OPEN ACCESS

Evaluation of the Nallasamy formula: a stacking ensemble machine learning method for refraction prediction in cataract surgery

Tingyang Li [ID],[1] Joshua Stein,[2,3,4] Nambi Nallasamy [ID] [1,2]

ABSTRACT

Aims To develop a new intraocular lens power selection method with improved accuracy for general cataract patients receiving Alcon SN60WF lenses.
Methods and analysis A total of 5016 patients (6893 eyes) who underwent cataract surgery at University of Michigan's Kellogg Eye Center and received the Alcon SN60WF lens were included in the study. A machine learning-based method was developed using a training dataset of 4013 patients (5890 eyes), and evaluated on a testing dataset of 1003 patients (1003 eyes). The performance of our method was compared with that of Barrett Universal II, Emmetropia Verifying Optical (EVO), Haigis, Hoffer Q, Holladay 1, PearlDGS and SRK/T.
Results Mean absolute error (MAE) of the Nallasamy formula in the testing dataset was 0.312 Dioptres and the median absolute error (MedAE) was 0.242 D. Performance of existing methods were as follows: Barrett Universal II MAE=0.328 D, MedAE=0.256 D; EVO MAE=0.322 D, MedAE=0.251 D; Haigis MAE=0.363 D, MedAE=0.289 D; Hoffer Q MAE=0.404 D, MedAE=0.331 D; Holladay 1 MAE=0.371 D, MedAE=0.298 D; PearlDGS MAE=0.329 D, MedAE=0.258 D; SRK/T MAE=0.376 D, MedAE=0.300 D. The Nallasamy formula performed significantly better than seven existing methods based on the paired Wilcoxon test with Bonferroni correction (p<0.05).
Conclusions The Nallasamy formula (available at https://lenscalc.com/) outperformed the seven other formulas studied on overall MAE, MedAE, and percentage of eyes within 0.5 D of prediction. Clinical significance may be primarily at the population level.

Key message

What is already known on this topic
⇒ Despite numerous recent advances, there is still room for improvement in intraocular lens (IOL) power selection.

What this study adds
⇒ The Nallasamy formula is an ensemble machine learning (ML)-based method for IOL power selection that outperformed seven existing IOL formulas, including Barrett, Emmetropia Verifying Optical and PearlDGS.

How this study might affect research, practice or policy
⇒ The Nallasamy formula offers cataract surgeons an accurate, ML-based method for IOL power selection and has the potential to help improve refractive outcomes after cataract surgery.

INTRODUCTION

Cataract surgery is the most commonly performed surgical procedure in the United States (approximately 4 million/year) and worldwide (approximately 23 million/year). The appropriate selection of intraocular lens (IOL) power based on accurate prediction of postoperative refraction is necessary for achieving a favourable refractive outcome and is closely associated with patient satisfaction. An inappropriate IOL power was found to be the indication for approximately 20% of cataract surgery cases that required secondary intervention, lens removal or lens exchange, according to analyses of records between 2002 and 2017.[1,2]

Various generations of IOL power calculation formulas have been published since the 1960s. From the earliest regression formulas (Binkhorst formula, SRK formula) to the fourth and fifth generation of vergence formulas which established the effective lens position (ELP) as a function of the axial length, lens thickness (LT) and keratometry, the accuracy of IOL power calculation has been substantially improved. Among existing formulas, the Barrett Universal II formula[3] is widely used and several publications have demonstrated that Barrett Universal II has greater accuracy than other traditional formulas.[4,5] In addition to the above-mentioned formulas, a number of new IOL formulas have been published recently, such as the Emmetropia Verifying Optical (EVO) formula,[6] which is a theoretical thick lens formula, and the PearlDGS formula,[7,8] which is a machine learning (ML)-based thick lens calculation method.

Although the methodology for IOL power selection has been studied for decades, patient expectations for refractive outcomes continue to rise and room remains for improvement in refraction prediction performance. ML and artificial intelligence have proven to be successful in many medical applications, including ophthalmology.[9,10] Researchers have begun to incorporate ML into IOL power calculations in recent years.

However, key limitations exist among recently-published ML-based IOL calculation methods: (1) performance comparisons limited to older generation formulas,[11] (2) failure to achieve statistically significant improvement over current generation

formulas,[12] and (3) small datasets that leave the robustness and generalisability of methods in question.[13]

With a goal of advancing the understanding of IOL power selection for general cataract patients and improving refraction prediction accuracy, in this study, we developed a novel ML-based IOL power calculation method, the Nallasamy formula, based on a large dataset of 5016 cataract patients. In this model, we employed ensemble ML methods and novel data augmentation methods. The performance of our method was compared with that of Barrett Universal II, EVO, Haigis, Hoffer Q, Holladay 1, PearlDGS, and SRK/T on an unseen testing dataset of 1003 patients.

MATERIALS AND METHODS
Data collection and preprocessing
This study focused on a subset of patients receiving care at the University of Michigan between 25 August 2015 and 27 June 2019. The preoperative biometry records were obtained from Lenstar LS 900 optical biometers (Haag-Streit USA, EyeSuite software V.i9.1.0.0) at University of Michigan's Kellogg Eye Center. Patient demographics (including patient age, gender and ethnicity) and cataract surgery information were obtained via the Sight Outcomes Research Collaborative (SOURCE) Ophthalmology Data Repository. SOURCE is a data repository that tracks the electronic health record data of all patients receiving any eye care at participating academic medical institutions. The information deposited in SOURCE includes patient demographics, diagnoses identified based on International Classification of Diseases codes, procedures based on Current Procedural Terminology (CPT) codes, and structured and unstructured (free-text) data from all clinical encounters (clinic visits, operative reports, etc). Various studies using data from SOURCE were published.[14–18] Manifest refractions were performed at the end of the first postoperative month by trained technicians employed by University of Michigan's Kellogg Eye Center. Manifest refraction data was obtained through the SOURCE repository.

The inclusion criteria for the cases were as follows: (1) Cataract surgery was performed (CPT code=66982 or 66984); (2) An Alcon SN60WF one-piece acrylic monofocal lens was implanted, (3) No refractive surgery was performed before the cataract surgery; (4) No additional surgery was performed at the time of cataract surgery. Cases with any CPT code other than 66982 or 66984 were excluded; (5) Visual acuity was 20/40 or better and (6) Data were complete and was not out of bounds for any of the formulas with which performance was compared.

Stacking ensemble ML framework
After all preprocessing steps, we obtained a clean tabular dataset of 5016 patients wherein each eye had a complete profile of preoperative biometry, patient demographics (patient gender and age), the power of the implanted IOL and the postoperative refraction. Preoperative biometry included the axial length (AL), crystalline LT, anterior chamber depth (ACD), aqueous depth, astigmatism, white-to-white, central corneal thickness, and keratometry (K1 and K2, $K = (K1 + K2)/2$). The postoperative refraction was calculated from the spherical component (SC) and the cylindrical component (CC) with an adjustment with regard to the lane length at Kellogg Eye Center (10 ft, 3.048 m): *spherical equivalent* (SE) *refraction* = $(SC - 0.1614) + 0.5CC$ according to Simpson and Charman's recommendation.[19]

The prediction task was framed as a regression problem where the goal was to build an ML algorithm that predicts the postoperative refraction using available information. The value to be predicted is referred to as the target value (represented as Y in figure 1) and the inputs that are used to make the predictions are referred to as features or predictors (represented as X in figure 1). The dataset was randomly split into a training/validation set with 4013 patients (5890 eyes) which was 80% of all patients, and a testing set with 1003 patients which was 20% of all patients (figure 1). In order to make sure all samples in the testing set were independent, one eye was selected at random and dropped from the dataset for all patients with both eyes available in the dataset. The training/validation set was used for cross-validation and hyperparameter selection of the ML model. The testing set was used for performance comparison between the existing formulas and our ML-based method.

Ensemble learning is a technique that involves combining the predictions from base learners with the goal of reducing variance and achieving improved prediction performance. An ensemble model is usually believed to outperform individual learners in most cases.[20] Stacking (or stacked generalisation) is one of the most commonly used meta-learning paradigms, where a number of base-learners are trained using the raw training data and a single meta-learner is trained to combine the predictions from the base-learner.[21] The reason for using an ensemble ML model in this study is to take advantage of different classes of ML
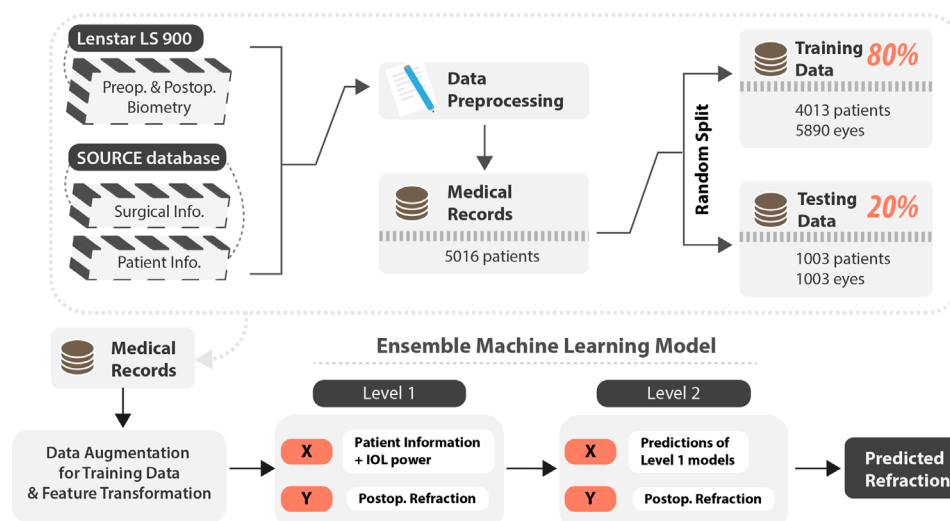

Figure 1 The overall method pipeline. IOL, intraocular lens.

Table 1 The optimised lens constants

Formula	Constant	Value
Barrett	Lens factor	1.94
EVO	A constant	119.0
Haigis	a0, a1, a2	−0.739, 0.234, 0.217
Hoffer Q	Personalised ACD	5.727
Holladay 1	Surgeon factor	1.860
PearlDGS	IOL A constant	119.1
SRK/T	A constant	119.082

ACD, anterior chamber depth; EVO, Emmetropia Verifying Optical; IOL, intraocular lens.

Table 2 Summary of patient demographics

Characteristic	Training set (mean±SD)	Testing set (mean±SD)
Count	5890 eyes, 4013 patients	1003 eyes, 1003 patients
Gender	Male: 2573 eyes (43.7%), Female: 3317 eyes (56.3%)	Male: 433 eyes (43.2%), Female: 570 eyes (56.8%)
Age at surgery (years)	71.00±9.43	70.73±9.50
Preoperative K (D)	43.87±1.54	43.88±1.56
Preoperative AL (mm)	24.17±1.34	24.15±1.35
Preoperative LT (mm)	4.53±0.44	4.53±0.45
Preoperative ACD (mm)	3.26±0.41	3.25±0.41
Postoperative refraction (D)	−0.55±0.85	−0.59±0.93

ACD, anterior chamber depth; AL, axial length; D, dioptre; K, keratometry; LT, lens thickness.

algorithms and improve the overall performance of the model. The stacking model consists of two layers. In the first layer a group of level-1 learners was trained based on the raw data (preoperative patient data and the postoperative refraction). The second layer consists of the metamodel which uses the output of the level-1 learners as the input features. Therefore, the number of input features for the level-2 model equals the number of level-1 models. The output from the level-2 meta-model is the final prediction result (figure 1).

Lens constant optimisation of existing IOL formulas
The existing formulas Haigis, Hoffer Q, Holladay 1, SRK/T were implemented in Python based on their specific equations.[22–29] The results obtained were validated against printouts from Haag-Streit USA, EyeSuite software V.i9.1.0.0. The prediction results of Barrett Universal II,[3] EVO (V2.0)[6] and PearlDGS[7 8] were obtained through their online calculators. The constants of the corresponding formulas were optimised based on the cases in the training dataset (4013 patients). The most optimal constant was selected by zeroing the mean prediction error. The optimised constants are listed in table 1.

Cross-validation and hyperparameter tuning
During the development of the ML model, we performed model evaluation and selection through five-fold cross-validation. During the cross-validation, 4013 training/validation cases were divided into training sets and validation datasets. A random eye was removed for patients with both eyes available in the validation dataset. The optimisation of hyperparameters of the ML models, the combination of the level-1 models, and the selection of the level-2 model were performed by minimising the averaged mean absolute error (MAE) based on the cross-validation results.

Performance comparison on the testing set
To compare the performance between our method and existing IOL formulas, we trained the ML-model with the entire training dataset (5890 eyes) and made predictions on the testing dataset. We calculated the mean arithmetic error (ME), MAE, median absolute error (MedAE) of the postoperative refraction predictions and the SD of the prediction error. We also calculated the number and percentage of patients with an absolute prediction error of less than or equal to 0.25 D, 0.50 D, 0.75 D and 1.00 D, and evaluated the statistical significance of the difference between formulas with Cochran's Q test. The statistical significance of the difference between the testing set performance of the IOL formulas was assessed using a Friedman test followed by a paired Wilcoxon test with Bonferroni correction. To investigate the performance of our method in cases with different axial lengths, we calculated the SD, ME, MAE, and MedAE for

patients in the short AL group (AL<22 mm), medium AL group (22 mm ≤AL ≤ 26 mm) and long AL group (AL>26 mm). In addition to the above metrics, we calculated the slope of the correlation between the arithmetic error and AL as m. Using the above variables, we computed the IOL Formula Performance Index (FPI) as recommended by Hoffer and Savini[30] for each formula as follows, where n is the percentage of eyes with an absolute error within 0.5 D. Higher FPI means better accuracy.

$$FPI = \frac{1}{SD + MedAE + 10 * abs(m) + 10 * (n10)^{-1}}$$

To investigate the effect of the size of the training data on the performance of the ML model, we randomly sampled 10%, 20%, ..., 90% of the training data, then retrained and compared the alternative models' results on the testing set. The proportions of training cases were adjusted before the application of data augmentation and data transformation techniques. All other configurations and hyperparameters were kept the same for alternative models except for the number of training cases.

In this study, the refraction prediction error was defined as follows. The criterion for statistical significance was p<0.05. All statistical analyses were scripted with Python V.3.9.5.

$$error = true\ postoperative\ refraction - predicted\ postoperative\ refraction$$

RESULTS
Out of 5016 patients, 4013 patients (5890 eyes) were assigned to the training/validation dataset, and 1003 cases were isolated

Table 3 Performance summary in the testing set

	MAE	MedAE	ME	SD	m	AE ≤0.5 D	FPI	P value
Barrett	0.328	0.256	0.038	0.437	0.307	78.3%	0.198	<0.05
EVO	0.322	0.251	0.043	0.427	0.128	79.8%	0.312	<0.05
Haigis	0.363	0.289	0.024	0.469	0.226	74.7%	0.230	<0.05
Hoffer Q	0.404	0.331	0.009	0.518	0.951	70.3%	0.085	<0.05
Holladay 1	0.371	0.298	0.021	0.487	0.773	74.0%	0.101	<0.05
PearlDGS	0.329	0.258	0.044	0.438	0.408	77.7%	0.165	<0.05
SRK/T	0.376	0.300	0.014	0.485	0.486	73.2%	0.143	<0.05
Our method	0.312	0.242	0.015	0.418	−0.033	80.2%	0.447	/

SD: the standard deviation of the signed prediction error; m: the axial length bias, computed as the slope of the correlation between the arithmetic error and al; AE ≤0.5 D: percentage of eyes with an AE less than or equal to 0.5 D. The unit for the errors is dioptres (D). Wilcoxon test p<0.05 indicates the statistical significance of the difference between the performance of our method and an existing method.
AE, absolute error; D, dioptre; EVO, Emmetropia Verifying Optical; FPI, Formula Performance Index; MAE, mean absolute error; ME, mean error; MedAE, median absolute error.
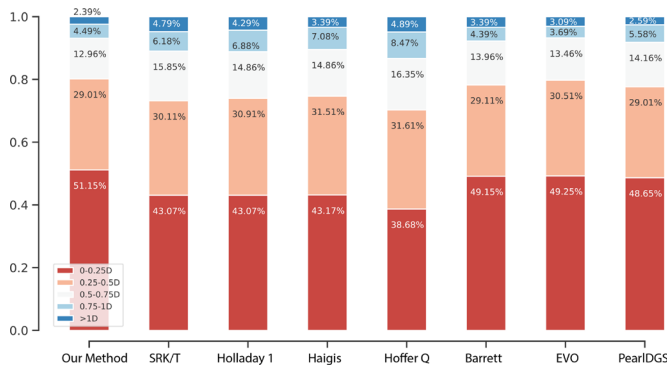
Figure 2 The percentage of patients in each error category for each formula, calculated based on the results in the testing dataset. EVO, Emmetropia Verifying Optical.

Figure 3 The mean prediction errors in the testing set grouped based on axial lengths. Each dot represents the mean prediction error of eyes with an axial length between a specific range. EVO, Emmetropia Verifying Optical.

as a hold-out testing dataset for performance comparison. A summary of the patient demographics in the training and testing sets is shown in table 2. A total of 49 surgeons performed the surgeries included in dataset. The distribution of data is shown in online supplemental figure S1.

The performance of our method and existing methods is shown in table 3. According to the Wilcoxon test, our method performed significantly better than all the other seven methods with an MAE of 0.312 D, which was 4.9% lower than that of Barrett (0.328 D) and 3.1% lower than that of EVO (0.322 D). The specific p values can be found in online supplemental table S1. Our method also achieved the highest FPI.

The percentage of patients with an absolute error less than or equal to 0.25 D, 0.50 D, 0.75 D and 1.00 D is shown in figure 2. Our method resulted in a larger percentage of patients in the absolute error ≤0.5 D group (80.2%) compared with Barrett (78.3%), EVO (79.8%) and PearlDGS (77.7%), and a larger percentage of patients in the absolute error ≤1.0 D group (97.6%) compared with Barrett (96.6%), or EVO (96.9%) and PearlDGS (97.4%). Overall, our method achieved the highest percentage in the absolute error ≤0.5 D group among all eight formulas, and was statistically better than all other formulas except EVO (Cochran's Q test p values were shown in online supplemental table S2) on this metric.

We compared the performance of the tested formulas among patients with different axial lengths in table 4. Numerically, our method achieved the lowest MAEs and SDs among all eight formulas in all 3 AL groups. The relationship between the
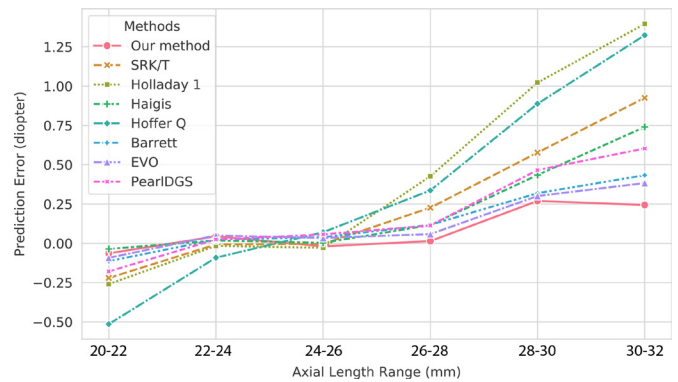
prediction errors and the ALs is shown in figure 3. The errors of our method remained close to zero across the whole span of ALs.

When the model was trained with different proportions of the training data (figure 4), the corresponding performance on the testing set displayed a trend towards improving performance (decreased MAE) with increasing training set sizes.

DISCUSSION

We have presented here a new ML-based IOL power calculation method which performs statistically significantly better than Barrett Universal II, EVO (V.2.0) and PearlDGS on a large unseen testing dataset. We chose an ensemble ML framework for this particular problem, and this choice allows the method to compensate for the potential biases of individual learners. During the development of the model, we designed and applied several data augmentation methods to enhance prediction performance. Data augmentation methods are not only beneficial for enlarging the dataset size, but also to address natural imbalances in clinical datasets. The biometry measures are not uniformly distributed as shown in online supplemental figure S1. For example, the axial length has more instances in the medium group (between 22 mm and 26 mm) compared with the long and short AL groups. The postoperative refractions and the implanted IOL powers were not uniformly distributed either. All IOL powers in the dataset were manually selected by surgeons with a particular target refraction in mind, typically between 0 D and −3 D. Data

Table 4 The postoperative refraction prediction performance of existing formulas and our method in short/medium/long al groups in the testing set

	Short AL (<22.0 mm, n=32)				Medium AL (≥22.0 and ≤26.0 mm, n=878)				Long AL (>26.0 mm, n=93)			
Method	MAE	MedAE	ME	SD	MAE	MedAE	ME	SD	MAE	MedAE	ME	SD
Barrett	0.401	0.259	−0.108	0.543	0.325	0.255	0.032	0.434	0.332	0.259	0.142	0.402
EVO	0.390	0.277	−0.088	0.514	0.319	0.250	0.043	0.424	0.330	0.269	0.086	0.410
Haigis	0.429	0.380	−0.029	0.551	0.360	0.285	0.012	0.466	0.369	0.295	0.156	0.437
Hoffer Q	0.653	0.577	−0.494	0.623	0.381	0.310	−0.015	0.486	0.532	0.442	0.407	0.520
Holladay 1	0.520	0.500	−0.241	0.592	0.346	0.280	−0.021	0.452	0.549	0.489	0.505	0.473
PearlDGS	0.426	0.402	0.170	0.564	0.320	0.250	0.040	0.426	0.387	0.320	0.154	0.468
SRK/T	0.513	0.491	−0.198	0.592	0.368	0.294	−0.005	0.476	0.402	0.322	0.273	0.430
Our method	**0.380**	**0.314**	**−0.050**	**0.512**	**0.310**	**0.238**	**0.015**	**0.417**	**0.312**	**0.269**	**0.041**	**0.389**

AL, axial length; EVO, Emmetropia Verifying Optical; MAE, mean absolute error; ME, mean error; MedAE, median absolute error; n, number of eyes in each group; SD, standard deviation of the signed prediction error.

augmentation helps to account for the scarcity of extreme cases and biases introduced by clinical decision-making process.

In this study, we used a relatively large dataset of 6893 eyes. Evaluation of the relationship between the proportion of the available training data used and MAE demonstrated the expected inverse relationship. This trend continued even as the training set was increased from 90% to 100% of the available training data (figure 4), indicating the potential for further improvement as the same model is exposed to larger datasets.

We achieved lower MAEs than Barrett Universal II, PearlDGS, and EVO in all three axial length groups. Our method yielded 80.2% of eyes with a predicted refraction within $\pm 0.5\,D$ of the true refraction, which was approximately 2% more than that of Barrett (78.3%) (p=0.04). The Nallasamy formula also achieved 51.2% of eyes within $\pm 0.25\,D$, which was approximately 2% more than that of all other methods (next closest was EVO at 49.3%). Due to sheer volume of cataract surgery worldwide—23 million cataract surgeries each year—achieving an additional 2% of patients with refractive error less than 0.25 D would likely be clinically relevant at a population level. At the same time, the difference in MAE between our method and the next closest (EVO) of 0.010 D is not likely to be of clinical significance for the average patient. This discrepancy in clinical relevance appears to arise from the difference between the average patient and the overall population. Table 4 demonstrates that the differences in MAE are smaller in the medium axial length group than in the short and long axial length groups. Since there are far more patients in the medium axial length group than in the short and long axial length groups, the reported MAE reflects the smaller difference in errors in the more common medium axial length group. The overall difference in percentage of patients with errors less than 0.25 D is reflective of larger errors typically seen in the short and long axial length groups. Figure 3 highlights the divergence in prediction error of the Nallasamy formula and other methods at the limits of axial length.

Recently, Hoffer *et al* proposed in Ophthalmology the use of the FPI as a means of evaluating and ranking the performance of IOL power calculation methods.[30] Higher values of the FPI indicate higher performance. Our method strongly outperformed the existing formulas on FPI, achieving a 0.447 FPI while the existing formulas ranged from 0.085 to 0.312 (table 3). The FPI takes into account the (1) SD of the prediction error, (2) the MedAE, (3) the AL bias, and (4) the percentage of eyes with refraction predictions within 0.5 D of true refractions. Our method demonstrated superior performance on each of these
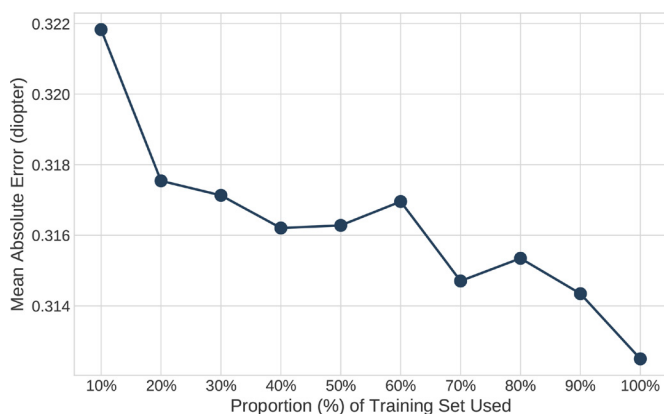
individual metrics, as summarised in table 3. Of particular note is our method's superior SD of the prediction error, which Holladay *et al* recently referred to as "the single best parameter to characterise the performance of an IOL power calculation formula."[31]

Also of interest is the AL bias, which is calculated as the slope of the correlation of the AL and the prediction error for a given formula. The existing IOL formulas demonstrate strong correlations between AL and the prediction error, as depicted in figure 3. ML-based methods such as ours, on the other hand, have the potential to better capture the nonlinearity of the relationship between biometric variables, IOL power, and postoperative refraction, resulting in substantially smaller AL bias (eg, −0.03 for Nallasamy vs 0.31 for Barrett). This translates to improved performance across AL categories (short, medium, and long), and should obviate the need for using different formulas based on axial length.

We are aware of multiple limitations of our study. Our method has not yet been validated on a dataset from a different medical institution. Performance analysis on external datasets will be a focus of future work as we begin to apply our approach to different populations around the world. Another limitation is that we were not able to compare our performance with a few formulas such as Hill-RBF because of a lack of access. However, prior studies indicate that Barrett Universal II is a good reference point for top-tier IOL formulas.[4 5 32] An additional limitation is that at present, our method has been customised for the Alcon SN60WF lens, and additional data will be needed to adjust the method for additional lens models. We were not able to test the Nallasamy formula's performance on eyes with extremely long or extremely short axial lengths due to a lack of available data in our dataset. Considering the Nallasamy formula was not trained with those eyes either, we believe the Nallasamy formula is currently not suitable to be used for extreme eyes. The online Nallasamy formula calculator (available at https://lenscalc.com/) displays a warning message if AL is outside the range of 21 mm - 31.5 mm. Similarly, a warning is displayed if the K readings are outside the range of 37 D–52 D.

An intrinsic difference between ML-based methods and the vergence formulas is that vergence formulas estimate the ELP as a vital variable during the calculation of the postoperative refraction, but ML-based methods usually take a one-step approach for prediction, unless the model is specifically designed to predict both the ELP and the postoperative refraction. In previously published work, we reported the development of an ML-based method for postoperative ACD estimation.[16 17] However, the method presented here does not rely on prediction of a postoperative ACD or ELP as an intermediate variable, unlike the vergence formulas. This approach may allow the ML method to avoid the propagation of errors (however small) introduced during the prediction of the postoperative ACD or ELP.

While the theoretical optics-based methods remain crucial for special cases, ML offers improved performance for large populations through the identification of latent patterns in historical data that can go unrecognised by existing methods. To that end, we have reported here the successful development and testing of an ML-based approach to IOL power calculation for cataract surgery that outperforms Barrett Universal II, PearlDGS and EVO on all broadly accepted metrics of IOL calculation performance. The Nallasamy formula is now freely available to the public to use online at https://lenscalc.com/.

Figure 4 The change of the mean absolute prediction error in the testing set when the machine learning method uses 10%, 20%, …, 100% of the training data.

method development, and writing of the manuscript. NN is responsible for the overall content as the guarantor.

ORCID iDs
Tingyang Li http://orcid.org/0000-0003-2559-184X
Nambi Nallasamy http://orcid.org/0000-0001-7501-7198

REFERENCES

1 Jones JJ, Jones YJ, Jin GJC. Indications and outcomes of intraocular lens exchange during a recent 5-year period. *Am J Ophthalmol* 2014;157:154–62.
2 Goemaere J, Trigaux C, Denissen L, *et al*. Fifteen years of IOL exchange: indications, outcomes, and complications. *J Cataract Refract Surg* 2020;46:1596–603.
3 Barrett universal II formula V1.05. Available: https://calc.apacrs.org/barrett_universal2105/ [Accessed 31 Aug 2021].
4 Kane JX, Van Heerden A, Atik A, *et al*. Accuracy of 3 new methods for intraocular lens power selection. *J Cataract Refract Surg* 2017;43:333–9.
5 Melles RB, Holladay JT, Chang WJ. Accuracy of intraocular lens calculation formulas. *Ophthalmology* 2018;125:169–78.
6 EVO formula. Available: https://www.evoiolcalculator.com/ [Accessed 20 Jan 2022].
7 Debellemanière G, Dubois M, Gauvin M, *et al*. The PEARL-DGS formula: the development of an open-source machine Learning-based thick IOL calculation formula. *Am J Ophthalmol* 2021;232:58–69.
8 PearlDGS formula. Available: https://iolsolver.com/ [Accessed 29 Jan 2022].
9 Akkara J, Kuriakose A. Role of artificial intelligence and machine learning in ophthalmology. *Kerala J Ophthalmol* 2019;31:150–60.
10 Armstrong GW, Lorch AC. A(eye): A Review of Current Applications of Artificial Intelligence and Machine Learning in Ophthalmology. *Int Ophthalmol Clin* 2020;60:57–71.
11 Clarke GP, Kapelner A. The Bayesian additive regression trees formula for safe machine Learning-Based intraocular lens predictions. *Front Big Data* 2020;3:46.
12 Yamauchi T, Tabuchi H, Takase K, *et al*. Use of a machine learning method in predicting refraction after cataract surgery. *J Clin Med* 2021;10. doi:10.3390/jcm10051103. [Epub ahead of print: 06 03 2021].
13 Carmona González D, Palomino Bautista C. Accuracy of a new intraocular lens power calculation method based on artificial intelligence. *Eye* 2021;35:517–22.
14 Stein JD, Rahman M, Andrews C, *et al*. Evaluation of an algorithm for identifying ocular conditions in electronic health record data. *JAMA Ophthalmol* 2019;137:491–7.
15 Bommakanti NK, Zhou Y, Ehrlich JR, *et al*. Application of the sight outcomes research collaborative ophthalmology data Repository for triaging patients with glaucoma and clinic appointments during pandemics such as COVID-19. *JAMA Ophthalmol* 2020;138:974–80.
16 Li T, Stein J, Nallasamy N. AI-powered effective lens position prediction improves the accuracy of existing lens formulas. *Br J Ophthalmol* 2022;106:1222–6.
17 Li T, Yang K, Stein JD, *et al*. Gradient boosting decision tree algorithm for the prediction of postoperative intraocular lens position in cataract surgery. *Transl Vis Sci Technol* 2020;9:38.
18 Zhang Y, Li T, Reddy A, *et al*. Gender differences in refraction prediction error of five formulas for cataract surgery. *BMC Ophthalmol* 2021;21:183.
19 Simpson MJ, Charman WN. The effect of testing distance on intraocular lens power calculation. *J Refract Surg* 2014;30:726.
20 Dieterich TG. Ensemble Methods in Machine Learning. In: *Multiple classifier systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000: 1–15.
21 Wolpert DH. Stacked generalization. *Neural Networks* 1992;5:241–59.
22 Retzlaff JA, Sanders DR, Kraff MC. Development of the SRK/T intraocular lens implant power calculation formula. *J Cataract Refract Surg* 1990;16:333–40.
23 Hoffer KJ. The Hoffer Q formula: a comparison of theoretic and regression formulas. *J Cataract Refract Surg* 1993;19:700–12.
24 Haigis W, Lege B, Miller N, *et al*. Comparison of immersion ultrasound biometry and partial coherence interferometry for intraocular lens calculation according to Haigis. *Graefes Arch Clin Exp Ophthalmol* 2000;238:765–73.
25 Holladay JT, Prager TC, Chandler TY, *et al*. A three-part system for refining intraocular lens power calculations. *J Cataract Refract Surg* 1988;14:17–24.
26 Correction. *J Cataract Refract Surg* 1994;20:677.
27 Erratum. *J Cataract Refract Surg* 1990;16:528.
28 Zuberbuhler B, Morrell AJ. Errata in printed Hoffer Q formula. *J Cataract Refract Surg* 2007;33:2.
29 Hoffer KJ. Reply: Errata in printed Hoffer Q formula. *J Cataract Refract Surg* 2007;33.
30 Hoffer KJ, Savini G. Update on intraocular lens power calculation study protocols: the better way to design and report clinical trials. *Ophthalmology* 2021;128.
31 Holladay JT, Wilcox RR, Koch DD. Re: Hoffer et al.: update on intraocular lens power calculation study protocols: the better way to design and report clinical trials. *Ophthalmology* 2021;128:e20.
32 Kane JX, Van Heerden A, Atik A, *et al*. Intraocular lens power formula accuracy: comparison of 7 formulas. *J Cataract Refract Surg* 2016;42:1490–500.