scientific reports

OPEN

Anti-VEGF treatment outcome prediction based on optical coherence tomography images in neovascular age-related macular degeneration using a deep neural network

Jeong Mo Han[1,2,9], Jinyoung Han[3,8,9], Junseo Ko[3,9], Juho Jung[3], Ji In Park[4], Joon Seo Hwang[5], Jeewoo Yoon[3], Jae Ho Jung[2] & Daniel Duck-Jin Hwang[6,7✉]

Age-related macular degeneration (AMD) is a major cause of blindness in developed countries, and the number of affected patients is increasing worldwide. Intravitreal injections of anti-vascular endothelial growth factor (VEGF) are the standard therapy for neovascular AMD (nAMD), and optical coherence tomography (OCT) is a crucial tool for evaluating the anatomical condition of the macula. However, OCT has limitations in accurately predicting the degree of functional and morphological improvement following intravitreal injections. Artificial intelligence (AI) has been proposed as a tool for predicting the treatment response of nAMD based on OCT biomarkers. Our study focuses on the development and assessment of an AI model utilizing the DenseNet201 algorithm. The model aims to predict anatomical improvement based on OCT images before, and during anti-VEGF therapy. The training process involves two scenarios: (1) using only preinjection OCT images and (2) utilizing both OCT images before and during anti-VEGF therapy for model training. The outcomes of our investigation, involving 2068 images from a cohort of 517 Korean patients diagnosed with nAMD, indicate that the AI model we introduced surpassed the predictive performance of ophthalmologists. The model exhibited a sensitivity of 0.915, specificity of 0.426, and accuracy of 0.820. Notably, its predictive capabilities were further enhanced with the inclusion of additional OCT images taken after the first and second injections during the loading phase. The treatment prediction performance of the model was the highest when using all input modalities (before injection, and after the first and second injections) and concatenation-based fusion layers. This study highlights the potential of AI in assisting individualized and tailored nAMD treatment.

Age-related macular degeneration (AMD) is a leading cause of blindness in developed countries. The number of patients with AMD is increasing worldwide, thus further intensifying the global treatment burden[1,2]. Intravitreal anti-vascular endothelial growth factor (VEGF) injections are administered on a monthly three-loading basis as the standard therapy for neovascular AMD (nAMD)[3,4]. Although optical coherence tomography (OCT) is one of the most important modalities for evaluating the anatomical condition of the macula in vivo, it cannot precisely predict the functional improvement in visual acuity following intravitreal injections and the degree of morphological improvement of the macula[5].

[1]Kong Eye Hospital, Seoul, Korea. [2]Department of Ophthalmology, Seoul National University College of Medicine, Seoul, Korea. [3]Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul, Korea. [4]Department of Medicine, Kangwon National University Hospital, Kangwon National University School of Medicine, Chuncheon, Gangwon-do, Korea. [5]Seoul Plus Eye Clinic, Seoul, Korea. [6]Department of Ophthalmology, Hangil Eye Hospital, 35 Bupyeong-daero, Bupyeong-gu, Incheon 21388, Korea. [7]Lux Mind, Incheon, Korea. [8]Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, Korea. [9]These authors contributed equally as first authors: Jeong Mo Han, Jinyoung Han, and Junseo Ko. ✉email: daniel.dj.hwang@gmail.com

The treatment plan is predominantly determined based on the OCT findings. In the treat-and-extend method, which has been considered a standard treatment in recent years[6], the presence of new or persistent intraretinal fluid (IRF), new or persistent subretinal fluid (SRF), enlargement of retinal pigment epithelial detachment (PED), or presence of subretinal hyperreflective material (SHRM) indicates that nAMD is in the active stage. The response to anti-VEGF therapy varies based on the location of the fluid on OCT. The persistence rate of PED following anti-VEGF injections is reportedly higher than that of IRF or SRF[7]. Furthermore, functional improvement varies depending on the location of the fluid, and IRF is particularly associated with poor visual prognosis[8].

The investigation of these OCT biomarkers using artificial intelligence (AI) has been proposed as a method for predicting the treatment response to anti-VEGF therapy in nAMD. As the treatment response varies based on the location of the fluid on OCT, an attempt was made to acquire the SRF, IRF, and PED volumes by automated segmentation and predict the degree of vision improvement following anti-VEGF injection using each computed volume of the fluid[9]. In addition, a machine-learning algorithm was developed to assess the treatment burden associated with AMD using OCT images during the early treatment period (OCT images before treatment and at monthly loading times)[10–12]. A model that predicts the OCT status following anti-VEGF injection by learning pretreatment OCT images using a generative adversarial network (GAN) technique has been postulated as an alternative approach[13,14]. These trials indicate that implementation of AI could be useful in informing patients about their conditions after treatment.

This study aimed to determine whether an AI model can predict anatomical improvement prediction using OCT images taken prior to anti-VEGF therapy. If complete remission forecasting of the fluid following three intravitreal anti-VEGF injections is possible, patients could receive a more thorough explanation of the post-treatment status of the macula and a future treatment plan at the time of initial therapy. We also investigated whether the predictive performance of the model could be improved using OCT images after the first and second treatments, as well as pre-injection images. In addition, the performance of the ophthalmologists was investigated and compared with the performance of the model.

Results

We included 2068 images from 517 eyes of 517 patients with nAMD. The mean age was 71.4 ± 9.0 (range, 65–78; median 72) years. Table 1 summarizes the baseline characteristics of the enrolled patients. As the drug used for intravitreal injection, ranibizumab (Lucentis®; Novartis AG, Basel, Switzerland and Genentech Inc. San Francisco,

Variables	Neovascular AMD (N = 517)
Age, years (IQR)	71.4 (65–78, median 72)
Sex, n (%)	
Male	311 (60)
Female	206 (40)
Eye treated, n (%)	
Right	272 (53)
Left	245 (47)
Underlying disease, n (%)	
Hypertension	245 (46)
Diabetes	107 (21)
Subtypes of wet AMD, n (%)	
PCV	167 (32)
Type 1 MNV, except PCV	110 (22)
Type 2 MNV	178 (34)
RAP (type 3 MNV)	62 (12)
Treatment response, n (%)*	
Dry macula	409 (79)
SRF remained	85 (16)
IRF remained	23 (5)
Visual acuity (logMAR, IQR)†	
Pre-injection	0.43 (0.15–0.52, median 0.30)
1-month after first injection	0.36 (0.10–0.52, median 0.30)
1-month after second injection	0.31 (0.10–0.40, median 0.22)
1-month after third injection	0.29 (0.10–0.40, median 0.22)

Table 1. Baseline characteristics. *AMD* age-related macular degeneration, *IQR* interquartile range, *PCV* polypoidal choroidal vasculopathy, *MNV* macular neovascularization, *RAP* retinal angiomatous proliferation, *SRF* subretinal fluid, *IRF* intraretinal fluid, *logMAR* logarithm of the minimum angle of resolution. *After administering injections three monthly injections, the evaluation was conducted one month after the last injection using OCT (optical coherence tomography). †Only the 203 individuals for whom visual acuity data could be obtained were listed.

Fusion layer	Inputs	Precision	Accuracy	Specificity	Sensitivity
Concatenate	A	0.4495	0.7786	0.0111	**0.9704**
Concatenate	A + B	0.6383	0.7828	0.3029	0.8987
Concatenate	A + B + C	**0.7203**	0.8201	**0.4260**	0.9148

Table 2. Model performance metrics between different modalities. A, B, and C denote the patient's spectral-domain optical coherence tomography images captured before injection, after the first injection, and after the second injection, respectively. Significant values are in [bold].

Fusion layer	Inputs	Precision	Accuracy	Specificity	Sensitivity
Average	A + B + C	0.6171	0.8083	0.1181	0.9839
Attention	A + B + C	0.5809	0.7952	0.0767	0.9691
LSTM	A + B + C	0.6979	0.8035	**0.5498**	0.8685
Concatenate	A + B + C	**0.7203**	**0.8201**	0.4260	**0.9148**

Table 3. Model performance comparison between the fusion baselines. A, B, and C denote the patient's spectral-domain optical coherence tomography captured before injection, after the first injection, and after the second injection, respectively. The baselines used all images (A, B, and C) as inputs. Significant values are in [bold].

CA, USA) was used in 71% (365 eyes) of cases and aflibercept (Eylea®, Regeneron, Tarrytown, NY, USA and Bayer HealthCare, Berlin, Germany) in 29% (152 eyes) of cases.

Table 2 presents the performances of the different input modalities. A concatenation fusion layer was used for the baseline fusion method. The model with the A + B + C input modality exhibited the highest performance in terms of precision (0.7203), accuracy (0.8201), and specificity (0.4260). The proposed model demonstrates an ability to comprehend inter-sequence dependencies and utilize the full information in the input data to accurately predict treatment outcomes.

Table 3 lists the performance metrics for the fusion modalities. We used the same input modalities and fused each modality to concatenate the average, attention, and long-short term memory (LSTM)-based methods. Concatenation-based fusion demonstrated the highest precision (0.7203), accuracy (0.8201), and sensitivity (0.9148). As shown in Table 3, the concatenation fusion layer exhibited the highest performance compared to the other baselines. The average, attention, and LSTM models aggregated the feature vectors into small vector sizes to reduce the total number of parameters. However, the concatenated fusion layer does not reduce the vector size; moreover, it prevents information loss. Therefore, the concatenation fusion layer demonstrated an approximately 2% higher accuracy.

Table 4 presents the performance metrics comparing our proposed AI model with six ophthalmologists (three ophthalmology residents and three retina specialists). The conditions A, B, and C correspond to different sets of spectral-domain (SD)-OCT images provided for the prediction tasks. For the resident group, the Fless' kappa values were 0.0383 for Condition A, 0.1868 for Condition B, and 0.4086 for Condition C. For the specialists, the Fless' kappa values were 0.3562 for Condition A, 0.4302 for Condition B, and 0.6649 for Condition C. These results demonstrate that as more images were provided (from Condition A to C), intergrader agreement improved, with retina specialists achieving higher consistency than residents. Importantly, the AI model consistently outperformed both resident and specialist groups, demonstrating approximately 7%, 2%, and 5% higher accuracy than the average ophthalmologist performance across each condition. This experiment highlights the robustness of the design, as increased image availability led to better human performance, yet the AI model maintained a superior level of accuracy, underscoring its potential clinical value.

Discussion

This study presents the development and evaluation of an AI model that predicts the outcome of intravitreal anti-VEGF injections in nAMD based on the OCT images. This study also investigated whether the incorporation of OCT images acquired after the first and second anti-VEGF injections, as well as the images acquired prior to treatment, improved the predictive value of the model. The results revealed that the AI model outperformed ophthalmologists in treatment outcome prediction, which further improved with additional OCT images during the loading phase.

The application of AI in ophthalmology has demonstrated great potential for diagnosing diseases, predicting treatment outcomes, and developing treatment policies. Many studies have demonstrated the high accuracy of AI in aiding diagnosis[15,16]. However, few studies have predicted the treatment outcomes or recommended individualized and tailored treatment[17,18].

In a study that predicted the treatment outcomes of patients with macular disease treated with anti-VEGF injections using AI, Gallardo et al. attempted to predict the treatment burden using OCT images and demographic information, and developed an AI model that classified whether the average treatment interval between injections was low, high, or moderate[11]. Liu et al. constructed a model utilizing a GAN to predict the effects of a single treatment based on OCT images obtained before treatment in patients with typical nAMD.

Accuracy	A	A + B	A + B + C
Ophthalmology residents			
1	0.4897	0.5714	0.6530
2	0.6734	0.7346	0.6938
3	0.5306	0.5714	0.7142
Intergrader Agreement (Kappa)	0.0383	0.1868	0.4086
Retina specialists			
1	0.6326	0.8163	0.8163
2	0.7142	0.6938	0.8136
3	0.7142	0.7551	0.8775
Intergrader Agreement (Kappa)	0.3562	0.4302	0.6649
Ophthalmologists average	**0.6257**	**0.6904**	**0.7614**
Our proposed model	**0.6939**	**0.7142**	**0.8163**

Table 4. Model performance comparison between the ophthalmologists. The patients' spectral-domain optical coherence tomography images captured before injection, after the first injection, and after the second injection are denoted as A, B, and C, respectively. Both ophthalmologists and our proposed model used these images as inputs to predict the results. Fleiss' kappa scores were used to calculate intergrader agreement, revealing that agreement improved as more OCT images were provided. The AI model consistently demonstrated higher predictive accuracy compared to both resident and specialist groups across all conditions. Additionally, the p-value confirmed that the observed improvements in agreement were statistically significant, with a p-value of 0.05 or lower for intergrader agreements exceeding 0.1. Significant values are in [bold].

The accuracy of predicting the wet or dry macular state by doctors assessing OCT images generated by AI following treatment has been examined[13]. To predict the outcome of AMD treatment, a study was conducted to estimate the prognosis of treatment based on OCT images obtained by a conditional GAN using OCT images prior to treatment and after three loading treatments[14]. A total of 90% of the synthetic OCT images produced by this model revealed pathological lesions similar to the actual post-treatment images. Based on these OCT images, clinicians assessed the treatment effect. The dry-up prediction sensitivity and specificity of IRF and SRF were 33.3% and 95.1%, and 21.2% and 95.1%, respectively. The addition of fluorescein angiography (FA) and indocyanine green angiography (ICGA) images improved the IRF and SRF to 33.3% and 98.4%, and 24.2% and 99.0% in the SRF.

In our study, pre-injection OCT images were used to predict the inactive state after three anti-VEGF injections. In contrast to previous studies[13,14], we utilized a convolutional neural network (CNN) instead of a GAN. The clinicians did not evaluate the pathological lesions in the GAN-generated images, and the rate of the inactive state was immediately depicted quantitatively. Specifically, the pre-injection images and the OCT image during the loading injection treatment were learned using the AI model in various ways (average, attention, LSTM, and concatenation), and the prediction value was enhanced. We chose fusion methods (concatenation, average, attention, LSTM) based on clinical considerations to enhance OCT image interpretability in nAMD treatment. Each method has a specific role: LSTM captures temporal information, attention focuses on critical regions, and average assesses overall disease severity. These choices ensure the model interprets temporal changes, highlights crucial areas, and assesses disease severity accurately. Most significantly, the concatenation fusion layer was selected to preserve detailed information without reducing the vector size, preventing potential information loss. This approach allows the model to capture and utilize a richer set of features for improved performance in predicting treatment outcomes for nAMD, demonstrating superior results.

Our study is the first study to attempt to compare the prediction results of AI with those of ophthalmologists and retinal specialists regarding whether the nAMD status would become inactive after treatment, as well as predict the results after three loading injections using not only images before treatment but also images taken during the treatment process. Upon predicting the treatment outcomes after three injections, retinal specialists demonstrated only approximately 80% accuracy in their predictions, as cases with slight SRF or IRF remain, rendering determination of treatment effectiveness challenging. The use of AI also did not demonstrate a significantly better performance than the experts in this complex task.

This study has many limitations. First, this was a retrospective study with a small number of patients and was conducted in only one hospital using only one OCT examination device. Second, the use of ranibizumab and aflibercept was not differentiated and the AMD subtypes were not categorized separately. Third, the study was conducted only in the Korean population, and the treatment response and performance of the patients of other races, especially in Western countries, should be investigated in future studies. Fourth, only anatomical responses were evaluated using OCT without assessing functional improvements such as visual acuity. Fifth, recently developed drugs such as brolucizumab and faricimab were not included in this study. Although the differences in these drugs could potentially lead to variations in anatomical outcomes, this study did not investigate differences between the drugs. Finally, the study did not investigate whether the performance of the model could be improved by using other modalities, such as FA or ICGA, in addition to OCT. Future directions such as functional improvement assessment, exploring other imaging modalities, and conducting multicenter trials with long-term follow-up would be invaluable in investigating the AI model's predictive accuracy.

In conclusion, this study developed an AI model that predicts the dry-up status after three loading treatments using OCT images before treatment in patients with nAMD and compared the model's performance with that of ophthalmologists. The model demonstrated a higher mean performance than ophthalmologists, and the model's treatment prediction performance further improved with additional OCT images during the loading phase. Future studies should address the limitations of the present study to improve the generalizability and clinical applicability of this model.

Methods

Ethics statement

This study adhered to the principles of the Declaration of Helsinki and was approved by the Institutional Review Board of Kong Eye Hospital (KIRB-202202-HR-001-01). The Institutional Review Board of Kong Eye Hospital waived the requirement for obtaining informed consent given that this was a retrospective observational study of medical records and was retrospectively registered.

Data collection and labelling

We conducted a thorough examination of the medical documents of individuals who were diagnosed with nAMD at the Kong Eye Hospital from January 2015 to June 2021[19]. Only patients who had not had any previous treatment for nAMD were included in the study. All participants were given three monthly injections of either ranibizumab or aflibercept. If both eyes were treated, a single eye was chosen at random. The exclusion criteria for this study included the following: patients with extrafoveal nAMD or non-exudative AMD; those who had more than a 6-week interval between three loading injections; individuals who had received prior treatment in the study eye with photodynamic therapy, subfoveal focal laser photocoagulation, or vitrectomy; patients who had received anti-VEGF injections other than ranibizumab and aflibercept; individuals with macular degeneration such as epiretinal membrane and macular hole; those with retinal vascular diseases such as retinal vein occlusion, retinal artery occlusion, and diabetic retinopathy; patients who had missed the OCT examination, and individuals who had undergone cataract surgery within 3 months.

All patients were assessed for their age, sex, presence of underlying diseases such as hypertension and diabetes, and any history of ocular surgery. Visual acuity, intraocular pressure, and fundus examinations were conducted, and the presence of neovascularization in the macula was verified using FA and ICGA. Optical coherence tomography (OCT) was conducted during each visit to assess changes in the macula. FA was conducted using the Heidelberg Retina Angiograph (HRA; Heidelberg Engineering, Heidelberg, Germany), whereas OCT was conducted using the Heidelberg Spectralis (Heidelberg Engineering, Heidelberg, Germany).

OCT scans were conducted prior to injection therapy and at every 4-week visit during injection therapy; treatment response was evaluated using OCT scans 4 weeks after the third injection. The procedure was conducted only for wet age-related macular degeneration located in the macula. Therefore, a 6×6 mm raster scan was performed four times around the macula. Among the 25 images obtained, one image where the lesion was centered was selected. The OCT images taken before the injection treatments were used as references to evaluate the response to subsequent treatments, utilizing the same areas in the OCT. A single picture was taken from a 25-volume scan that contained the AMD lesions. OCT images of the identical area were then acquired after the first, second, and third injections, utilizing a reference image to assure consistent scanning of the same area during each visit. The macular fluids were identified as IRF, SRF, and PED. A dry macula is characterized by the absence of IRF and SRF. Fluid located under the RPE was not taken into account for determining the presence of a dry macula, unless there was an observable increase in the volume of the PED since the previous visit. Regarding the treatment outcomes, a good response was assessed when both the IRF and SRF totally disappeared, indicating a successful dry-up. Conversely, a poor response was determined when the SRF and IRF persisted, with residual fluid still visible. When just the PED remained and no additional fluid was detected, it was determined to be completely dried. A dry macula was not taken into consideration when there were instances of new macular hemorrhage or increased macular edema shown on OCT.

Ground truth labels for treatment outcomes were established by two retinal specialists (DDH, JMH) with over 15 years of clinical experience, who determined whether the patients' conditions improved following anti-VEGF injections. Another retinal specialist (JSH) evaluated the discrepancy in cases where the two specialists had different opinions. Any discrepancies were resolved by consensus. They did not participate in the evaluation of the physicians' accuracy, ensuring the fairness of the performance comparison.

We utilized several data augmentation techniques to enhance the robustness of our model, including horizontal flipping, 20-degree rotation, and translation with width and height adjustments (width = 0.2, height = 0.2).

Model architecture

We proposed a deep learning-based model for predicting the nAMD treatment outcomes. The model architecture consisted of an OCT record encoder (ORE), sequence fusion layer, and treatment predictor. The ORE analyzes the SD-OCT image and extracts an image representation vector with 2,048 values from a single record of the SD-OCT image. We used a CNN-based model for ORE. At this stage, we did not apply any segmentation techniques, and the raw OCT images were directly processed by the CNN model. The sequence fusion layer fuses the representation vectors. Four baseline fusion methods are employed for the sequence fusion layer. If our model received a single record of an SD-OCT image, the fusion layer was eliminated. Figure 1 shows the details of the fusion architecture. The treatment predictor predicted the result after triple injections. The predictor is a multilayer perceptron (MLP).
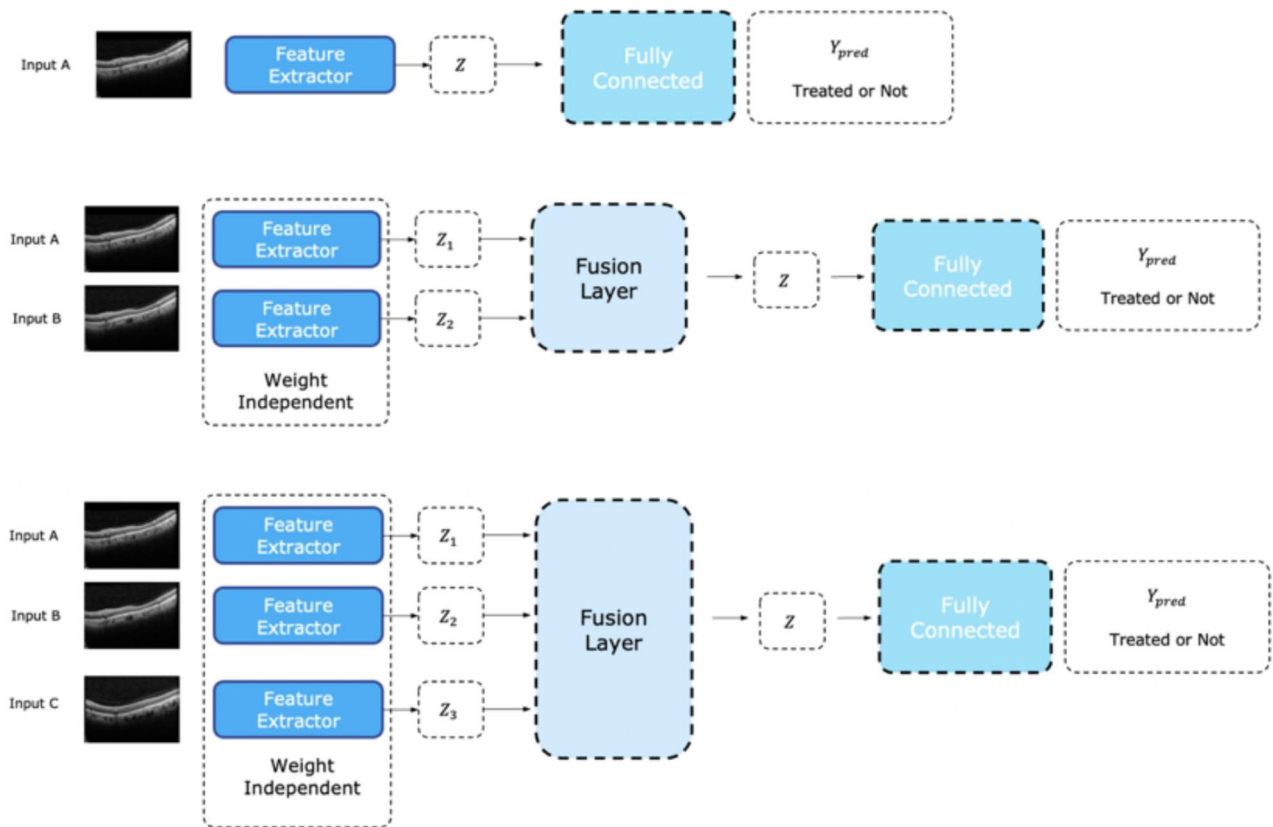
Fig. 1. Model Architecture consisting of optical coherence tomography (OCT) record encoder, sequence fusion layer, and treatment predictor. The OCT record encoder generates an image representation vector for each slice of the SD-OCT images. The sequence fusion layer is chosen from four different types of layers, namely, average fusion, concatenation fusion, attention fusion, and long-short term memory (LSTM), which is responsible for aggregating the image representation vectors from each slice that is generated by the OCT record encoder. Finally, the last fully connected layer predicts the treatment for each patient based on the aggregated image representation vectors.

In this study, we utilize original OCT images with a resolution of 764×490 pixels. To address GPU memory limitations and ensure robustness in our model, these images are resized to 224×224 pixels before being input into the network. For each input SD-OCT image, the ORE generates z_k from the input SD-OCT x_k. We adopted the DenseNet[20] architecture with 201 layers for ORE, which consists of four dense blocks and three transition layers. The decision to utilize DenseNet is driven by its unique feature reuse mechanism, which involves concatenating feature-maps learned from all previous layers with the feature-map learned from the current layer[18]. This approach significantly reduces the number of parameters and enhances parameter utilization. The adoption of DenseNet is grounded in its potential to capture richer hierarchical features from OCT images, thereby improving interpretability and predictive performance in predicting anatomical improvement during anti-VEGF therapy for nAMD. The dense block $F(x_{input})$ has 6–48 convolution blocks with 1×1 and 3×3 convolution layers. The bottleneck architecture was applied to a single dense block, in which the input of layer x_{input} was concatenated as follows: $y = [x_{input}, F(x_{input})]$. To improve information flow, dense connectivity is utilized where the layer x_l is calculated following $x_l = H_l([x_0, x_1, \ldots, x_{l-1}])$. The transition block is referred to as that between the dense blocks. It consists of a 1×1 convolution layer and a 2×2 average pooling layer with a stride of 2. The purpose of the layer is to reduce the size of the feature map and number of channels to avoid overfitting. ORE used 18,321,984 trainable parameters. To process the entire SD-OCT record after injection, we utilized a triple ORE without sharing model parameters. Figure 2 illustrates the details of ORE.

The sequence fusion layer combines the description vectors from ORE according to the sequence of the records. We investigated four fusion baselines as sequence fusion layers: *concat*, *average*, *attention*, and *long-short term memory (LSTM)*. The *concat* method combines the description vector of SD-OCT by concatenating it as follows: $Z_{concat} = [Z_1, Z_2, \ldots Z_k]$. The *average* method aggregates all the description vectors by averaging as: $Z_{average} = \frac{Z_1 + Z_2 + \cdots + Z_k}{k}$. The *attention* method summarizes all the description vectors, considering the importance of each vector. It stacks the k instance vectors $Z \in \mathbb{R}^{(k \times 2048)}$ from ORE and attention score $A \in \mathbb{R}^{(k \times 1)}$ after linear projection and *softmax* activation. LSTM is a deep-learning-based fusion layer that captures the features of a time sequence. In the sequence fusion layer, the LSTM considers.
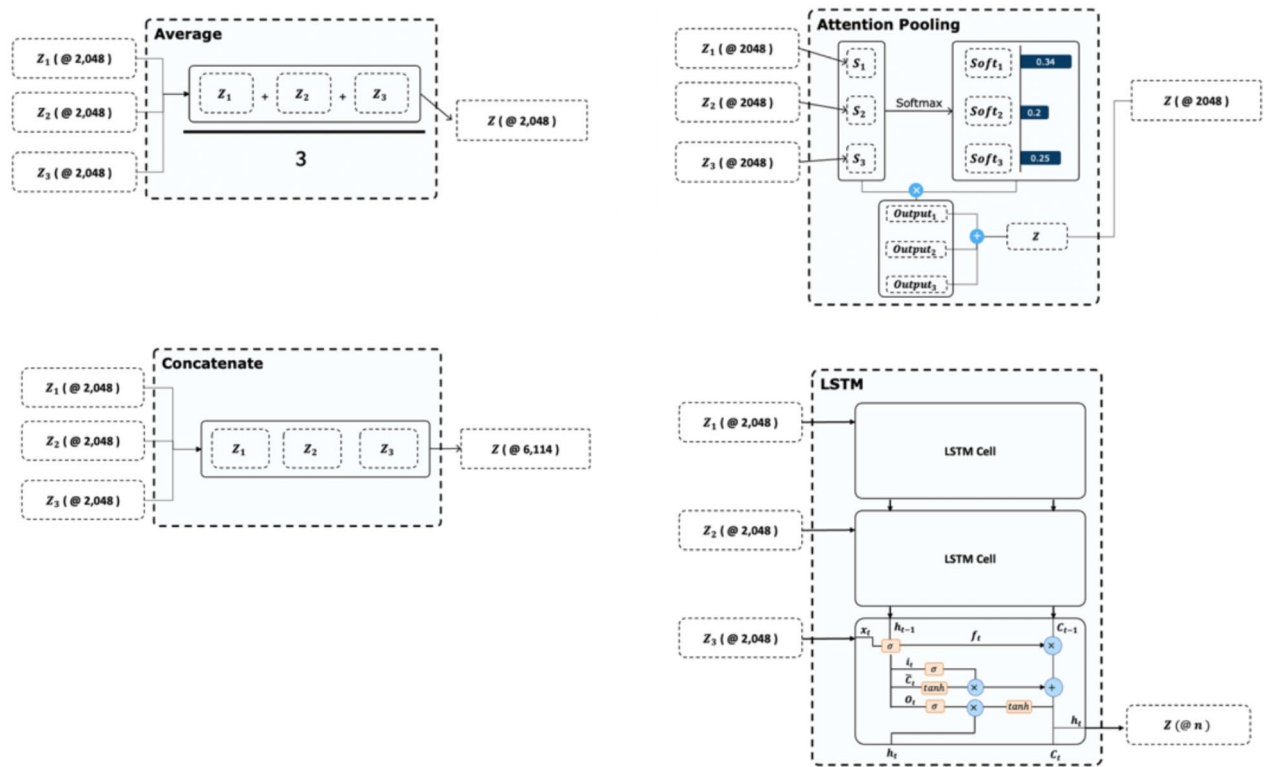
Fig. 2. Optical coherence tomography (OCT) record encoder. The architecture of the proposed model, which comprises several dense blocks that contain multiple convolutional layers with dense connectivity, is illustrated. Specifically, each layer within the same dense block is connected to every other layer, resulting in a highly connected network. Following each dense block, a transition layer performs spatial compression of the feature maps by applying batch normalization, ReLU activation, a 1×1 convolution layer, and a 2×2 average pooling layer. The output of the last dense block is fed into a global average pooling layer, which aggregates the feature maps and produces a fixed-size representation of the input data. This representation serves as a latent variable Z for next layer.

The treatment predictor predicted the results after three injection trials using latent vector records. The predictor consists of a dropout layer with 0.3 ratio and an output linear layer. The logit of the treatment predictor can be computed as $\widehat{y} = sigmoid\left(F\left(Dropout_{0.3}\left(Z\right)\right)\right)$. The $F\left(\cdot\right)$ is traditional MLP, which can be computed as: $F\left(x\right) = \sum_i w_i \cdot x_i + b$. The dropout layer prevented overfitting of our model. Sigmoid activation changes the output value according to the Bernoulli distribution, which ranges from zero to one where $\widehat{y}$ denotes the probability of treatment after three injections.

Experiment setup

Ten-fold cross-validation was performed to train and evaluate the proposed model. We split the entire dataset into ten different folds, trained the model with nine folds, and tested it with the remaining folds. Each fold had a distribution similar to that of the original dataset. We split the nine-fold dataset into training and validation datasets: eight folds are used for training the model and the remaining one-fold is for validation. We trained the model using 30 epochs with binary cross-entropy loss and selected the best epoch based on the validation set. An Adam optimizer with a learning rate of 0.0003 and batch size of four was utilized to optimize the proposed model. Data augmentation was applied to all the baseline models to avoid overfitting and build a robust model applicable to a variety of input images. We employed python3 and kerasAPI to train and evaluate the model. To evaluate our model from a clinical perspective, the prediction results for the test set were compared with those obtained by six ophthalmologists, including three ophthalmology residents and three experts, each with more than 10 years of clinical experience at an academic ophthalmology center. The conditions A, B, and C correspond to different sets of SD-OCT images provided for the prediction tasks. Specifically, Condition A provides only the OCT image taken before the first injection, Condition B provides the OCT images taken before the first injection and after the first injection, and Condition C provides all three OCT images, before the first injection, after the first injection, and after the second injection. This stepwise image masking allowed us to assess the impact of additional image data on both physician and model performance. To evaluate the model's clinical relevance, we selected the best-performing fold from the ten-fold cross-validation and provided the data to the ophthalmologists for binary classification of treatment outcomes.

Statistical analysis

All the statistical analyses were performed using the Scikit-learn modules. True positives, true negatives, false positives, and false negatives were calculated and compared with the predictions of the proposed model and true labels. We set the good responder classes as positive classes and the bad responder classes as false cases. The accuracy, precision, specificity, and sensitivity are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP},$$
$$Precision = \frac{TP}{TP + FP},$$
$$Specificity = \frac{TN}{TN + FP},$$
$$Sensitivity = \frac{TP}{TP + FN}.$$

TP, TN, FN, and FP denote true positives, true negatives, false negatives, and false positives, respectively. Additionally, agreement between ophthalmologists was measured using Fleiss' kappa[21] scores.

Data availability

The data are not available for public access because of patient privacy concerns, but are available from the corresponding author on reasonable request.

References

1. Wong, W. L. et al. 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**, e106-116. https://doi.org/10.1016/s2214-109x(13)70145-1 (2014).
2. Zou, M. et al. Variations and trends in global disease burden of age-related macular degeneration: 1990–2017. *Acta Ophthalmol.* **99**, e330–e335. https://doi.org/10.1111/aos.14589 (2021).
3. Arnold, J. J. et al. Two-year outcomes of "treat and extend" intravitreal therapy for neovascular age-related macular degeneration. *Ophthalmology* **122**, 1212–1219. https://doi.org/10.1016/j.ophtha.2015.02.009 (2015).
4. Holz, F. G. et al. Safety and efficacy of a flexible dosing regimen of ranibizumab in neovascular age-related macular degeneration: The SUSTAIN study. *Ophthalmology* **118**, 663–671. https://doi.org/10.1016/j.ophtha.2010.12.019 (2011).
5. Fung, A. E. et al. An optical coherence tomography-guided, variable dosing regimen with intravitreal ranibizumab (Lucentis) for neovascular age-related macular degeneration. *Am. J. Ophthalmol.* **143**, 566–583. https://doi.org/10.1016/j.ajo.2007.01.028 (2007).
6. American Society of Retina Specialists. Global trends in retina 2019. https://www.asrs.org/content/documents/2019-global-trends-survey-for-website.pdf.
7. Guymer, R. H. et al. Tolerating subretinal fluid in neovascular age-related macular degeneration treated with ranibizumab using a treat-and-extend regimen: FLUID study 24-month results. *Ophthalmology* **126**, 723–734. https://doi.org/10.1016/j.ophtha.2018.11.025 (2019).
8. Waldstein, S. M. et al. Morphology and visual acuity in aflibercept and ranibizumab therapy for neovascular age-related macular degeneration in the VIEW trials. *Ophthalmology* **123**, 1521–1529. https://doi.org/10.1016/j.ophtha.2016.03.037 (2016).
9. Fu, D. J. et al. Predicting incremental and future visual change in neovascular age-related macular degeneration using deep learning. *Ophthalmol. Retina* **5**, 1074–1084. https://doi.org/10.1016/j.oret.2021.01.009 (2021).
10. Romo-Bucheli, D., Erfurth, U. S. & Bogunovic, H. End-to-end deep learning model for predicting treatment requirements in neovascular AMD from longitudinal retinal OCT imaging. *IEEE J. Biomed. Health Inform.* **24**, 3456–3465. https://doi.org/10.1109/JBHI.2020.3000136 (2020).
11. Gallardo, M. et al. Machine learning can predict anti-VEGF treatment demand in a treat-and-extend regimen for patients with neovascular AMD, DME, and RVO associated macular edema. *Ophthalmol. Retina* **5**, 604–624. https://doi.org/10.1016/j.oret.2021.05.002 (2021).
12. Bogunovic, H. et al. Prediction of anti-VEGF treatment requirements in neovascular AMD using a machine learning approach. *Investig. Ophthalmol. Vis. Sci.* **58**, 3240–3248. https://doi.org/10.1167/iovs.16-21053 (2017).
13. Liu, Y. et al. Prediction of OCT images of short-term response to anti-VEGF treatment for neovascular age-related macular degeneration using generative adversarial network. *Br. J. Ophthalmol.* **104**, 1735–1740. https://doi.org/10.1136/bjophthalmol-2019-315338 (2020).
14. Lee, H., Kim, S., Kim, M. A., Chung, H. & Kim, H. C. Post-treatment prediction of optical coherence tomography using a conditional generative adversarial network in age-related macular degeneration. *Retina* **41**, 572–580. https://doi.org/10.1097/IAE.0000000000002898 (2021).
15. Hwang, D. D. et al. Distinguishing retinal angiomatous proliferation from polypoidal choroidal vasculopathy with a deep neural network based on optical coherence tomography. *Sci. Rep.* **11**, 9275. https://doi.org/10.1038/s41598-021-88543-7 (2021).
16. Yoon, J. et al. Classifying central serous chorioretinopathy subtypes with a deep neural network using optical coherence tomography images: A cross-sectional study. *Sci. Rep.* **12**, 422. https://doi.org/10.1038/s41598-021-04424-z (2022).
17. Ferrara, D., Newton, E. M. & Lee, A. Y. Artificial intelligence-based predictions in neovascular age-related macular degeneration. *Curr. Opin. Ophthalmol.* **32**, 389–396. https://doi.org/10.1097/ICU.0000000000000782 (2021).
18. Jung, J. et al. Prediction of neovascular age-related macular degeneration recurrence using optical coherence tomography images with a deep neural network. *Sci. Rep.* **14**, 5854 (2024).
19. Jung, J. et al. Prediction of neovascular age-related macular degeneration recurrence using optical coherence tomography images with a deep neural network. *Sci. Rep.* **14**, 5854. https://doi.org/10.1038/s41598-024-56309-6 (2024).
20. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.
21. Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **76**, 378 (1971).

Acknowledgements

Author contributions

D.D.H. and J.M.H. conceived and designed the study described here. D.D.H., J.M.H. and J.H. wrote the main manuscript text and J.M.H., J.H., J. K., J.J. designed the algorithm and experiments. J.M.H. collected the data and D.D.H., J.Y and J.S.H. verify the data. J.M.H., J.K., J.J. and J.Y. performed statistical analysis. All authors reviewed the manuscript. J.I.P., J.H.J, and J.S.H. advised in study design, data analysis methods. All authors reviewed and revised the manuscript prior to submission.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.D.-J.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.