

Optical coherence tomography-based diabetic macula edema screening with artificial intelligence

De-Kuang Hwang^{a,b}, Yu-Bai Chou^{a,b}, Tai-Chi Lin^{a,b}, Hsin-Yu Yang^a, Zih-Kai Kao^c, Chung-Lan Kao^{d,e}, Yi-Ping Yang^{c,f}, Shih-Jen Chen^{a,b}, Chih-Chien Hsu^{a,b,*}, Ying-Chun Jheng^{c,d,e,*}

^aDepartment of Ophthalmology, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; ^bDepartment of Ophthalmology, School of Medicine, National Yang-Ming University, Taipei, Taiwan, ROC; ^cDepartment of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; ^dDepartment of Physical Medicine & Rehabilitation, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; ^eDepartment of Physical Medicine & Rehabilitation, School of Medicine, National Yang-Ming University, Taipei, Taiwan, ROC; ^fDepartment of Pharmacology, Taipei Veterans General Hospital, Taipei, Taiwan, ROC

Abstract

Background: Optical coherence tomography (OCT) is considered as a sensitive and noninvasive tool to evaluate the macular lesions. In patients with diabetes mellitus (DM), the existence of diabetic macular edema (DME) can cause significant vision impairment and further intravitreal injection (IVI) of anti-vascular endothelial growth factor (VEGF) is needed. However, the increasing number of DM patients makes it a big burden for clinicians to manually determine whether DME exists in the OCT images. The artificial intelligence (AI) now enormously applied to many medical territories may help reduce the burden on clinicians.

Methods: We selected DME patients receiving IVI of anti-VEGF or corticosteroid at Taipei Veterans General Hospital in 2017. All macular cross-sectional scan OCT images were collected retrospectively from the eyes of these patients from January 2008 to July 2018. We further established AI models based on convolutional neural network architecture to determine whether the DM patients have DME by OCT images.

Results: Based on the convolutional neural networks, InceptionV3 and VGG16, our AI system achieved a high DME diagnostic accuracy of 93.09% and 92.82%, respectively. The sensitivity of the VGG16 and InceptionV3 models was 96.48% and 95.15%, respectively. The specificity was corresponding to 86.67% and 89.63% for VGG16 and InceptionV3, respectively. We further developed an OCT-driven platform based on these AI models.

Conclusion: We successfully set up AI models to provide an accurate diagnosis of DME by OCT images. These models may assist clinicians in screening DME in DM patients in the future.

Keywords: Artificial intelligence; Optical coherence tomography; Vascular endothelial growth factor

1. INTRODUCTION

Although the idea that artificial intelligence (AI) could replace the human brain began to appear in the minds of scientists in 1950s, there was no significant progress in the development of AI until 2010s. The tremendous advances in recent development of AI were due to the popularity and improvement of graphic processing units, which could make parallel computing faster,

cheaper, and more powerful. Deep learning is developed with multilayered artificial neural networks. It achieves the highest agreement with the input “answer” by repeatedly modifying the weight of each synapse. Among all of the deep learning programs, convolutional neural networks (CNNs) are commonly used for analyzing visual imagery in various field. One of the possible AI applications in ophthalmologic field is to provide screening and diagnostic aid for patients who live in area which is in a lack of ophthalmologists or well-trained optometrists.

Diabetes mellitus (DM) is one of the most prevalent global diseases which may comorbid with sight-threatening retinopathy. Early detection and intervention are the most effective ways to prevent blindness caused by diabetic retinopathy. Diabetic macular edema (DME) is one of the major comorbidities of diabetic retinopathy which marked impaired patients' sights. Patients with DME may suffered from blurred vision, image distortion, or dark areas in the field of view. In recent years, the development and evolution of optical coherence tomography (OCT) have changed the clinical practice in ophthalmology. It is a noninvasive medical imaging diagnostic technology which now be used for the diagnosis of DME.¹ Several cross-sectional images of optical anatomy of retina could be obtained within seconds. Clinical treatment strategy

*Address correspondence. Dr. Chih-Chien Hsu, Department of Ophthalmology, Taipei Veterans General Hospital, 201, Section 2, Shi-Pai Road, Taipei 112, Taiwan ROC. E-mail address: chihchienym@gmail.com (C.-C. Hsu); Dr. Ying-Chun Jheng, Department of Medical Research, Taipei Veterans General Hospital, 201, Section 2, Shi-Pai Road, Taipei 112, Taiwan, ROC. E-mail address: cycom1220@gmail.com (Y.-C. Jheng).

Conflicts of interest: The authors declare that they have no conflicts of interest related to the subject matter or materials discussed in this article.

Journal of Chinese Medical Association. (2020) 83: 1034-1038.

Received April 23, 2020; accepted April 23, 2020.

doi: 10.1097/JCMA.0000000000000351.

Copyright © 2020, the Chinese Medical Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

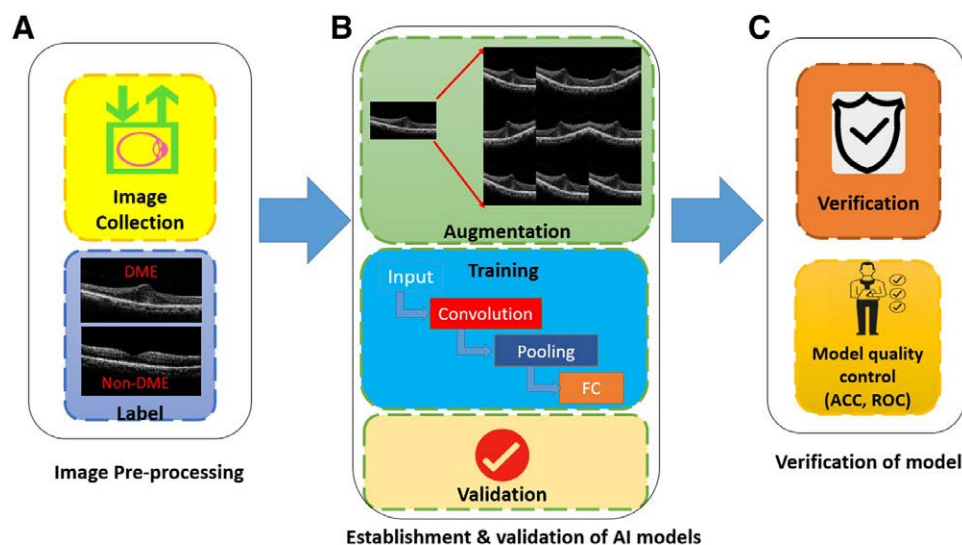


Fig. 1 A, The workflow of artificial intelligence (AI) development in this study. Optical coherence tomography image collection and labeling were the first step of all (Image pre-processing stage). B, During the establishment and validation stage, the image database was augmented by randomly shearing, zooming, rotating, or horizontally flipping. An AI program containing the convolution layer, pooling layer, and fully connection (FC) layer, and validation were designed. C, After the development of the program, verification and model quality controls were evaluated for the performance of this model. ACC = accuracy; DME = diabetic macular edema; ROC = receiver operating characteristic.

of DME is usually made based on the interpretation of these OCT image. In this study, we are aiming to establish machine learning based in training AI systems to identify DME through analyzing OCT images, especially for differentiating the patients with disease progression with the requirement of further anti-VEGF treatment.

2. METHODS

The workflow of our study is shown in Fig. 1. There were three major stages: image preprocessing, establishment of AI models, and verification of AI models. The detailed processes are described below.

2.1. Data collection

Patients with DME who received intravitreal injections (IVIs) of either anti-vascular endothelial growth factor (VEGF) or corticosteroid in Taipei Veterans General Hospital during January of 2017 to December of 2017 were enrolled in the study. All OCT macula cross-section scanning images between January of 2008 and July of 2018 from both eyes of these patients were retrospectively collected from the Department of Ophthalmology, Taipei Veterans General Hospital, Taiwan, with the approval from the Institutional Review Board of Taipei Veterans General Hospital.

One senior vitreoretinal surgeon and one experienced ophthalmologist performed the DME labeling of collected OCT images. The OCT definition of DME used in our study is based on international clinical DME disease severity scale² and OCT patterns of DME,³ as indicated in Fig. 2, including mild DME, moderate DME, severe DME, DME with serous retinal detachment (SRD), DME with posterior hyaloid traction (PHT), and DME with traction retinal detachment (TRD). OCT images presented with above patterns of DME were all labeled as DME, and the rest images were labeled as non-DME. The human labeling was masked to the AI grading results. In case of discrepancy between the labeling, another senior retinal specialist was consulted and all three doctors determined the final label together. After human labeling, the OCT images have been removed discernible information and classified as two types, DME and non-DME.

2.2. Data set augmentation

In this work, the images in training data set have been executed the data augmentation in each epoch to overcome the limitation of data size and reinforce the performance of the AI model. The augmented method was applied only for perturbing the training data set, but not for validation or verification data set. The following augmentation parameters have been applied and the detail value of those parameters has been randomized by AI:

1. Random shear: sheared in counter-clockwise direction with random angle degrees
2. Random zoom: randomly zoom in and zoom out
3. Random rotation: images have been randomly rotated while preserving their shape
4. Random horizontal flip: rotates the image with randomly radians

2.3. Establishment and validation of AI Models

VGG16 and InceptionV3 have been chosen as the main architectures for developing the AI model and the trained AI models have been validated with validation data set aimed to evaluate whether the models required further modification and retraining. The AI model executed image classification as binary results; thus, the loss function was set as the Sigmoid Cross-Entropy loss. The Root Mean Square Propagation (RMSprop) was selected as the optimizer function in order to speed up the network function converge. All of those architectures include some hyperparameters that could be adjusted for enhancing the recognition accuracy, such as batch size, epoch, learning rate, and optimizer. Transfer learning has been applied to decrease the study time and achieve the satisfactory results. Moreover, the AI models were established using the Google cloud platform with two-core vCPU, 7.5GB RAM and an NVIDIA Tesla K80 GPU card; meanwhile, the software is the CentOS7 with Keras 2.2.4 and tensorflow-gpu 1.6.0 for training and validation.

2.4. Verification of AI Models and Statistical analysis

Several metrics have been used for image recognition. The confusion matrix was used to present the results of clinical verification

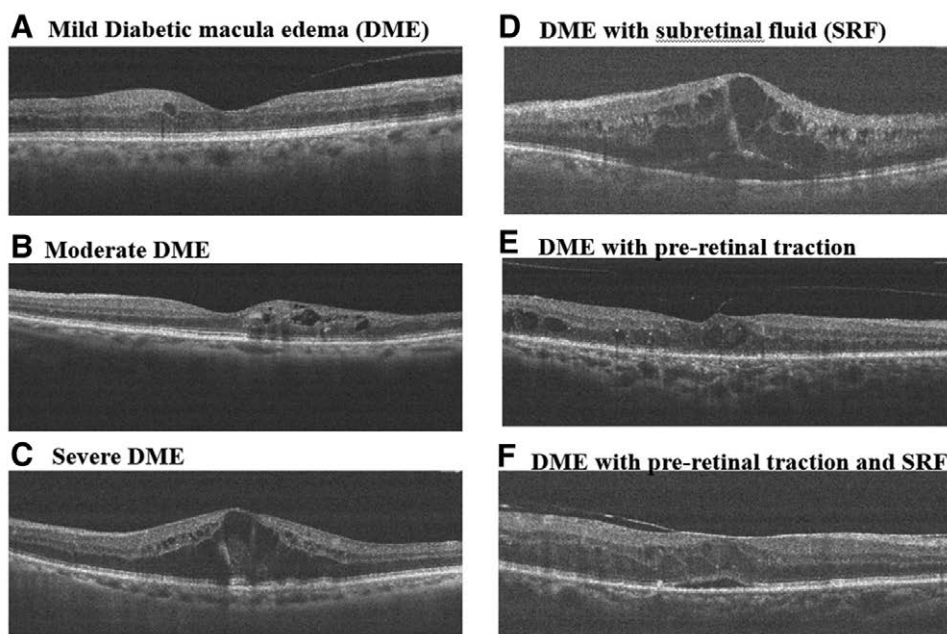


Fig. 2 Optical coherence tomography (OCT) definitions of diabetic macular edema (DME). The definition of DME was based on international clinical DME disease severity scale and OCT patterns of DME. A, Mild DME, identified by retinal thickening or hard exudates in the posterior pole but distant from the center of the macula. B, Moderate DME, characterized by retinal thickening or hard exudates approaching the center of the macula, but not affecting the center. C, Severe DME, characterized by retinal thickening or hard exudates affecting the center of the macula. D, DME with serous retinal detachment (SRD), identified by the presence of subretinal fluid with DME. E, DME with posterior hyaloidal traction (PHT), characterized by DME with preretinal membrane attached to vitreous. F, DME with traction retinal detachment (TDR), identified by DME with the presence of preretinal traction and subretinal fluid.

and compare the predictions of the AI models. The confusion matrix used predict result (positive, P and negative, N) and ground truth (true, T and false, F) to integrate a 2*2 matrix which include four categories, true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Moreover, for biomedical image recognition, sensitivity, specificity, and accuracy were standard parameters. We define those parameters as the following equations:

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

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

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

3. RESULTS

3.1. Image collection

The OCT images were collected from 173 diabetic patients. All images were randomly selected in the database based on different patients. A total number of 4932 OCT images were collected. Of them, 3495 passed the image quality control and were used in this study. Among these, 80% (2768) of images were randomly selected (based on different patients) as the training data set to establish our AI models, 365 images were selected as validation data set, and 362 images were selected as verification data set. The images in training data set have been executed the data augmentation, which was only applied for perturbing the training data set, but not for validation or verification data set.

3.2. Establishing AI models

Two CNN architectures (InceptionV3 and VGG16) have been applied to establish the AI models. The CNN architecture includes three main layers, convolution layer, pooling layer, and fully connected layer. The images have been adjusted to the same size and the skill of data augmentation has been used to enhance the efficiency of our model. The batch size of the training layers has been set as 16 images per step and used the RMSprop optimizer with a learning rate of $1e-4$. Each model with 100 epochs for training, the more iterative the epoch was, and the more accurate the AI model was (Fig. 3A). The best models with the minimal value of loss were selected for the verification. These models' accuracy for recognizing the validation data set was corresponding to 93.15% and 93.42% for VGG16 and InceptionV3, respectively (Fig. 3B).

3.3. Verification of the final model

The performance of each AI model (InceptionV3 and VGG16) has been verified by the validation data set, which contained 227 DME and 135 non-DME OCT images. The accuracy of the AI model based on VGG16 and InceptionV3 architectures was 92.82% and 93.09%, respectively. The sensitivity of the VGG16 and InceptionV3 AI model was 96.48% and 95.15%. The specificity was corresponding to 86.67% and 89.63% for VGG16 and InceptionV3, respectively.

4. DISCUSSION

Recently, there were enormous AI model architectures have been applied to execute the classification task, such as VGG, Inception, and ResNet. In 2014, the Inception applied the batch normalization skill to accelerate the training task and the Inception architecture won the champion.^{4,5} At the same time, VGGNet built more layers of models, reaching 16 hidden layers (VGG16), which was the runner-up in 2014 ImageNet's

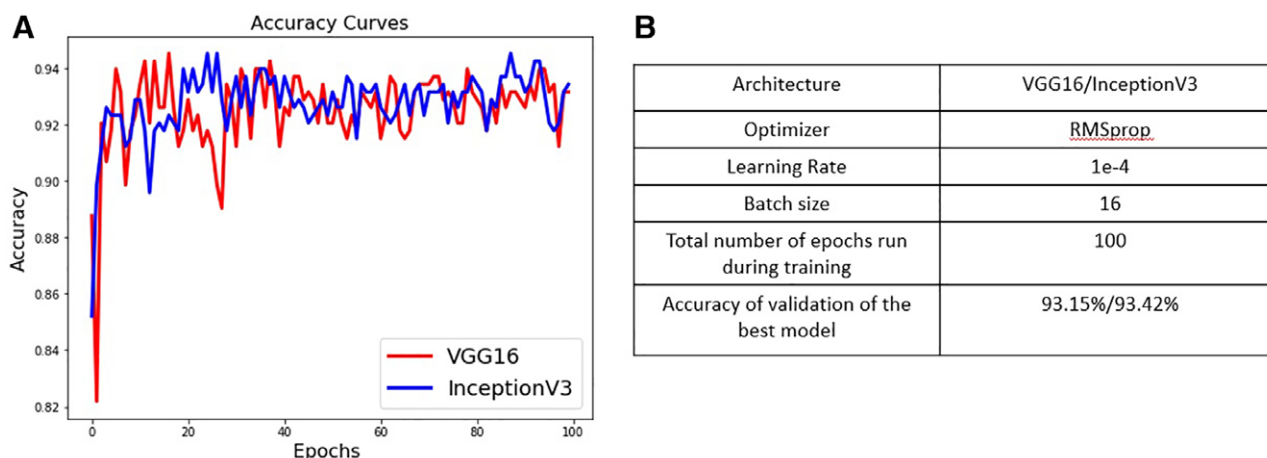


Fig. 3 A, Validation curves for each convolutional neural network (CNN)-based artificial intelligence (AI) models. Our trained models have been re-examined (VGG16 and InceptionV3) for each CNN-based AI model ensured that the training process did not overfit. As the training epoch increased, the accuracy increased. B, The final model has been trained by RMSprop optimizer with a learning rate set as $1e-4$, batch size with 16 and the total epoch number was 100. The accuracy of the validation data set of the final model of VGG16 and InceptionV3 was 93.15% and 93.42%, respectively.

annual competition (ILSVRC).^{6,7} Moreover, the VGG16 and InceptionV3 have been enormously applied to assist medical image identification, such as taxonomy of the CT image with lung cancer to classify the pathological types,⁸ to identify the endoscopy images with different lesions,⁹ and assisted ophthalmologist to analyze the variation of the OCT images.^{10,11}

DME is a type of diabetic retinopathy, caused by metabolic effects of hyperglycemia, which results in retinal vascular changes and subsequent retinal inflammation and ischemia. Since the macula is located at the center of the retina and plays the most important role in fine vision, once the microvascular leakage or the exudation increases, the central vision of the macula decreases significantly. DME may present at any stage of diabetic retinopathy and should be diagnosed and treated as early as possible. However, the early symptoms are difficult to detect, and patients may not be aware of them, resulting in irreversible visual defects by the time of medical remedy.

Currently, there are several approaches to diagnose the DME, which include indirect ophthalmoscopy, color retinal photography, fluorescein angiography (FAG), and OCT. Among them, indirect ophthalmoscopy and color retinal photography are the primary exams, which can provide good screen for diabetic retinopathy, such as retinal hemorrhage, exudates, and cotton-wool spots. However, for detection of early or mild DME, especially central thickness $<300 \mu\text{m}$, even experienced retinal specialists could find some difficulties in making a diagnosis with indirect ophthalmoscopy or color retinal photograph alone. Under the suspicion of advanced diabetic retinopathy and DME, FAG is often indicated to further determine the area of vascular changes, the degree of macular edema, and the distribution of ischemic areas. However, FAG is an invasive diagnostic method that carries the risk of a severe allergic reaction.¹² Nowadays, OCT, a noninvasive, safe, quick, and reliable medical imaging diagnostic technology, has been developed and widely applied to diagnose DME.

The global DM prevalence in 2019 is estimated to be 9.3%, rising to 10.2% by 2030 and 10.9% by 2045. This means that by 2045, there will be approximately 700 million diabetic patients, and in 2019, there were only 463 million diabetic patients.¹³ However, patients with DME often need regular OCT examinations to determine whether DME exists. The increasing number of DM patients makes it a big burden for clinicians to manually determine whether DME exists in the

OCT images. An AI that can help to screen may decrease the loading for the ophthalmologists. Moreover, Funakoshi et al¹⁴ found the odds of having diabetic retinopathy are higher among those who have lower educational levels, receiving public assistance, with irregular or no employment. Patients in remote islands or mountains require far-reaching traffic to seek medical examination because not all places have ophthalmologists. The transportation cost may be a great burden for these patients. A simple OCT device in the local area and passing the image to the AI with high accuracy in determining macula conditions can be of great help. In this study, we have demonstrated our OCT-based AI screening platform for DME. The high diagnostic accuracy of our AI models based on VGG16 and InceptionV3 architectures for detecting DME is 92.82% and 93.09%, respectively. Although our AI model has high accuracy in determining DME, it is worth noting that for other retinal diseases that can cause macular edema, such as Irvine Gass syndromes (cystoid macular edema after cataract surgery), our AI system might make the wrong diagnosis. Hence, our AI system should be used in patients currently receiving treatment for DME. For those patients without past history of DME, they still need to attend ophthalmology to seeking medical exams and remedies.

In conclusion, we build our OCT-based AI system with the commonly used CNN architectures such as VGG16 and Inception V3 and obtain a good diagnosis rate of DME. This system can reduce the burden of clinical ophthalmologists and even can be used in remote areas in the future. However, when using it, we must pay attention to the choice of the patients so that we can avoid misdiagnosis.

ACKNOWLEDGMENTS

We thank Ying-Hsuan Wu, Feng-Yuan Yang for data collection, and thank Hsin-Yi Huang for serving as a scientific advisor of the project. This study was assisted in part by Big Data Center of Taipei Veterans General Hospital. This research was funded by the Taiwan Ministry of Science and Technology (MOST-108-2314-B-010-042-MY3, MOST-108-2811-B-010-541, MOST-108-2314-B-075-055), Taipei Veterans General Hospital (CI-109-19).

REFERENCES

1. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, et al. Optical coherence tomography. *Science* 1991;254:1178–81.
2. Wilkinson CP, Ferris FL 3rd, Klein RE, Lee PP, Agardh CD, Davis M, et al.; Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;110:1677–82.
3. Kim BY, Smith SD, Kaiser PK. Optical coherence tomographic patterns of diabetic macular edema. *Am J Ophthalmol* 2006;142:405–12.
4. Ioffe S, Szegedy CJA. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015. Available at <https://arxiv.org/abs/1502.03167>.
5. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna ZB. Rethinking the inception architecture for computer vision. 2015. Available at <https://arxiv.org/abs/1512.00567>.
6. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. Available at <https://arxiv.org/abs/1409.1556>.
7. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* 2015;115:211–52.
8. Wang S, Dong L, Wang X, Wang X. Classification of pathological types of lung cancer from CT Images by deep residual neural networks with transfer learning strategy. *Open Med (Wars)* 2020;15:190–7.
9. Mori Y, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. *Ann Intern Med* 2018;169:357–66.
10. Chan GCY, Muhammad A, Shah SAA, Tang TB, Lu C, Meriaudeau F, editors. Transfer learning for diabetic macular edema (DME) detection on optical coherence tomography (OCT) images. Paper presented at: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA). September 12–14, 2017; Kuching, Malaysia.
11. Hwang DK, Hsu CC, Chang KJ, Chao D, Sun CH, Jheng YC, et al. Artificial intelligence-based decision-making for age-related macular degeneration. *Theranostics* 2019;9:232–45.
12. Stein MR, Parker CW. Reactions following intravenous fluorescein. *Am J Ophthalmol* 1971;72:861–8.
13. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al.; IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract* 2019;157:107843.
14. Funakoshi M, Azami Y, Matsumoto H, Ikota A, Ito K, Okimoto H, et al. Socioeconomic status and type 2 diabetes complications among young adult patients in Japan. *PLoS One* 2017;12:e0176087.