefficient pertaining to the best model were 93.9%, 0.910, 0.969, and 0.911, respectively.

In the second validation scheme, the Inception-v3 model was trained using the final training dataset and validated using the test dataset. The training process required approximately 150 h for 250 epochs with fine-tuning for the proposed model. In our CycleGAN-based DL model, the accuracy of diagnosis,
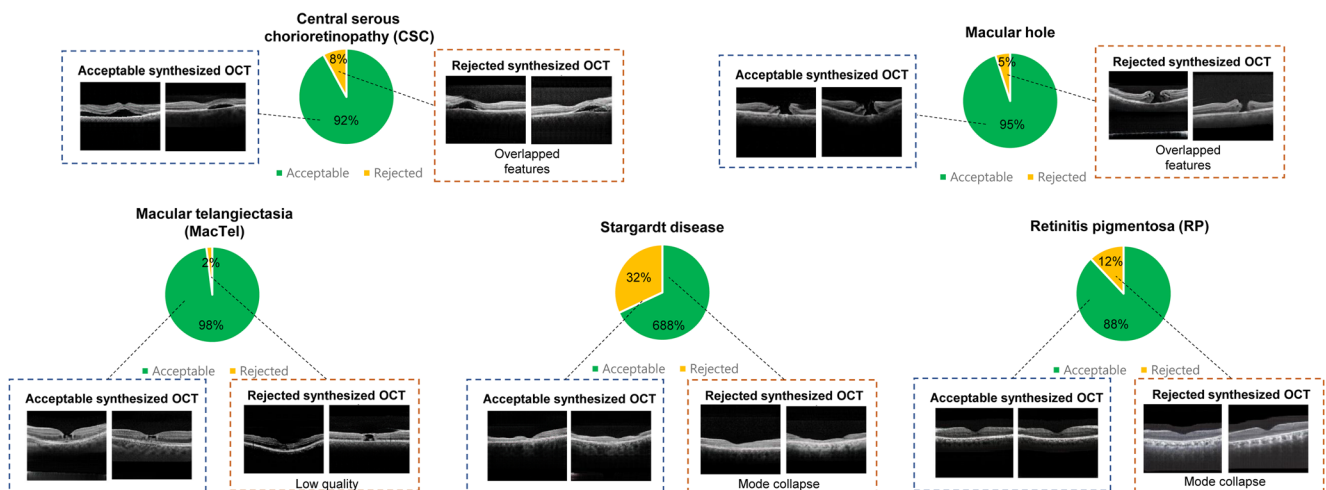

Fig. 6 The acceptance rate for synthetic OCT images to train the deep learning model. The images were generated by the highly tuned CycleGAN models for each rare disease. For each group, 100 synthesized samples were extracted randomly for evaluation by two ophthalmologists
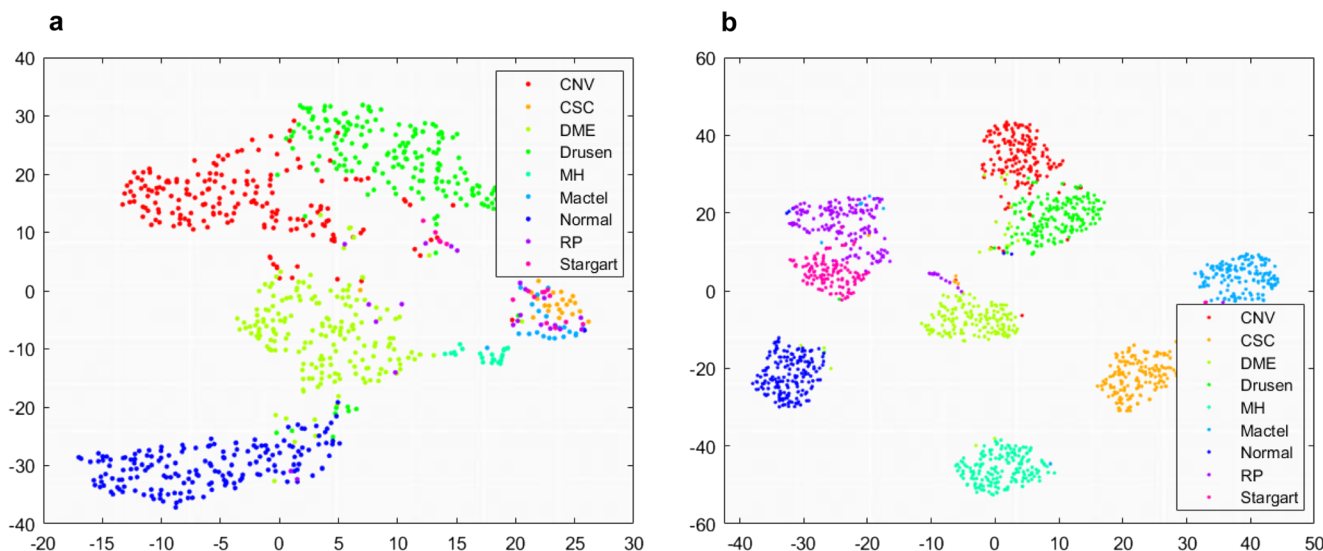
a

b

Fig. 7 The feature space visualized using the 3D t-SNE technique. **a** t-SNE visualization of transfer learning without data augmentation. **b** t-SNE visualization demonstrating the effect of the CycleGAN-based augmentation

Cohen's κ, RCI, and Matthews correlation coefficient were 92.1%, 0.896, 0.983, and 0.897 for the test dataset, respectively (Table 3). Our proposed model demonstrated superior performance in comparison with the other FSL techniques. Regarding accuracy, the Siamese network and prototypical network showed lower classification performance than the transfer learning methods. A similar tendency was observed

for Cohen's κ, RCI, and Matthews correlation coefficient values, demonstrating that our proposed model outperforms the other models in terms of multi-class classification. The accuracy of diagnosis, Cohen's κ, RCI, and Matthews correlation coefficient of the ophthalmologist without prior knowledge were 97.5%, 0.967, 0.956, and 0.968, respectively, and the diagnostic performance of the human expert was better
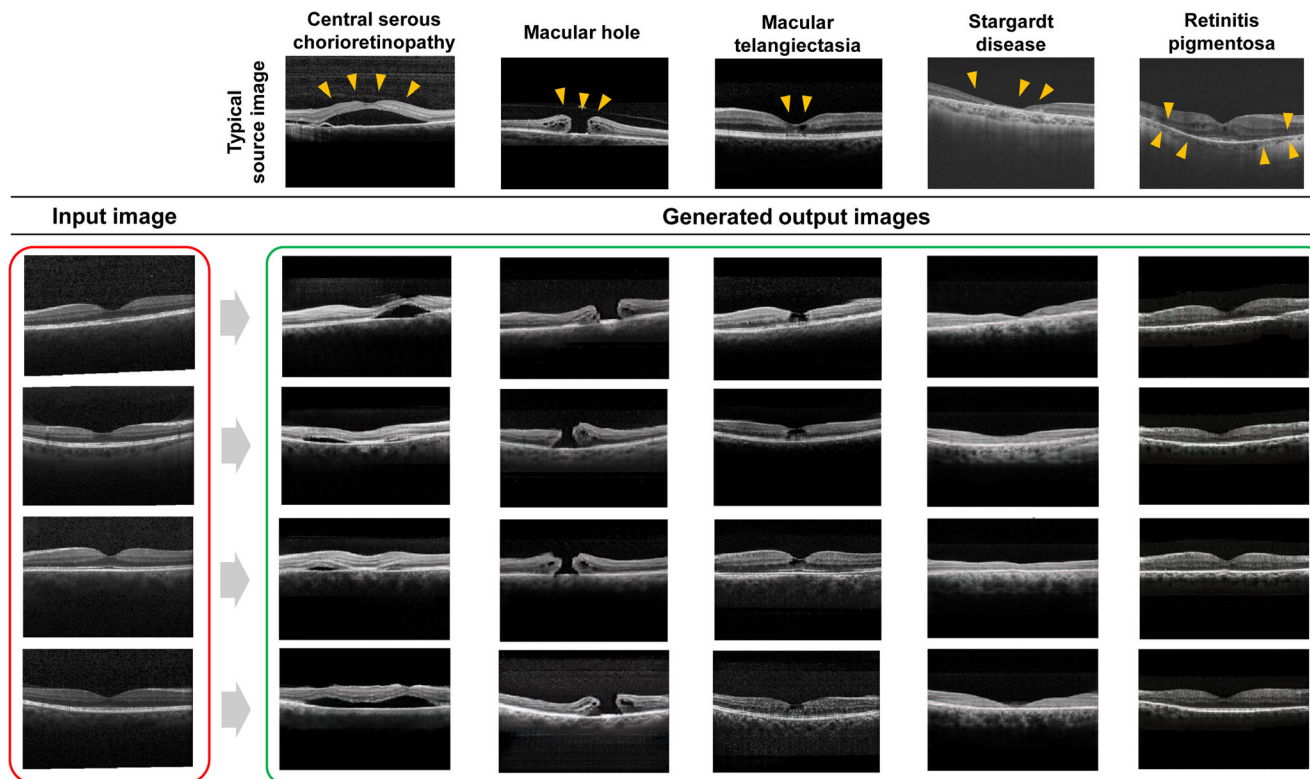

Fig. 8 Examples of pathological OCT images with rare diseases generated by the CycleGAN. The rare disease classes include central serous chorioretinopathy (CSC), macular telangiectasia (MacTel), macular hole, Stargardt disease, and retinitis pigmentosa

Table 2 Multiclass performance results pertaining to the nine-class classification of retinal diseases in the five-fold cross-validation using the whole data set

	Accuracy (%) (mean±standard deviation)	Unweighted Cohen's κ (mean±standard deviation)	RCI (mean±standard deviation)	Matthews correlation (mean±standard deviation)
DL without augmentation	88.3 ± 4.8	0.833 ± 0.084	0.960 ± 0.039	0.835 ± 0.081
DL with basic augmentation	91.5 ± 4.4	0.876 ± 0.070	0.965 ± 0.028	0.878 ± 0.068
Proposed DL model using GAN	93.9 ± 4.5	0.910 ± 0.065	0.969 ± 0.028	0.911 ± 0.062
Convolutional Siamese neural network	83.4 ± 4.7	0.768 ± 0.127	0.943 ± 0.038	0.773 ± 0.122
Prototypical network	80.4 ± 5.1	0.748 ± 0.098	0.931 ± 0.040	0.750 ± 0.094
KNN with feature extraction ($K=3$)	90.9 ± 4.3	0.881 ± 0.071	0.966 ± 0.029	0.882 ± 0.062

DL, deep learning; *KNN*, K-nearest neighbor; *RCI*, relative classifier information

than that of the FSL models. However, Table 4 shows that the human expert conducted frequent misclassification of rare diseases, considering the true positive rates per class. The ophthalmologist's true positive rates per class for diagnosing central serous chorioretinopathy, macular hole, macular telangiectasia, retinitis pigmentosa, and Stargardt disease were 1.00, 1.00, 0.50, 0.25, and 0.50, respectively, whereas those of our proposed model were 1.00, 1.00, 1.00, 0.75, and 0.75, respectively.

The detection performance of each disease was evaluated using the receiver operating characteristic curves (Fig. 9). The AUCs of the DL models without augmentation, with only basic augmentation, and with the proposed GAN-based augmentation are not distinguishable in the major classes. In the detection of rare diseases, the individual performance of the DL models showed a significant improvement with our proposed GAN-based augmentation. We also generated a saliency map using the Grad-CAM technique by successfully visualizing the characteristic pathological features for the predicted evidence (Fig. 10).

Additionally, we performed experiments to evaluate the dataset imbalance using the test dataset. After GAN-based data augmentation, under-sampling was performed by random selection to control the data distribution. Figure 11 shows that

controlling the distribution of the dataset did not have a significant impact on the classification results after data augmentation. The MobileNet-v2 and ResNet models demonstrated similar classification performance to that of the Inception-v3 model, which is used in this study (Supplementary Materials).

3.3 Additional experiments using other types of GAN models

Because our method requires a limited amount of data to train the CycleGAN model, it is expected to be highly applicable in the segmentation of OCT images of rare diseases. To determine the feasibility of our approach in a segmentation task, we also trained the CycleGAN model using 72 manually segmented OCT images and 1000 normal images (Fig. 12(a)). By considering the mean Dice score, data augmentation using 50 ground truth segmentation images could generate enhanced OCT images highlighting the pathological features with the mean Dice score of 0.784 (Fig. 12(b)). Although the training dataset includes few-shot ground-truth segmentation images of rare diseases, the results indicate that the pathological features, such as sub-retinal fluid, intra-retinal cyst, and pigmented epithelial detachment, were segmented

Table 3 Multiclass performance results pertaining to the nine-class classification of retinal diseases in the independent test dataset validation

	Accuracy (%)	Unweighted Cohen's κ	RCI	Matthews correlation
DL without augmentation	88.4	0.847	0.916	0.848
DL with basic augmentation	91.3	0.886	0.953	0.887
Proposed DL model using GAN	92.1	0.896	0.983	0.897
Convolutional Siamese neural network	80.9	0.755	0.915	0.756
Prototypical network	80.8	0.753	0.933	0.755
KNN with feature extraction ($K=3$)	91.2	0.885	0.972	0.886
Human expert*	97.5	0.967	0.956	0.968

DL, deep learning; *KNN*, K-nearest neighbor; *RCI*, relative classifier information

* Categorization was performed by an expert ophthalmologist without previous knowledge of the probable disease

Table 4 True positive rate per class pertaining to the nine-class classification of retinal diseases in the independent test dataset validation

	Normal	CNV	DME	Drusen	CSC	Macular hole	MacTel	Retinitis pigmentosa	Stargardt disease
DL without augmentation	0.876	0.924	0.920	0.880	0.200	0.200	0.000	0.000	0.250
DL with basic augmentation	0.932	0.940	0.888	0.924	0.800	0.800	0.250	0.250	0.500
Proposed DL model using GAN	0.912	0.948	0.932	0.888	1.000	1.000	1.000	0.750	0.750
Convolutional Siamese neural network	0.784	0.860	0.807	0.784	1.000	1.000	1.000	0.500	0.750
Prototypical network	0.872	0.916	0.832	0.820	1.000	1.000	1.000	0.750	0.750
KNN with feature extraction ($K=3$)	0.944	0.936	0.916	0.904	0.600	0.800	0.000	0.000	0.000
Human expert*	1.000	0.988	0.980	0.956	1.000	1.000	0.500	0.250	0.500

CNV, choroidal neovascularization; *CSC*, central serous chorioretinopathy; *DL*, deep learning; *DME*, diabetic macular edema; *KNN*, K-nearest neighbor; *MacTel*, macular telangiectasia; *OCT*, optical coherence tomography; *RCI*, relative classifier information

*Categorization was performed by an expert ophthalmologist without previous knowledge of the probable disease
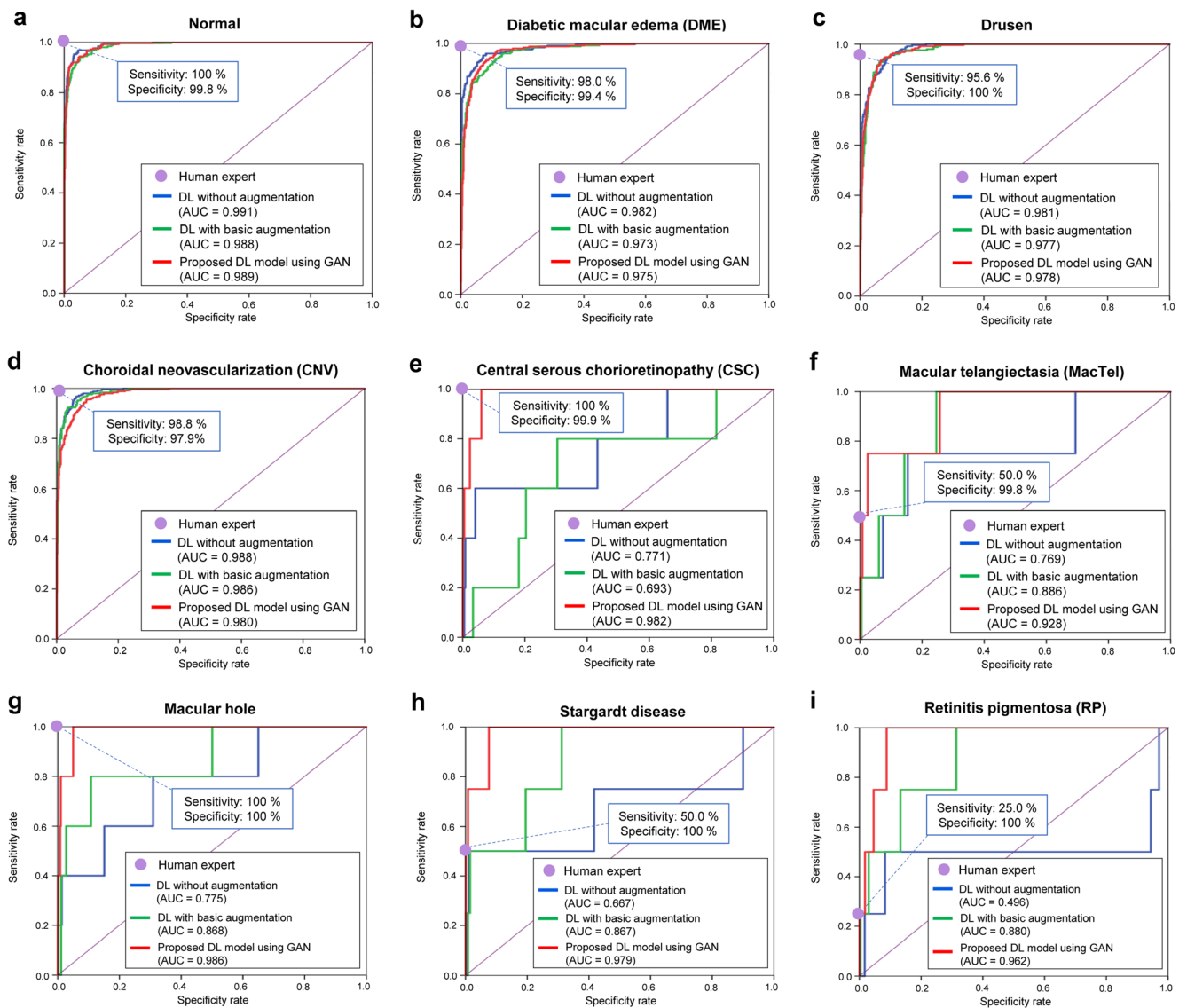

Fig. 9 Comparison of the area under the AUC values of deep learning (DL) without augmentation, DL with basic augmentation, and proposed DL with CycleGAN-based rare disease augmentation using the Inception-v3 model outputs for each disease. **a** Normal versus other classes. **b** Diabetic macular edema versus other classes. **c** Drusen versus other classes. **d** Choroidal neovascularization versus other classes. **e** Central serous chorioretinopathy versus other classes. **f** Macular telangiectasia versus other classes. **g** Macular hole versus other classes. **h** Stargardt disease versus other classes. **i** Retinitis pigmentosa versus other classes
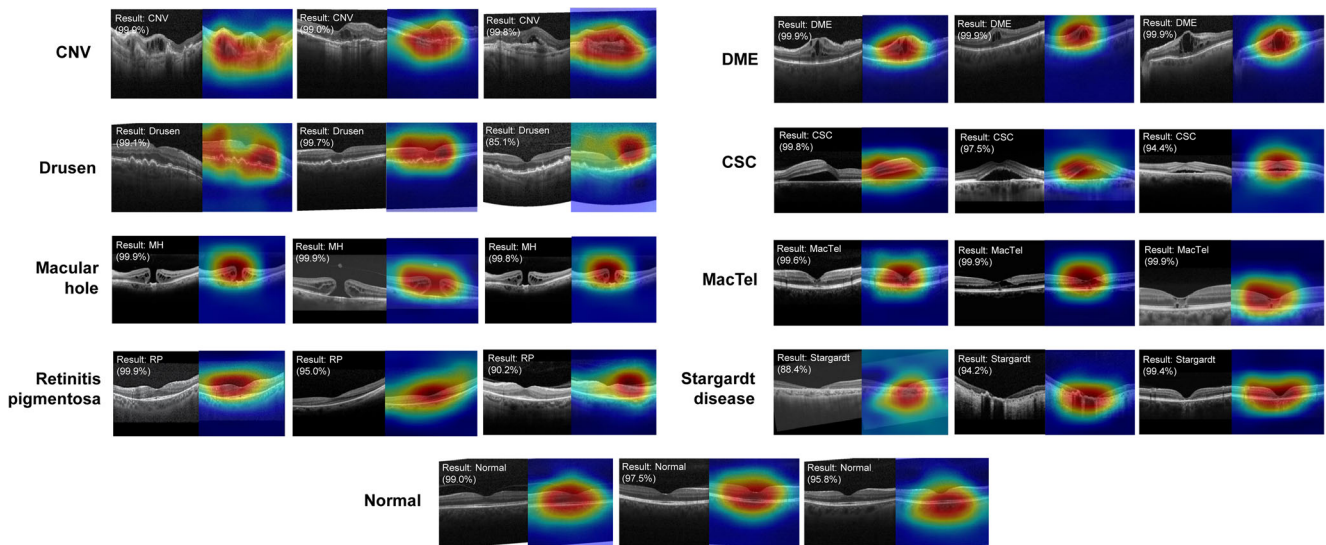
Fig. 10 Example of pathological OCT images with saliency map using the Grad-CAM technique

successfully in central serous chorioretinopathy and macular telangiectasia (Fig. 12(c)).

4 Discussion

In this study, we investigated the feasibility of DL with a GAN technique for accurately detecting rare retinal diseases using OCT images. We found that CycleGAN-based augmentation could improve the diagnostic accuracy of rare diseases using a conventional DL model with an interpretable explanation via Grad-CAM. In addition, this GAN technique can be extended to segmentation tasks using small datasets. To the best of our knowledge, this is the first experimental study to construct a few-shot DL model for OCT images considering rare disease diagnosis using GAN-based augmentation.

A recent study emphasized the large amount of OCT data required to train a DL model but did not investigate the feasibility of FSL in OCT imaging [40]. To address the limitations of traditional DL models, we first performed an experiment to explore the feasibility of FSL in the OCT imaging domain.

We found that FSL could be a valuable tool for detecting rare retinal diseases. Our FSL model using GAN-based data augmentation performed better than an expert without prior knowledge in diagnosing rare diseases considering the true positive rate per class. This result strongly illustrated the feasibility of applying FSL to improve the diagnostic accuracy of rare diseases. Because there are less noisy features compared to other image domains such as skin [15] and fundus photographs [18], OCT appears to be more suitable for image synthesis and few-shot learning. However, it is important to note that all the many synthetic images generated by the GAN models were not acceptable for use. Therefore, considerable effort and time to select acceptable images are needed to build an accurate DL model. Moreover, it will be a huge challenge to improve the diagnostic accuracy of both major and rare diseases to a very accurate level for real clinical application.

This study aimed to increase the accuracy of DL in diagnosing rare retinal diseases while maintaining the diagnostic performance for major diseases. Several previous studies have focused on building DL models for the diagnosis of rare retinal diseases, including macular hole [41], retinitis pigmentosa
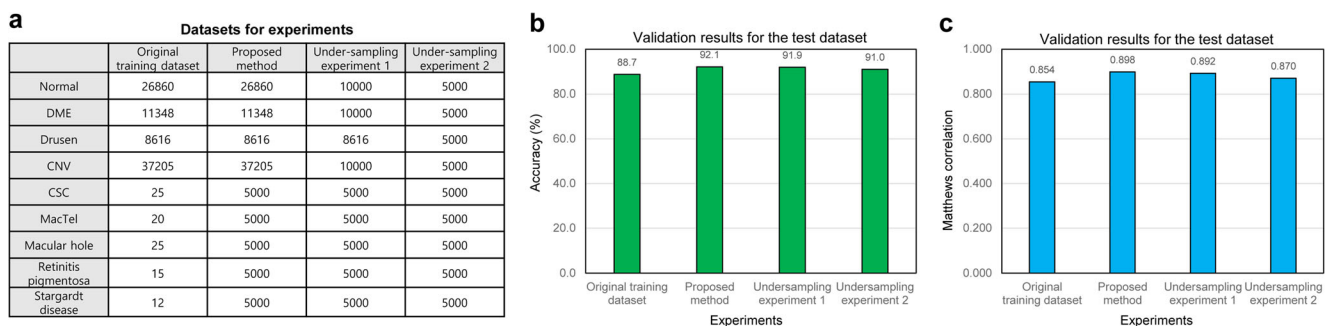

Fig. 11 Experiments to evaluate dataset imbalance using the test dataset. Under-sampling was performed by random selection. **a** The numbers of OCT images used in each experiment. **b** The validation accuracies for each experiment. **c** The validation results of Matthews correlation coefficient for each experiment
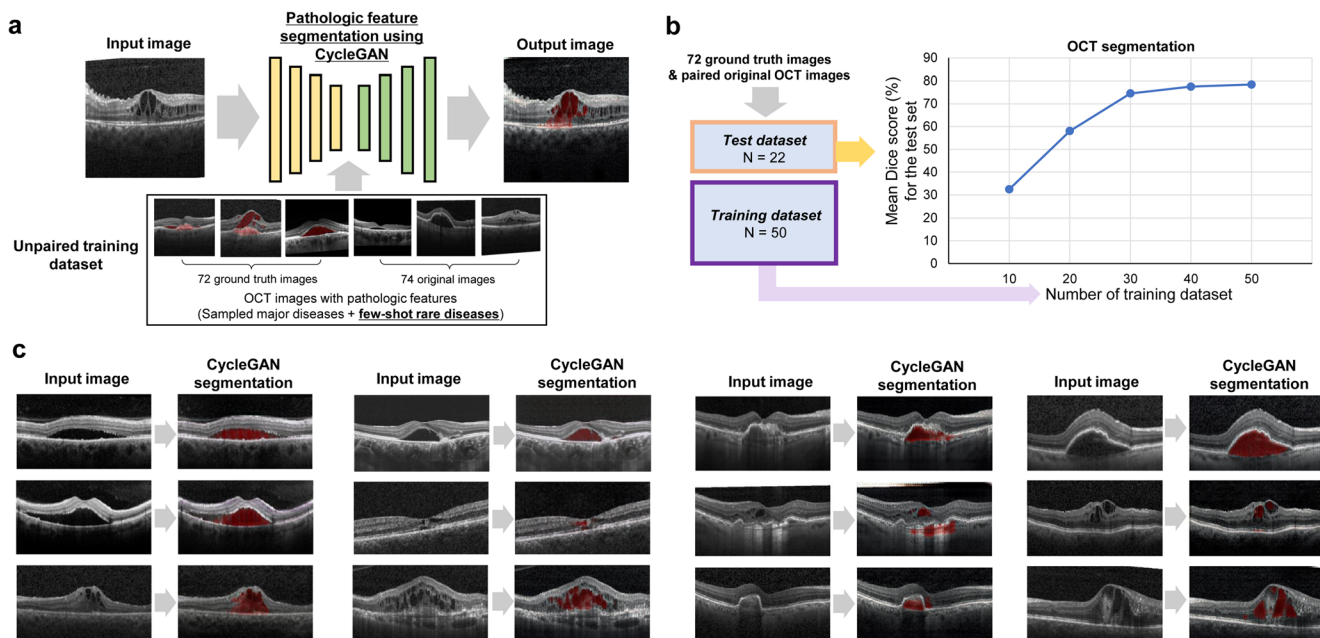
Fig. 12 Additional CycleGAN model for pathological feature segmentation; the training process is based on randomly sampled pathological images. **a** Flow-diagram of the OCT dataset and the CycleGAN model. **b** The segmentation accuracy (mean Dice score) according to the number of training images. **c** Example of pathological feature segmentation results generated by CycleGAN. Pathological images include central serous chorioretinopathy, macular telangiectasia, diabetic macular edema, and choroidal neovascularization

[42, 43], and Stargardt disease [4]. However, these DL models were designed for binary classification using normal and pathological image data. Therefore, a multiclass classification DL model is necessary to detect not only rare diseases but also major diseases such as diabetic retinopathy and age-related macular degeneration [10, 44]. One study demonstrated that CNN could classify five classes of OCT images using a large dataset without augmentation [45]. A recent study using both segmentation and multiclass classification networks improved the performance using affine and elastic transformations [35]. Another study using fundus photographs demonstrated the applicability of the FSL model based on principal component analysis and k-nearest neighbor [12]; however, this approach was limited by the lack of sufficient interpretability. This study established that the accuracy of DL models and the quality of the images generated using the few-shot setting decreases significantly with a decrease in the amount of available data. We succeeded in improving the accuracy of OCT diagnosis of rare diseases by using the GAN technique.

The main limitation of DL models in diagnosing rare retinal diseases is the inability to generalize decision boundaries from a very small number of datasets. DL using the FSL technique enables the model to learn a new task with limited information from a few instances by incorporating prior knowledge [14]. FSL relieves the burden of collecting a large amount of labeled data on rare diseases. In the medical field, FSL can learn even from extremely imbalanced disease data distribution using prior knowledge [12]. To solve this problem, several methods such as meta-learning, metric learning, and data augmentation have been proposed [31]. As most FSL methods are based on pre-trained DL networks, they generally lack interpretability regarding their operation [46]. Previous studies have demonstrated that GAN can improve FSL models by generating training situations to learn better decision boundaries between categories [14]. Recent studies using CT and MRI datasets have shown that the GAN-based data augmentation technique significantly improves the performance of machine learning models [47, 48]. GAN has also been successfully applied to cancer cell classification with insufficient training data [49]. CycleGAN has been used to improve the breast mass classification accuracy using a small dataset [50]. Consistent with previous studies using GAN-based augmentation, the accuracy of diagnosing rare retinal diseases was significantly improved using the CycleGAN model in the OCT domain.

Unlike the studies aiming at developing new GAN-based CNN models to accommodate the limited number of datasets [49], we used a standard CNN model that utilizes CycleGAN-based augmentation. This method is advantageous because researchers can easily check the output images of CycleGAN to assess the accuracy of the DL model. Synthetic OCT images can generalize rare disease classes based on a variety of normal OCT images and can guide the CNN model to avoid over-fitting to specific images [51]. In addition, the trained standard CNN model can be easily combined with Grad-CAM to improve interpretability. Previous studies have shown that CycleGAN is effective in generating synthetic images with morphologic feature transformation and

in performing the segmentation task using a small number of datasets [51, 52]. However, we established that synthetic images contain several artifacts; therefore, future studies should be directed at increasing the quality of synthetic images generated by GAN with few-shot setting. Further clinical validation of the resulting synthetic images using real-world data from clinics is also necessary.

This study has several limitations. First, the OCT images generated by the CycleGAN model have a low resolution of 256×256 pixels. This is because CycleGAN incurs a high computational cost for training networks for high-resolution applications. The low resolution may affect the classification results of the DL model [53]. Second, this study does not include a volumetric analysis for OCT. A recent study demonstrated that there is a lack of standardization in the OCT acquisition and analysis protocol [40]. Future studies should consider the variations in OCT images and devices. Third, the dataset includes a limited number of rare disease classes. Although we attempted to collect rare disease data from web-based sources, we could not include all the retinal diseases that have been reported in the existing literature. A recent study demonstrated that the conventional DL model can classify over 100 disease classes if the data is prepared for training [54]. We believe that our CycleGAN-based augmentation for rare diseases can be adopted to address similar classification problems with a large number of classes.

5 Conclusions

In summary, our DL model using GAN was useful in improving the accuracy of OCT diagnosis of rare retinal diseases while maintaining the diagnostic performance for major diseases. In particular, the CycleGAN-based augmentation was effective for the generalization of few-shot OCT images of rare diseases to avoid over-fitting. Thus, by increasing the accuracy of diagnosing rare retinal diseases via FSL, clinicians can avoid neglecting rare diseases with DL assistance, thereby reducing diagnosis delay and social burden of patients.

Declarations

References

1. Schieppati A, Henter J-I, Daina E, Aperia A (2008) Why rare diseases are an important medical and social issue. Lancet Lond Engl 371:2039–2041. https://doi.org/10.1016/S0140-6736(08)60872-7

2. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD (2019) Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet J Rare Dis 14:69. https://doi.org/10.1186/s13023-019-1040-6

3. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, Niemeijer M (2016) Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 57:5200–5206. https://doi.org/10.1167/iovs.16-19964

4. Shah M, Ledo AR, Rittscher J (2020) Automated classification of normal and Stargardt disease optical coherence tomography images using deep learning. Acta Ophthalmol (Copenh) n/a. https://doi.org/10.1111/aos.14353

5. Islam MS, Wang J-K, Johnson SS, Thurtell MJ, Kardon RH, Garvin MK (2020) A deep-learning approach for automated OCT en-face retinal vessel segmentation in cases of optic disc swelling using multiple en-face images as input. Transl Vis Sci Technol 9: 17–17. https://doi.org/10.1167/tvst.9.2.17

6. Kermany DS, Goldbaum M, Cai W, et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172:1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

7. Caixinha M, Nunes S (2017) Machine learning techniques in clinical vision sciences. Curr Eye Res 42:1–15. https://doi.org/10.1080/02713683.2016.1175019

8. Yoo TK, Ryu IH, Lee G, Kim Y, Kim JK, Lee IS, Kim JS, Rim TH (2019) Adopting machine learning to automatically identify candidate patients for corneal refractive surgery. Npj Digit Med 2:59. https://doi.org/10.1038/s41746-019-0135-8

9. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316:2402–2410. https://doi.org/10.1001/jama.2016.17216

10. Choi JY, Yoo TK, Seo JG, Kwak J, Um TT, Rim TH (2017) Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database. PLoS One 12:e0187336. https://doi.org/10.1371/journal.pone.0187336

11. Barbedo JGA (2018) Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput Electron Agric 153:46–53. https://doi.org/10.1016/j.compag.2018.08.013

12. Quellec G, Lamard M, Conze P-H, Massin P, Cochener B (2020) Automatic detection of rare pathologies in fundus photographs

using few-shot learning. Med Image Anal 61:101660. https://doi.org/10.1016/j.media.2020.101660

13. Feng S, Duarte MF (2019) Few-shot learning-based human activity recognition. Expert Syst Appl 138:112782. https://doi.org/10.1016/j.eswa.2019.06.070

14. Zhang R, Che T, Ghahramani Z et al (2018) Metagan: an adversarial approach to few-shot learning. Advances in Neural Information Processing Systems, In, pp 2365–2374

15. Prabhu V, Kannan A, Ravuri M, et al (2018) Prototypical clustering networks for dermatological disease diagnosis. ArXiv181103066 Cs

16. Argüeso D, Picon A, Irusta U, Medela A, San-Emeterio MG, Bereciartua A, Alvarez-Gila A (2020) Few-shot learning approach for plant disease classification using images taken in the field. Comput Electron Agric 175:105542. https://doi.org/10.1016/j.compag.2020.105542

17. Mahajan K, Sharma M, Vig L (2020) Meta-DermDiagnosis: few-shot skin disease identification using meta-learning. Pp 730–731

18. Burlina P, Paul W, Mathew P, Joshi N, Pacheco KD, Bressler NM (2020) Low-shot deep learning of diabetic retinopathy with potential applications to address artificial intelligence bias in retinal diagnostics and rare ophthalmic diseases. JAMA Ophthalmol 138:1070–1077. https://doi.org/10.1001/jamaophthalmol.2020.3269

19. Zhong F, Chen Z, Zhang Y, Xia F (2020) Zero- and few-shot learning for diseases recognition of Citrus aurantium L. using conditional adversarial autoencoders. Comput Electron Agric 179:105828. https://doi.org/10.1016/j.compag.2020.105828

20. Yoo TK, Choi JY, Jang Y, Oh E, Ryu IH (2020) Toward automated severe pharyngitis detection with smartphone camera using deep learning networks. Comput Biol Med 125:103980. https://doi.org/10.1016/j.compbiomed.2020.103980

21. Lai Y, Li G, Wu D, Lian W, Li C, Tian J, Ma X, Chen H, Xu W, Wei J, Zhang Y, Jiang G (2020) 2019 Novel coronavirus-infected pneumonia on CT: a feasibility study of few-shot learning for computerized diagnosis of emergency diseases. IEEE Access 8:194158–194165. https://doi.org/10.1109/ACCESS.2020.3033069

22. Varma R, Bressler NM, Doan QV, Gleeson M, Danese M, Bower JK, Selvin E, Dolan C, Fine J, Colman S, Turpcu A (2014) Prevalence of and risk factors for diabetic macular edema in the United States. JAMA Ophthalmol 132:1334–1340. https://doi.org/10.1001/jamaophthalmol.2014.2854

23. Jonasson F, Arnarsson A, Sasaki H, Peto T, Sasaki K, Bird AC (2003) The prevalence of age-related maculopathy in Iceland: Reykjavik eye study. Arch Ophthalmol Chic Ill 1960 121:379–385. https://doi.org/10.1001/archopht.121.3.379

24. Murthy RK, Haji S, Sambhav K, Grover S, Chalam KV (2016) Clinical applications of spectral domain optical coherence tomography in retinal diseases. Biom J 39:107–120. https://doi.org/10.1016/j.bj.2016.04.003

25. Rim TH, Kim HS, Kwak J, Lee JS, Kim DW, Kim SS (2018) Association of corticosteroid use with incidence of central serous chorioretinopathy in South Korea. JAMA Ophthalmol 136:1164–1169. https://doi.org/10.1001/jamaophthalmol.2018.3293

26. Bitner H, Schatz P, Mizrahi-Meissonnier L, et al (2012) Frequency, genotype, and clinical spectrum of best vitelliform macular dystrophy: data from a National Center in Denmark. Am J Ophthalmol 154:403-412.e4. https://doi.org/10.1016/j.ajo.2012.02.036

27. Sen P, Bhargava A, George R, Ramesh SV, Hemamalini A, Prema R, Kumaramanickavel G, Vijaya L (2008) Prevalence of retinitis pigmentosa in south Indian population aged above 40 years. Ophthalmic Epidemiol 15:279–281. https://doi.org/10.1080/09286580802105814

28. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, le Cam Y, Rath A (2020) Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet 28:165–173. https://doi.org/10.1038/s41431-019-0508-0

29. Aung KZ, Wickremasinghe SS, Makeyeva G et al (2010) The prevalence estimates of macular telangiectasia type 2: the Melbourne collaborative cohort study. RETINA 30:473–478. https://doi.org/10.1097/IAE.0b013e3181bd2c71

30. Williams RE, Beeby M, Logie J (2012) Prevalence of diagnosed macular hole, macular pucker, vitreomacular adhesions/traction, retinal tear/detachment, and pterygium in US health care claims databases. Invest Ophthalmol Vis Sci 53:5221–5221

31. Wang Y, Yao Q (2019) Few-shot learning: a survey. ArXiv Prepr ArXiv190405046

32. Zhang F, Zhang T, Mao Q, Xu C (2018) Joint pose and expression modeling for facial expression recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, In, pp 3359–3368

33. Yu Y, Liu G, Odobez J-M (2019) Improving few-shot user-specific gaze adaptation via gaze redirection synthesis. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, In, pp 11937–11946

34. Liu M-Y, Huang X, Mallya A et al (2019) Few-shot unsupervised image-to-image translation. Proceedings of the IEEE International Conference on Computer Vision, In, pp 10551–10560

35. De Fauw J, Ledsam JR, Romera-Paredes B et al (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 24:1342–1350. https://doi.org/10.1038/s41591-018-0107-6

36. Yoo TK, Ryu IH, Choi H, Kim JK, Lee IS, Kim JS, Lee G, Rim TH (2020) 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 9:8–8

37. Figueroa-Mata G, Mata-Montero E (2020) Using a convolutional Siamese Network for image-based plant species identification with small datasets Biomim Basel Switz 5. https://doi.org/10.3390/biomimetics5010008

38. Yoo TK, Choi JY, Seo JG, Ramasubramanian B, Selvaperumal S, Kim DW (2019) 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 57:677–687. https://doi.org/10.1007/s11517-018-1915-z

39. Gorodkin J (2004) Comparing two K-category assignments by a K-category correlation coefficient. Comput Biol Chem 28:367–374. https://doi.org/10.1016/j.compbiolchem.2004.09.006

40. Yanagihara RT, Lee CS, Ting DSW, Lee AY (2020) Methodological challenges of deep learning in optical coherence tomography for retinal diseases: a review. Transl Vis Sci Technol 9:11–11. https://doi.org/10.1167/tvst.9.2.11

41. Nagasawa T, Tabuchi H, Masumoto H, Enno H, Niki M, Ohsugi H, Mitamura Y (2018) Accuracy of deep learning, a machine learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting idiopathic macular holes. PeerJ 6:e5696. https://doi.org/10.7717/peerj.5696

42. Masumoto H, Tabuchi H, Nakakura S, Ohsugi H, Enno H, Ishitobi N, Ohsugi E, Mitamura Y (2019) Accuracy of a deep convolutional neural network in detection of retinitis pigmentosa on ultrawide-field images. PeerJ 7:e6900. https://doi.org/10.7717/peerj.6900

43. Wang Y-Z, Galles D, Klein M, Locke KG, Birch DG (2020) Application of a deep machine learning model for automatic measurement of EZ width in SD-OCT images of RP. Transl Vis Sci Technol 9:15–15. https://doi.org/10.1167/tvst.9.2.15

44. Arunkumar R, Karthigaikumar P (2017) Multi-retinal disease classification by reduced deep learning features. Neural Comput Appl 28:329–334. https://doi.org/10.1007/s00521-015-2059-9

45. Lu W, Tong Y, Yu Y, Xing Y, Chen C, Shen Y (2018) Deep learning-based automated classification of multi-categorical

abnormalities from optical coherence tomography images. Transl Vis Sci Technol 7:41–41. https://doi.org/10.1167/tvst.7.6.41

46. Xian Y, Sharma S, Schiele B, Akata Z (2019) F-VAEGAN-D2: a feature generating framework for any-shot learning. Pp 10275–10284

47. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H (2018) GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 321:321–331. https://doi.org/10.1016/j.neucom.2018.09.013

48. Han C, Rundo L, Araki R, Furukawa Y, Mauri G, Nakayama H, Hayashi H (2020) Infinite brain MR images: PGGAN-based data augmentation for tumor detection. In: Esposito A, Faundez-Zanuy M, Morabito FC, Pasero E (eds) Neural approaches to dynamics of signal exchanges. Springer, Singapore, pp 291–303

49. Rubin M, Stein O, Turko NA, Nygate Y, Roitshtain D, Karako L, Barnea I, Giryes R, Shaked NT (2019) TOP-GAN: stain-free cancer cell classification using deep learning with a small training set. Med Image Anal 57:176–185. https://doi.org/10.1016/j.media.2019.06.014

50. Muramatsu C, Nishio M, Goto T, Oiwa M, Morita T, Yakami M, Kubo T, Togashi K, Fujita H (2020) Improving breast mass classification by shared data with domain transformation using a generative adversarial network. Comput Biol Med 119:103698. https://doi.org/10.1016/j.compbiomed.2020.103698

51. Sandfort V, Yan K, Pickhardt PJ, Summers RM (2019) Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Sci Rep 9:16884. https://doi.org/10.1038/s41598-019-52737-x

52. Yoo TK, Choi JY, Kim HK (2020) A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease. Comput Biol Med 103628:103628. https://doi.org/10.1016/j.compbiomed.2020.103628

53. Yip MYT, Lim G, Lim ZW, Nguyen QD, Chong CCY, Yu M, Bellemo V, Xie Y, Lee XQ, Hamzah H, Ho J, Tan TE, Sabanayagam C, Grzybowski A, Tan GSW, Hsu W, Lee ML, Wong TY, Ting DSW (2020) Technical and imaging factors influencing performance of deep learning systems for diabetic retinopathy. Npj Digit Med 3:1–12. https://doi.org/10.1038/s41746-020-0247-1

54. Han SS, Park I, Chang SE, et al (2020) Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders. J Invest Dermatol 0: https://doi.org/10.1016/j.jid.2020.01.019

Tae Keun Yoo is an ophthalmologist and a researcher of the research center in Aerospace Medical Center, Korean Air Force, and works as an ophthalmologist at Aerospace Medical Center. He received his diploma in Mechanical Engineering from Seoul National University and received the Doctor of Medicine from Yonsei University. His research interest concerns ophthalmology and machine learning.

Joon Yul Choi is currently doing his post-doc at Epilepsy Center in Cleveland Clinic. He received Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, Korea. His research interest focuses on deep learning and MRI data analysis.

Hong Kyu Kim is a retinal specialist, and is a professor of the department of ophthalmology, Dankook University Hospital, Dankook University College of Medicine. He received the Doctor of Medicine from Yonsei University and is a PhD student at Yonsei University. His areas of research interest are big data and data analysis.