Medical & Biological Engineering & Computing (2019) 57:677–687
https://doi.org/10.1007/s11517-018-1915-z

ORIGINAL ARTICLE

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

Tae Keun Yoo[1] · Joon Yul Choi[2] · Jeong Gi Seo[1] · Bhoopalan Ramasubramanian[3] · Sundaramoorthy Selvaperumal[3] · Deok Won Kim[4]

Received: 11 December 2017 / Accepted: 9 October 2018 / Published online: 22 October 2018

Abstract

Recently, researchers have built new deep learning (DL) models using a single image modality to diagnose age-related macular degeneration (AMD). Retinal fundus and optical coherence tomography (OCT) images in clinical settings are the most important modalities investigating AMD. Whether concomitant use of fundus and OCT data in DL technique is beneficial has not been so clearly identified. This experimental analysis used OCT and fundus image data of postmortems from the Project Macula. The DL based on OCT, fundus, and combination of OCT and fundus were invented to diagnose AMD. These models consisted of pre-trained VGG-19 and transfer learning using random forest. Following the data augmentation and training process, the DL using OCT alone showed diagnostic efficiency with area under the curve (AUC) of 0.906 (95% confidence interval, 0.891–0.921) and 82.6% (81.0–84.3%) accuracy rate. The DL using fundus alone exhibited AUC of 0.914 (0.900–0.928) and 83.5% (81.8–85.0%) accuracy rate. Combined usage of the fundus with OCT increased the diagnostic power with AUC of 0.969 (0.956–0.979) and 90.5% (89.2–91.8%) accuracy rate. The Delong test showed that the DL using both OCT and fundus data outperformed the DL using OCT alone (P value < 0.001) and fundus image alone (P value < 0.001). This multimodal random forest model showed even better performance than a restricted Boltzmann machine (P value = 0.002) and deep belief network algorithms (P value = 0.042). According to Duncan's multiple range test, the multimodal methods significantly improved the performance obtained by the single-modal methods. In this preliminary study, a multimodal DL algorithm based on the combination of OCT and fundus image raised the diagnostic accuracy compared to this data alone. Future diagnostic DL needs to adopt the multimodal process to combine various types of imaging for a more precise AMD diagnosis.

Keywords Age-related macular degeneration · Multimodal deep learning · OCT · Fundus photograph

1 Introduction

Age-related macular degeneration (AMD) is the most common form of progressive retinal disorder that causes visual impairment in developed countries [1]. Patients with early AMD have the drusen and retinal pigment epithelial abnormality. Progression of AMD with neovascularization and geographic atrophy of retina may be associated with visual loss. It can be challenging to establish the clinical diagnosis of AMD for patients with decreased visual acuity. Classic diagnostic and grading system for AMD were based on color fundus images [2]. Deep learning for fundus image classification has been recently developed [3]. This technique was able to detect subtle pathologic lesions of retina with high accuracy [4], and it has been applied to diagnose and

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

✉ Tae Keun Yoo
eyetaekeunyoo@gmail.com

[1] Institute of Vision Research, Department of Ophthalmology, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea

[2] Department of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea

[3] Department of Electronics and Communication Engineering, Syed Ammal Engineering College, Ramanathapuram, Tamil Nadu, India

[4] Department of Medical Engineering Seoul, South Korea, Yonsei University College of Medicine, Seoul, South Korea

grade AMD using the National Institutes of Health AREDS dataset [5]. Diagnostic accuracy was reported to be nearly 90%, which was comparable to human physicians [6]. Deep learning also showed a good performance to predict exudative AMD with the wide-field fundus images [7]. An ensemble model using 6 different popular deep learning techniques predicted the 13 classes of AMD severity and outperformed human graders [8]. However, diagnosis using fundus images without other types of examination is limited due to its two-dimensional nature.

Optical coherence tomography (OCT) is the light wave-based diagnostic tool providing 3-dimensional structural information, analogous to ultrasonography showing cross-sectional images. OCT has proved useful in evaluating retinal pathology in difficult diagnostic cases obtaining cross-sectional lesion with neovascularization and surrounding tissue. Nowadays, OCT is routinely used in clinic practice to determine the activity of AMD [9]. The current deep learning technique classified retinal OCT images with good performance for AMD and diabetic macular edema diagnosis [10]. Deep learning model using OCT images from AMD patients showed the good diagnostic accuracy [11]. Particularly, exudative AMD was classified by a deep learning method in a recent study [12]. Other deep learning research group goes one step farther to choose the patient needed to be treated with anti-vascular endothelial factor injection [13]. Based on large OCT data, transfer learning using pre-trained deep learning was applied to classify diabetic macular edema, drusen, and advanced AMD [14]. Retinal segmentation and feature extraction using deep learning were also successfully applied to diagnose AMD [15].

Generally, ophthalmologists consider two image domains including fundus and OCT to diagnose AMD in the clinic. However, to our knowledge, the diagnostic value of the combination of OCT and fundus images has not been emphasized previously in the literature studying machine learning. The purpose of this experimental research was to investigate the value of adding fundus imaging to OCT imaging when distinguishing between normal and AMD.

2 Methods

2.1 Image dataset

It is difficult to acquire OCT and fundus data from real AMD patients due to the absence of open database including both OCT and matched fundus images. Therefore, publically available OCT and retinal image database at the Project Macula (available at http://projectmacula.cs.uab.edu) was used in order to evaluate the combination of fundus and OCT imaging [16]. The Project Macula aimed to develop to investigate age-related macular degeneration (AMD) in patients and in postmortem specimens. This dataset provides the diagnosis of AMD confirmed by pathohistological examination.

Fundus photography is referred to as the process of acquiring a two-dimensionally projected representation of the three-dimensional retina by means of reflected light. Whereas a cross-sectional visualization of OCT provides excellent detail for evaluating the vitreo-retinal interface, neurosensory retinal morphology, and the retinal layer complex. Figure 1 demonstrates the differences of image acquisition position and intensity distribution between fundus and OCT images. When subretinal fluid or inner-retinal hemorrhage develops in an AMD eye, fundus images can represent three-channel color changes at the lesion, and OCT images can show a disruption of retinal layers and retinal thickness changes in an intensity map.

The initial dataset included three categories: normal (50 eyes), non-neovascular AMD (19 eyes), and neovascular AMD (40 eyes). The two ophthalmologists (T.K.Y and J.G.S) manually reviewed and cropped fundus images to remove the artifacts such as alignment pin and scale around fundus image. OCT raster scans were obtained by using a SD-OCT system with image averaging (Spectralis, Heidelberg-Engineering, Heidelberg, Germany). Geographic atrophy was excluded due to the small dataset (3 eyes). The authors reviewed all OCT scan images for each patient to determine the quality of scans. Since the OCT and fundus images were irregular in size and showed very poor quality, the two ophthalmologists (T.K.Y and J.G.S) manually selected a square region of interest including foveal retina in each image, which is the most important lesion of macula pathophysiology. After a quality check, we used 27 normal, 18 non-neovascular AMD, and 38 neovascular AMD eyes for analysis. Figure 2 presents examples of each disease status. The experimental process was in compliance with the Declaration of Helsinki. This study did not need ethics committee approval; instead, researchers used public database.

Data augmentation was performed by increasing data by oversampling images with translation, rotation, brightness change, and additive Gaussian noise in an effort to train deep learning models properly. Data augmentation is a widely used approach to improve generalization of deep learning models [17]. We randomly retrieved transformed 1000 fundus images per each disease class (total 3000 fundus images) because of the imbalanced data problems. The OCT images, which matched to fundus images, were also augmented (total 3000 OCT images). We obtained samples with translation with ranges of [− 10%, + 10%] of the image width, with rotation [− 15°, + 15°], and with brightness change with ranges of [− 10%, + 10%]. Additive Gaussian noise has a uniformly collected sigma from [0, 0.04]. All images including OCT and fundus images were organized according to the size of the input pre-trained model (224 × 224 pixels) in the course of oversampling.

2.2 Description of the automated method

We used the deep convolutional neural network (CNN) model with 19 layers (VGG-19) along with MatConvNet

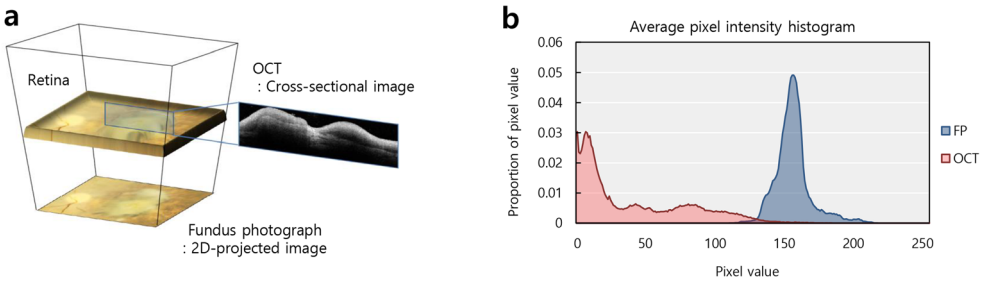

Fig. 1 The difference between OCT and fundus images. **a** Illustration of OCT scan location and fundus photography scan location. **b** The histogram of fundus photography and OCT image pixel values

(available at http://www.vlfeat.org/matconvnet). VGG-19 consisted of 16 convolutional layers and 3 fully connected layers [18]. A machine learning model can put accumulated knowledge into a new task domain by applying a transfer learning technique [10]. Transfer learning by using pre-trained CNN enables to avoid problems in relation to a few datasets in medicine. We used VGG-19 pre-trained with images from ImageNet. When the CNN served as a feature extractor, multiclass random forest (RF) models were operated by utilizing the 4096 input features from the last covered layer of the pre-trained VGG-19 model [3].

We compared five different deep learning models (Fig. 3): (1) transfer learning with RF based on OCT imaging alone; (2) transfer learning with RF based on fundus imaging alone; (3) transfer learning with RF based on the combination of OCT and fundus imaging; (4) transfer learning with multimodal restricted Boltzmann machine (RBM) [19]; (5) transfer learning with multimodal deep belief network (DBN) [20]. The current popular deep CNN techniques widely used did not provide the training and validation schemes that combine two different imaging domains simultaneously. Therefore, we combined the VGG-19 feature vectors from each OCT and fundus image using RF, RBM, and DBN. RF is an ensemble technique based on bootstrap and random selection of features. It refers to a robust and powerful multiclass classification [21]. In the multimodal RF setting, local features including OCT and fundus images are integrated into a z-score normalization process in each image modality as follows:

$$I_{\text{multimodal}} = \left(\frac{V_{\text{OCT}} - \mu_{\text{OCT}}}{\sigma_{\text{OCT}}}, \frac{V_{\text{fundus}} - \mu_{\text{fundus}}}{\sigma_{\text{fundus}}} \right)$$

Where $I_{\text{multimodal}}$ is an input vector (including 8192 feature values), and V is the feature vector (including 4096 feature values) from pre-trained VGG-19. In addition, μ and σ were the mean and the standard deviation of the given feature vector in each image modality. Finally, an input channel of multimodal RF comprised of output feature layers of VGG-19, which were individually computed by using OCT and fundus image. To fully utilize the information in each image of two examinations, the model should learn to aggregate features from all of those images. Then the model is supposed to complete the recognition with the use of the aggregated features. RBM consisted of hidden, visible layer, and bias nodes. The connections between the visible and hidden layers are undirected and fully connected. RBM can help to determine appropriate initial parameters. DBN is an architecture stacking multiple RBM. There was an unsupervised layer-wise pre-training followed by supervised fine tuning using the gradient descent method in the training process. For multiclass classification, one-versus-all technique was adopted for RBM and DBN. In RBM and DBN learning, we updated the parameters with a learning rate of 0.001 and momentum of 0.1 for 1000 epochs. Different sets of parameters (grid search) for the tunable variables were evaluated to identify the optimal set of parameters.

MATLAB 2017a (Mathworks, Natick, MA, USA) was adopted to perform the algorithms. MedCalc 12.3 (MedCalc, Mariakerke, Belgium) was used to conduct an analysis of the Receiver Operating Characteristic (ROC) curve. The NVIDIA GEFORCE GTX1060 3GB GPU for transfer learning and GTX980 6GB for SGD with Intel core i7 processor were employed to train deep learning models more rapidly.

2.3 Statistical evaluation

The measurement of classification problems was based on area under the curve (AUC), accuracy, and relative classifier information (RCI) [22]. The AUC is the most commonly used

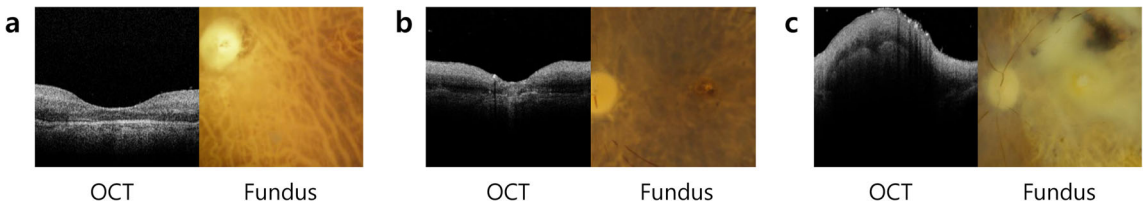

Fig. 2 Examples of OCT and fundus images. **a** The normal eye. **b** The eye with non-neovascular age-related macular degeneration. **c** The eye with neovascular age-related macular degeneration. The image sources are available in the Project MACULA website (http://projectmacula.cis.uab.edu)

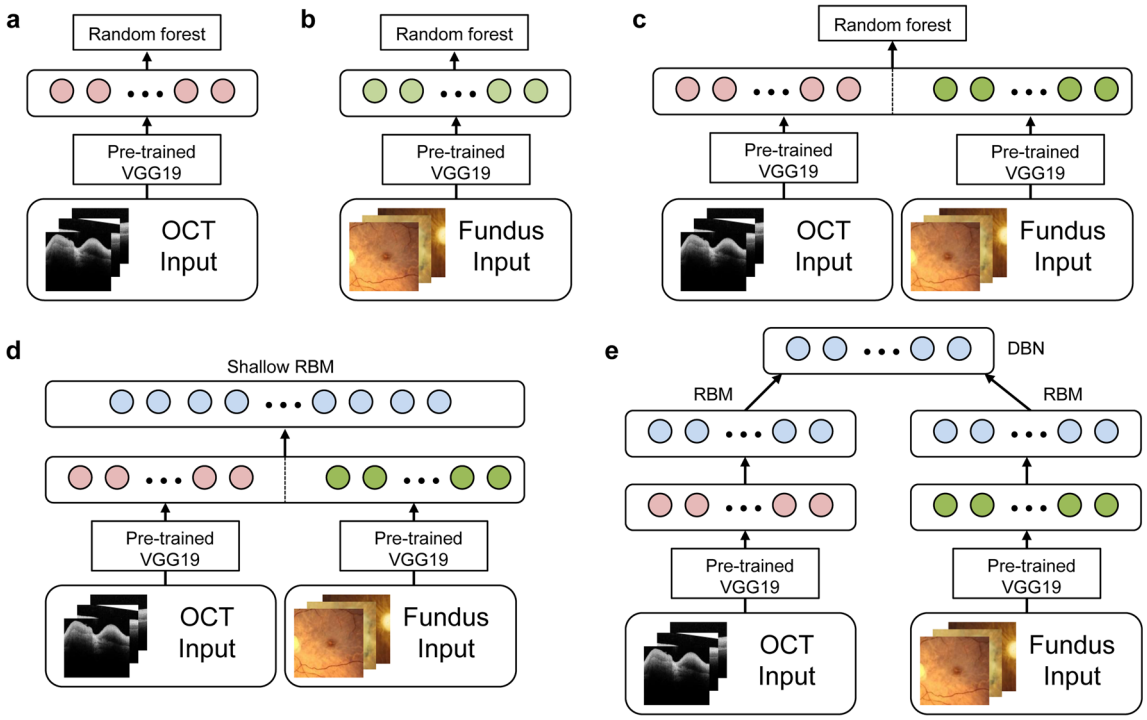

Fig. 3 Illustration of deep learning models in this study. **a** Transfer learning based on random forest using OCT image alone. **b** Transfer learning based on random forest using fundus image alone. **c** Transfer learning based on random forest using the combination of OCT and fundus image. **d** Multimodal restricted Boltzmann machine. **e** Multimodal deep belief network

index of binary diagnostic performance. We calculated the AUC by trapezoidal integration of the ROC curve. Accuracy is a standard metric for evaluation of a classifier. It is defined as follows:

$$\text{Accuracy} = \frac{\sum_i q_{ii}}{\sum_{ij} q_{ij}}$$

where the element q_{ij}, refers to the number of test, and test input actually labeled C_i is C_j noted by the classifier. These elements organize the confusion matrix. However, the classification details of how the samples of each class are separated from that of the others cannot be revealed by the accuracy [23]. Therefore, several studies have used the measure of RCI to evaluate multiclass classifier performances [3, 24]. RCI is defined as follows:

$$RCI = \sum_i -\frac{\sum_j q_{ij}}{\sum_{ij} q_{ij}} log\left(\frac{\sum_j q_{ij}}{\sum_{ij} q_{ij}}\right) - \sum_j \left(\frac{\sum_i q_{ij}}{\sum_{ij} q_{ij}} \times \sum_i -\frac{q_{ij}}{\sum_i q_{ij}} log\left(\frac{q_{ij}}{\sum_i q_{ij}}\right)\right)$$

RCI represents the performance of unbalanced classes capable of distinguishing among different misclassified distributions. In the multi-class problems, the average of the AUCs on the pairwise binary classification derived was also evaluated. The pairwise binary classifications were performed on the one-versus-all and one-versus-one approach [25]. This process needed additional training for calculating each AUC of pairwise binary classifications. When we trained binary classifiers, we maximized Youden's index [26]. Youden's index J is defined as follows:

$$J = Sensitivity + Specificity - 1$$

This value is used frequently to determine the optimal cut-off values of the diagnostic tests in imbalanced dataset.

Comparison between AUCs used the non-parametric empirical method of Delong [27]. When a cross-validation was performed, Duncan's multiple range test was available to obtain detailed information about the differences between classifiers. This test presented the subsets of adjacent means that are different or not with a given level of significance ($\alpha < 0.05$ in this study) [28].

2.4 Additional experiments

Generally, routine clinical assessments using fundus photograph and OCT were conducted by physician's judgments based on detection of the segmented abnormal lesion (hemorrhage, exudates, and drusen) and retinal thickness of OCT. To show the effectiveness of the proposed method, we compared the proposed models with three classic methods close to routine clinical assessments including regression, detection of abnormal lesions, and measurement of retinal thickness. A regression using the least absolute shrinkage and selection operator (LASSO) was used to aggregate features from pre-

trained VGG-19. Recently, LASSO has been reported as the best regression method to handle clinical data features [29]. Fundus segmentation using geometric background leveling and threshold selection was performed to detect drusen, exudates, and hemorrhage according to Smith's method [30]. An artificial neural network with two hidden layers containing 100 and 50 neurons was trained using these segmentation results to predict AMD. Measurement of retinal thickness was conducted manually at the fovea, since foveal thickness well reflects the severity of AMD [31]. Additionally, the proposed multimodal models were compared to the Inception-v3 using fundus imaging alone, which is the most popular convolutional deep learning architecture [32]. The stochastic gradient descent was employed to train the inception-v3 model. The inception-v3 model was built in the Tensorflow framework using the open source GitHub package.

To validate the proposed models in a clinical setting, other experiments were performed using non-postmortem OCT and fundus dataset. Images including both OCT and matched fundus photograph was crawled from Google Image Search using keywords related to normal retina, AMD, and OCT. The dataset was built by harvesting images manually for normal and AMD categories. Finally, 44 normal eyes and 75 AMD eyes including both OCT and matched fundus images were obtained. The proposed methods and inception-v3 were evaluated using the leave-one-out cross-validation without data augmentation.

3 Experimental results

The grid search demonstrated that 500 trees and 3 predictors for each node for RF were found to be optimal for OCT images. A thousand of trees and five predictors of RF were set for fundus images and the multimodal architecture. In RBM, we set the number of hidden units 500 for optimal performance. Four layers of 1000 hidden units determined the optimal model of DBN. More than four hidden unit layers of DBN showed no appreciable improvement in performance.

The five-fold cross-validation binary classification (normal vs AMD) performance of deep learning models is summarized in Table 1. When we assigned a diagnosis based on OCT image alone, the AUC and accuracy of deep learning model were 0.906 (95% confidence interval [CI], 0.891–0.921) and 82.6% (95% CI, 81.0–84.3%), respectively. When we applied fundus image alone for the diagnosis of deep learning, the AUC and accuracy were 0.914 (95% CI, 0.900–0.928) and 83.5% (95% CI, 81.8–85.0%), respectively. When deep learning evaluated the combined fundus and OCT images, the AUC and accuracy turned out to be 0.969 (95% CI, 0.956–0.979) and 90.5% (95% CI, 89.2–91.8%), respectively. The Delong test showed that the multimodal RF using both OCT and fundus data outperformed the single-modal RF using OCT alone (P value < 0.001) and fundus image alone (P value < 0.001). Multimodal RBM and DBN models showed the AUC of 0.940 (95% CI, 0.926–0.953) and 0.956 (95% CI, 0.943–0.968), respectively. Duncan's multiple range test showed two significant subgroups under AUCs. The subgroup of multimodal methods using both OCT and fundus data performed better than that of single-modal methods.

Figure 4 shows ROC curves of the deep learning models of each training scheme, when datasets were divided into training (70%) and test (30%) dataset. The ROC curves showed that the deep learning combined with fundus and OCT imaging was more efficient than fundus imaging alone and OCT imaging alone. Table 2 demonstrates the detailed performance of proposed methods and three classic methods. The multimodal RF model showed the AUC, accuracy, sensitivity, and specificity of 0.981 (95% CI, 0.938–0.967), 94.6% (95% CI, 92.9–96.0%), 95.5% (95% CI, 93.5–97.0%), and 92.7% (95% CI, 89.1–95.4%), respectively. Considering AUC as a performance metric, the multimodal RF model was not significantly superior to the multimodal RBM model (P value = 0.294), whereas it showed a statistically significant difference compared to the multimodal DBN (P value = 0.008), single-modal RF using OCT alone (P value < 0.001), and RF using fundus image alone (P value = 0.001). Three classic methods including LASSO regression, detection of abnormal lesions, and measurement of retinal thickness showed the AUC of 0.950 (95% CI, 0.934–0.963), 0.911 (95% CI, 0.891–0.929), and 0.827 (95% CI 0.801–0.851), respectively.

In the problem of multi-class AMD classification (normal, non-neovascular AMD, and neovascular AMD), the classification results of deep learning models were presented in Fig. 5. When a diagnosis based on OCT image alone was assigned, the five-fold cross-validation accuracy and RCI of deep learning model were 67.3% (95% CI, 65.8–68.8%) and 0.220 (95% CI, 0.193–0.247), respectively. After fundus images were applied into the diagnosis of deep learning, the accuracy rate and RCI were 69.5% (95% CI, 68.1–71.0%) and 0.248 (95% CI, 0.227–0.269), respectively. When the multimodal RF model combined fundus and OCT images, the accuracy and RCI turned out to be 73.2% (95% CI, 71.8–74.6%) and 0.302, respectively. Multimodal RBM and DBN demonstrated the accuracy of 72.0% (95% CI, 70.6–73.4%) and 71.7% (95% CI, 70.3–73.1%), and the RCI of 0.284 (95% CI, 0.265–0.303) and 0.279 (95% CI, 0.259–0.298), respectively. The average of the AUCs on the pairwise binary classification revealed that the multimodal RF model showed the value of 0.852 (95% CI, 0.828–0.867) in the one-versus-all and 0.892 (95% CI, 0.876–0.902) in the one-versus-one setting. According to Duncan's multiple range test, the multimodal RF model significantly improved the performance obtained by the single-modal methods.

Since the small samples of the original data were analyzed, we conducted the leave-one-out cross-validation on patient

Table 1 Diagnostic performance of deep learning models using five-fold cross-validation in training set

	AUC (95% CI)	Accuracy (%) (95% CI)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	*P* value*	Duncan subgroup†
RF - OCT image alone	0.906 (0.891–0.921)	82.6 (81.0–84.3)	83.2 (81.2–85.1)	81.6 (78.5–84.4)	< 0.001	B
RF - Fundus image alone	0.914 (0.900–0.928)	83.5 (81.8–85.0)	83.4 (81.4–85.3)	83.6 (80.6–86.2)	< 0.001	B
RF - OCT+ Fundus image	0.969 (0.956–0.979)	90.5 (89.2–91.8)	91.0 (89.4–92.5)	89.6 (87.2–91.9)	–	A
RBM - OCT+ Fundus image	0.940 (0.926–0.953)	86.5 (84.9–87.9)	86.0 (84.1–87.8)	87.5 (84.7–89.8)	0.002	A
DBN - OCT+ Fundus image	0.956 (0.943–0.968)	88.9 (87.4–90.1)	88.0 (86.2–89.7)	90.5 (88.0–92.5)	0.042	A

*Comparison of ROC curves with the deep learning method using RF – OCT + Fundus image according to the Delong test

† Subgroups were separated by Duncan's multiple range test using the AUCs

AUC, area under curve; *DBN*, deep belief network; *RBM*, restricted Boltzmann machine; *RF*, random forest

level instead of image-level. During the leave-one-out cross-validation, training and test process were conducted without data augmentation. Figure 6 shows the patient-level validation results of deep learning models, which classified AMD status with patient level maintaining data augmentation during the training process. Consistent results were observed; as a result, multimodal RF model yielded an accuracy of 72.3% (95% CI, 69.8–74.5%) and RCI of 0.379 (95% CI, 0.361–0.400) in the patient level analysis. The single-modal Inception-v3 using fundus images showed the accuracy of 69.9% (95% CI, 67.5–72.1%) and RCI of 0.322 (95% CI, 0.305–0.342).

The proposed models were also evaluated in the non-postmortem fundus and OCT images of normal and AMD eyes, which were crawled from Google Image Search (Fig. 7). Therefore, training and validation were performed renewedly. In a binary classification setting, the leave-one-out cross-validation was conducted. The experiment demonstrated a consistent result with the previous finding with post-mortem data. The multimodal RF model showed the AUC and accuracy of 0.994 (95% CI, 0.959–1.000) and 97.3% (95% CI, 92.8–99.5%), respectively. Multimodal RBM and DBN demonstrated the AUC of 0.941 (95% CI, 0.882–0.975) and 0.969 (95% CI, 0.920–0.992) and accuracy of 89.2% (95% CI, 82.4–94.5%) and 94.6% (95% CI, 89.3–98.1%), respectively. The Duncan's multiple range test in AUCs showed that the multimodal RF using both OCT and fundus data performed better than the single-modal RF models.

4 Discussion

To the best of our knowledge, this is the first experimental study to construct multimodal deep learning models that consider both fundus and OCT images at the same time. Fundus images provide information on the number and area of drusen or atrophic lesion of AMD. Measurement of OCT is closely associated with the aggressiveness of subretinal and intraretinal fluid lesion. The tissue property and retinal thickness, which affected by choroidal neovascularization and leakage of vessel, can be measured by using OCT. A previous study suggested that fundus images can exhibit changes of the size of drusen and pigmentary lesion, yet incapable of

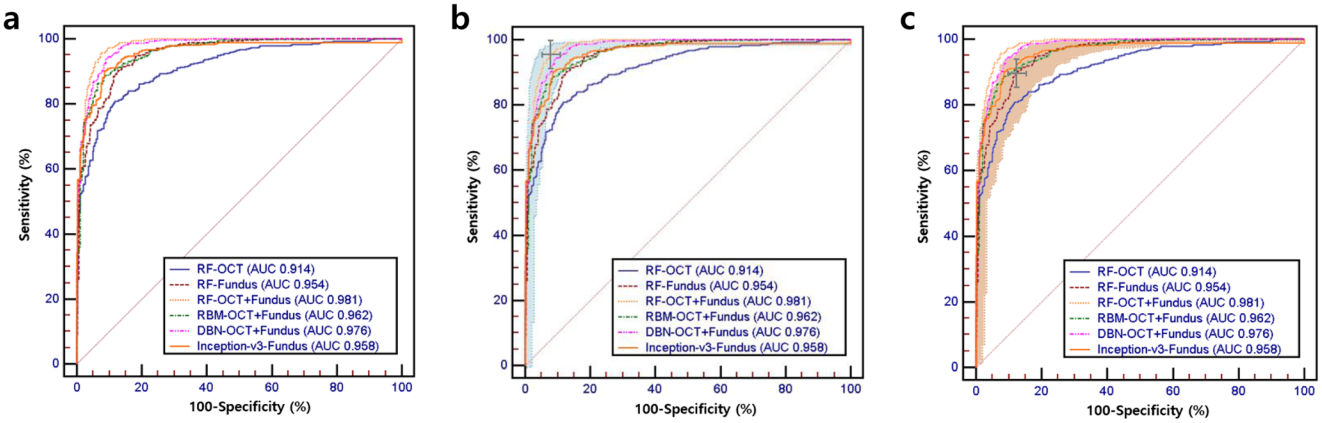

Fig. 4 Receiver operating characteristic (ROC) curves of transfer learning based on random forest (RF), multimodal restricted Boltzmann machine (RBM), and multimodal deep belief network (DBN) for the purpose of conducting binary classification (normal versus age-related macular degeneration). The convolutional deep neural network for transfer learning consisted of pre-trained VGG-19. **a** The original ROC curves. **b** ROC curves with a 95% confidence interval band of multimodal transfer learning based RF using both OCT and fundus image. **c** ROC curves with a 95% confidence interval band of transfer learning based RF using fundus image alone

Table 2 Diagnostic performance of trained deep learning models in test set

	AUC (95% CI)	Accuracy (%) (95% CI)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	*P* value*
RF - OCT image alone	0.914 (0.894–0.932)	83.3 (80.7–85.7)	80.8 (77.4–83.9)	88.3 (84.1–91.7)	< 0.001
RF - Fundus image alone	0.954 (0.938–0.967)	89.2 (87.0–91.2)	90.0 (87.3–92.3)	87.7 (83.4–91.2)	0.001
RF - OCT + Fundus image	0.981 (0.970–0.989)	94.6 (92.9–96.0)	95.5 (93.5–97.0)	92.7 (89.1–95.4)	–
RBM - OCT+ Fundus image	0.976 (0.964–0.985)	93.1 (91.3–94.7)	94.2 (92.0–95.9)	91.0 (87.2–94.0)	0.294
DBN - OCT + Fundus image	0.961 (0.947–0.973)	89.7 (87.5–91.6)	88.8 (86.0–91.2)	91.3 (87.6–94.3)	0.008
Inception-v3 - Fundus image alone	0.958 (0.942–0.971)	89.2 (87.0–91.2)	90.0 (87.3–92.3)	87.7 (83.4–91.2)	0.004
LASSO regression - OCT + Fundus image	0.950 (0.934–0.963)	88.5 (86.2–90.5)	88.7 (85.9–91.1)	88.2 (83.8–91.5)	0.001
ANN - Fundus segmentation	0.911 (0.891–0.929)	84.1 (81.5–86.4)	85.2 (82.1–87.9)	82.0 (77.2–86.2)	< 0.001
OCT foveal thickness	0.827 (0.801–0.851)	75.9 (73.0–78.7)	76.0 (72.4–79.4)	75.5 (70.4–80.4)	< 0.001

*Comparison of ROC curves with the deep learning method using RF - OCT + Fundus image according to the Delong test

AUC, area under curve; *DBN*, deep belief network; *LASSO*, least absolute shrinkage and selection operator; *RBM*, restricted Boltzmann machine; *RF*, random forest

detecting choroidal neovascularization precisely [33]. On the other hand, OCT showed a highly sensitive ability to detect choroidal neovascularization, with the possibility of missing changes of drusen and retinal pigment epithelium [34]. Therefore, these two imaging modalities offer complementary information on retina (Fig. 8). Since OCT is not yet routinely used for clinical screening AMD, most previous researches using machine learning have reported the performance of fundus image classification [6]. Currently, OCT was recognized as a useful tool to monitor patients with AMD in previous studies [34]. Recent research using OCT image alone through deep learning technique reported to have a high diagnostic performance [13]. The present findings showed that adding fundus image to OCT image can increase the accuracy in deep learning model for diagnosis of AMD. The result revealed a statistically significant improvement of this multimodal approach to prediction AMD. This study suggests that the multimodal deep learning is needed to combine multiple types of

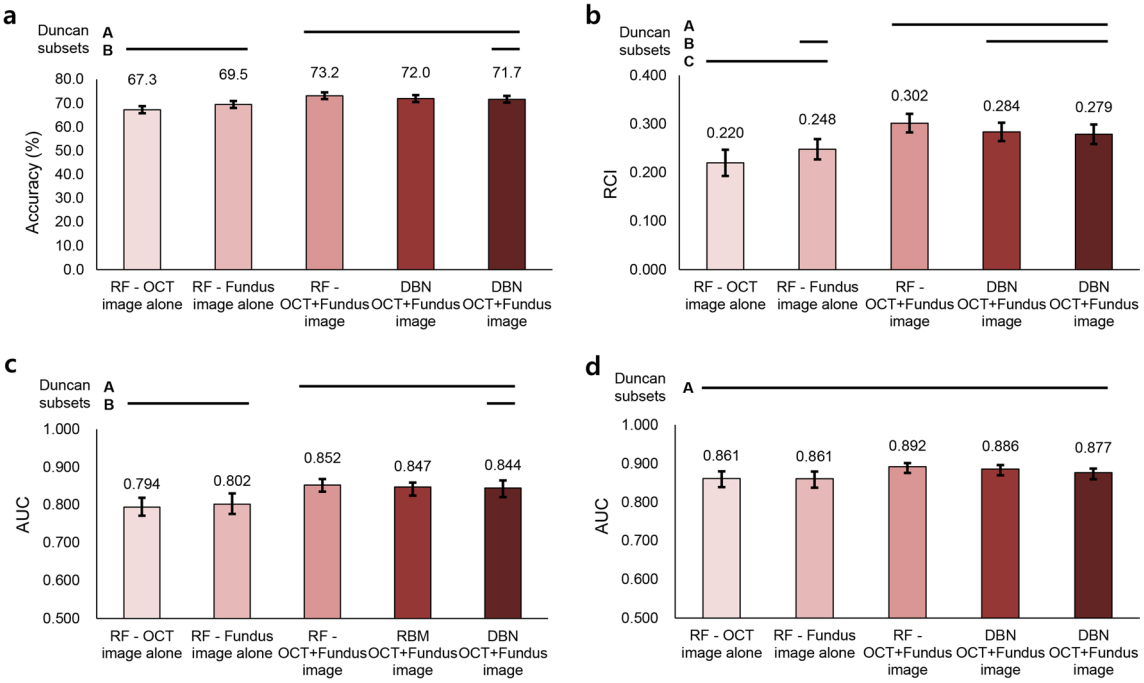

Fig. 5 Three-class (normal, non-neovascular age-related macular degeneration, and neovascular age-related macular degeneration) classification results using transfer learning based on random forest (RF), multimodal restricted Boltzmann machine (RBM), and multimodal deep belief network (DBN). The convolutional deep neural network for transfer learning consisted of pre-trained VGG-19. **a** The accuracy of classification models. **b** The relative classifier information (RCI) of classification models. **c** The average of the AUCs on the one-versus-all pairwise binary classification. **d** The average of the AUCs on the one-versus-one pairwise binary classification. The error bars represent 95% confidence intervals

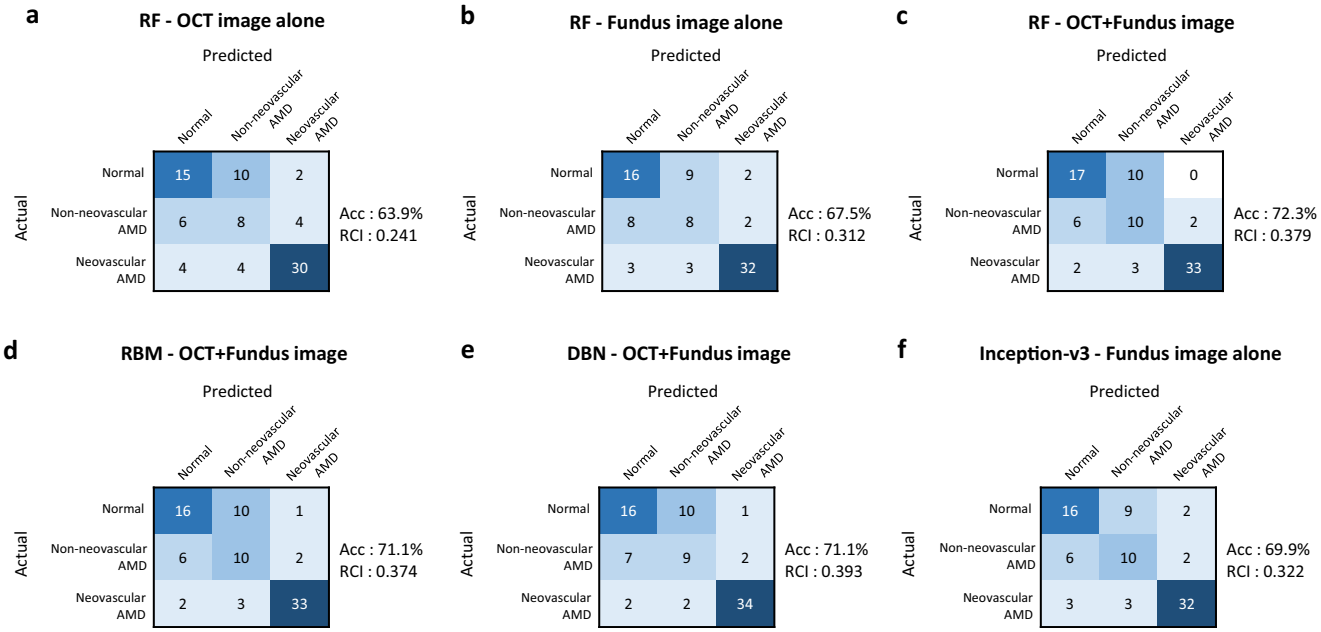

Fig. 6 The confusion matrix from five multi-class models based on the leave-one-out cross-validation at the patient level without data augmentation. **a** Transfer learning based on random forest using OCT image alone. **b** Transfer learning based on random forest using fundus image alone. **c** Transfer learning based on random forest using the combination of OCT and fundus image. **d** Multimodal restricted Boltzmann machine. **e** Multimodal deep belief network. **f** Inception using fundus image alone

image information to reach a more precise diagnosis. However, the findings are limited for the group studied with methodological limitations.

In general, the inferential statistics showed the significant differences between multimodal and single-modal deep learning methods in this study. Notably, multimodal RF predicted AMD with better performance than LASSO regression and other classic methods. The flexible and deep structure of non-linear multi-modal machine learning was more appropriate than LASSO, which is one of the linear regression methods, to model complex systems, thus has a great potential to make a full use of data observed in fundus and OCT images. Generally, the recent imaging methods with regression, segmentation, and retinal thickness measurement were included in routine clinical assessments [35]. The method of combining OCT with fundus image in this paper could provide a better prediction of AMD than using current imaging methods.

The combination of multiple biomarkers or multiple image modalities may increase the diagnostic accuracy [36]. In other clinical imaging area, several studies have reported that the combination of multiple modalities and imaging techniques could improve the diagnostic performance [37]. In the field

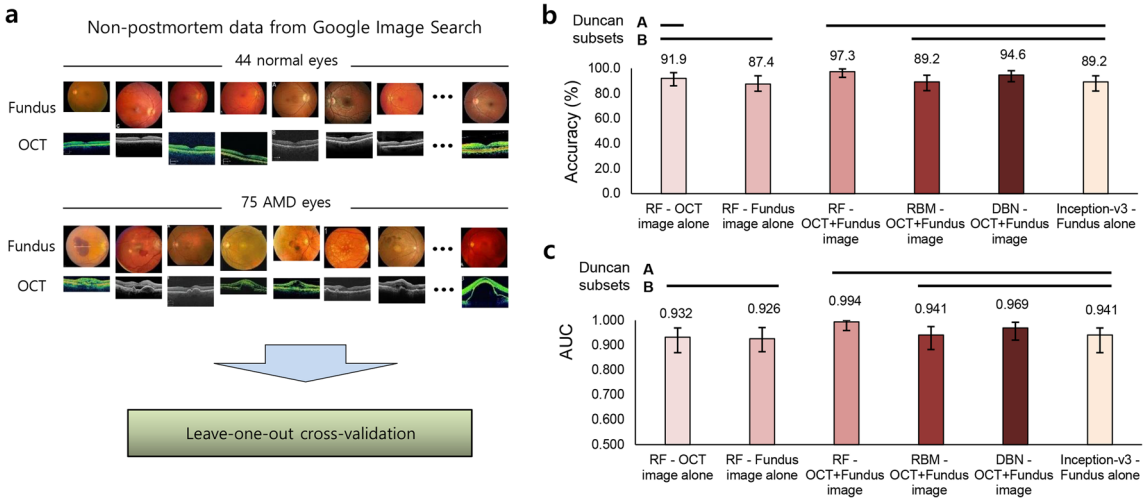

Fig. 7 The leave-one-out classification using non-postmortem data. **a** Non-postmortem fundus and OCT images of normal and AMD eyes were crawled from Google Image Search. **b** The accuracy of classification models. **c** The AUCs of classification models. The error bars represent 95% confidence intervals

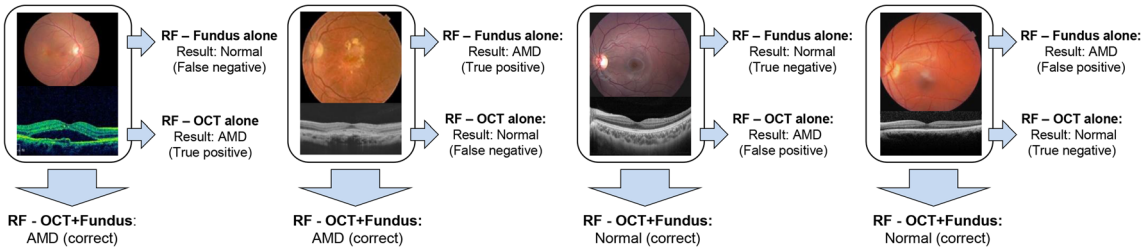

Fig. 8 Four examples of correctly classified retinal images by the multimodal transfer learning based on random forest using the combination of OCT and fundus image

of machine learning, the multimodal learning has been considered as an important problem [38]. A multimodal machine learning method which combined MRI, FDG-PET, and cerebrospinal fluid data, for example, offers utility in predicting the probability of the prevalence of Alzheimer's disease [39]. Fundus and OCT images can be recognized as the most influential biomarkers of retinal diseases. However, previous literatures paid little attention to the function of the diagnostic model that combines fundus with OCT images. Since the most widely used deep learning techniques have been developed to handle a single image modality at one time, previous researchers ignored the efficacy of concomitant use of fundus and OCT data. Therefore, we modified transfer learning process to combine the feature vectors from different imaging domain such as fundus and OCT images by using RF. Findings from this study show that future computer-aid systems should incorporate all clinical data from various examination modalities for precise diagnosis. Thus, this work demonstrated the potential of multimodal deep learning systems more similar to a real clinician.

In this preliminary experiment, RF with multimodal setting worked better than RBM and DBN. RF was found to be a robust and accurate machine learning classifier in the previous literatures [3]. As a nonparametric statistical method, RF can deal with nonlinearity, interactions between predictors, and heterogeneity of predictors [40]. One previous study reported that logistic regression, RF, RBM, DBN, and recurrent neural network showed comparable performance for the purpose of churn prediction [41]. A slight difference is that RF functioned best compared to other machine learning methods. This result was consistent with this finding that RF showed its optimal performances in multimodal imaging problems. RF using multimodal MRI imaging was introduced for the segmentation of stroke lesion [40]. Another example is the application of RF to predict tumor grade by using multimodal MRI imaging [42].

Several limitations should be noted. First, we included a few study eyes. Deep learning generally requires a large size of samples for proper training [43]. Although we used data augmentation to solve this problem, a relatively small number of dataset is involved in a fundamental problem. Second, we used the dataset from postmortem studies. Fundus photographs and OCT images showed postmortem changes and artifacts [44]. Due to these postmortem characteristics, the present deep learning model is not able to apply to the real patients in the clinic. However, in spite of these limitations, this deep learning model emphasized effectively the usefulness of adding fundus imaging to OCT imaging to diagnose AMD. Moreover, these postmortem data have an advantage of being confirmed by pathohistological examination for accurate AMD diagnosis. We tried to make up for this shortcoming by training and validation in non-postmortem OCT and fundus images from Google Image Search. Third, the input images are cropped manually according to ophthalmologists' subjective analysis. Further research is needed for automated selection of the important region of OCT and fundus images. Fourth, we did not train the whole deep convolutional neural networks for the multimodal setting. Previous papers used a stochastic gradient descent method to train large-scaled deep neural network such as AlexNet, VGG, and GoogleNet [8]. However, our previous research revealed that transfer learning using RF was more effective than trained neural network using stochastic gradient descent when the small database was adopted [3].

5 Conclusion

This experimental study demonstrated the benefits of the concomitant use of fundus and OCT data in deep learning technique for AMD diagnosis. However, this result is limited by significant limitations, which not allow the application to real patients in the clinic. Since fundus and OCT imaging provide unique and complementary information on the retina, additional studies should be conducted to determine the optimal deep learning technique to incorporate OCT images and other clinical information into the fundus photograph diagnosis system. Future diagnostic deep learning models need to adopt the multimodal process to combine different types of imaging.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng CY, Wong TY (2014) Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. Lancet Glob Health 2:106–116. https://doi.org/10.1016/S2214-109X(13)70145-1
2. Ferris FL, Wilkinson CP, Bird A et al (2013) Clinical classification of age-related macular degeneration. Ophthalmology 120:844–851. https://doi.org/10.1016/j.ophtha.2012.10.036
3. 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
4. Lam C, Yu C, Huang L, Rubin D (2018) Retinal lesion detection with deep learning using image patches. Invest Ophthalmol Vis Sci 59:590–596. https://doi.org/10.1167/iovs.17-22721
5. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM (2017) Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. JAMA Ophthalmol 135:1170–1176. https://doi.org/10.1001/jamaophthalmol.2017.3782
6. Burlina P, Pacheco KD, Joshi N, Freund DE, Bressler NM (2017) Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis. Comput Biol Med 82:80–86. https://doi.org/10.1016/j.compbiomed.2017.01.018
7. Matsuba S, Tabuchi H, Ohsugi H, Enno H, Ishitobi N, Masumoto H, Kiuchi Y (2018) Accuracy of ultra-wide-field fundus ophthalmoscopy-assisted deep learning, a machine-learning technology, for detecting age-related macular degeneration. Int Ophthalmol. https://doi.org/10.1007/s10792-018-0940-0
8. Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, Peters A, Heid IM, Palm C, Weber BHF (2018) A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology 125:1410–1420. https://doi.org/10.1016/j.ophtha.2018.02.037
9. Wilde C, Patel M, Lakshmanan A, Amankwah R, Dhar-Munshi S, Amoaku W, Medscape (2015) The diagnostic accuracy of spectral-domain optical coherence tomography for neovascular age-related macular degeneration: a comparison with fundus fluorescein angiography. Eye 29:602–609. https://doi.org/10.1038/eye.2015.44
10. Karri SPK, Chakraborty D, Chatterjee J (2017) Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. Biomed Opt Express 8:579–592. https://doi.org/10.1364/BOE.8.000579
11. Lee CS, Baughman DM, Lee AY (2017) Deep learning is effective for classifying normal versus age-related macular degeneration optical coherence tomography images. Ophthalmol Retina 1:322–327. https://doi.org/10.1016/j.oret.2016.12.009
12. Treder M, Lauermann JL, Eter N (2017) Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. Graefes Arch Clin Exp Ophthalmol 256:259–265. https://doi.org/10.1007/s00417-017-3850-3
13. Prahs P, Radeck V, Mayer C, Cvetkov Y, Cvetkova N, Helbig H, Märker D (2017) OCT-based deep learning algorithm for the evaluation of treatment indication with anti-vascular endothelial growth factor medications. Graefes Arch Clin Exp Ophthalmol 256:91–98. https://doi.org/10.1007/s00417-017-3839-y
14. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J, Prasadha MK, Pei J, Ting MYL, Zhu J, Li C, Hewett S, Dong J, Ziyar I, Shi A, Zhang R, Zheng L, Hou R, Shi W, Fu X, Duan Y, Huu VAN, Wen C, Zhang ED, Zhang CL, Li O, Wang X, Singer MA, Sun X, Xu J, Tafreshi A, Lewis MA, Xia H, Zhang K (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172:1122–1131. https://doi.org/10.1016/j.cell.2018.02.010
15. Fang L, Cunefare D, Wang C, Guymer RH, Li S, Farsiu S (2017) Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. Biomed Opt Express 8:2732–2744. https://doi.org/10.1364/BOE.8.002732
16. Schaal KB, Freund KB, Litts KM, Zhang Y, Messinger JD, Curcio CA (2015) Outer retinal tubulation in advanced age-related macular degeneration: optical coherence tomographic findings correspond to histology. Retina 35:1339–1350. https://doi.org/10.1097/IAE.0000000000000471
17. Tran T, Pham T, Carneiro G, et al (2017) A Bayesian data augmentation approach for learning deep models. In: Advances in Neural Information Processing Systems. pp 2794–2803
18. Wallis TSA, Funke CM, Ecker AS, Gatys LA, Wichmann FA, Bethge M (2017) A parametric texture model based on deep convolutional features closely matches texture appearance for humans. J Vis 17:5. https://doi.org/10.1167/17.12.5
19. Wang Y, Zeng J (2013) Predicting drug-target interactions using restricted Boltzmann machines. Bioinformatics (Oxford England) 29:126–134. https://doi.org/10.1093/bioinformatics/btt234
20. Ngiam J, Khosla A, Kim M, et al (2011) Multimodal deep learning. In: Proceedings of the 28th international conference on machine learning (ICML-11). pp 689–696
21. Breiman L (2001) Random forests. Mach Learn 45:5–32
22. Sindhwani V, Bhattacharya P, Rakshit S (2001) Information theoretic feature crediting in multiclass support vector machines. In: Proceedings of the 2001 SIAM international conference on data mining. Society for Industrial and Applied Mathematics, pp 1–18
23. Wei JM, Yuan XJ, Hu QH, Wang SQ (2010) A novel measure for evaluating classifiers. Expert Syst Appl 37:3799–3809. https://doi.org/10.1016/j.eswa.2009.11.040
24. Statnikov A, Aliferis CF, Tsamardinos I, Hardin D, Levy S (2005) A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. Bioinformatics (Oxford England) 21:631–643. https://doi.org/10.1093/bioinformatics/bti033
25. Hand DJ, Till RJ (2001) A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach Learn 45:171–186. https://doi.org/10.1023/A:1010920819831
26. Schisterman EF, Faraggi D, Reiser B, Hu J (2008) Youden index and the optimal threshold for markers with mass at zero. Stat Med 27:297–315. https://doi.org/10.1002/sim.2993
27. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837–845
28. Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn 30:1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2
29. Oh E, Yoo TK, Park EC (2013) Diabetic retinopathy risk prediction for fundus examination using sparse learning: a cross-sectional study. BMC Med Inform Decis Mak 13:106. https://doi.org/10.1186/1472-6947-13-106
30. Smith RT, Chan JK, Nagasaki T, Ahmad UF, Barbazetto I, Sparrow J, Figueroa M, Merriam J (2005) Automated detection of macular drusen using geometric background leveling and threshold selection. Arch Ophthalmol 123:200–206. https://doi.org/10.1001/archopht.123.2.200
31. Chen CY, Wong TY, Heriot WJ (2007) Intravitreal bevacizumab (Avastin) for neovascular age-related macular degeneration: a short-term study. Am J Ophthalmol 143:510–512. https://doi.org/10.1016/j.ajo.2006.10.004
32. 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
33. Mokwa NF, Ristau T, Keane PA, Kirchhof B, Sadda SR, Liakopoulos S (2013) Grading of age-related macular degeneration: comparison between color fundus photography, fluorescein angiography, and spectral domain optical coherence tomography. J Ophthalmol 2013: 385915–385916. https://doi.org/10.1155/2013/385915
34. Castillo MM, Mowatt G, Elders A, Lois N, Fraser C, Hernández R, Amoaku W, Burr JM, Lotery A, Ramsay CR, Azuara-Blanco A (2015) Optical coherence tomography for the monitoring of neovascular age-related macular degeneration: a systematic review. Ophthalmology 122:399–406. https://doi.org/10.1016/j.ophtha.2014.07.055
35. Yang Q, Reisman CA, Wang Z, Fukuma Y, Hangai M, Yoshimura N, Tomidokoro A, Araie M, Raza AS, Hood DC, Chan K (2010) Automated layer segmentation of macular OCT images using dual-scale gradient information. Opt Express 18:21293–21307
36. Liu C, Liu A, Halabi S (2011) A min-max combination of biomarkers to improve diagnostic accuracy. Stat Med 30:2005–2014. https://doi.org/10.1002/sim.4238
37. Yabuuchi H, Matsuo Y, Kamitani T, Setoguchi T, Okafuji T, Soeda H, Sakai S, Hatakenaka M, Nakashima T, Oda Y, Honda H (2008) Parotid gland tumors: can addition of diffusion-weighted MR imaging to dynamic contrast-enhanced MR imaging improve diagnostic accuracy in characterization? Radiology 249:909–916. https://doi.org/10.1148/radiol.2493072045
38. Srivastava N, Salakhutdinov RR (2012) Multimodal learning with deep boltzmann machines. In: Advances in neural information processing systems. pp 2222–2230
39. Zhang D, Shen D, Alzheimer's Disease Neuroimaging Initiative (2012) Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 59:895–907. https://doi.org/10.1016/j.neuroimage.2011.09.069
40. Mitra J, Bourgeat P, Fripp J, Ghose S, Rose S, Salvado O, Connelly A, Campbell B, Palmer S, Sharma G, Christensen S, Carey L (2014) Lesion segmentation from multimodal MRI using random forest following ischemic stroke. NeuroImage 98:324–335. https://doi.org/10.1016/j.neuroimage.2014.04.056
41. Prashanth R, Deepak K, Meher AK (2017) High accuracy predictive modelling for customer churn prediction in telecom industry. In: International conference on machine learning and data mining in pattern recognition. Springer, pp 391–402
42. Fellah S, Caudal D, De Paula AM et al (2013) Multimodal MR imaging (diffusion, perfusion, and spectroscopy): is it possible to distinguish oligodendroglial tumor grade and 1p/19q codeletion in the pretherapeutic diagnosis? AJNR Am J Neuroradiol 34:1326–1333. https://doi.org/10.3174/ajnr.A3352
43. Larochelle H, Bengio Y, Louradour J, Lamblin P (2009) Exploring strategies for training deep neural networks. J Mach Learn Res 10:1–40
44. Yun YS, Kwon OW (1993) Postmortem change of adhesive forces between the retina and the retinal pigment epithelium. J Korean Ophthalmol Soc 34:111–116

Tae Keun Yoo works as an ophthalmologist at Severance Hospital. He received his diploma in Mechanical Engineering from Seoul National University and received the Doctor of Medicine from Yonsei University. His research interest concerns ophthalmologic epidemiology and machine learning.

Joon Yul Choi is a Ph.D. candidate at Seoul National University. He is a Junior Researcher at the Department of Electrical and Computer Engineering, Seoul National University. His research interest focuses on deep learning and MRI data analysis.

Jeong Gi Seo works as an ophthalmologist at Severance Hospital, Yonsei University College of Medicine. He received the Doctor of Medicine from Yonsei University. His current research area is intraocular lens power calculation.

B. Ramasubramanian is currently an Assistant Professor in the Department of Electronics & Communication Engineering at Syed Ammal Engineering College, Ramanathapuram. He received the B.E. degree in Electronics and Communication Engineering from Anna University in 2008 and M.Tech. degree in Communication Systems from B.S. Abdur Rahman University, Chennai, Tamil Nadu, in 2012, respectively and doing Ph.D. degree in Information and Communication Engineering at Anna University, Chennai, Tamilnadu. He had published more than five journals and one book all over the world. His areas of interest are Image Processing, Signal Processing and Embedded Systems.

S. Selvaperumal is currently a Professor and Head in the Department of Electrical & Electronics Engineering at Syed Ammal Engineering College, Ramanathapuram. He had published articles in reputed journals like IET and IEEE Transactions. He had published more than 40 journals and 10 books all over the world. He acted as reviewer for more than 20 journals including IEEE & IET journals. He organized two international conferences and as a recognized supervisor for Ph.D guidance in Anna University and guiding 12 Ph.D Scholars and produced 3 Ph.D scholar in the domain of Electrical Engineering. His areas of interest are Power Electronics, Embedded Systems and control system design.

Deok Won Kim is a professor in the Medical Engineering Department, Yonsei University College of Medicine, Seoul, Korea, where he has been since 1987. He received the B.S degree from Seoul National University, Seoul, Korea in 1976, the M.S. degree in electrical engineering from Northwestern University, Evanston, IL, USA, and the Ph.D. degree in biomedical engineering from the University of Texas, Austin, TX, USA, in 1980 and 1986, respectively. His areas of research interest are electromagnetic field hazards and multiclass classification using machine learning in medicine.