Article

Development of a Web-Based Ensemble Machine Learning Application to Select the Optimal Size of Posterior Chamber Phakic Intraocular Lens

Eun Min Kang[1,*], Ik Hee Ryu[1,2,*], Geunyoung Lee[3], Jin Kuk Kim[1,2], In Sik Lee[1], Ga Hee Jeon[1,2], Hojin Song[1,2], Kazutaka Kamiya[4], and Tae Keun Yoo[1,5]

[1] B&VIIT Eye Center, Seoul, South Korea
[2] VISUWORKS, Seoul, South Korea
[3] MediWhale, Seoul, South Korea
[4] Visual Physiology, School of Allied Health Sciences, Kitasato University, Kanagawa, Japan
[5] Department of Ophthalmology, Aerospace Medical Center, Republic of Korea Air Force, Cheongju, South Korea

Correspondence: Tae Keun Yoo, Department of Ophthalmology, Aerospace Medical Center, Republic of Korea Air Force, Cheongju, South Korea, 635 Danjae-ro, Namil-myeon, Cheongwon-gun, Chungcheongbuk-do, 363-849, South Korea.
e-mail: eyetaekeunyoo@gmail.com

Purpose: Selecting the optimal lens size by predicting the postoperative vault can reduce complications after implantation of an implantable collamer lens with a central-hole (ICL with KS-aquaport). We built a web-based machine learning application that incorporated clinical measurements to predict the postoperative ICL vault and select the optimal ICL size.

Methods: We applied the stacking ensemble technique based on eXtreme Gradient Boosting (XGBoost) and a light gradient boosting machine to pre-operative ocular data from two eye centers to predict the postoperative vault. We assigned the Korean patient data to a training ($N = 2756$ eyes) and internal validation ($N = 693$ eyes) datasets (prospective validation). Japanese patient data ($N = 290$ eyes) were used as an independent external dataset from different centers to validate the model.

Results: We developed an ensemble model that showed statistically better performance with a lower mean absolute error for ICL vault prediction (106.88 μm and 143.69 μm in the internal and external validation, respectively) than the other machine learning techniques and the classic ICL sizing methods did when applied to both validation datasets. Considering the lens size selection accuracy, our proposed method showed the best performance for both reference datasets (75.9% and 67.4% in the internal and external validation, respectively).

Conclusions: Applying the ensemble approach to a large dataset of patients who underwent ICL implantation resulted in a more accurate prediction of vault size and selection of the optimal ICL size.

Translational Relevance: We developed a web-based application for ICL sizing to facilitate the use of machine learning calculators for clinicians.

Introduction

Selection of the proper size of the EVO Implantable Collamer Lens (ICL with KS-aquaport; STAAR Surgical, Monrovia, CA, USA) is a fundamental component of posterior phakic lens implantation surgery.[1] In clinical practice, optimal ICL sizing provides a safe postoperative ICL vault, which is a gap between the ICL and the crystalline lens.[2] The general consensus is that the ideal ICL vault is approximately 500 μm and should not exceed 1000 μm.[3] A higher vault after implantation of ICL without a center hole is associated with complications of angle-closure glaucoma and abnormally large pupils, and a lower vault is a risk factor for anterior subcapsular cataracts.[4]

Implantation of a recently developed ICL with a center hole (aquaport) was reported to have a low risk of anterior subcapsular cataract and angle-closure glaucoma.[5] However, it is associated with several other complications, such as increased intraocular pressure, abnormal pupil size, or lens dislocation. In cases with severe complications, ICL exchange or removal can be indicated. The manufacturer provides a nomogram for ICL sizing based on the corneal size (white-to-white [WTW]) and anterior chamber depth (ACD).[6] However, these two measurements do not accurately reflect the size of the posterior chamber space in which the lens will be placed. Several approaches have attempted to estimate the optimal ICL size using new technologies, such as anterior segment optical coherence tomography (AS-OCT) and ultrasound bio-microscopy (UBM). Using these new measurements, several statistical regression methods have been proposed to improve the accuracy of vault prediction for ICL sizing.[7] However, measurement error and structural variance may guide the implantation of inappropriate ICL size, yet there is no standardized method for selecting the optimal ICL size.

Recently, machine learning has enabled more accurate inference based on the large training data for medical diagnosis.[8] Several studies have proposed novel machine learning models to predict postoperative outcomes in various medical fields.[9] Machine learning models can consider a large number of features and can minimize human variation for clinical decision making.[10] A recent study showed that traditional machine learning methods are expected to improve the accuracy of ICL vault prediction.[11] However, all ICL surgery cases cannot be the gold standard for training machine learning models. Several cases may be considered as outliers with incorrect lens selection and measurement noises, and these factors can lead to the incorrect fitting of a given machine learning model and reduce the performance in the validation sets. To solve this problem, ensemble machine learning approaches have been introduced because of their robustness and ability to enhance the performance of machine learning models.[12] Ensemble learning-based approaches have shown very good performance in many recent machine learning competitions. They can reduce the effect of overfitting and outliers by combining the results of several machine learning models.

Here, we propose a novel ensemble learning-based ICL with a center-hole (Visian ICL with KS-aquaport) size selection framework that analyzes clinical ocular measurements. In this retrospective study, we sought to build a web-based machine learning application that incorporates clinical measurements to predict the postoperative ICL vault and select the optimal ICL size. The validation performance of the machine learning model was improved by using an ensemble learning framework. We report the ICL sizing performance of our models on a local South Korean dataset and an external Japanese dataset.

Methods

Overview

The objective of this study was to build a web-based calculator application with ensemble machine learning that predicts the postoperative vault and selects the optimal ICL size (Fig. 1). Because four types of ICL sizes (12.1, 12.6, 13.2, and 13.7 mm) are available, the surgeon would have to select one that achieves the best surgical outcome. In this study, we set a postoperative vault of 500 μm to be the optimal result. We retrospectively collected pre-operative and postoperative ocular data that was used to develop the machine learning model. The ICL vault calculator was generated based on the ocular measurements from a large dataset. This study adhered to the tenets of the Declaration of Helsinki. The study protocol was approved by the institutional review board (IRB) of the Korean National Institute for Bioethics Policy (P01-202008-21-017). This retrospective data collection for external validation was approved by the IRB at Kitasato University Hospital. Protected personal health information was removed for the purpose of the study.

Datasets

The Korean patients underwent refractive surgery with posterior phakic intraocular lens implantation using ICL (V4c and V5 Visian ICL with KS-AquaPORT) in the B&VIIT Eye Center (Seoul, South Korea) from January 2018 to June 2020. The inclusion criteria for ICL implantation at the B&VIIT Eye Center were: age between 17 and 55 years old (inclusive = $17 \leq age \leq 55$), stable refraction, 0 to −18.0 diopters (D) of myopia with astigmatism of 6.5 D or less, anterior chamber depth ≥ 2.5 mm, and endothelial cell density ≥ 1800 cells/mm². The Japanese patients underwent ICL implantation using the same types of lenses at Kitasato University Hospital (Kanagawa, Japan). The inclusion criteria at the Kitasato University Hospital were: age between 18 and 55 years old (inclusive = $18 \leq age \leq 55$), stable refraction, −3.0 to −19.0 D of myopia with astigmatism of 5.0 D or less, anterior chamber depth ≥ 2.8 mm, and endothelial cell density ≥ 1800 cells/mm². An ICL was implanted in the posterior chamber via a 3 mm temporal clear corneal

a Definition of postoperative ICL vault
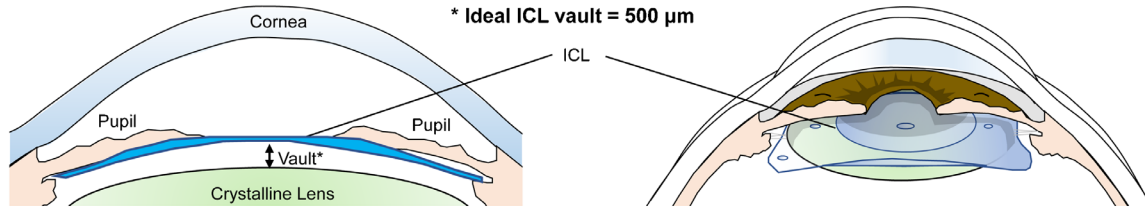
b Machine learning model for vault prediction and ICL sizing
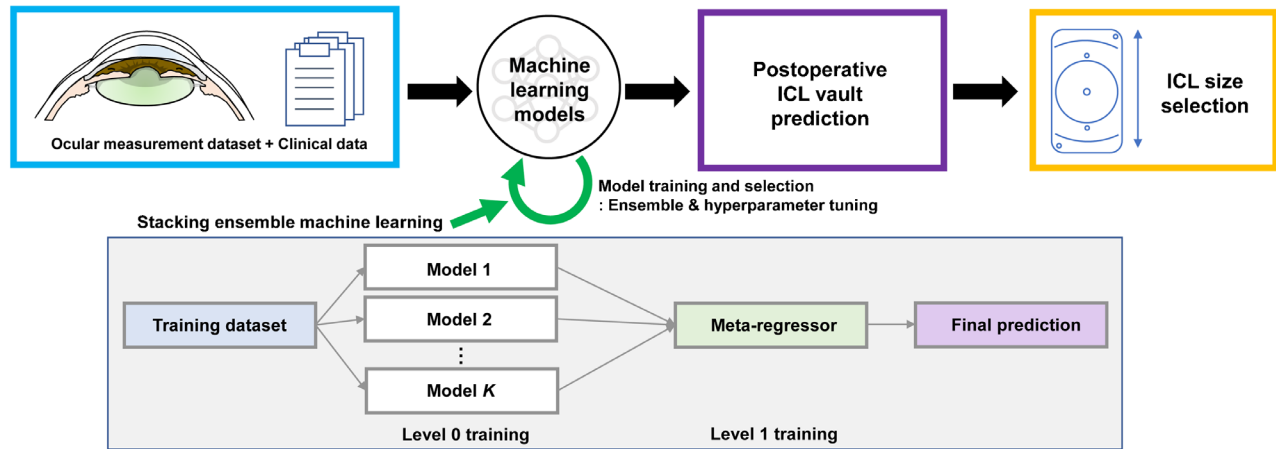
Figure 1. Schematic diagram of our proposed machine learning model for ICL sizing. (**A**) Definition of the postoperative ICL vault. (**B**) Machine learning model for vault prediction and ICL sizing.

incision. The selection of ICL size was assigned based on the clinical decision obtained from a full evaluation by five experts. Before surgery, the experts decided on the manufacturer's nomogram and ocular measurements as well as on the lens size for each patient. All experts were board-certified ophthalmologists with an average experience of 5 years in ICL surgery.

In the development and validation of this machine learning model, we used anonymized medical records and ocular measurement data. Study subjects included 3506 eyes from 1753 Korean patients and 290 eyes from 145 Japanese patients (Fig. 2). In the Korean patients, 57 eyes were excluded due to missing values, and the final dataset consisted of 3449 eyes. The training and internal validation datasets were split by calendar time to validate the machine learning model prospectively according to the design of previous studies.[13,14] We assigned the Korean patient data before December 2019 ($N = 2756$ eyes) to the training dataset and those after January 2020 ($N = 693$ eyes) were used as the internal validation dataset (prospective validation). In the training process, we performed fivefold cross-validation, which currently corresponds to the preferred technique for assessing performance and optimizing the prediction models. The dataset of Japanese patients was used to validate the model as

an independent external dataset, from different centers ($N = 290$ eyes).

All patients underwent pre-operative measurements of corrected distance visual acuity, manifest refraction, slit-lamp examination, and dilated fundus examination. Anterior-segmental measurements of the preoperative WTW, angle-to-angle (ATA), ACD, anterior chamber width (ACW), crystalline lens rise (CLR), central corneal thickness (CCT), and postoperative central vaults were obtained using swept-source AS-OCT using CASIA-2 (Tomey, Nagoya, Japan). Pupil size was measured using Keratograph 4 (Oculus GmbH, Wetzlar, Germany). Ophthalmologic examinations, including AS-OCT, were performed postoperatively at 1 month to measure the postoperative ICL vault. The input features for ICL vault prediction included age, sex, pre-operative spherical equivalent, ICL refractive power, type of ICL (toric lens or not), WTW, ATA, ACD, ACW, CLR, CCT, pupil size, and lens size.

Model Development

We built a machine learning-based framework to construct postoperative ICL vault predictions using
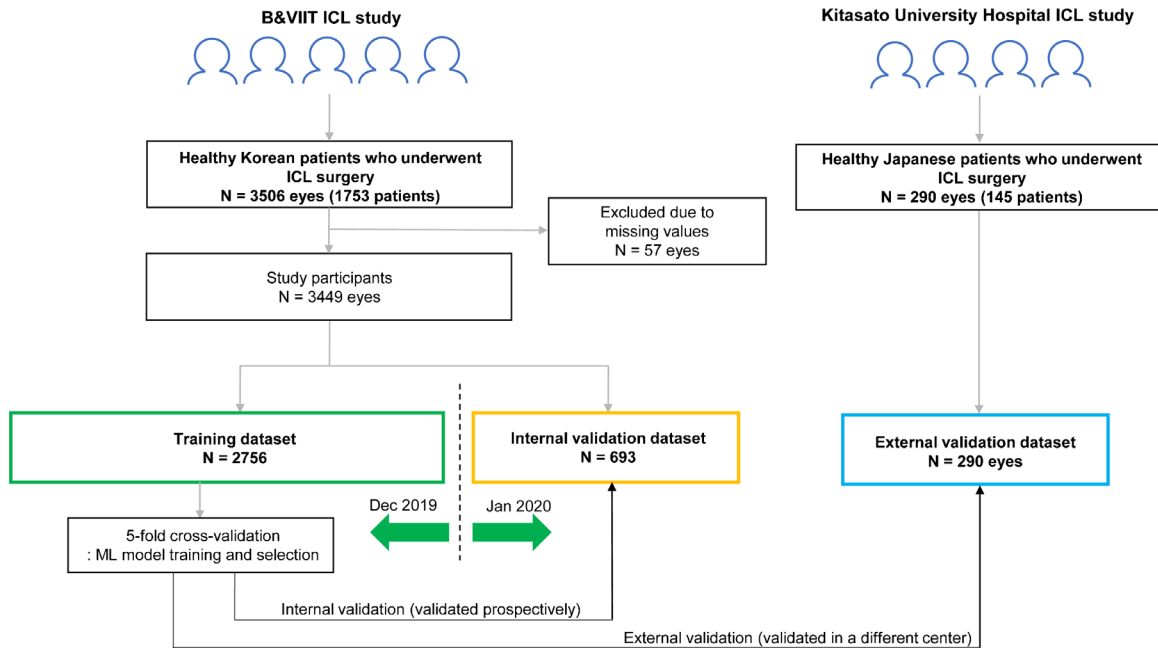
Figure 2. Workflow for data management and development of machine learning model for ICL sizing.

an ensemble regression model. After predicting the postoperative ICL vault using four ICL sizes (12.1, 12.6, 13.2, and 13.7 mm), we selected the optimal size that was closest to the ideal postoperative vault of 500 µm. Our study was focused on the stacking ensemble model owing to its superior performance compared with classic techniques. Recently, researchers have shown that tree-based meta-algorithms outperform the other techniques, we adopted eXtreme Gradient Boosting (XGBoost) and light gradient boosting machine (lightGBM) for the independent low-level regressors of the ensemble method. This method works by generalizing multiple low-level classifiers to increase the predictive power of the models. In the last layer of our stacking ensemble model, after obtaining the outputs of 12 XGBoost and 8 lightGBM models, we performed linear regression processing to obtain the final prediction of the ICL vault. The number of single models in the esemble model was determined empirically by the machine learning engineer. We also used other representative machine learning methods, including average ensemble, single XGBoost, random forest, support vector machine, and linear regression. To find the optimal hyperparameters and features for each machine learning technique, we adopted recursive feature elimination and grid search (Cartesian method), in which a range of parameter values was tested via the fivefold cross validation.[15] Finally, a risk calculator using a stacking ensemble was developed for the web-based interface (http://loocus-iolcalc.ai,

see Supplementary Materials). Our web-based model also deals with missing values because XGBoost can automatically impute each sample with missing values. It provides the effective confidence range of the postoperative ICL vault using the distribution of outputs from each XGBoost and lightGBM models in the ensemble architecture.

To obtain an interpretation of the features from the prediction model result, we applied the SHapley Additive exPlanations (SHAP) technique to our ensemble model. The SHAP package provides a decision-tree-based estimation of the SHAP value, with which the SHAP value from our ensemble model could be calculated based on XGBoost and lightGBM by summing the values of each model. The scikit-learn Python library and R studio version 3.5.1 (The Comprehensive R Archive Network; http://cran.r-project.org) were used to implement the machine learning and SHAP algorithms. We used the SHAP and XGBoost packages available in the GitHub repository (https://github.com/slundberg/shap and https://github.com/pablo14/shap-values).

To evaluate the prediction performance, we used the mean absolute error (MAE), median absolute prediction error (MedAE), and root mean square error (RMSE) in the fivefold cross validation, internal validation, and external validation. We also calculated the percentage of eyes that showed a prediction error of ±50, 100, 150, and 200 µm compared with the targeted ICL vault. All cases were included in the

validation procedure when the ICL vault prediction errors were calculated. However, we were unable to confirm that all our surgery cases had the best outcome. It should be noted that there is no gold standard validation dataset for ICL size selection. Therefore, when the multiclass lens selection performance was calculated, we excluded the outliers and only chose the cases with good outcomes (patients with 400 µm ≤ achieved ICL vault ≤ 600 µm and 300 µm ≤ achieved ICL vault ≤ 700 µm) to build a reference standard validation dataset. We used accuracy and Cohen's κ to evaluate lens selection ability.[16] We also compared the two classic lens selection methods: the manufacturer's nomogram based on WTW and ACD, and the NK formula based on ACW and CLR. These methods were built using linear regression models for predicting the postoperative ICL vault. All statistical analyses were conducted using R studio version 3.5.1. Differences in the distribution of variables between datasets were assessed using the χ^2 test for categorical variables and Student's t-test for continuous variables.

Results

The clinical characteristics of the study subjects with ICL implantation included in the training dataset (Korean patients, $N = 2,756$), internal validation dataset (Korean patients, $N = 693$), and external validation dataset (Japanese patients, $N = 290$) are shown in Table 1. The distribution of the internal validation dataset was similar to that of the training dataset, but

the external validation was not. All factors (but for ATA) showed significant differences between the training and external validation datasets. Very few cases required ICL with a size of 13.7 mm for surgery in both Korean and Japanese datasets. We used the t-distributed stochastic neighbor embedding (t-SNE) method to project features into a 3-dimensional space. The feature space shows that there might be considerable boundaries between lens sizes and meaningful distribution according to the ICL vault values, although the data distribution is not linear but complex (Fig. 3).

Using our training dataset, we performed the fivefold cross validation to optimize the machine learning model parameters. For this step, six candidate machine learning algorithms (stacking ensemble, average ensemble, single XGBoost, random forest, support vector machine, and linear regression) were trained and assessed on the corresponding hold-out sets. We identified that all input features were required to achieve the best performance for all the machine learning regressors. Therefore, we used all the input features to develop the ICL vault prediction model in this study. The fivefold cross validation performance of the optimized models for each algorithm is shown in Table 2. The stacking ensemble using 12 XGBoost and 8 lightGBM showed better ICL vault MAE than single XGBoost, random forest, support vector machine, and linear regression, achieving 99.67 µm on MAE, 84.72 µm on MedAE, and 125.73 µm on RMSE. There was no significant difference between the stacking and average ensemble methods. The

Table 1. Pre-Operative Demographics and Postoperative ICL Vaults of the Study Participants

	Training Dataset	Internal Validation Dataset (Korean Patients)	External Validation (Japanese Patients)	*P* Value for Training Versus Internal Validation	*P* Value for Training Versus External Validation
	(N = 2756 eyes)	(N = 693 eyes)	(N = 290 eyes)		
Age (y)	24.82 ± 5.79	25.05 ± 5.89	32.45 ± 7.56	0.457	<0.001
Gender, female (%)	1717 (62.3)	442 (63.8)	161 (55.5)	0.472	<0.001
Spherical equivalent (Diopters)	−8.98 ± 2.08	−8.97 ± 2.25	−7.11 ± 3.47	0.914	<0.001
Axial length (mm)	27.05 ± 1.34	26.99 ± 1.62	−	0.628	−
White-to-white (mm)	11.69 ± 0.34	11.71 ± 0.33	11.93 ± 0.44	0.255	<0.001
Angle-to-angle (mm)	11.77 ± 0.35	11.81 ± 0.35	11.78 ± 0.38	0.072	0.723
Anterior chamber depth (mm)	3.36 ± 0.22	3.36 ± 0.23	3.26 ± 0.27	0.847	<0.001
Anterior chamber width (mm)	11.90 ± 0.44	11.91 ± 0.43	11.80 ± 0.37	0.514	<0.001
Crystalline lens rise (µm)	−75.48 ± 166.02	−60.61 ± 174.28	64.97 ± 194.63	0.096	<0.001
Central corneal thickness (µm)	528.66 ± 35.22	525.53 ± 34.05	534.3 ± 30.14	0.068	0.020
Pupil size (mm)	6.64 ± 0.71	6.63 ± 0.70	3.10 ± 0.52	0.896	<0.001
ICL power (Diopters)	−10.84 ± 2.27	−10.81 ± 2.34	−7.79 ± 3.49	0.839	<0.001
Toric ICL (%)	1072 (38.9)	286 (41.3)	135 (46.6)		0.015
Achieved ICL size				0.414	<0.001
12.1 mm (%)	1279 (46.4)	335 (48.3)	99 (34.1)		
12.6 mm (%)	1301 (47.2)	321 (46.3)	152 (52.4)		
13.2 mm (%)	175 (6.3)	37 (5.2)	38 (13.1)		
13.7 mm (%)	1 (0.0)	1 (0.1)	1 (0.3)		
Postoperative achieved ICL vault (µm)	515.48 ± 170.73	514.39 ± 174.85	476.56 ± 249.06	0.905	0.004

ICL, implantable collamer lens.
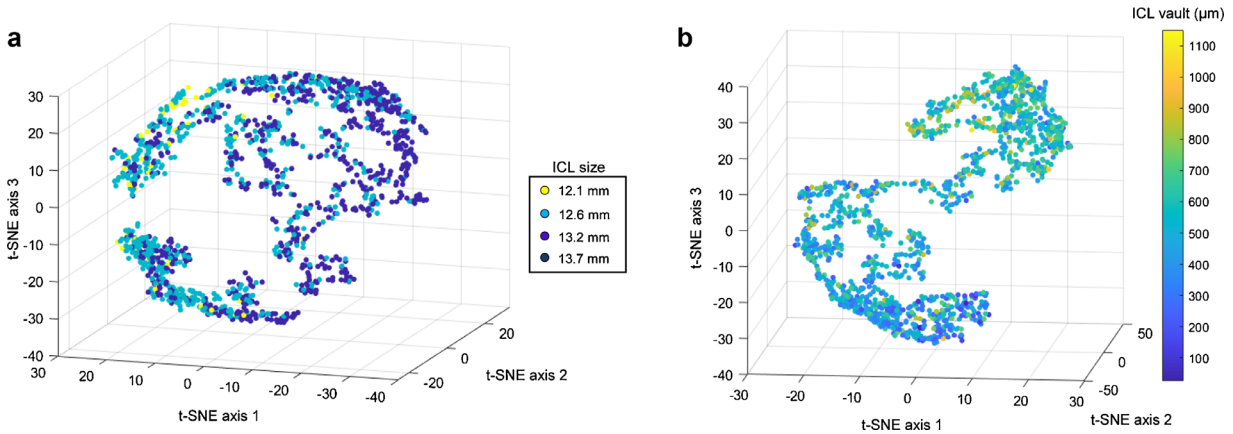
Figure 3. **The feature space visualized using the 3D t-SNE technique.** (**A**) The feature space without embedded ICL size to show data distribution labeled by ICL size. (**B**) The feature space with embedded ICL size to show data distribution labeled by the postoperative ICL vault.

Table 2. Postoperative ICL Vault Prediction Performance of Machine Learning and Conventional Models Via Fivefold Cross Validation

|		Mean Vault $\pm$ SD (µm)	MAE $\pm$ SD (µm)	MedAE (µm)	RMSE (µm)	*P* Value for MAE
Achieved ICL vault (target value)		514.39 $\pm$ 174.85	–	–	–	–
Predicted ICL vault	Stacking ensemble (XGBoost + LightGBM)	517.47 $\pm$ 125.45	99.67 $\pm$ 76.69	84.72	125.73	Reference
	Average ensemble (XGBoost + LightGBM)	513.96 $\pm$ 126.21	100.46 $\pm$ 75.53	86.12	125.65	0.227
	XGBoost (single model)	509.78 $\pm$ 116.46	104.54 $\pm$ 78.70	89.85	130.82	<0.001
	Random forest	511.36 $\pm$ 129.53	104.50 $\pm$ 78.47	87.61	130.65	<0.001
	Support vector machine	511.17 $\pm$ 99.20	109.68 $\pm$ 87.06	92.92	140.00	<0.001
	Linear regression	513.64 $\pm$ 120.76	106.63 $\pm$ 78.81	91.75	132.56	<0.001
	Manufacturer's nomogram (WTW + ACD)	509.34 $\pm$ 77.95	125.49 $\pm$ 92.10	110.12	155.62	<0.001
	NK formula[22] (ACW + CLR)	516.42 $\pm$ 77.22	123.58 $\pm$ 93.07	105.76	154.67	<0.001

ACD, anterior chamber depth; ACW, anterior chamber width; CLR, crystalline lens rise; ICL, implantable collamer lens; Light-GBM, light gradient boosting machine; MAE, mean absolute prediction error; MedAE, median absolute prediction error; RMSE, root mean square error; SD, standard deviation; WTW, white-to-white.

stacking ensemble also outperformed the manufacturer's nomogram and the NK formula ($P < 0.001$). A similar tendency was observed for MedAE and RMSE, demonstrating that the stacking ensemble was the best among the ICL vault prediction methods.

We examined the feature importance identified by the SHAP technique and averaged across all predictions of the stacking ensemble model in the fivefold cross validation. Because the ICL size is not only an input feature but also a prediction target, we evaluated the machine learning models with and without ICL size to predict ICL vault (Figs. 4A, 4B). In the model with ICL size, the SHAP values showed that ICL size was the most important predictor of ICL vault. Among the remaining important features, ACW, ACD, ICL power, CLR, ATA, pupil size, WTW, age, and the toric lens had a strong effect on the trained model. When the ICL size was excluded from the machine learning model, ACD was the most important factor among the input features (Figs. 4C, 4D).

Box plots according to the absolute error are shown in Figure 5A. The distributional characteristics of vault prediction demonstrate that machine learning methods are better than classic lens selection methods. Figure 5B shows the percentage of eyes within a given range of the ICL vault prediction errors. In the fivefold cross validation, the stacking ensemble model showed the highest predictability based on the discrete error range analysis. The correlation coefficient between the achieved vault and prediction using the stacking ensemble model was 0.678 with a *P* value < 0.001 (Fig. 6A). When we used the stacking ensemble model in the fivefold cross validation, 62% of the eyes were within the range of -99 to 100 µm.

Validation of the stacking ensemble model on the internal (Table 3) and external (Table 4) validation datasets also yielded better performance than the classic methods (*P* value < 0.001), achieving MAEs of 106.88 µm in the internal validation and 143.69 µm in the external validation. These results showed that the
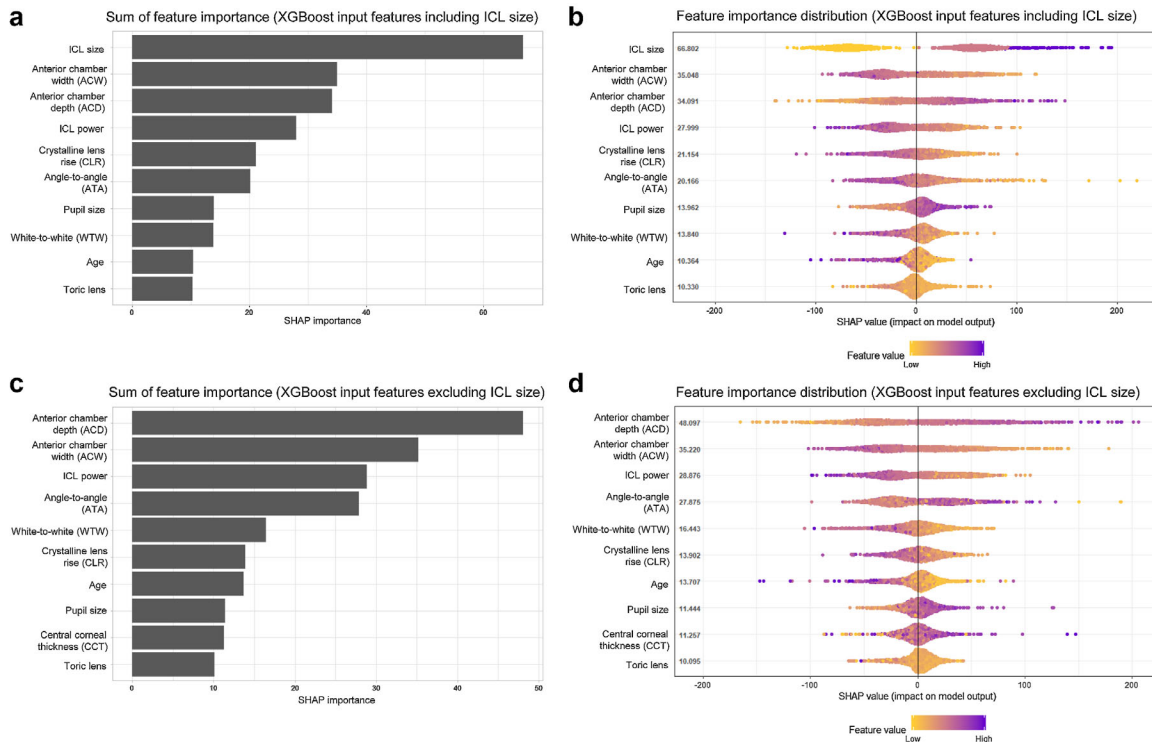
Figure 4. **Global feature importance estimates selected by the SHAP technique using the proposed ensemble model.** The model is based on XGBoost and lightGBM to predict postoperative ICL vault. (**A**) Total sum of SHAP importance from the ensemble model with ICL size. (**B**) The summary plot showing SHAP feature importance distributions. (**C**) Total sum of SHAP importance from the ensemble model without ICL size.
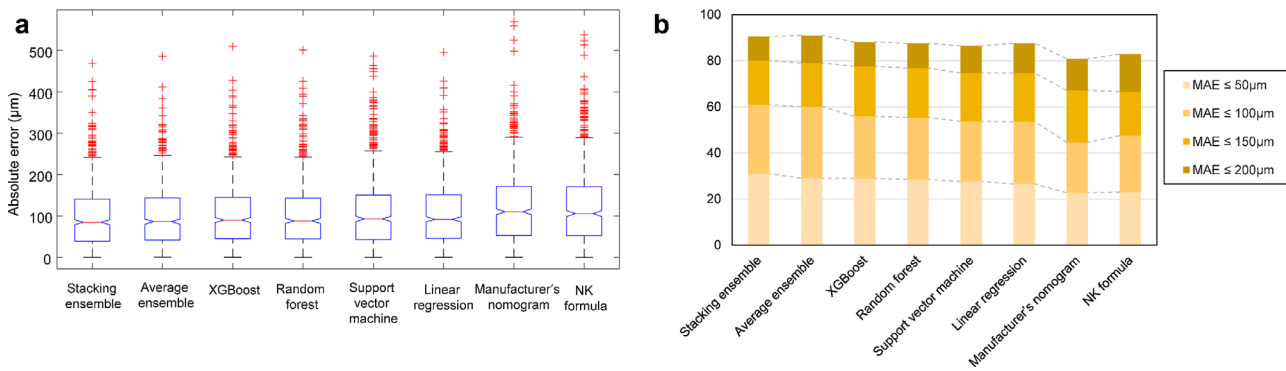
Figure 5. **Comparison of postoperative ICL vault prediction performance of machine learning and conventional models via fivefold cross validation.** (**A**) Box plot of the absolute error values for the predicted vault. (**B**) Proportions of eyes within a given range of absolute errors for the predicted vault.

MAEs of the machine learning methods decreased in the validation datasets. In Figure 7, the box plots of the prediction methods in the internal validation showed that the model performed reasonably well at predicting the ICL vault compared with the classic methods. In the internal and external validation datasets, the Bland-Altman plots revealed that our stacking ensemble model slightly overestimated when actual vault values were low and slightly underestimated when vault values were high (Fig. 8). Table 5 shows the multiclass classi-

fication results for the internal and external validation datasets when the ICL size was selected using the postoperative vault prediction methods. In this analysis, we excluded the outliers where the cases could not be confirmed as reference standards. The stacking ensemble model showed better multiclass classification performance than that of other methods. Considering patients with 400 μm ≤ achieved ICL vault ≤ 600 μm, the multiclass accuracies of the proposed ensemble method for ICL sizing were 75.9% (Cohen's
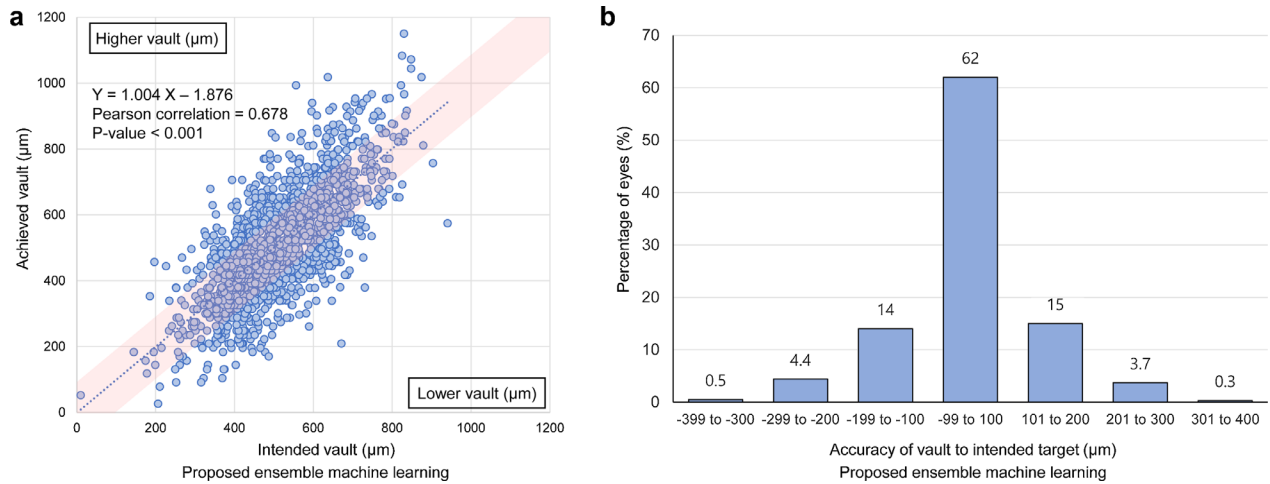
Figure 6. **Performance of the stacking ensemble machine learning model via fivefold cross validation.** (**A**) Distribution of the achieved vault against the predicted vault. (**B**) Distribution of the postoperative vault error to show the accuracy of the predicted vault to the intended target vault.

Table 3. Postoperative ICL Vault Prediction Performance of Machine Learning and Conventional Models in the Internal Validation Dataset From the Korean Patients (Internal Validation)

|		Mean Vault ± SD (µm)	MAE ± SD (µm)	MedAE (µm)	RMSE (µm)	*P* Value for MAE
Achieved ICL vault (target value)		516.82 ± 195.98	–	–	–	–
Predicted ICL vault	Stacking ensemble (XGBoost + LightGBM)	517.76 ± 134.52	106.88 ± 90.67	82.91	140.14	Reference
	Average ensemble (XGBoost + LightGBM)	517.18 ± 127.31	107.40 ± 98.49	83.09	145.69	0.678
	XGBoost (single model)	514.72 ± 124.08	110.33 ± 100.31	84.50	149.08	0.018
	Random forest	517.45 ± 123.89	110.74 ± 100.35	85.05	149.42	0.008
	Support vector machine	518.92 ± 103.71	123.78 ± 114.30	97.54	168.44	<0.001
	Linear regression	517.10 ± 115.21	112.31 ± 99.67	86.78	150.14	<0.001
	Manufacturer's nomogram (WTW + ACD)	520.11 ± 79.91	138.16 ± 114.89	118.54	179.65	<0.001
	NK formula[22] (ACW + CLR)	515.48 ± 79.98	133.12 ± 121.12	107.56	179.93	<0.001

ACD, anterior chamber depth; ACW, anterior chamber width; CLR, crystalline lens rise; ICL, implantable collamer lens; LightGBM, light gradient boosting machine; MAE, mean absolute prediction error; MedAE, median absolute prediction error; RMSE, root mean square error; SD, standard deviation; WTW, white-to-white.

Table 4. Postoperative ICL Vault Prediction Performance of Machine Learning and Conventional Models in the External Validation Dataset from the Japanese Patients (External Validation)

		Mean Vault ± SD (µm)	MAE ± SD (µm)	MedAE (µm)	RMSE (µm)	*P* Value for MAE
Achieved ICL vault (target value)		476.56 ± 249.06	–	–	–	–
Predicted ICL vault	Stacking ensemble (XGBoost + LightGBM)	473.04 ± 164.94	143.69 ± 118.76	118.68	186.29	Reference
	Average ensemble (XGBoost + LightGBM)	473.57 ± 162.70	144.07 ± 138.89	105.89	199.95	0.927
	XGBoost (single model)	468.22 ± 143.11	144.11 ± 141.72	100.01	201.94	0.923
	Random forest	473.41 ± 144.97	145.22 ± 141.38	108.35	202.51	0.723
	Support vector machine	474.56 ± 107.32	166.15 ± 163.48	134.49	232.90	0.002
	Linear regression	476.38 ± 139.56	146.58 ± 138.91	108.74	201.78	0.500
	Manufacturer's nomogram (WTW + ACD)	522.97 ± 93.06	179.36 ± 150.69	156.28	234.09	<0.001
	NK formula[22] (ACW + CLR)	456.12 ± 83.09	167.45 ± 169.90	127.45	238.35	0.002

ACD, anterior chamber depth; ACW, anterior chamber width; CLR, crystalline lens rise; ICL, implantable collamer lens; LightGBM, light gradient boosting machine; MAE, mean absolute prediction error; MedAE, median absolute prediction error; RMSE, root mean square error; SD, standard deviation; WTW, white-to-white.

$\kappa = 0.572$) and 67.4% (Cohen's $\kappa = 0.417$) in the internal and external validation, respectively. Considering patients with 300 µm ≤ achieved ICL vault ≤ 700 µm, which is a more generous criterion that involves more patients, the accuracies were 75.6% (Cohen's $\kappa = 0.567$

and 65.0% (Cohen's $\kappa = 0.366$) in the internal and external validations, respectively.

The developed model was applied to a preoperative case example from the internal validation dataset (Fig. 9). Using preoperative measurements, the
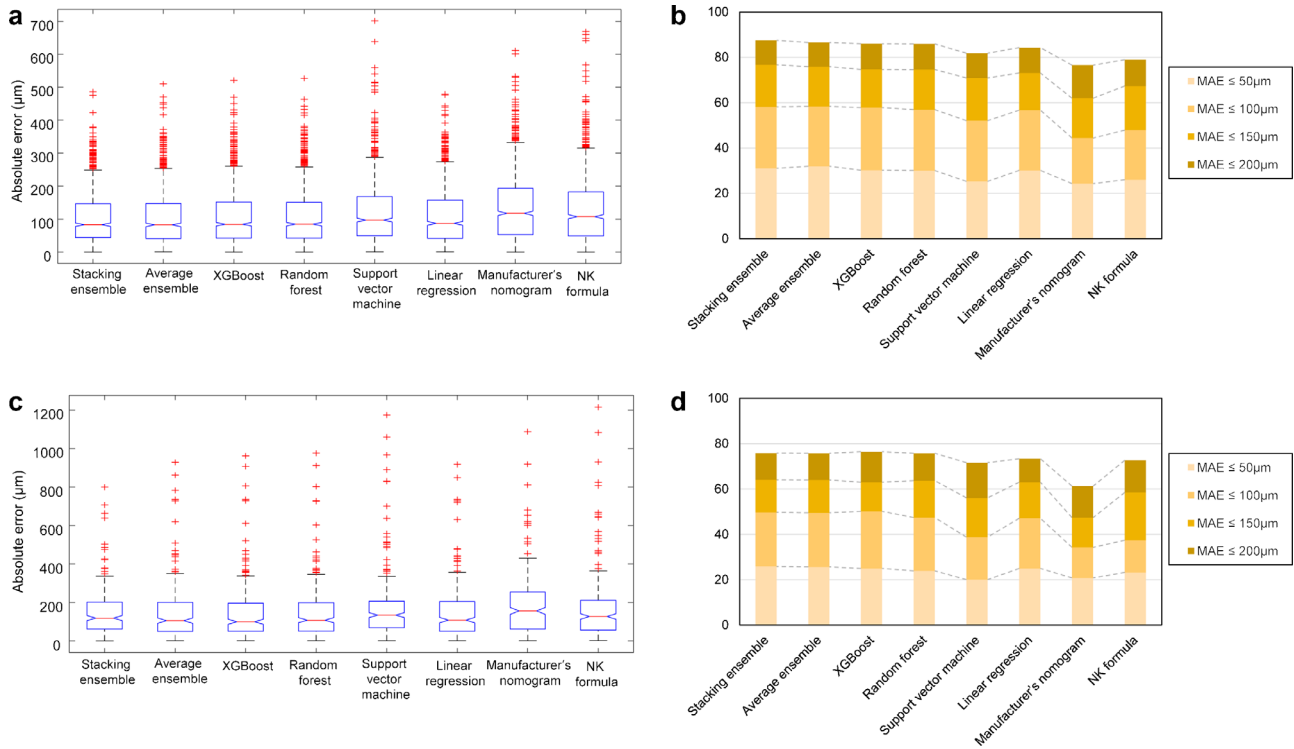
Figure 7. Comparison of postoperative ICL vault prediction performance in the internal and external validation datasets. (**A**) Box plot of the absolute error values for the predicted vault in the internal validation dataset. (**B**) Proportions of eyes within a given range of absolute errors for the predicted vault in the internal validation dataset. (**C**) Box plot of the absolute error values for the predicted vault in the external validation dataset. (**D**) Proportions of eyes within a given range of absolute errors for the predicted vault in the external validation dataset.
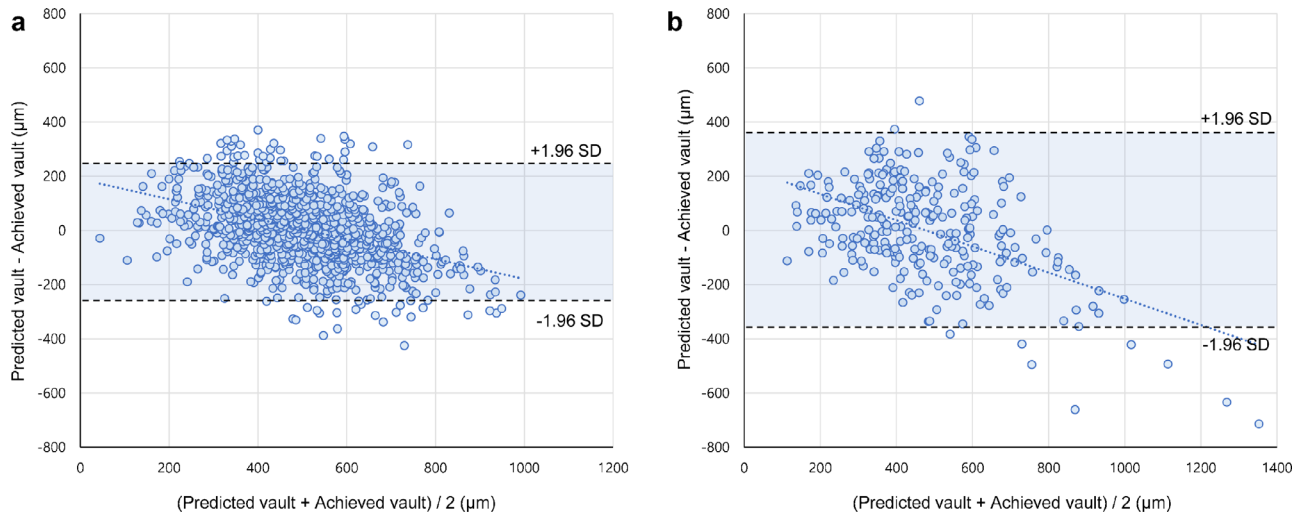
Figure 8. Bland-Altman plots for the achieved ICL vault and predicted vault using the ensemble machine learning model. (**A**) The result from the internal validation dataset. (**B**) The result from the external validation dataset.

NK formula, random forest, and stacking ensemble predicted the postoperative ICL vault to be 338 μm, 573 μm, and 642 μm, respectively, using a 12.6 mm ICL. All methods indicated that 12.6 mm was the optimal size for ICL implantation in this case. The force plot based on the explainable machine learning technique showed that the ICL size contributed to an increase in the predicted vault value, whereas ACD, ICL power, and ATA contributed to the decrease. Postoperative AS-OCT demonstrated that the stacking ensemble showed a lower error than the other techniques.
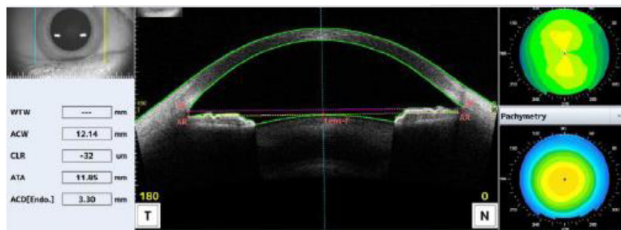
Table 5. Multiclass Classification Performance for ICL Size Selection Among the Cases With Good Outcomes in the Internal and External Validation Datasets

		Included Cases[*]			
		Cases of 400 µm ≤ Achieved ICL Vault ≤ 600 µm (Target Vault = 500 µm)		Cases of 300 µm ≤ Achieved ICL Vault ≤ 700 µm (Target Vault = 500 µm)	
Dataset	Model	Multiclass ICL Size Selection Accuracy (%)	Cohen's κ	Multiclass ICL Size Selection Accuracy (%)	Cohen's κ
Internal validation	Stacking ensemble (XGBoost + LightGBM)	75.9	0.572	75.6	0.567
	Random forest	74.1	0.542	73.8	0.564
	Manufacturer's nomogram (WTW + ACD)	41.4	0.177	38.9	0.109
	NK formula[22] (ACW + CLR)	57.4	0.337	52.8	0.266
External validation	Stacking ensemble (XGBoost + LightGBM)	67.4	0.417	65.0	0.366
	Random forest	65.3	0.390	64.4	0.354
	Manufacturer's nomogram (WTW + ACD)	48.4	0.217	36.1	0.026
	NK formula[22] (ACW + CLR)	64.2	0.416	61.1	0.349

[*]In this analysis, we excluded the outliers and only chose cases with good outcomes to build a reference standard validation dataset.

ACD, anterior chamber depth; ACW, anterior chamber width; CLR, crystalline lens rise; ICL, implantable collamer lens; Light-GBM, light gradient boosting machine; WTW, white-to-white.
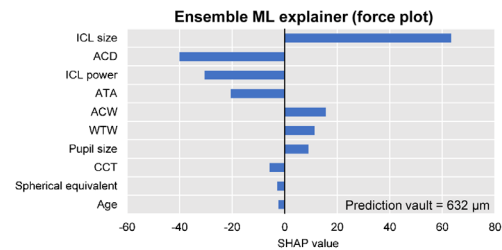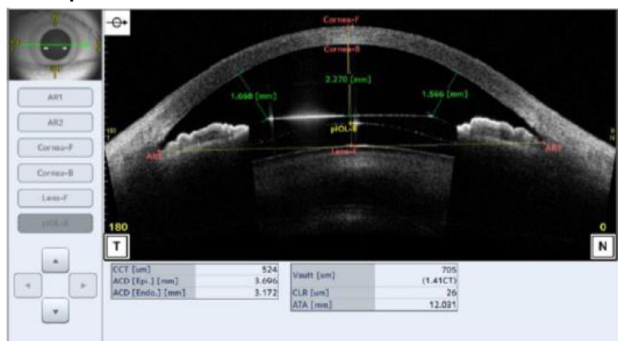

Figure 9. A case example of postoperative ICL vault prediction and lens size selection using the proposed web-based ensemble machine learning application.

Discussion

We applied the ensemble machine learning algorithm to clinical data to develop a postoperative ICL vault prediction model. The dataset was larger than those of any previous study and we provided a state-of-the-art comparison of vault prediction and ICL size selection methods. Our ensemble model based on XGBoost and lightGBM showed better MAE than the other machine learning techniques and classic prediction methods when applied to the internal (prospective design) and external validation datasets (other country datasets). A web-based application was

also developed for practice (http://loocus-iolcalc.ai), which made our ensemble model easily accessible to clinicians. Although the performance was not satisfactory for direct application without a clinician review, the proposed machine learning model significantly improved ICL size selection accuracy in the internal and external validation datasets.

Implantation of ICLs offers good results with safety, stability, and predictability. However, selecting the optimal ICL size by predicting the postoperative vault remains a very important issue. A previous study showed that approximately 1% of eyes with ICL implantation required lens extraction.[17] The most common reason for ICL extraction was cataract formation caused by mechanical contact because the vault gap was too small. Too large vaulting could result in angle-closure with elevated IOP from a nonpupillary block mechanism.[18] Oval-shaped pupil or endothelial cell loss may be associated with an abnormal postoperative ICL vault.[19] Clinicians can use proper ICL vault prediction models to select the optimal ICL size for successful vision correction without complications.

In current clinical practice, beginner clinicians rely on the predicted vault values provided by the manufacturer's nomogram or several formulas. The manufacturer's nomogram uses WTW and ACD; however, they do not reflect practical information about the space where the lens will fit.[7] A previous study has shown that the distance of the sulcus-to-sulcus measured by UBM could provide better vault predictability for selecting the optimal ICL size than the manufacturer's nomogram, but this measurement using UBM is subjective and time-consuming.[20] Recently, the measurements of ACW, ATA, or CLR using AS-OCT have been used to predict the postoperative vault with the linear regression technique.[21] After AS-OCT devices were widely used, the NK formula, which was developed based on ACW and CLR, has been commercially available and widely used for ICL sizing.[22] Researchers have updated the NK formula,[23] however, linear regression has a limitation of explaining the relationships between measurements. In our study, the t-SNE graphs show the nonlinear distribution of the dataset according to the ICL size or vault. Kamiya et al. developed the first machine learning model using the random forest algorithm to predict the postoperative vault using pre-operative measurements. However, the study did not compare their model to previous methods, and the developed model had practical limitations. In this study, we developed a web-based and easily accessible tool using an ensemble machine learning technique for ICL size prediction based on a larger dataset than in previous studies. Our machine learning calculator will effectively help both beginner and experienced clinicians to minimize surgical complications, such as anterior subcapsular cataracts or angle-closure glaucoma.

Recently, the ensemble technique has been successfully used in many complex machine learning tasks owing to its generalizability.[24,25] We also found its strength in this study because ICL sizing is also a nonlinear and complex problem.[26] This finding indicates that more accurate ICL vault prediction for ICL sizing can be obtained by combining machine learning models after extensive validation in other datasets. We adopted XGBoost and the stacking ensemble technique, which has been widely used to achieve better performance in many Kaggle competitions and research.[27,28] Recently, the combination of single XGBoost models using a stacking ensemble technique showed better performance than other machine learning algorithms did, which is consistent with our observations.[29] In our experience, the proposed ensemble technique showed superior performance to that of other state-of-the-art machine learning methods through trial and error. It is expected to be used in more areas in the future due to its ease of implementation and explainability.[14] We believe that our novel approach and large dataset will provide better performance for ICL sizing than previously proposed methods.

We found that all machine learning techniques showed reduced performance in the validation datasets, although they outperformed the classic methods. This decrease in the performance of machine learning models in independent validation sets has been commonly reported.[30] Although we were able to develop an accurate ICL vault prediction model from a large dataset of a single-center, the model could be overfitted to a training dataset. Overfitting to the training dataset may lead to a lack of generalizability in machine learning tasks. Therefore, testing on real-world data is required before the clinical use of machine learning systems.[31] To overcome this problem, we adopted the ensemble learning technique and validated the model using unseen datasets, including the prospectively designed dataset and completely independent external dataset. Our proposed method performed well across all datasets, including the Korean and Japanese datasets.

It should be noted that there are obstacles to applying the proposed machine model. Additional training might be necessary before our model is applied to other centers because there are several factors in ocular measurements that affect surgical outcomes. AS-OCT measurements were performed under stationary dark conditions in Korea and bright conditions in Japan. However, the developed model was based on

measurements under dark conditions. Some researchers have shown that the ICL vault varies significantly under bright conditions. According to a previous study, the mean postoperative vault range was 167 ± 70 μm, which is relatively large enough to affect the results of the study.[26] We noticed that we did not control the ambient illumination strictly during AS-OCT measurement in both the Korean and Japanese centers. The different light conditions in both centers certainly had a significant negative impact on performance in the external validation. The effects of the surgeon's skill may also have a significant impact on surgical outcomes, such as ICL vault. These factors indicate a potential limitation in the application of the trained model for out-of-distribution data.

A major advantage of the proposed method is that it is based on multimodal measurements from a large dataset. Inclusion of many pre-operative measurements can improve lens selection accuracy through machine learning. We built an ensemble machine learning model using XGBoost and lightGBM to create a sophisticated predictive model without overfitting. However, using many input parameters can be considered as a disadvantage because it is difficult for users to understand why vault prediction is performed. The classic methods are easy to understand because they use simple linear relationships with fewer pre-operative variables.

Our study has several limitations. First, the datasets consisted of East Asian populations, although two centers were involved in the study. A previous study demonstrated that there are ethnic differences between Asian and non-Asian eyes in the anterior segment of the eyes.[32] According to a study previously conducted on Caucasian eyes, the CLR values differ greatly from that of the East Asian eye dataset of this study.[33] Due to the anatomic difference in the anterior segment, the use of a large lens (13.7 mm) was extremely rare in our study. Therefore, it is not confirmed whether our proposed model can be applied to other ethnic groups when considering the differences between Asian and Caucasian eyes. Second, our proposed model was based on a single-center dataset. We anticipate that larger, more diverse ICL cases with multicenter studies should be performed to assess the feasibility of our method. Third, the analysis was based on a retrospective design, which had several defects due to medical considerations. Because the use of our application could influence clinicians' ICL size choice, intervention studies with randomized controlled trials are needed to confirm the effectiveness of the proposed machine learning model. Fourth, our study was conducted with a relatively short follow-up period; it is possible that the ICL vault could further change over time. However,

the stability of outcomes of ICL implantation was confirmed in observational studies.[34] Fifth, the distribution of ICL sizes in the datasets was imbalanced. In particular, the number of 13.7 mm was extremely small. Due to this distribution, we were unable to build a direct multiclass classification model without postoperative vault prediction. Sixth, the machine learning model was trained using AS-OCT measurements under dark conditions. Future studies should consider the lighting condition because it can significantly affect important parameters, such as vault, pupil diameter, iris thickness, and angle opening distance.[26,35] Last, endothelial cell loss was not evaluated, although it is a well-known complication of ICL implantation.

Conclusion

Applying the ensemble machine learning approach to a large dataset of patients with ICL (Visian ICL with KS-aquaport) implantation resulted in a more accurate prediction of vault size and selection of the optimal ICL size. We developed a web-based application to facilitate the use of the machine learning calculator for clinicians. After extensive validation with other datasets, a data-driven model could help both clinicians and patients minimize the risk of postoperative complications of ICL implantation.

Acknowledgments

or stock as part of the standard compensation package. The remaining authors declare no conflicts of interest.

References

1. Shimizu K, Kamiya K, Igarashi A, Shiratani T. Early clinical outcomes of implantation of posterior chamber phakic intraocular lens with a central hole (Hole ICL) for moderate to high myopia. *Br J Ophthalmol*. 2012;96(3):409–412.

2. Alfonso JF, Fernández-Vega L, Lisa C, Fernandes P, González-Meijome J, Montés-Micó R. Long-term evaluation of the central vault after phakic Collamer lens (ICL) implantation using OCT. *Graefes Arch Clin Exp Ophthalmol*. 2012;250(12):1807–1812.

3. Lovisolo CF, Zaldivar R. Complications of Posterior Chamber Phakic IOLs. In: Alio JL, Azar DT, eds. *Management of Complications in Refractive Surgery*. New York, NY: Springer International Publishing; 2018:289–309.

4. Fernandes P, González-Méijome JM, Madrid-Costa D, Ferrer-Blasco T, Jorge J, Montés-Micó R. Implantable collamer posterior chamber intraocular lenses: a review of potential complications. *J Refract Surg Thorofare NJ 1995*. 2011;27(10):765–776.

5. Gonzalez-Lopez F, Bouza-Miguens C, Tejerina V, Mompean B, Ortega-Usobiaga J, Bilbao-Calabuig R. Long-term assessment of crystalline lens transparency in eyes implanted with a central-hole phakic collamer lens developing low postoperative vault. *J Cataract Refract Surg*. Published online October 20, 2020, https://doi.org/10.1097/j.jcrs.0000000000000425.

6. Igarashi A, Shimizu K, Kato S, Kamiya K. Predictability of the vault after posterior chamber phakic intraocular lens implantation using anterior segment optical coherence tomography. *J Cataract Refract Surg*. 2019;45(8):1099–1104.

7. Ando W, Kamiya K, Hayakawa H, Takahashi M, Shoji N. Comparison of phakic intraocular lens vault using conventional nomogram and prediction formulas. *J Clin Med*. 2020;9(12):4090.

8. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. *Transl Vis Sci Technol*. 2020;9(2):14–14.

9. Ting DSJ, Foo VH, Yang LWY, et al. Artificial intelligence for anterior segment diseases: emerging applications in ophthalmology. *Br J Ophthalmol*. Published online June 12, 2020, https://doi.org/10.1136/bjophthalmol-2019-315651.

10. Achiron A, Gur Z, Aviv U, et al. Predicting refractive surgery outcome: machine learning approach with big data. *J Refract Surg Thorofare NJ 1995*. 2017;33(9):592–597.

11. Kamiya K, Ryu IH, Yoo TK, et al. Prediction of phakic intraocular lens vault using machine learning of anterior segment optical coherence tomography metrics: phakic lens vault prediction using machine learning. *Am J Ophthalmol*. Published online February 9, 2021, https://doi.org/10.1016/j.ajo.2021.02.006.

12. Zeng K, Pan Z, Xu Y, Qu Y. An ensemble learning strategy for eligibility criteria text classification for clinical trial recruitment: algorithm development and validation. *JMIR Med Inform*. 2020;8(7):e17832.

13. Ye C, Fu T, Hao S, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. *J Med Internet Res*. 2018;20(1):e22.

14. Yoo TK, Ryu IH, Choi H, et al. Explainable machine learning approach as a tool to understand factors used to select the refractive surgery technique on the expert level. *Transl Vis Sci Technol*. 2020;9(2):8.

15. Yadaw AS, Li Y, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *Lancet Digit Health*. 2020;2(10):e516–e525.

16. Choi JY, Yoo TK, Seo JG, Kwak J, Um TT, Rim TH. Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database. *PLoS One*. 2017;12(11):e0187336.

17. Hayakawa H, Kamiya K, Ando W, Takahashi M, Shoji N. Etiology and outcomes of current posterior chamber phakic intraocular lens extraction. *Sci Rep*. 2020;10(1):21686.

18. Khalifa YM, Goldsmith J, Moshirfar M. Bilateral explantation of visian implantable collamer lenses secondary to bilateral acute angle closure resulting from a non-pupillary block mechanism. *J Refract Surg*. 2010;26(12):991–994.

19. Guber I, Mouvet V, Bergin C, Perritaz S, Othenin-Girard P, Majo F. Clinical outcomes and cataract formation rates in eyes 10 years after posterior phakic lens implantation for myopia. *JAMA Ophthalmol*. 2016;134(5):487–494.

20. Lee D-H, Choi S-H, Chung E-S, Chung T-Y. Correlation between preoperative biometry and posterior chamber phakic visian implantable collamer lens vaulting. *Ophthalmology*. 2012;119(2):272–277.

21. Trancón AS, Manito SC, Sierra OT, Baptista AM, Serra PM. Determining vault size in implantable collamer lenses: preoperative anatomy and lens parameters. *J Cataract Refract Surg*. 2020;46(5):728–736.

22. Nakamura T, Isogai N, Kojima T, Yoshida Y, Sugiyama Y. Implantable collamer lens sizing method based on swept-source anterior segment optical coherence tomography. *Am J Ophthalmol*. 2018;187:99–107.

23. Nakamura T, Isogai N, Kojima T, Yoshida Y, Sugiyama Y. Optimization of implantable collamer lens sizing based on swept-source anterior segment optical coherence tomography. *J Cataract Refract Surg*. 2020;46(5):742–748.

24. Yoo TK, Ryu IH, Lee G, et al. Adopting machine learning to automatically identify candidate patients for corneal refractive surgery. *Npj Digit Med*. 2019;2(1):59.

25. Somasundaram SK, Alli P. A machine learning ensemble classifier for early prediction of diabetic retinopathy. *J Med Syst*. 2017;41(12):201.

26. Gonzalez-Lopez F, Mompean B, Bilbao-Calabuig R, Vila-Arteaga J, Beltran J, Baviera J. Dynamic assessment of light-induced vaulting changes of implantable collamer lens with central port by swept-source OCT: pilot study. *Transl Vis Sci Technol*. 2018;7(3):4.

27. Chen C, Zhang Q, Yu B, et al. Improving protein-protein interactions prediction accuracy using XGBoost feature selection and stacked ensemble classifier. *Comput Biol Med*. 2020;123:103899.

28. Bojer CS, Meldgaard JP. Kaggle forecasting competitions: An overlooked learning opportunity. *Int J Forecast*. Published online September 2, 2020, https://doi.org/10.1016/j.ijforecast.2020.07.007.

29. Massaoudi M, Refaat SS, Chihi I, Trabelsi M, Oueslati FS, Abu-Rub H. A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for short-term load forecasting. *Energy*. 2021;214:118874.

30. Rim TH, Lee AY, Ting DS, et al. Detection of features associated with neovascular age-related macular degeneration in ethnically distinct data sets by an optical coherence tomography: trained deep learning algorithm. *Br J Ophthalmol*. Published online September 9, 2020, https://doi.org/10.1136/bjophthalmol-2020-316984.

31. Lee AY, Yanagihara RT, Lee CS, et al. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. *Diabetes Care*. Published online January 5 2021, https://doi.org/10.2337/dc20-1877.

32. Chan SM, Svitova TF, Lin MC. Accounting for ethnicity-related differences in ocular surface integrity as a step toward understanding contact lens discomfort. *Eye Contact Lens Sci Clin Pract*. 2017;43(1):23–31.

33. Gonzalez-Lopez F, Bilbao-Calabuig R, Mompean B, Luezas J, Ortega-Usobiaga J, Druchkiv V. Determining the potential role of crystalline lens rise in vaulting in posterior chamber phakic collamer lens surgery for correction of myopia. *J Refract Surg Thorofare NJ 1995*. 2019;35(3):177–183.

34. Niu L, Miao H, Han T, Ding L, Wang X, Zhou X. Visual outcomes of Visian ICL implantation for high myopia in patients with shallow anterior chamber depth. *BMC Ophthalmol*. 2019;19(1):121.

35. Hirose F, Hata M, Ito S, Matsuki T, Kurimoto Y. Light–dark changes in iris thickness and anterior chamber angle width in eyes with occludable angles. *Graefes Arch Clin Exp Ophthalmol*. 2013;251(10):2395–2402.