Graefe's Archive for Clinical and Experimental Ophthalmology (2022) 260:1113–1123
https://doi.org/10.1007/s00417-021-05427-2

RETINAL DISORDERS

Prediction of postoperative visual acuity after vitrectomy for macular hole using deep learning–based artificial intelligence

Shumpei Obata[1] · Yusuke Ichiyama[1] · Masashi Kakinoki[1] · Osamu Sawada[1] · Yoshitsugu Saishin[1] · Taku Ito[1] · Mari Tomioka[1] · Masahito Ohji[1]

Received: 19 May 2021 / Revised: 20 August 2021 / Accepted: 19 September 2021 / Published online: 12 October 2021

Abstract

Purpose To create a model for prediction of postoperative visual acuity (VA) after vitrectomy for macular hole (MH) treatment using preoperative optical coherence tomography (OCT) images, using deep learning (DL)–based artificial intelligence.

Methods This was a retrospective single-center study. We evaluated 259 eyes that underwent vitrectomy for MHs. We divided the eyes into four groups, based on their 6-month postoperative Snellen VA values: (A) ≥ 20/20; (B) 20/25–20/32; (C) 20/32–20/63; and (D) ≤ 20/100. Training data were randomly selected, comprising 20 eyes in each group. Test data were also randomly selected, comprising 52 total eyes in the same proportions as those of each group in the total database. Preoperative OCT images with corresponding postoperative VA values were used to train the original DL network. The final prediction of postoperative VA was subjected to regression analysis based on inferences made with DL network output. We created a model for predicting postoperative VA from preoperative VA, MH size, and age using multivariate linear regression. Precision values were determined, and correlation coefficients between predicted and actual postoperative VA values were calculated in two models.

Results The DL and multivariate models had precision values of 46% and 40%, respectively. The predicted postoperative VA values on the basis of DL and on preoperative VA and MH size were correlated with actual postoperative VA at 6 months postoperatively ($P < .0001$ and $P < .0001$, $r = .62$ and $r = .55$, respectively).

Conclusion Postoperative VA after MH treatment could be predicted via DL using preoperative OCT images with greater accuracy than multivariate linear regression using preoperative VA, MH size, and age.

Keywords Artificial intelligence · Deep learning · Macular hole · Visual acuity · Prediction

Key messages

- Several factors have been reported to influence postoperative visual acuity after macular hole surgery, but the models using these predictive factors are not accurate enough to predict postoperative visual acuity.
- Our study is the first report of successful postoperative visual acuity prediction from preoperative optical coherence tomography images using deep learning-based artificial intelligence.
- Our deep learning - based artificial intelligence method showed better accuracy and stronger correlation with actual postoperative visual acuity than prediction of postoperative visual acuity using a multivariate linear regression model or an ordinal multinomial logistic regression model.

✉ Shumpei Obata
obata326@belle.shiga-med.ac.jp

[1] Department of Ophthalmology, Shiga University of Medical Science, 520 – 2192, Seta Tsukinowacho, Otsu, Shiga, Japan

Introduction

Macular holes (MHs) are full-thickness vertical retinal defects in the foveal neurosensory retina [1]. MHs cause visual loss and/or anorthopia. Internal limiting membrane (ILM) peeling and gas tamponade are surgical procedures considered important for MH treatment, both of which have been shown to improve the MH closure rate [2–4]. Patients with MHs can achieve improved visual acuity (VA) through surgical treatment, but it is difficult to predict postoperative VA in these patients. Several factors have been reported to influence postoperative VA, namely, patient age, MH size, symptom duration, preoperative VA, and foveal vessel density [5–11]. In clinical practice, preoperative VA is an important prognostic factor that is frequently considered the strongest predictor of postoperative VA [12–15]. However, preoperative VA alone is insufficient for the prediction of postoperative VA.

Deep learning (DL)–based artificial intelligence (AI) has recently attracted considerable global interest. DL uses a representation learning method with multiple levels of abstraction to process input data without the need for manual feature engineering; it automatically recognizes the complex structure of high-dimensional data by projecting these data onto a low-dimensional manifold [16]. DL has the potential to enable analysis of complex phenomena that are difficult for humans to analyze manually. In the field of ophthalmology, there have been several DL-based studies involving fundus photographs and optical coherence tomography (OCT). Major ophthalmic diseases have been investigated using OCT-focused DL technology; these include diabetic retinopathy, retinal vein occlusion, glaucoma, and age-related macular degeneration [17–24]. DL might enable the prediction of VA using OCT images in patients with age-related macular degeneration [25]. According to a PubMed search using the keywords "artificial intelligence, deep learning, prediction, retina" there are currently no published reports concerning the prediction of postoperative VA after vitrectomy for the treatment of MHs using DL analysis of preoperative OCT images. We hypothesized that the postoperative VA values predicted using DL-based AI would be more strongly correlated with actual VA at 6 months postoperatively, in comparison with postoperative VA values predicted using preoperative VA alone, a multivariate linear regression model based on preoperative factors, or an ordinal multinomial logistic regression model based on preoperative factors. In this study, we created a model for the prediction of postoperative VA from preoperative OCT images using DL-based AI and compared the results with predictions made using preoperative VA alone, a multivariate linear regression model based on preoperative factors, and an ordinal multinomial logistic regression model based on preoperative factors.

Methods

Patient population, inclusion criteria, and treatment protocol

This retrospective study protocol was approved by the Institutional Review Board/Ethics Committee Shiga University of Medical Science (Otsu, Japan). For this study, an opt-out consent process was used at Shiga University of Medical Science Hospital following approval by the Institutional Review Board. This study adhered to the tenets of the Declaration of Helsinki. We retrospectively reviewed the medical records of all patients who underwent vitrectomy for the treatment of idiopathic MH from January 2011 to April 2020 at Shiga University of Medical Science Hospital. Patients were excluded if they met any of the following criteria: presence of high myopia (axial length, $\geq$ 26.5 mm), history of previous vitreous surgery; presence of macular disease other than MH, and/or postoperative complications.

All patients underwent 23- or 25-gauge pars plana vitrectomy, performed by six retinal specialists. Phacoemulsification and intraocular lens implantation were performed before vitrectomy in nearly all phakic eyes at the time of MH surgery. Triamcinolone acetonide, alone or in combination with either indocyanine green or brilliant blue G, was sprayed gently onto the macular area to detect the ILM, which was then peeled away from the retina in the macular area. Finally, air or 20% sulfur hexafluoride (SF_6) gas tamponade was applied.

Deep learning model

Because of the lack of a large training dataset and considering the tradeoff between accuracy and computational speed, we built a deep neural network and trained it with the Adam optimizer [26]. We used batch normalization [25] (Fig. 1), which is a common approach for the acceleration of modern DL training methods. The batch size was 8, and the loss function was categorical cross-entropy. The learning rate was set to 0.0002. Our training data comprised horizontal and vertical scans of OCT images (taken by Cirrus OCT, Carl Zeiss Meditec, Dublin, CA, USA), coupled with postoperative VAs as ground truth. The center of the original image was cropped at 1/7 of the image height + 6 pixels from the top, 1/4 of the image height from the bottom, and 1/4 of the image width from both the left and right; it was resized using the OpenCV INTER_AREA algorithm. The resolution and size of the input images were 640 × 512 pixels and 169.35 × 135.47 mm, respectively. Data augmentation

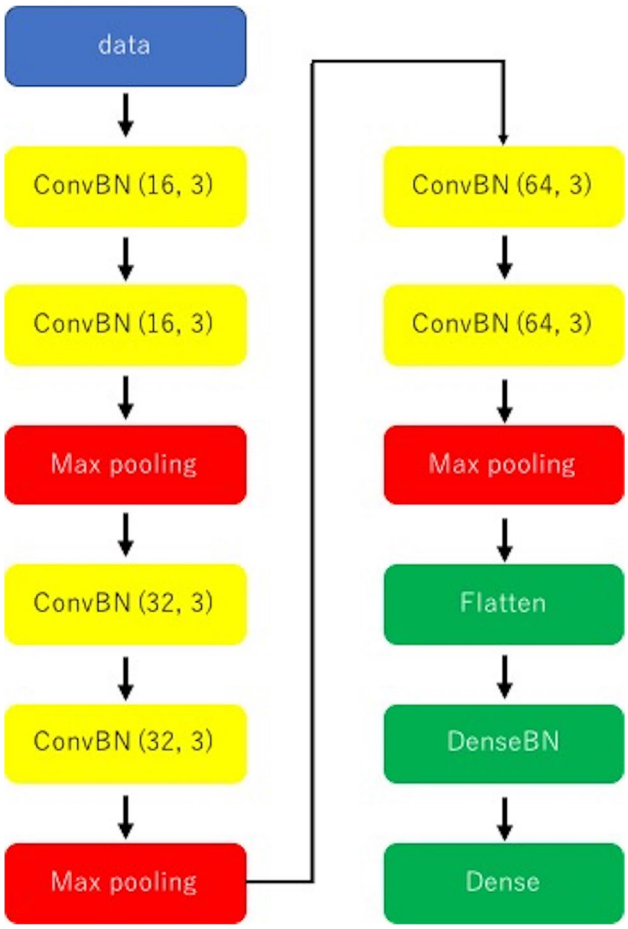

Fig. 1 Deep neural network architecture used in this study. ConvBN (16, 3): convolution batch normalization, 16 filters, 3×3 kernel. A method to stabilize and accelerate learning by allowing each layer to learn independently after the convolutional layer. The output of the convolutional layer will be 16 channels, with a kernel size of 3×3. Filtering on nearby nodes in the previous layer to obtain a "feature map." Max pooling: the feature map output from the convolutional layer is further reduced to create a new feature map, which extracts only the larger values from the filtered ones. Flatten: converting data dimension to one dimension. DenseBN: batch normalization after the dense layer. Dense: a layer where all the neurons in the front and back layers are connected

techniques (e.g., horizontal flip, scale, shift, rotate, shear, crop, gamma contrast, and multiply) were programmed using the Python module "imgaug"; these were then applied to the images in the training dataset to increase the amount of training data.

For the DL model, we divided the patients into four groups (A, B, C, and D), based on their postoperative decimal VA (Snellen VA) values: (A) ≥ 1.0 (20/20), (B) 0.9–0.7 (approximately 20/25–20/32), (C) 0.6–0.3 (approximately 20/32–20/63), and (D) ≤ 0.2 (20/100). These categories were defined as follows: ≥ 1.0 (20/20) was considered normal; ≥ 0.7 (approximately 20/32) is adequate for a driver's license in Japan; ≥ 0.3 (approximately 20/63) [27] is adequate for reading and writing; and ≤ 0.2 (20/100) is considered poor VA.

In this model, the input was the preoperative OCT images and the output was categories of postoperative VA (A, B, C, and D). Following categorization, the resulting predicted postoperative VAs were converted into a single predicted postoperative VA using regression analysis.

We used a simple but very successful approach to regularization, known as label smoothing [28]. If the actual postoperative VA was in group A, we weighted the data at 65% toward group A, 25% toward group B, and 10% toward group C, rather than 100% toward group A during training. If the actual postoperative VA was in group B, we weighted the data at 25% toward group A, 50% toward group B, and 25% toward group C, rather than 100% toward group B. If the actual postoperative VA was in group C, we weighted the data at 25% toward group B, 50% toward group C, and 25% toward group D, rather than 100% toward group C. If the actual postoperative VA was in group D, we weighted the data at 10% toward group B, 25% toward group C, and 65% toward group D, rather than 100% toward group D. Because groups A and D were the boundary/limiting groups in terms of actual postoperative decimal VA, our label smoothing could not include values at both ends of the true value of those groups; therefore, we modified the weighting values, compared with weighting values used for groups B and C. We also produced saliency heatmaps using gradient-weighted class activation mapping (Grad-CAM) [29].

The number of eyes in each group was as follows: (A) 68 (26%), (B) 81 (31%), (C) 85 (33%), and (D) 25 (10%). Group D had the lowest number of eyes, so the sizes of training and test datasets were established in accordance with the size of group D. OCT data were randomly divided into 80% (20 eyes/40 images) for training and 20% (5 eyes/10 images) for testing in group D. Because imbalances among categories are commonly regarded as potential reasons for unsatisfactory training results in machine learning studies [30], we used a similar number of training images ($n = 40$) for each group. Training data were randomly selected for all groups from among images of eyes that met the corresponding group criteria. Test data were randomly selected at the same proportions as those of each group in the total data: A, 14 (26%); B, 16 (31%); C, 17 (33%); and D, 5 (10%) (Table 1). The test data were used to evaluate the final performance of the trained model.

Preoperative VA model and multivariate linear regression model

Subsequently, a model for the prediction of postoperative VA from preoperative VA (i.e., a preoperative VA model) and a model for the prediction of postoperative VA from preoperative predictors of postoperative VA (i.e., a multivariate

Table 1 The number of eyes in each group A–D; total data, training data, testing data, and unused data

	Total data	Training data	Testing data	Unused data
Group A (eyes) Decimal VA: ≧ 1.0	68 (26%)	20	14 (27%)	34
Group B (eyes) Decimal VA: 0.9–0.7	81 (31%)	20	16 (31%)	45
Group C (eyes) Decimal VA: 0.6–0.3	85 (33%)	20	17 (33%)	48
Group D (eyes Decimal VA: ≦ 0.2	25 (10%)	20	5 (10%)	0
Total (eyes)	259 (100%)	80	52(100%)	127

linear regression model) were created using the 80 eyes that had been included as training data in the DL model. For the preoperative VA model, we calculated the correlation coefficients between the preoperative and postoperative VA; predicted postoperative VA values were calculated using the correlation equation obtained from this calculation. Several factors have been reported to influence postoperative VA: patient age, MH size, symptom duration, preoperative VA, and foveal vessel density [5–11]. For the multivariate linear regression model, preoperative VA, MH size, and patient age were used as independent variables, for which information was available in the current study.

Ordinal multinomial logistic regression model

We also created an ordinal multinomial logistic regression model for the prediction of postoperative VA groups from preoperative predictors of postoperative VA (i.e., an ordinal multinomial logistic regression model). Several factors have been reported to influence postoperative VA: patient age, MH stage and size, symptom duration, preoperative VA, and foveal vessel density [5–11]. For the ordinal multinomial logistic regression model, we used the following information from the 52 eyes that had been included as test data in the DL model: patient age, preoperative VA, and minimum MH size. After the exclusion of variables with a variance inflation factor > 10, preoperative VA and minimum MH size remained independent variables in the ordinal multinomial logistic regression model. The precision (positive predictive value) and recall (sensitivity) of this ordinal multinomial logistic regression model were evaluated using the 52 eyes that had been included as test data in the DL model.

Model evaluation

We evaluated the accuracy of these four models, as well as the precision and recall in each group, using the 52 eyes that had been included as test data in the DL model. The accuracy is the proportion of correct predictions (both true positives and true negatives) among the total number of eyes. We also calculated the correlation coefficient and mean absolute error (MAE), performed Bland–Altman analysis, and compared area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) in each group between the predicted postoperative VA and the actual postoperative VA in the DL model, preoperative VA model, and multivariate linear regression model. The ordinal multinomial logistic regression model was excluded from these comparisons because it could not predict postoperative VA values.

Finally, to determine whether the VA in an eye could be accurately predicted, the test data were divided into two groups: eyes with correctly predicted VA and eyes with incorrectly predicted VA. Differences in preoperative and intraoperative factors (e.g., age, sex, right or left side, preoperative VA, axial length, MH stage and estimated duration, surgical time, ILM peeling or inverted treatment, dye liquid, and tamponade substance) were compared between eyes with correctly predicted VA and eyes with incorrectly predicted VA.

Statistical analyses

The best-corrected visual acuity (BCVA) was measured using a Landolt C chart in this study. The decimal VA was converted to Snellen VA and the logarithm of the minimum angle of resolution (logMAR) for statistical analyses. The statistical analyses were performed using EZR version 1.52 (Saitama Medical Center, Jichi Medical University, Saitama, Japan) [31], which is a graphical user interface for R version 4.03 (The R Foundation for Statistical Computing, Vienna, Austria), and GraphPad Prism 9 software (GraphPad Software, Inc., La Jolla, CA, USA). Results were expressed as mean ± standard deviation for continuous variables and as proportion (percentage) for categorical variables. The Mann–Whitney *U* test was used for two-group comparisons. Spearman correlation coefficients were calculated to determine the relationships between actual and predicted postoperative VA values from the DL, preoperative VA, and multivariate linear regression models. Differences with $P < 0.05$ were considered to be statistically significant.

A

		Predicted postoperative VA					
		A	B	C	D	Recall	n
Actual Postoperative VA	A	6	4	4	0	43%	14
	B	1	6	9	0	35%	16
	C	1	4	10	2	63%	17
	D	0	0	3	2	40%	5
	Precision	75%	43%	38%	50%		

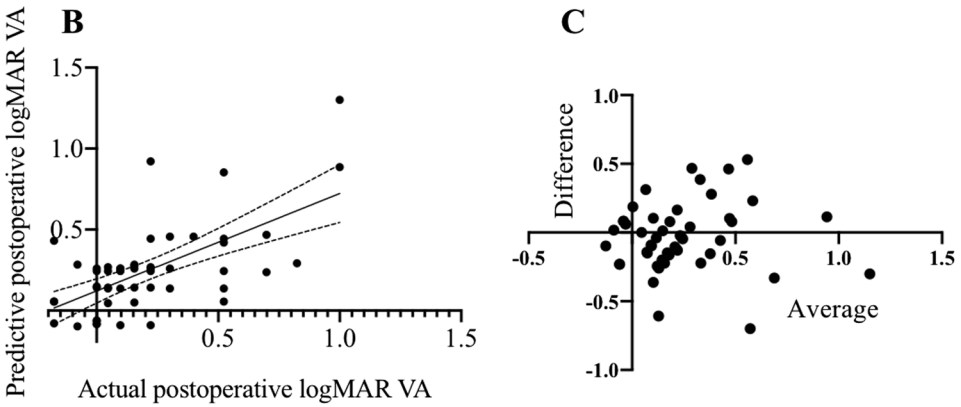

Fig. 2 Deep learning (DL) model. (**A**) Precision and recall. (**B**) Correlation coefficients between the deep learning–based predicted postoperative logarithm of the minimum angle of resolution (logMAR) visual acuity (VA) and actual postoperative VA (test data: 52 eyes/104 images). (**C**) Bland–Altman plot

Results

During the study period, 259 eyes with idiopathic MHs that were treated surgically met the abovementioned inclusion/exclusion criteria. In total, 140 eyes were in female patients (54%) and the median overall patient age was 67.5 years (range, 30–85 years). Of the total, 126 eyes (48%) were right eyes. ILM peeling was performed in 251 eyes (97%) and an inverted ILM flap procedure was performed in 8 eyes (3%). SF_6 tamponade was performed in 239 eyes (92%) and air tamponade was performed in 20 eyes (8%). Twenty-one eyes (8%) were pseudo-phakic at the time of surgery and 238 eyes (92%) were phakic. Cataract surgery was performed during MH surgery in 236 of the 238 phakic eyes (99%), and lens-sparing surgery was performed in the other 2 eyes (1%). The mean preoperative BCVA was 0.59 ± 0.28 logMAR; the mean postoperative BCVA significantly improved to 0.21 ± 0.26 logMAR ($P < 0.0001$). MHs closed in 250 eyes (97%) following the first surgery and closed in all 259 eyes (100%) following additional surgeries.

DL model for prediction of postoperative VA

The model precision was 46%, according to analysis of the test data. The precision and recall in each group are shown in Fig. 2. The predicted postoperative VA based on DL was correlated with actual postoperative VA at 6 months postoperatively in training data ($P < 0.0001$, $r = 0.81$, MAE = 0.15) and in test data ($P < 0.0001$, $r = 0.62$, MAE = 0.186) (Fig. 2). The AUROC values in each group (A through D) were 0.75, 0.57, 0.65, and 0.85, respectively. The AUPRC values in each group (A through D) were 0.53, 0.52, 0.49, and 0.52, respectively. Only random errors were apparent, according to Bland–Altman analysis (Fig. 2). We could not determine the regions of OCT images that were the focus of our DL-based AI predictions (Fig. 3). The training curve is shown in Fig. 4.

Preoperative VA model for prediction of postoperative VA

The model precision was 40%, according to analysis of the test data. The precision and recall in each group are shown in Fig. 5. The actual postoperative VA at 6 months

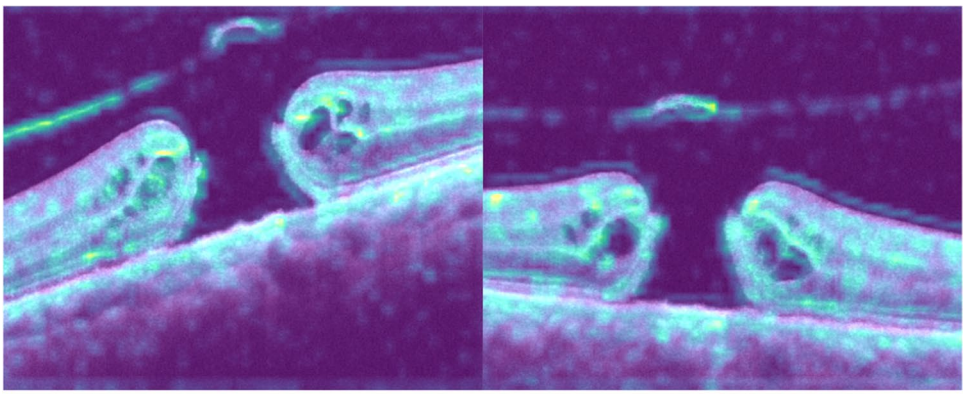

Fig. 3 Saliency heatmaps using gradient-weighted class activation mapping

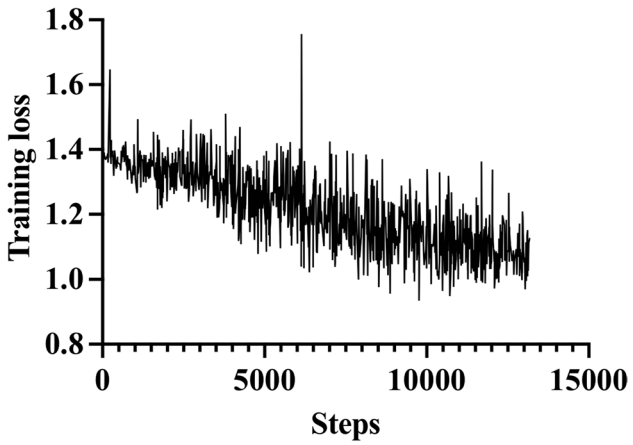

Fig. 4 Training curve of deep learning (DL) model. Training loss: output value of a loss function during training. Loss function is a function defined to evaluate discrepancy between prediction calculated using current trained parameters and ground-truth values during model training. Training loss keeps decreasing until the training process is convergent in theory. Although there is no guarantee, the model obtained by a well-convergent training usually has good accuracy

postoperatively was correlated with preoperative VA ($P < 0.0001$, $r = 0.65$). The predicted postoperative VA based on preoperative VA was correlated with actual VA at 6 months postoperatively in the test data ($P = 0.0003$, $r = 0.48$, MAE = 0.19) (Fig. 5). The AUROC values in each group (A through D) were 0.70, 0.54, 0.62, and 0.74, respectively. The AUPRC values in each group (A through D) were 0.52, 0.50, 0.59, and 0.43, respectively.

Multivariate linear regression model for prediction of postoperative VA

Multivariate linear regression showed that preoperative VA, minimum MH size, and patient age were significantly associated with postoperative VA ($P < 0.0001$, $P = 0.0009$, and $P = 0.08$, respectively; Table 2). The model precision was 40%, according to the analysis of test data. The precision and recall in each group are shown in Fig. 6. The predicted postoperative VA based on preoperative VA was correlated with actual VA at 6 months postoperatively, according to analysis of the test data ($P < 0.0001$, $r = 0.55$, MAE = 0.194) (Fig. 6). The AUROC values in each group (A through D) were 0.67, 0.57, 0.58, and 0.88, respectively. The AUPRC values in each group (A through D) were 0.42, 0.45, 0.38, and 0.52, respectively.

Ordinal multinomial logistic regression model for prediction of postoperative VA

The model precision was 37%, according to the analysis of test data. The precision and recall in each group are shown in Table 3. Variance inflation factor assessment was used to check for multicollinearity. None of the variance inflation factor values was > 10, implying no collinearity in the model. Residual deviance was 121.9 and Akaike's information criterion was 131.9. The odds ratio, confidence interval, and *P*-value of each independent variable are shown in Table 4.

Differences in preoperative and intraoperative factors between eyes with correctly predicted VA and eyes with incorrectly predicted VA are shown in Table 5. None of the tested factors differed significantly between eyes with correctly predicted VA and eyes with incorrectly predicted VA.

Discussion

We are unaware of any previous reports on prediction of postoperative VA after vitrectomy for MHs from preoperative OCT images using DL-based AI; moreover, we could find no references to such reports in PubMed. Thus, we presume that this is the first report on prediction of postoperative VA from preoperative OCT using DL-based AI.

A

		Predicted postoperative VA					
		A	B	C	D	Recall	n
Actual Postoperative VA	A	5	2	7	0	36%	14
	B	2	4	10	0	25%	16
	C	1	4	12	0	71%	17
	D	0	1	4	0	0%	5
	Precision	63%	36%	36%			

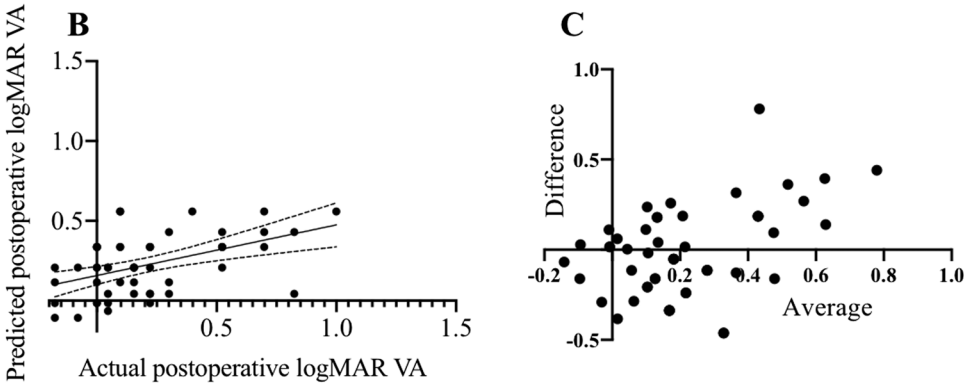

Fig. 5 Preoperative visual acuity (VA) model. (**A**) Precision and recall. (**B**) Correlation coefficients between the predicted postoperative logarithm of the minimum angle of resolution (logMAR) visual acuity (VA) from preoperative VA and actual postoperative VA (test data: 52 eyes/104 images). (**C**) Bland–Altman plot

Table 2 Multivariate linear regression analysis

	β value	SE	95% CI	*t* value	*P* value
Preoperative visual acuity	0.52	0.11	0.31 to 0.74	4.9	< 0.0001
Age (years)	0.009	0.005	− 0.0009 to 0.018	1.8	0.08
Minimum MH size (μm)	0.0005	0.00013	0.0002 to 0.0008	3.5	0.0009

SE, standard error; *CI*, confidence interval

Notably, our predictions were good despite the small number of eyes in this study. Preoperative VA is the strongest predictor of postoperative VA according to several previous reports [12–15]. Additionally, in comparison with predictions made on the basis of preoperative VA, a multivariate linear regression model constructed using preoperative factors, or an ordinal multinomial logistic regression model constructed using preoperative factors, our DL-based AI model predicted postoperative VA with similar or slightly better precision; the correlation coefficient of the DL model was better than the correlation coefficients of the preoperative VA model or the multivariate linear regression model.

To efficiently train data, first, we used the Adam optimization algorithm and convolutional batch normalization; for each group, the data augmentation methods (horizontal flip, scale, shift, rotate, shear, crop, gamma contrast, and multiply) and number of eyes were identical. Second, we used a simple but very successful approach to regularization, known as label smoothing (detailed in the "Methods" section) [28]. Label smoothing can help to avoid overconfident models while enhancing overall accuracy and learning speed [28]. Finally, we focused the analysis on the center part of each image, which was expected to be most important for the prediction of postoperative VA. Thus, we successfully created a model for the prediction of postoperative VA after vitrectomy for MHs, despite the small number of included images.

In the present study, the DL model showed the highest accuracy among the tested models. In the DL model, precision and recall tended to be relatively high in most groups.

A

		Predicted postoperative VA					
		A	B	C	D	Recall	n
Actual Postoperative VA	A	2	8	4	0	14%	14
	B	3	6	7	0	38%	16
	C	2	3	12	0	71%	17
	D	0	0	4	1	20%	5
	Precision	29%	35%	44%	100%		

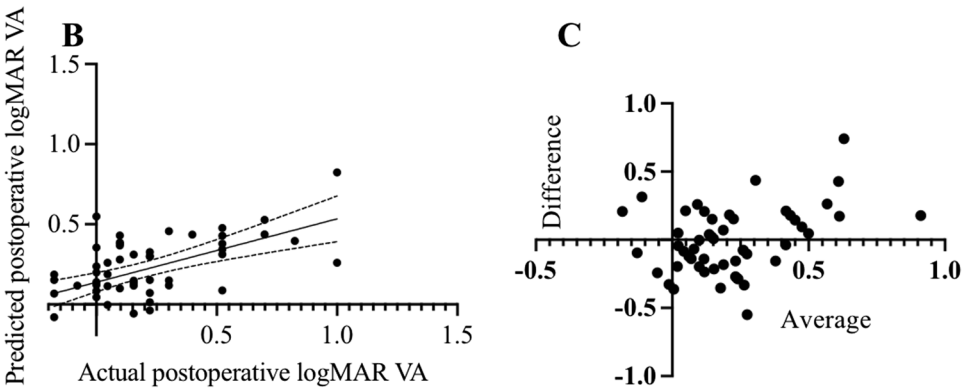

Fig. 6 Multivariate linear regression model. (**A**) Precision and recall. (**B**) Correlation coefficients between the predicted postoperative logarithm of the minimum angle of resolution (logMAR) visual acuity (VA) from preoperative VA and minimum size of macular hole, and actual postoperative VA (test data: 52 eyes/104 images). (**C**) Bland–Altman plot

Table 3 The ordinal multinomial logistic regression model using preoperative VA and minimum size of MH

		Predicted postoperative VA					
		A	B	C	D	Recall	Total
Actual postoperative VA	A (eyes)	5	6	3	0	36%	14
	B (eyes)	8	1	7	0	6.3%	16
	C (eyes)	4	1	12	0	71%	17
	D (eyes)	0	0	4	1	20%	5
	Precision	29%	13%	48%	100%		

Table 4 Ordinal multinomial logistic regression analysis

Independent variables	Odds ratio	95% CI	*P* value
Preoperative VA	1	1–1.01	0.02
Minimum MH size	11.6	1.38–113	0.03

CI, confidence interval

In contrast, the preoperative VA model had worse recall in groups B and D, the multivariate linear regression model had worse recall in groups A and D, and the ordinal multinomial logistic regression model had worse recall in groups B and D. Bland–Altman analyses showed that the DL model had random error, indicating that this model was ideal. In eyes with poor actual VA and poor predicted VA, the predicted VA tended to be better than the actual VA; in eyes with good actual VA and good predicted VA, the predicted VA tended to be worse than the actual VA. The AUROC and AUPRC tended to be greater in the DL model than in the other models. Because the ordinal multinomial logistic regression model could not calculate a single postoperative VA, it could not be included in the Bland–Altman analysis or in analyses of AUROC and AUPRC.

The DL model precision was 46%, according to the analysis of test data; however, postoperative VA could not be successfully predicted in 54% of eyes. To investigate whether specific eyes exhibited features that facilitated predictability, we divided the eyes into those with correctly

Table 5 The differences in the preoperative and intraoperative factors between eyes predicted correctly and incorrectly

	Eyes predicted correctly 24 eyes	Eyes predicted incorrectly 28 eyes	*P* value
Females/males (eyes)	16/9	15/12	0.58
Age (years)	65.2 ± 9.1	68.8 ± 5.4	0.43
Right/left (eyes)	12/13	14/13	> 0.99
Preoperative logMAR VA	0.56 ± 0.28	0.54 ± 0.22	0.85
Axial length (mm)	23.9 ± 1.4	23.6 ± 1.1	0.58
Stage of MH (1/2/3/4, eyes)	1/11/8/4	0/14/7/5	0.69
Estimated duration of MH (days)	126 ± 113	113 ± 179	0.72
Phakic/intraocular lens	23/1	25/5	0.21
Combined cataract surgery (eyes)	22	23	0.43
Surgical time (min)	36 ± 11	38 ± 16	0.85
ILM peeling/inverted flap (eyes)	23/1	28/0	0.46
Dyeing liquid (BBG/TA/ICG, eyes)	18/5/1	18/10/0	0.31
Tamponade substance (SF6/air, eyes)	22/2	28/0	0.20
Initial MH closure rate	96%	93%	> 0.99
Final MH closure rate	100%	100%	> 0.99

predicted VA and those with incorrectly predicted VA. We then examined differences in preoperative and intraoperative factors between the two groups. However, we found no significant differences between the groups; thus, it remains unclear why VA was not correctly predicted in some eyes.

Because the lack of accurate prediction may have involved cases in which the MH did not close during the first surgery, we examined these cases in detail. The preoperative and postoperative characteristics of MHs that did not close after the first surgery are shown in Table 6. Of the 9 eyes with unclosed MHs, 2 eyes were assigned to the training data, 3 eyes to the testing data, and 4 eyes were assigned to neither the training nor the testing data. Of the 3 eyes assigned to the testing data, 1 case was correctly predicted and 2 were incorrectly predicted; the prediction accuracy was 33%. However, due to the small number of cases, we could not conclude whether the prediction accuracy was poor in cases where the MH did not close after the first surgery. The predicted and actual VA for each eye in which the MH did not close after the first surgery are listed below. In case 1, the predicted decimal VA was 0.6–0.3 and the actual decimal VA was 0.3; postoperative VA was correctly predicted in this case. In case 2, the predicted decimal VA was 0.6–0.3; however, the actual decimal VA was 0.15. In case 3, the predicted decimal VA was less than 0.2 and the actual decimal VA was 0.3. Postoperative VA was incorrectly predicted in cases 2 and 3. In a previous report, poor VA and large MH were risk factors for MH not closing [32]. In our study, the reasons why some MHs did not close after the first surgery were also considered to be poor preoperative VA (mean preoperative logMAR VA: 0.82) and large MH (mean MH size: 423 μm).

Table 6 Preoperative and postoperative characteristics of macular holes that did not close after the first surgery

	9 eyes
Training/testing/unused	2/3/4
Female/male (no. eyes)	5/4
Age (years)	68.2 ± 7.2
Right / Left (eyes)	2/7
Preoperative logMAR VA	0.82 ± 0.2
Axial length (mm)	23.4 ± 0.9
Stage of MH: 1/2/3/4 (no. eyes)	0/3/3/3
Estimated duration of MH (days)	233 ± 220
Combined cataract surgery/already underwent cataract surgery (no. eyes)	6/3
Surgical time (minutes)	41.0 ± 15
ILM peeling/inverted flap (no. eyes)	9/0
Dye: BBG/TA/ICG (no. eyes)	6/3/0
Tamponade substance: SF_6/air, (no. eyes)	7/2
Minimum MH size (μm)	422.9 ± 143
Postoperative logMAR VA	0.52 ± 0.32

Abbreviations: VA, visual acuity; *logMAR*, logarithm of the minimum angle of resolution; *MH*, macular hole; *ILM*, internal limiting membrane; *BBG*, brilliant blue G; *TA*, triamcinolone acetonide; *ICG*, indocyanine green; *SF_6*, sulfur hexafluoride

A notable strength of this study is that it appears to be the first report of successful postoperative VA prediction from preoperative OCT images using DL-based AI. However, this study had some limitations. First, we included a small number of eyes. To improve prediction accuracy, this

method should be validated using a large amount of training data in a future study. In particular, a large training dataset will enable us to rebuild our DL network with deeper architecture, which may lead to much more accurate predictions. Most DL research thus far indicates that a deeper network often provides better performance; a disadvantage of this is the exponential increase in the amount of training data required. Second, eyes with high myopia (axial length ≥ 26.5 mm) were excluded because the morphology of the retina in OCT images of such eyes differs considerably from that in eyes with a normal axial length owing to the presence of staphyloma; moreover, high myopia can cause vision loss. Finally, we could not determine the regions of OCT images that were the focus of our DL-based AI predictions. Further DL studies may elucidate these regions providing a more useful clinical tool. Overall, a highly accurate model would help patients to more fully consider providing preoperative informed consent and may be useful for clinical decision making.

Conclusion

Our study is the first report of successful postoperative VA prediction from preoperative OCT images using DL-based AI. Our DL-based AI method showed slightly better accuracy and stronger correlation with actual postoperative VA than prediction of postoperative VA based on preoperative VA alone (simplest and strongest predictive factor for postoperative VA). Furthermore, our DL-based AI method showed slightly better accuracy and stronger correlation with actual postoperative VA than prediction of postoperative VA using a multivariate linear regression model or an ordinal multinomial logistic regression model. Our DL-based AI method using OCT images may be a useful tool for the prediction of postoperative VA. Further studies with a larger number of eyes are warranted.

Acknowledgements The analysis using AI based on DL was outsourced to and conducted in collaboration with Tiwaki Co., Ltd. We thank Yoshiki Ohashi, Taro Watasue, Tomohiro Nakagawa, and Xiang Ruan for contributions to our deep learning analysis. We thank Ryan Chastain-Gross, Ph.D., from Edanz Group (https://en-author-services.edanz.com/ac) for editing a draft of this manuscript. The Clinical Research and Advanced Center at Shiga University of Medical Science assisted with statistical analysis.

Funding Funding for this study was provided by Shiga University of Medical Sciences. The Department of Ophthalmology (Shiga University of Medical Science) provides financial support to Tiwaki Co., Ltd. The funding body had no further input into the collection, analysis, and interpretation of the data, or manuscript preparation.

Declarations

Ethics approval All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the Institutional Review Board (IRB)/Ethics Committee Shiga University of Medical Science (Otsu, Japan).

Consent to participate For this type of study, formal consent was not required.

Conflict of interest The authors declare no competing interests.

References

1. Gass JD (1988) Idiopathic senile macular hole. Its early stages and pathogenesis. Arch Ophthalmol 106:629–639. https://doi.org/10.1001/archopht.1988.01060130683026
2. Kelly NE, Wendel RT (1991) Vitreous surgery for idiopathic macular holes. Results of a pilot study. Arch Ophthalmol 109:654–659. https://doi.org/10.1001/archopht.1991.01080050068031
3. Park DW, Sipperley JO, Sneed SR, Dugel PU, Jacobsen J (1999) Macular hole surgery with internal-limiting membrane peeling and intravitreous air. Ophthalmology 106:1392–1398. https://doi.org/10.1016/s0161-6420(99)00730-7
4. Brooks HL (2000) Macular hole surgery with and without internal limiting membrane peeling. Ophthalmology 107:1939–1948. https://doi.org/10.1016/s0161-6420(00)00331-6
5. Itoh Y, Inoue M, Rii T, Hiraoka T, Hirakata A (2012) Correlation between length of foveal cone outer segment tips line defect and visual acuity after macular hole closure. Ophthalmology 119:1438–1446. https://doi.org/10.1016/j.ophtha.2012.01.023
6. Baba T, Kakisu M, Nizawa T, Oshitari T, Yamamoto S (2017) Superficial foveal avascular zone determined by optical coherence tomography angiography before and after macular hole surgery. Retina 37:444–450. https://doi.org/10.1097/IAE.0000000000001205
7. Amram AL, Mandviwala MM, Ou WC, Wykoff CC, Shah AR (2018) Predictors of visual acuity outcomes following vitrectomy for idiopathic macular hole. Ophthalmic Surg Lasers Imaging Retina 49:566–570. https://doi.org/10.3928/23258160-20180803-03
8. Kim SH, Kim HK, Yang JY, Lee SC, Kim SS (2018) Visual recovery after macular hole surgery and related prognostic factors. Korean J Ophthalmol 32:140–146. https://doi.org/10.3341/kjo.2017.0085
9. Kumagai K, Ogino N, Demizu S, Atsumi K, Kurihara H, Iwaki M, Ishigooka H, Tachi N (2000) Factors related to initial success in macular hole surgery. Nippon Ganka Gakkai Zasshi 104:792–796
10. Kokame GT, de Bustros S (1995) Visual acuity as a prognostic indicator in stage I macular holes. Am J Ophthalmol 120:112–114. https://doi.org/10.1016/s0002-9394(14)73768-7
11. Ullrich S, Haritoglou C, Gass C, Schaumberger M, Ulbig MW, Kampik A (2002) Macular hole size as a prognostic factor in macular hole surgery. Br J Ophthalmol 86:390–393. https://doi.org/10.1136/bjo.86.4.390
12. Scott RA, Ezra E, West JF, Gregor ZJ (2000) Visual and anatomical results of surgery for long standing macular holes. Br J Ophthalmol 84:150–153. https://doi.org/10.1136/bjo.84.2.150
13. Jaycock PD, Bunce C, Xing W, Thomas D, Poon W, Gazzard G, Williamson TH, Laidlaw DA (2005) Outcomes of macular hole surgery: implications for surgical management and clinical

governance. Eye (Lond) 19:879–884. https://doi.org/10.1038/sj.eye.6701679
14. Gupta B, Laidlaw DA, Williamson TH, Shah SP, Wong R, Wren S (2009) Predicting visual success in macular hole surgery. Br J Ophthalmol 93:1488–1491. https://doi.org/10.1136/bjo.2008.153189
15. Larsson J, Holm K, Lovestam-Adrian M (2006) The presence of an operculum verified by optical coherence tomography and other prognostic factors in macular hole surgery. Acta Ophthalmol Scand 84:301–304. https://doi.org/10.1111/j.1600-0420.2006.00672.x
16. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539
17. Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY (2017) Deep-learning based, automated segmentation of macular edema in optical coherence tomography. Biomed Opt Express 8:3440–3448. https://doi.org/10.1364/BOE.8.003440
18. Lu D, Heisler M, Lee S, Ding GW, Navajas E, Sarunic MV, Beg MF (2019) Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network. Med Image Anal 54:100–110. https://doi.org/10.1016/j.media.2019.02.011
19. Alsaih K, Yusoff MZ, Tang TB, Faye I, Meriaudeau F (2020) Deep learning architectures analysis for age-related macular degeneration segmentation on optical coherence tomography scans. Comput Methods Programs Biomed 195:105566. https://doi.org/10.1016/j.cmpb.2020.105566
20. Asaoka R, Murata H, Hirasawa K, Fujino Y, Matsuura M, Miki A, Kanamoto T, Ikeda Y, Mori K, Iwase A, Shoji N, Inoue K, Yamagami J, Araie M (2019) Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. Am J Ophthalmol 198:136–145. https://doi.org/10.1016/j.ajo.2018.10.007
21. Russakoff DB, Mannil SS, Oakley JD, Ran AR, Cheung CY, Dasari S, Riyazzuddin M, Nagaraj S, Rao HL, Chang D, Chang RT (2020) A 3D deep learning system for detecting referable glaucoma using full OCT macular cube scans. Transl Vis Sci Technol 9:12. https://doi.org/10.1167/tvst.9.2.12
22. Nagasato D, Tabuchi H, Masumoto H, Enno H, Ishitobi N, Kameoka M, Niki M, Mitamura Y (2019) Automated detection of a nonperfusion area caused by retinal vein occlusion in optical coherence tomography angiography images using deep learning. PLoS ONE 14:e0223965. https://doi.org/10.1371/journal.pone.0223965
23. Sonobe T, Tabuchi H, Ohsugi H, Masumoto H, Ishitobi N, Morita S, Enno H, Nagasato D (2019) Comparison between support vector machine and deep learning, machine-learning technologies for detecting epiretinal membrane using 3D-OCT. Int Ophthalmol 39:1871–1877. https://doi.org/10.1007/s10792-018-1016-x
24. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Askham H, Glorot X, O'Donoghue B, Visentin D, van den Driessche G, Lakshminarayanan B, Meyer C, Mackinder F, Bouton S, Ayoub K, Chopra R, King D, Karthikesalingam A, Hughes CO, Raine R, Hughes J, Sim DA, Egan C, Tufail A, Montgomery H, Hassabis D, Rees G, Back T, Khaw PT, Suleyman M, Cornebise J, Keane PA, Ronneberger O (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 24:1342–1350. https://doi.org/10.1038/s41591-018-0107-6
25. Kawczynski MG, Bengtsson T, Dai J, Hopkins JJ, Gao SS, Willis JR (2020) Development of deep learning models to predict best-corrected visual acuity from optical coherence tomography. Transl Vis Sci Technol 9:51. https://doi.org/10.1167/tvst.9.2.51
26. Kingma D BJ Adam: (2014) A method for stochastic optimization. arXiv preprint arXiv:14126980
27. Lovie-Kitchin JE, Whittaker SG (1999) Prescribing near magnification for low vision patients. Clin Exp Optom 82:214–224. https://doi.org/10.1111/j.1444-0938.1999.tb06651.x
28. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2818–2826, https://doi.org/10.1109/CVPR.2016.308.) Rethinking the Inception Architecture for Computer Vision,
29. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2019:618e626) Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis
30. Chawla NV, Japkowicz N, Kotcz A (2004) Editorial ACM SIGKDD Explorations Newsletter 6:1–6. https://doi.org/10.1145/1007730.1007733
31. Kanda Y (2013) Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant 48:452–458. https://doi.org/10.1038/bmt.2012.244
32. Wendel RT, Patel AC, Kelly NE, Salzano TC, Wells JW, Novack GD (1993) Vitreous surgery for macular holes. Ophthalmology 100:1671–1676. https://doi.org/10.1016/s0161-6420(93)31419-3