Artificial Intelligence-Based Clinical Decision-Making System for Cataract Surgery

by

Tingyang Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in the University of Michigan
2022

Doctoral Committee:

Assistant Professor Nambi Nallasamy, Chair
Professor Jun Z. Li
Professor Kayvan Najarian
Associate Professor Xu Shi
Assistant Professor Michael W Sjoding
Associate Professor Bradford L Tannen

Tingyang Li

tyli@umich.edu

ORCID iD: 0000-0003-2559-184X

Dedication

To my cats Biggie and Tiny.

Acknowledgements

I would like to thank my advisor Dr. Nambi Nallasamy for his guidance on research and his assistance in all aspects of my Ph.D. life over the past several years. It would have been impossible for me to complete my Ph.D. and achieve what I have accomplished without his patience and support. I felt incredibly lucky and undeserved to have met the greatest, smartest, and kindest advisor in the world.

I would also like to thank my committee members Dr. Bradford Tannen, Dr. Jun Li, Dr. Kayvan Najarian, Dr. Michael Sjoding, and Dr. Xu Shi for their invaluable guidance and support throughout my dissertation research. I am thankful to Dr. Joshua Stein for helping us gather the datasets for the projects. Additionally, I would like to thank the students and friends with whom I have shared happy times.

Finally, I would like to thank my pets especially for providing emotional and physical support to me, and for spending a quarter of their lives with me.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Cataracts are a common eye condition characterized by the clouding of the natural crystalline lens that impairs vision. In cataract surgery, the natural lens of the eye is replaced by an intraocular lens (IOL) implant to restore vision to the patient. A cataract surgery procedure involves the removal of the natural cloudy lens by means of ultrasound phacoemulsification, followed by the replacement of the lens with an IOL implant chosen specifically for the patient. The power of the IOL should be selected appropriately in order to attain optimal postoperative vision. IOL power calculation has evolved over multiple generations, starting with regression and theoretical optics-based formulas to machine learning (ML) formulas, as well as hybrid formulas. Nevertheless, there is still room for improvement in terms of prediction accuracy.

At University of Michigan's Kellogg Eye Center, we gathered a large collection of medical records of cataract patients and retrieved demographic and surgical information from the Sight Outcomes Research Collaborative (SOURCE) database. In this dissertation, I present research projects in which we used artificial intelligence-based approaches to improve the functionality and accuracy of tools assisting the clinical decision-making procedure. Through a series of studies, we developed and investigated methodologies for achieving better refractive results with cataract surgery with the aid of artificial intelligence algorithms. Predicting where the intraocular lens resides within the eye after cataract surgeries is a critical step in the determination of the IOL power and several other applications (e.g., ray tracing). In Chapter 2, we gathered a dataset of 847 patients and developed methods for predicting the postoperative IOL position (postoperative anterior chamber depth) which outperformed existing methods

including Haigis, Hoffer Q, Holladay I, Olsen, and SRK/T. Further, we explored whether the ML-predicted lens position can be used to improve the performance of existing IOL formulas. In Chapter 3, we combined the ML-based postoperative ACD prediction method described in Chapter 2 with existing optics-based methods including Haigis, Hoffer Q, Holladay, and SRK/T. These methods use theoretical Gaussian optics-based equations to estimate the effective lens position (ELP). When combine with our ML-predicted lens position, all four formulas achieved significantly better prediction performance. In Chapter 4, we used our ML-predicted postoperative ACD with a ray tracing-based IOL formula (OKULIX), and significantly improved its performance. In Chapter 5, we described a new machine learning-based formula, Nallasamy formula, that predicts the most appropriate IOL power based on the preoperative biometry alone. We showed that this new formula outperformed existing formulas including the Barrett Universal II formula, and the Emmetropia Verifying Optical (EVO) formula. In Chapter 6, we demonstrate the risks of using standard evaluation metrics for ML-based IOL formulas, and present two new metrics for more robust evaluation of IOL power prediction formulas. In Chapter 7, we summarize the main findings and discuss future directions.

Together, the projects presented in this dissertation examined and proved the possibility of utilizing artificial intelligence in cataract surgery decision-making. These studies are among the pioneers of the use of ML in cataract surgery, and they will lay the foundation for the next era of cataract surgery planning.

Chapter 1 Introduction

1.1 Cataract and Cataract Surgery

Cataract is a pathological condition wherein the transparent crystalline lens of the eye develops opacifications, causing blurry, cloudy, dim vision which could negatively affect the patient's quality of life, and may eventually lead to blindness if not treated properly. Cataract formation, or the cataractogenesis process, has various contributing factors including ultraviolet radiation, trauma, systemic diseases such as diabetes and hypertension, genetic mutations, and chemical toxicity induced by drugs.[1,2] Among all factors, the most common cause of visually significant cataract is aging.[3] According to a survey of the National Eye Institute in 2010, 68.30% of people over 80 years old in the United States had cataract, and the overall prevalence rate of cataract in the United States was 17.11%.[4] Globally, senile cataract (age-related cataract) prevalence varies significantly among different regions, possibly due to variations in genetic and environmental factors.[3,5]

The earliest documentation of cataract dated back to the 5th dynasty of Egypt (2457-2467 B.C.).[6,7] Cataract was first treated around 800 B.C. by “couching”, a technique for dislodging the cataract into the vitreous cavity. [6,7] Today, cataract patients can safely restore their vision through surgical extraction of the clouded lens, and implantation of a transparent artificial lens. The high prevalence of cataract makes cataract surgery one of the most commonly performed surgery in the United States. It is estimated that about 4 million cataract surgeries are performed each year in the United States alone, and about 23 million cataract surgeries are performed each

year worldwide. With the advancement of cataract surgery over centuries, it is now a safe and effective procedure to restore sight to patients.

Prior to the cataract surgery, the surgeon decides which IOL type and which power is appropriate for the patient based on the patient's preference and the surgeon's experience. Three main types of IOLs are available: monofocal lenses, multifocal lenses, and toric lenses. Monofocal lenses correct vision at only one focal length (far, intermediate, or near), so the patient may require glasses for near vision if the lens corrects distance vision only, and vice versa. Multifocal lenses correct both near and distance vision. Toric lenses corrects astigmatism. There are other premium IOLs such as aspheric IOLs which are monofocal lenses that correct for spherical aberration; accommodating IOLs which move with the eye's ciliary muscles to provide clear vision at all distances; and trifocal IOLs which correct near, far, and intermediate vision. The most common type of lens used in cataract surgery is the monofocal lens. It is also called the standard IOL and is typically covered by medical insurance in the United States. The studies in this dissertation were focused on the Alcon AcrySof SN60WF lens which is a single-piece monofocal blue-light filtering lens. It is currently one of the most commonly implanted lenses in the US and the most commonly implanted IOL at University of Michigan's Kellogg Eye Center.

The power of the IOL is calculated using the surgeon's preferred IOL formulas based on the patient's preoperative biometry and the target refraction, which is the desired postoperative refraction. For monofocal lenses, such as the Alcon SN60WF lens, patients and surgeons usually prefer to aim for a distance target (emmetropia) to obtain a broader range of usable vision than with a near target. This is because the +/-0.25D depth of field of typical monofocal IOLs translates to a broader real-world depth of field at emmetropia than at near. The target refraction is typically set at a slightly myopic value (e.g. -0.25D), primarily to reduce the risk of hyperopic

surprise. A hyperopic result after cataract surgery can leave a patient without any point in space that is in focus without the use of corrective eyewear (e.g., glasses or contact lenses). Some patients prefer a near target (e.g., intentional myopia of -2.5D to -3.0D) instead of emmetropia, for example those with high or moderate myopia [8], those who spend more time performing near-distance tasks (such as reading, using computers and mobile phones), and those who intend to have monovision (one eye set for distance and one eye set for near).

The surgical procedure can be summarized as follows. The surgeon creates incisions in the cornea to gain access to the crystalline lens. After filling the anterior chamber with a viscoelastic device that provides anterior chamber stability and protection to intraocular structures, the surgeon carefully creates a small circular opening in the anterior capsule of the crystalline lens. The surgeon then uses an ultrasonic handpiece to divide, emulsify, and aspirate the cataractous lens material. This technique is named Phacoemulsification, first developed by Charles Kelman in 1967 and popularized in the 1980s. Phacoemulsification is now the method of choice for most cataract surgery.[9] After removal of lens cortical material, an artificial lens is then implanted into the capsular bag. Due to the thin profile of standard intraocular lens implants relative to the natural lens, cataract surgery causes deepening of the anterior chamber depth (ACD) and decreases the intraocular pressure.[10] The corneal power is also subject to surgical change due to the incisions made in the cornea. However, there is currently no reliable method of predicting such postoperative changes in these biometric characteristics.

During the patient's postoperative visits, the technician measures the manifest refraction and the best spectacle-corrected visual acuity (BSCVA) of the patient with a manual or automatic phoropter. The spherical equivalent (SE) is typically used as an estimate of the eye's refractive error, which combines the spherical component (SC) which measures nearsightedness

or farsightedness, and cylindrical component (CC) which measures astigmatism, with the following equation: $SE = SC + 0.5CC$. The visual acuity measures the sharpness of vision at a certain distance. It has been proven that factors such as the lens position, refraction, and visual acuity of the operated eye should stabilize about a month after the surgery.[11,12] In the event that the refractive outcome of the surgery is not satisfactory, additional surgical procedures such as IOL exchange may be required.[13]

1.2 Intraocular Lens Power Calculation

1.2.1 Optical Biometry Measurements

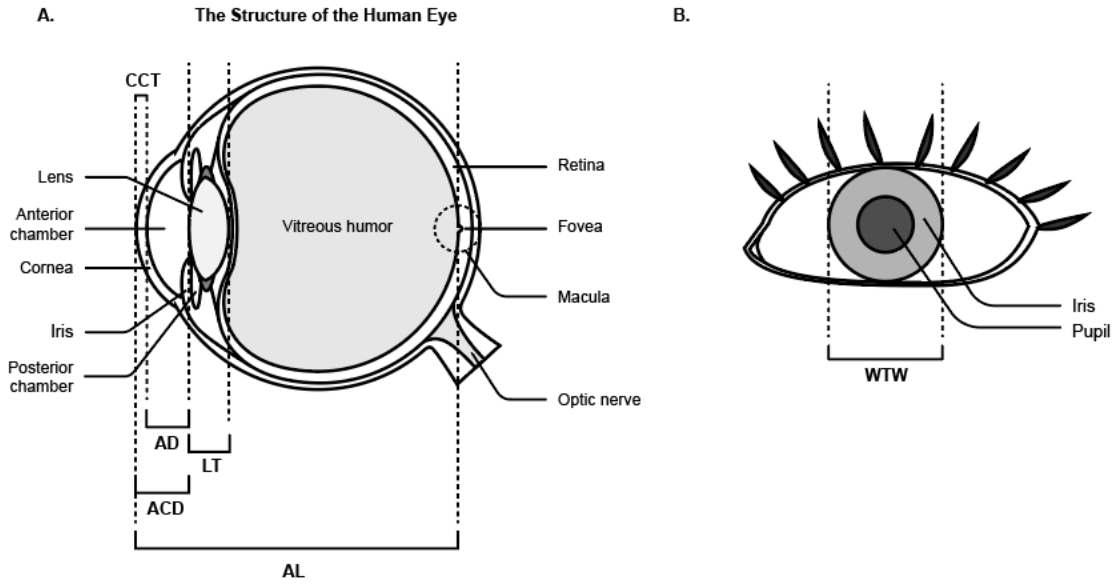

Figure 1.1 Eye structure.

The goal of an IOL formula is to predict an optimum IOL power so that the recipient of the IOL can reach a desired postoperative refraction. To make such predictions, preoperative biometry measurements of patients' eyes are taken into consideration. The following are the typical biometry measurements provided by standard optical biometers (**Figure *1.1***): the axial length (AL, measured in mm) which is the distance from the corneal vertex to the macula; the

lens thickness (LT, measured in mm), which measures the central thickness of the crystalline lens (or the IOL of a pseudophakic eye); the central corneal thickness (CCT, measured in μm) measures the central thickness of the cornea; the flat corneal power (K1, measured in diopter) and steep corneal power (K2, measured in diopter); astigmatism (AST, measured in diopter) is the degree of astigmatism, which is the difference between K2 and K1 ($AST = K2 - K1$). White to white (WTW, measured in mm) is the horizontal corneal diameter; the aqueous depth (AD, measured in mm) is the distance between the corneal endothelium and anterior lens surface; the anterior chamber depth (ACD, measured in mm) is the distance between the front surface of the cornea and anterior lens surface ($ACD = AD + CCT$). In pseudophakic eyes, the lens position can be represented by the ACD, and the LT is the thickness of the IOL.

Two types of biometry measurement devices (biometers) are available: optical biometers which use infrared light, and ultrasound biometers which use ultrasound. As optical biometers allow for noninvasive and noncontact measurements, they are used for routine cataract patients. Contact ultrasound biometers are typically used for patients with dense cataracts, an inability to follow fixation instructions, or positioning issues. Examples of most used optical biometers are the Lenstar series (Haag-Streit AG, Switzerland) which uses Partial Coherence Interferometry (PCI), and the IOLMaster series (Carl Zeiss AG, Germany) which uses optical low-coherence reflectometry (OLCR). Previous studies have shown that these devices have high repeatability and reproducibility with low intraobserver and interobserver variations.[10,14–16]. Biometers are operated through bundled software installed on a computer, which can be used to record patient information and biometry measurements, as well as calculate the IOLs powers with built-in formulas. The Pentacam (OCULUS, Germany) is another commonly used device that takes cross-sectional images of the anterior segment of the eye with a rotating Scheimpflug camera. A

distinguishable difference between Pentacam and optical biometers is that it recognizes both anterior and posterior surface of the cornea, which enables measurement of the true total refractive power of the cornea. By contrast, in the Lenstar and IOLMaster, the keratometry is an estimation of the corneal power based only on the measured radius of curvature of the anterior corneal surface. The computation of the keratometry considers a preset keratometric refraction index, and a best-fit ellipsoid built based on projected light reflections.[17] The difference in biometry measurements from IOLMaster 500, IOLMaster 700, and Lenstar LS 900 are believed to be clinically insignificant.[18–20] The difference between predicted IOL powers based on Lenstar or IOLMaster were also found to be clinically insignificant. [21] The keratometry readings from Pentacam were found to be significantly different from the ones obtained by Lenstar or IOLMaster [22–24]. Examples of biometry measurement devices are shown in **Table *1.1***.

Device	Manufacturer	Technology
Anterion	Heidelberg Engineering Ltd., Hertfordshire, United Kingdom	Swept Source (SS) OCT
Argos	Alcon Inc, Texas, United States	Swept Source (SS) OCT
Galilei G6	Ziemer, Port, Switzerland	All-in-one: dual Scheimpflug tomography, Placido topography and optical biometer
IOL Master 500	Carl Zeiss Meditec AG, Jena, Germany	Partial coherence interferometry (PCI)
IOLMaster 700	Carl Zeiss Meditec AG, Jena, Germany	Swept Source (SS) OCT
Lenstar LS 900	Haag-Streit AG, Koeniz, Switzerland	Low-coherence optical reflectometry (LCOR)
Pentacam AXL	OCULUS, Germany	Scheimpflug imaging

Table 1.1 Examples of biometry measurement devices.
OCT: optical coherence tomography.

1.2.2 Existing IOL Formulas

Existing IOL formulas can be broadly classified into these categories: theoretical Gaussian optics-based, empirical regression-based, ray tracing-based, and machine learning-based IOL formulas. The earliest Gaussian optics-based IOL formulas (published between the 1950s and 1980s), such as the Binkhorst formula, Thijssen formula, Colenbrander formula, Fyodorov formula shared a common form as shown below, even though they were independently derived (**Table *1.2***). The main difference is in the values of adjustment factors (h_i and h_c). [25] The derivation of the thin lens formula based on a paraxial eye model is demonstrated in **Appendix A**. These first-generation thin lens formulas contain an adjustment factor which can be adjusted to fit different IOL types. This adjustment factor represents an estimate of postoperative ACD that is constant across all patients.

$$P = \frac{N}{AL - C - h_i} - \frac{N}{\frac{N}{K} - C - h_c}$$

P = IOL power for emmetropia in diopters; AL = axial length in meters; K = corneal power; C = estimated postoperative anterior chamber depth in meters; N = refractive index of vitreous and aqueous humor; h_i and h_c = corrections for the principal planes of the IOL and the cornea, respectively.

Different from the optics-based formulas, the SRK formula (devised by Sanders, Retzlaff and Kraff in 1980) is a linear regression function fitted using a historical dataset.

The SRK II formula, which is a second generation formula, improved on the estimation of postoperative ACD by adjusting it with AL. Third-generation formulas introduced the concept of effective lens position (ELP), which represents an imaginary location of the thin lens (in the thin lens approximation), but not its actual physical location.[26,27] ELP was modeled as a function of the preoperative AL and K. It was later proved that the errors in the estimation of the ELP is a vital source of errors (35.5%) in refraction prediction according to Norrby et al.[28]

The fourth generation and newer generation formulas improved the estimation of ELP and the prediction accuracy further.

In addition to standard biometry, recent IOL formulas have started to include patient demographics, such as age and gender in IOL calculation. There is increasing evidence supporting the inclusion of demographic data. The influence of age on the prediction errors has been previously described. [29–32] Gender- and race-related variability in biometry of cataract patients was also observed.[33–36] Findings of these publications suggest that the demographics of a patient may play a role in IOL power prediction.

The above-mentioned IOL formulas are dedicated to serve general cataract patients, with physiologic, prolate corneas. IOL power calculation of difficult cases such as patients who had previous corneal refractive surgery should follow different protocols. The primary source of error in IOL power selection using standard formulas for post-refractive surgery eyes can be attributed to the inaccurate measurement of the central corneal power (K).[37] This is because standard optical biometers do not measure the refractive power of the cornea in diopters directly, but instead they estimate the total corneal power based on the measured radius of curvature of the anterior corneal surface using the following equation.[37–39]

$$K = \frac{n_c - 1}{R_{anterior}}$$

In this equation, K is the estimated corneal power in diopters; $R_{anterior}$ is the measured radius of curvature of the anterior corneal surface in meters; n_c is the keratometric index (typically set to 1.3315, 1.336, 1.3375 or 1.338). The detailed formulation of this equation has been described in previous publications,[40] as well as demonstrated in **Appendix B**. The estimation of the total corneal power makes assumptions about (1) the refractive index of the

cornea, (2) the refractive index of aqueous humor, (3) the average anterior corneal curvature, and (4) the anterior-posterior corneal curvature ratio.

$$\text{AP ratio} = \frac{R_{anterior}}{R_{posterior}}$$

Here $R_{anterior}$ is the radius of curvature of the anterior corneal surface, and $R_{posterior}$ is the radius of curvature of the posterior corneal surface. The keratometric index represents a mixture of the presumed constants. Corneal refractive surgery alters the shape of the cornea (anterior and posterior curvature), so the keratometric index will no longer be accurate in these cases, and the K measurements for these eyes become inaccurate. [39,41] Several publications have reviewed the methods and formulas (e.g. Barrett True-K, Haigis-L[42] and OCT-based formulas) for selecting IOL powers for post-refractive eyes.[41,43,44] Currently, there is no universally accepted gold standard, and it is generally recommended to take into consideration the results from different calculation methods for a better selection of the IOL powers for eyes that had previous refractive surgery.[44]

Generation	Formula	Equation	Reference
1st	Binkhorst	IOL power P to achieve emmetropia: $P = \frac{1336(4R - AL)}{(AL - C)(4R - C)}$ R is the radius of corneal curvature	[45]
	Colenbrander	IOL power P to achieve emmetropia: $P = \frac{1336}{AL - C - 0.05} - \frac{1336}{\frac{1336}{K} - C - 0.05}$	[46]
	Fyodorov	IOL power P to achieve emmetropia: $P = \frac{1336 - AL * K}{(AL - C)(1 - \frac{CK}{1336})}$	[47]
	Thijssen	IOL power P to achieve emmetropia: $P = \frac{1336}{AL - C + Const1} - \frac{1336}{\frac{1336}{K} - C + Const2}$	[48]
	SRK	IOL power P to achieve emmetropia: $P = A - 2.5AL - 0.9K$, A is the A-constant	[49,50]

2nd	SRK II	IOL power P to achieve emmetropia: $P = A1 - 2.5AL - 0.9K,$ A1 is a factor calculated based on the A-constant and AL	[51]

Table 1.2 Equations of first and second generation IOL formulas.

P: IOL power (D); AL: axial length (in mm); K: average keratometry (D); C = estimated constant postoperative anterior chamber depth (mm).

Generation	Formula	Principal	Input preoperative biometry	Reference
1st	SRK	Regression	AL, K	[49,50]
	Thijssen	Gaussian optics (thin lens)	AL, K	[48]
2nd	SRK II	Regression	AL, K	[51]
	Hoffer	Gaussian optics (thin lens)	AL, K	[52]
3rd	SRK/T	Gaussian optics (thin lens)	AL, K	[53,54]
	Hoffer Q	Gaussian optics (thin lens)	AL, K	[52,55]
	Holladay 1	Gaussian optics (thin lens)	AL, K	[56]
4th	Barrett Universal II	Gaussian optics (thick lens)	AL, K, ACD, LT*, WTW*	[57]
	Olsen	Ray tracing	AL, K, ACD, LT	[58]
	Haigis	Gaussian optics (thin lens)	AL, K, ACD	[59]
	Holladay 2	Gaussian optics (thin lens)	AL, K, ACD, LT*, age*, WTW*, preoperative refraction*	[60]
New	T2	Gaussian optics (thin lens)	AL, K	[61]
	Ladas Super Formula (LSF)	Amalgam of existing formulas	AL, K, ACD*	[62–64]
	OKULIX	Ray tracing	AL, K, ACD*, LT*, posterior K*	[65]
	Kane	Gaussian optics and artificial Intelligence	AL, K, ACD, LT*, CCT*, gender	[66]
	Hill-RBF v3.0	Artificial Intelligence	AL, K, ACD, LT*, WTW*, CCT*	[67]
	EVO v2.0	Gaussian optics (thick lens)	AL, K, ACD, LT*, CCT*	[68]
	PearlDGS	Artificial Intelligence and Gaussian optics (thick lens)	AL, K, ACD, LT, CCT, WTW	[69]
	VRF-G	Gaussian optics, regression, and ray tracing	AL, K, ACD, WTW, LT, CCT, gender, preoperative refraction	[70]
	K6	Gaussian optics (thin lens and thick lens)	AL, K, ACD, WTW, LT, CCT	[71]

Table 1.3 Examples of IOL formulas of different generations.

Measurements marked with * are optional.

1.2.3 Artificial Intelligence and Machine Learning

Machine learning (ML) is a discipline within the broader artificial intelligence (AI) area that uses statistical methods to recognize patterns in data. Machine learning can be broadly categorized into three types of tasks: supervised learning, where the ground truth is known, unsupervised learning, where the cases are not labeled, and reinforcement learning, where the model learns interactively through trial and error. The problem of supervised learning can be further divided into two types: regression tasks involving cases with continuous numerical observations as the ground truth, and classification tasks involving discrete labels as the ground truth.

This section introduces state-of-the-art supervised ML algorithms that are most relevant to this dissertation. Support Vector Machines (SVMs) use support vectors to establish margins when building predictive models based on the data. Using it in conjunction with nonlinear kernels can map feature vectors into higher-dimensional spaces, enabling better case differentiation. In addition, decision trees-based ensemble machine learning algorithms have gained significant popularity because they are efficient and powerful, such as the random forest algorithm which applies bootstrap aggregation (i.e., bagging) to a set of decision trees, and gradient boosted decision trees (GBDT) which applies boosting as the meta-algorithm to ensemble weak learners (decision trees) and uses gradient descent as the optimization algorithm. Artificial neural networks (ANN) are originally introduced in the 1940s and 1950s. In recent years, it has gained increasing popularity as a result of performance improvements in Central Processing Units (CPUs) and Graphics Processing Units (GPUs), as well as a data explosion. Deep learning refers to the use of neural networks with a large number of hidden layers. A deep network allows for a multitude of parameters and thus the ability to learn very complex patterns.

Therefore, deep learning is particularly useful for complex tasks such as image recognition and labeling, text recognition, speech recognition, etc.

ML algorithms typically involve a variety of hyperparameters which play a critical role in the model's prediction performance. For instance, the tree depth for decision trees, the learning rate for gradient descent, and the number of hidden layers and neurons for ANN. Additionally, model selection is also necessary to decide on the appropriate data augmentation and feature engineering methods. Together, the optimization of ML models can be performed via cross-validation. The basic workflow of supervised ML consists of steps shown in **Figure *1.2***. Many machine learning toolboxes (Keras[72], TensorFlow[73], PyTorch[74], Scikit-Learn[75], XGBoost[76], LightGBM[77]) have been developed in recent years, which greatly facilitate and streamline the development of ML models.

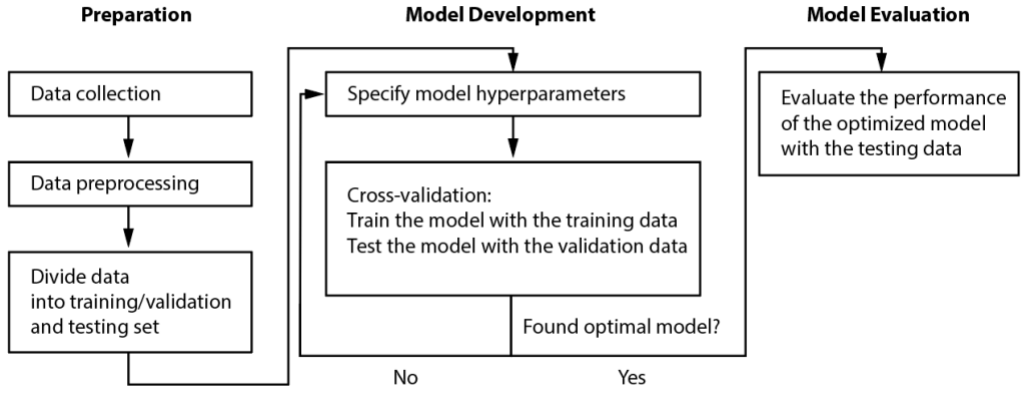

Figure 1.2 Supervised machine learning workflow.

1.2.4 AI-Based IOL Power Prediction

Ophthalmology is one of many fields in which machine learning has been used, including but not limited to recognition and diagnosis of ophthalmic diseases such as glaucoma, diabetic

retinopathy, macular edema and age-related macular degeneration.[78,79] IOL power calculation is in line with this trend. As shown in **Table *1.3***, existing AI-based IOL formula include the Hill-RBF formula, the Kane formula, and the PearlDGS formula. The Hill-RBF (Radial Basis Function) formula calculator was released in 2016 and is currently at version 3.0. The Kane formula was developed by Jack Kane in 2017, based on a dataset with nearly 30000 cases. This formula combines theoretical optics and artificial intelligence, and was found to outperform other existing formulas including Barrett Universal II. [70,80] The PearlDGS formula was published in 2021 claiming to outperform existing formulas including K6, Olsen, EVO v2.0, Hill-RBF v3.0 and Barrett Universal II.[69] Ladas Super Formula (LSF) AI,[64] which is currently unreleased, combines AI with the original Ladas Super Formula. Apart from the above-mentioned formulas, in the recent years, a growing number of machine learning-based methods for postoperative refraction prediction have been published and await further validation (**Table *1.4***).

ML has also been applied to postoperative ACD (or ELP) estimation.[81] ML-based prediction of ELPs using OCT measurements has also been explored. [82]

Method	Data Size	Function	Reference
Support vector machine regression model (SVM-RM) and multilayer neural network ensemble model (MLNN-EM)	1168 patients	Standard IOL calculation	[83]
Support vector regression (SVR), XGBoost and artificial neural network (ANN)	1107 patients	Adjust existing IOL formulas	[62]
Bayesian additive regression trees (BART)	3276 patients	Standard IOL calculation	[84]
Karmona method: an ensemble of support vector machines (SVMs) with radial basis function (RBF) and multivariate adaptive regression spline (MARS) with second-order polynomials	260 patients	Standard IOL calculation	[85]

Support vector regression (SVR), random forest regression (RFR), gradient boosting regression (GBR), and neural network (NN)	2010 patients	Standard IOL calculation	[86]
MM formula: combines thin lens geometric optics and machine learning (ensemble-based model)	681 patients	Standard IOL calculation	[87]

Table 1.4 Summary of recently (2020-2022) published artificial intelligence-based methods for IOL power prediction for cataract surgery.

1.2.5 Evaluation of IOL Formulas

The evaluation of the prediction performance of IOL formulas usually follows the following protocols [88–91].

To investigate the accuracy of IOL formulas designed for general cataract patients, eyes that had prior refractive surgery should be excluded. Patients with limited vision acuity (worse than 20/40) should be excluded because the postoperative refraction measurement of these cases may be inaccurate. In analyzing the formulas' performance, only one eye per patient should be included since most statistical tests assume the samples are independent.

The formula-specific constants should be optimized using previous patients' data to accommodate for systematic errors introduced by differences in lens material, lens geometry, surgical techniques of surgeons, biometry measurement devices, patient population, etc.[92,93] The most optimal constant should zero out the mean prediction error of postoperative refraction in the dataset. As an alternative to optimizing the constants with historical data, the surgeon may choose to use the constants provided by User Group for Laser Interference Biometry (ULIB).[94] After calculating predictions with preoperative biometry and appropriate lens constants, the prediction performance of IOL formulas can be measured and compared with the following metrics: mean error (ME), mean absolute error (MAE), median absolute error (MedAE), standard deviation (SD) of the arithmetic error of the refraction prediction, and the ratio of eyes

that obtain an error within $\pm 0.25\ D$, $\pm 0.5\ D$, $\pm 0.75\ D$, and $\pm 1\ D$. To investigate the stability of prediction accuracy in different AL groups, the performance of the IOL formulas in short (AL < 22 mm), medium (22 mm ≤ AL ≤ 26 mm), and long (AL > 26 mm) AL groups should be computed and compared. Based on previous publications, predictions of existing formulas for extremely short or long eyes are usually less accurate compared to eyes of medium axial length, especially the short eyes.[95–98]

Many retrospective studies have been conducted to compare the prediction performance of existing IOL formulas in general cataract patients.[70,80,96,99–105] The Barrett Universal II formula is well-established and historically proved to be one of the most effective formulas for general cataract patients, before the introduction of newer formulas. In some studies, newer formulas or newer version of existing formulas, such as Kane, Hill-RBF v3.0, VRF-G and EVO v2.0 have shown superior performance compared to traditional formulas, despite their relatively short existence.[70,80,96]

The performance rankings of formulas in long and short eyes are inconsistent across publications, especially as better-performing IOL formulas emerge. See 6.6 for more discussions on this topic.

1.3 Dissertation Outline

The goal of this dissertation is to describe a series of research focusing on the application of machine learning in postoperative refraction and lens position prediction for cataract surgery. This dissertation is structured as follows. In Chapter 2, I present our work on developing a machine learning-based method for predicting the postoperative lens position (i.e., the postoperative anterior chamber depth). In Chapter 3, I present a method to incorporate ML-predicted postoperative ACD into existing optics-based IOL formulas. In Chapter 4, I present

methods to incorporate ML-predicted postoperative ACD into an existing ray tracing based IOL formula. In Chapter 5, I present an ensemble machine learning-based model for predicting postoperative refractions for cataract surgery patient who received Alcon SN60WF lenses. In Chapter 6, I present two new metrics for evaluating the prediction performance of machine learning based IOL formulas. In Chapter 7, I summarize and discuss the findings and limitations of the presented studies, as well as future directions of research on AI-aided cataract surgery.

Chapter 2 Prediction of Postoperative Intraocular Lens Position in Cataract Surgery

2.1 Abstract

The purpose of this research was to develop a method for predicting postoperative anterior chamber depth (ACD) in cataract surgery patients based on preoperative biometry, demographics, and intraocular lens (IOL) power. Patients who underwent cataract surgery and had both preoperative and postoperative biometry measurements were included. Patient demographics and IOL power were collected from the Sight Outcomes Research Collaborative (SOURCE) database. A gradient boosting decision tree model was developed to predict the postoperative ACD. The mean absolute error (MAE) and median absolute error (MedAE) were used as evaluation metrics. The performance of the proposed method was compared to five existing formulas. A total of 847 patients were assigned randomly in a 4:1 ratio to a training/validation set (678 patients) and a testing set (169 patients). Using preoperative biometry and patient sex as predictors, the presented method achieved an MAE of 0.106 ± 0.098 (SD) on the testing set, and a MedAE of 0.082. MAE was significantly lower than that of the five existing methods ($p < 0.01$). When keratometry was excluded, our method attained an MAE of 0.123 ± 0.109, and a MedAE of 0.093. When IOL power was used as an additional predictor, our method achieved an MAE of 0.105 ± 0.091 and a MedAE of 0.080. In conclusion, the presented machine learning method achieved greater accuracy than previously reported methods for the prediction of postoperative ACD.

2.2 Introduction

As described in Chapter 1, the anterior chamber depth (ACD) indicates the location of the crystalline (natural) lens inside the phakic eye or that of an artificial lens in a pseudophakic eye. Since the natural lens is much thicker than current artificial lens implants, the postoperative ACD is usually deeper than the preoperative ACD. However, the exact change in the ACD after surgery is difficult to predict. Existing IOL formulas have taken different approaches to predict the postoperative lens position.

In first-generation IOL formulas, postoperative IOL axial position in the eye was modeled as a constant (4mm) in anterior chamber intraocular lens (ACIOL) power calculations. In second-generation formulas, Binkhorst introduced axial length as a predictor, while third-generation formulas involved both corneal power and axial length as predictors of postoperative IOL position. Olsen et al. introduced two additional variables, preoperative anterior chamber depth (ACD) and preoperative crystalline lens thickness as predictors for postoperative IOL position. In 1993, Holladay first proposed the term "expected lens position" or ELP to indicate the location of the lens as it relates to a given optical model of the eye.[106] The ELP estimates in SRK/T, Holladay1 and Hoffer Q are derived based on theoretical formulas. The ELP estimate in the Haigis formula is a simple linear combination of the axial length and the preoperative anterior chamber depth. Although ELP was initially intended to estimate the position of the IOL, ELPs in the aforementioned formulas were developed to account for different formula-specific assumptions and regression results.[26,28] In order to reflect the use of ELP to account for these formula-specific assumptions and regression results, the term ELP today refers to "effective lens position" rather than "expected lens position".

The importance of postoperative IOL position in IOL power calculations is due to the reliance of optical models of the eye on the distances between the optical components of the eye (the cornea and IOL) and the photoreceptors within the retina. Whether utilizing Gaussian optics or ray-tracing, optical models used in IOL power calculation require accurate estimates of postoperative IOL position to achieve useful results. Indeed, Norrby estimated in 2008 that estimates of IOL position were responsible for 36% of the error in IOL power predictions.[28]

While effective lens position (ELP) refers to the distance between the anterior surface of the cornea and the principal plane of the IOL resulting in the observed refraction in a given optical model, it is important to distinguish this entity from the postoperative ACD. The postoperative ACD is a measurable quantity, representing the distance between the anterior corneal surface and the anterior IOL surface along the visual axis in the postoperative eye. The relationship between ELP for a given optical model and postoperative ACD depends on the optical model itself, meaning that while postoperative ACD is a measurable quantity, ELP is only a computable quantity.

As Kriechbaum et al. pointed out in 2003, exact postoperative ACD prediction based on preoperative biometry data is, in principle, impossible because of the effect of several uncertain parameters including the shrinkage of the capsular bag.[107] There have, however, been reports of various preoperative features that may be predictive of postoperative ACD. For example, Plat in 2017 reported correlation between measurements of axial length (AL), horizontal white to white distance (WTW), and preoperative ACD with postoperative ACD.[108] Other approaches have added corneal power to improve postoperative ACD prediction.[26]

Methods utilizing measures from anterior segment optical coherence tomography (AS-OCT) have achieved high accuracy in postoperative ACD prediction.[109,110] However, these

approaches rely on angle-to-angle measurements that are not typically obtained in a standard cataract surgery preoperative workup. Furthermore, these measurements involve manual caliper-based measurements, introducing subjectivity and variability into the measurements while slowing the workflow of the cataract surgeon.

An ideal method for postoperative ACD estimation would utilize only data obtained from optical biometry. It would achieve high accuracy yet have minimal loss of accuracy in the absence of reliable keratometry data. Such a method would be able to integrate into existing workflows. It could also be used for patients who had previously undergone refractive surgery, and would lend itself well to integration into existing and novel methods for IOL power calculation.

Since it is not particularly common to obtain biometry both preoperatively and postoperatively, building a dataset large enough to accurately predict postoperative ACD can be a challenge. In this chapter, leveraging the Sight Outcomes Research Collaborative (SOURCE) repository, we describe the creation of a dataset including over 800 patients with both preoperative and postoperative biometry. Furthermore, we present here the development and testing of a machine learning approach to postoperative ACD prediction.

2.3 Materials and Methods

2.3.1 Data collection

Biometry records (including preoperative and postoperative biometry) between August 25, 2015 and June 27, 2019 were retrieved from Lenstar LS900 optical biometers (Haag-Streit USA Inc, EyeSuite software version i9.1.0.0) at University of Michigan's Kellogg Eye Center. Institutional review board approval was obtained for the study and it was determined that informed consent was not required because of its retrospective nature and the anonymized data

utilized in this study. The study was carried out in accordance with the tenets of the Declaration of Helsinki. Patient demographics and cataract surgery information (including date of surgery and implanted IOL power) were obtained via the Sight Outcomes Research Collaborative (SOURCE) Ophthalmology Data Repository, which captures electronic health record (EHR) data of all patients receiving any eye care at academic medical centers participating in this research collaborative. SOURCE captures information on patient demographics, diagnoses identified based on International Classification of Diseases (ICD) 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.). For this study, we focused on a subset of the SOURCE patients receiving care at the University of Michigan. Spherical equivalent manifest refractions from the postoperative month one visit were identified from the clinical record for all patients who underwent cataract surgery (CPT = 66984 or 66982) from the dataset. The power and model of the implanted intraocular lens for each surgery was collected as well. Only those surgeries involving the implantation of an Alcon SN60WF single-piece acrylic monofocal lens (Alcon, USA) were included in the study. Patients who had prior refractive surgeries were excluded from the dataset. Patients who had an additional surgery (e.g., endothelial keratoplasty) at the time of their cataract surgery were also excluded. Postoperative biometry records with outliers in IOL thickness were excluded to address the possibility of lens tilt affecting the assessment of ACD. The outliers were defined as records where the postoperative lens thickness fell greater than 1.96 standard deviations away from the mean thickness for a given IOL power.

2.3.2 Model development

After data collection, the raw data was reformatted so that each sample in the dataset consisted of a set of predictors and a target value that could be utilized by the machine learning model. Among the biometry records, it was possible for individual eyes to have multiple preoperative and postoperative sets of biometry measurements. In order to take advantage of these records, preoperative and postoperative biometry records of the same eye were matched in a way that accounted for all possible combinations. An eye with x preoperative records and y postoperative records had xy possible combinations. The inclusion of all possible preoperative and postoperative biometry record combinations represents a form of data augmentation, with the intention of increasing robustness to measurement variations while recognizing that the same eye can have varying lens thickness and preoperative anterior chamber depth due to natural cataract progression. At the end of data preprocessing (**Figure *2.1***, middle panel), a dataset of 4137 samples that involved 847 distinct patients was generated and used for the development of the machine learning model. Each sample consisted of (1) preoperative biometry: axial length (AL), central corneal thickness (CCT), anterior chamber depth (ACD), crystalline lens thickness (LT), flat keratometry K1, steep keratometry K2, $Km = \frac{K1+K2}{2}$, and horizontal white-to-white (WTW), (2) patient sex, (3) IOL power, and (4) postoperative ACD, where (1)-(3) were the predictors and (4) was the target variable in the machine learning model.

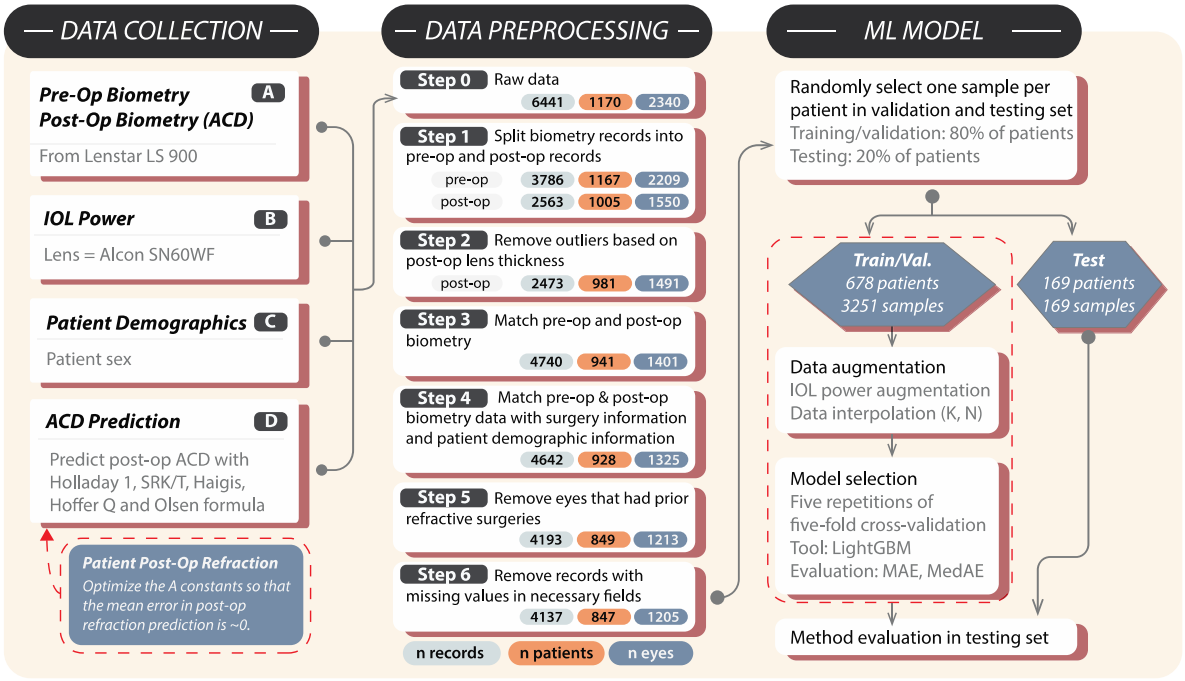

Figure 2.1 Method pipeline.

In the middle panel, "n records" refers to the total number of records or samples in the whole dataset; "n patients" refers to the total number of distinct patients; "n eyes" refers to the total number of distinct eyes. In the right panel, 3251 in the training/validation set is the total number of samples before selecting one sample per patient in the validation set.

Corneal power is one of the most important features in both postoperative ACD prediction and postoperative refraction prediction in cataract surgery. However, corneal power measurement is unreliable in patients with prior corneal refractive surgery. In order to evaluate applicability of our method to patients with prior corneal refractive surgery, we examined how well our method performed when corneal power was not available.

We also studied the effect of IOL power in postoperative ACD prediction, because even though IOL power is directly associated with IOL thickness, which could in turn affect postoperative ACD, IOL power, to our knowledge, has not been considered in existing formulas.

In summary, we examined the performance of three classes of models where different subsets of variables were used as predictors: (1) Base, which used AL, CCT, ACD, LT, K1, K2, Km, WTW, and patient sex as predictors, (2) Base + IOL, which added IOL power to "base" as an additional feature, and (3) Base – K, which removed K1, K2, and Km from "Base", using AL, CCT, ACD, LT, WTW, and patient sex as predictors.

LightGBM (2.2.3), which is a widely used framework for implementing the gradient boosted decision tree algorithm, was used to construct the machine learning model. During the training process, the training data were augmented through two methods (**Figure *2.1***, right panel) (1) IOL power augmentation, and (2) data interpolation. The purpose of using IOL power augmentation was to improve the prediction performance by incorporating the relationship between IOL power and IOL thickness into the training data. During IOL power augmentation, the implanted IOL power (IOL_{old}) was replaced by n_{IOL} randomly selected IOL powers, and the ground truth postoperative ACD was adjusted based on the selected IOL powers. Specifically, for each distinct patient, n_{IOL} synthetic IOL powers ($IOL_{new,1}, IOL_{new,2}$, ...) between $[IOL_{min}, IOL_{max}]$ were selected and the adjusted (new) postoperative ACD corresponding to each new IOL power was calculated as:

$$ACD_{new} = ACD_{old} - m(IOL_{new} - IOL_{old})$$

where $m \in [0,1]$ is a constant, $IOL_{min} \geq 6, IOL_{max} \leq 30$. The value of IOL_{min}, IOL_{max}, m, and n_{IOL} were optimized through cross-validation. In data interpolation, k samples were randomly picked, and the center of those k samples was calculated by averaging each dimension of the predictor vector X and the target value y. Categorical variables were treated as continuous variables. The number of samples, k, used to create each synthetic sample and the number of samples generated, n, were optimized through cross-validation.

2.3.3 Model evaluation

Five repetitions of five-fold cross-validation were used to perform a grid-search for the parameters in data augmentation (IOL_{min}, IOL_{max}, m, n_{IOL}, n, and k) and the hyperparameters in the machine learning model (the learning rate, number of estimators, maximum tree depth, and number of leaves). Cross-validation was also used to evaluate the performance of different subsets of features. Mean absolute error (MAE) in postoperative ACD prediction was used as the primary evaluation metric in cross-validation. The optimal models for three scenarios: (1) Base (2) Base + IOL (3) Base $-$ K were selected based on the mean of the MAEs in the cross-validation results.

We then tested the performance of our model on a hold-out testing dataset and compared the performance of our methods with five existing formulas: Haigis, Hoffer Q, Holladay I, Olsen, and SRK/T. These five existing formulas were implemented in Python 3 based on their publications.[52–56,58,59] The lens constants were optimized for each formula to eliminate systematic errors in refraction prediction using previously described methods.[58,90,97] The optimized constants were: 1.655 for Haigis, 5.844 for Hoffer Q, 1.990 for Holladay I, -0.225 for Olsen, and 119.303 for SRK/T. The corresponding mean errors in refraction are listed in **Table S *2.1***. We further compared our methods to two baseline prediction methods: (1) average postoperative ACD, which used the average postoperative ACD in the training/validation dataset as the predicted ACD for the testing set and (2) linear regression, which used AL, CCT, ACD, LT, K1, K2, Km and WTW as predictors. Data augmentation (i.e., interpolation and IOL augmentation) was not applied to the linear regression model.

Apart from the testing set described above, we also gathered a separate testing set consisting of 78 cataract patients (78 eyes) who had a history of prior refractive surgery. We

tested the Base-K model on this separate post-refractive testing set in order to compare the prediction performance of our method for patients with and without prior refractive surgery.

To investigate the degree to which dataset size affected prediction performance, the performance of our method and linear regression were compared as random subsets of the training dataset of varying size (20%, 40%, 60%, 80%, and 100%) were utilized. The subsampling of the training dataset was applied before data interpolation to better simulate the reduction in available raw data.

During the testing and validation process, one testing/validation sample was randomly selected for each patient to ensure that performance evaluation was not biased due to the varying number of records available per patient. In addition to the MAE, the median absolute error and Pearson correlation coefficient (r) were also calculated for the performance comparison in the testing set. To gain insights into the relative importance of predictors in the machine learning model, we calculated the total gain (total reduction in training loss) across splits in decision trees for each predictor in the model.

2.3.4 Statistical analysis

Statistical testing was performed to investigate relationships between variables in the dataset. A chi-square test was performed to evaluate the difference in the proportion of males and females among all patients. A two-tailed Student t-test was performed to evaluate for differences in the means of biometry values between males and females. The Pearson correlation coefficients and the p-values testing the significance of correlation were calculated between the postoperative ACD and the preoperative biometry measurements. To assess the difference in cross-validation results of different methods, a Wilcoxon sign-rank test was performed. The testing set results of different methods were compared based on the Friedman test followed by a post-hoc paired

Wilcoxon signed-rank test with Bonferroni correction. Performance of the keratometry-independent Base – K model was compared between the testing sets of patients with and without prior refractive surgery through an unpaired two-sample Wilcoxon rank-sum test (i.e., the Mann-Whitney U test). Statistical significance for all above tests was defined as p-value < 0.05. All statistical analysis and machine learning model construction scripts were written in Python 3.

2.4 Results

2.4.1 Data characteristics

In total, our dataset included the preoperative and postoperative biometry measurements and surgical records of 1205 eyes from 847 patients (**Figure *2.1***). These patients were split into training/validation and testing sets. The distributions of the biometry measurements were similar in the two sets (**Table *2.1***). There were significantly more females than males (chi-square test, $p < 0.01$). The postoperative anterior chamber depth (ACD) was positively correlated with preoperative AL, ACD, WTW, and CCT ($p < 0.01$ for each), and negatively correlated with preoperative LT and WTW ($p < 0.01$ for each). Postoperative ACD was not significantly correlated with preoperative Km ($p = 0.74$) (

Figure *2.2*A).

Figure *2.2*B shows the distribution of the power of the implanted IOL and the postoperative lens thickness ($r = 0.75$, $p < 0.01$). The scatter plot indicates a linear relationship between the IOL power and postoperative IOL thickness. The distributions of biometry measurements in male and female patients are shown in

Figure *2.2*C. The preoperative AL, preoperative ACD, and postoperative ACD in male patients were longer than those in female patients ($p < 0.01$ for each). Km in females was greater than that in males ($p < 0.01$).

Characteristic	Training/Validation Set (mean ± SD)	Testing Set (mean ± SD)
Gender	male: 283 (41.7%), female: 395 (58.3%)	male: 65 (38.5%), female: 104 (61.5%)
Age at surgery (years)	71.08 ± 10.50	71.02 ± 8.96
Preoperative Km (D)	43.78 ± 1.65	43.94 ± 1.70
Preoperative AL (mm)	23.98 ± 1.09	23.79 ± 1.09
Preoperative LT (mm)	4.52 ± 0.45	4.53 ± 0.44
Preoperative ACD (mm)	3.25 ± 0.41	3.24 ± 0.42
Postoperative ACD (mm)	4.66 ± 0.30	4.64 ± 0.30

Table 2.1 Patient demographics.

Since patients in our dataset had varied numbers of biometry exam records, we randomly selected one record/patient from the training/validation set and the testing set to calculate the summary statistics.

2.4.2 Model performance

As stated above, different subsets of features were tested to examine the performance of our machine learning model when (1) corneal power was not available and (2) IOL power was considered. **Figure S2** shows the cross-validation results of the alternative models with optimized parameters. The cross-validation results (i.e., the average MAE) of each alternative model were as follows: 0.121 mm for Base = biometry + patient sex, 0.120 mm for Base + IOL, 0.131 mm for Base − K. The addition of IOL power improved prediction performance, while IOL-based augmentation, which simulated the linear relationship between an IOL's power and its thickness further improved prediction performance, beyond the addition of IOL power alone. Base – K performed significantly worse compared to Base and Base + IOL ($p < 0.01$), as expected. For comparison purposes, we recalculated the cross-validation results using median

absolute error as the evaluation metric. The results were as follows: 0.100 mm for Base, 0.097 mm for Base + IOL, and 0.108 mm for Base – K. The prediction performance was consistent with the results obtained with MAE.

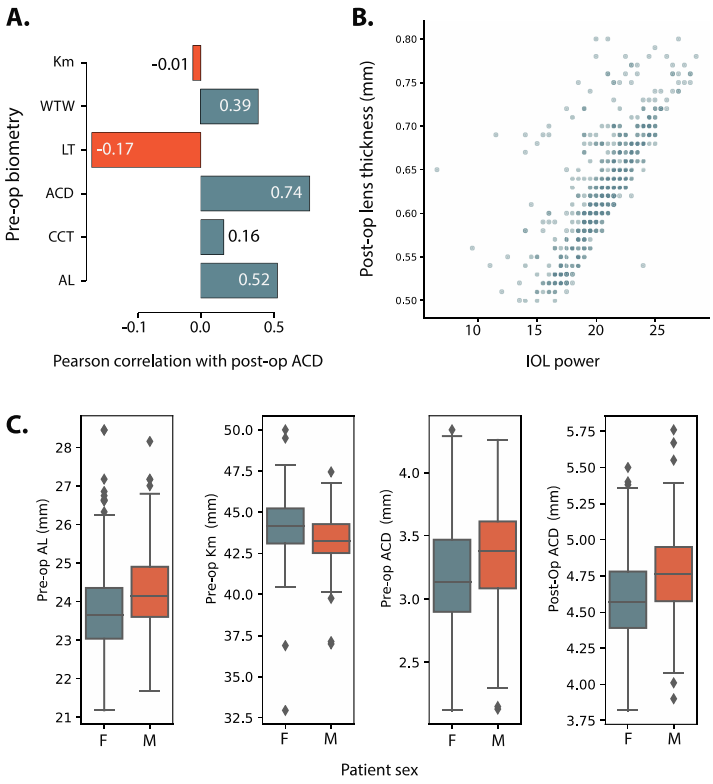

Figure 2.2 Baseline dataset characteristics.

A) Bar graph plotting the Pearson correlation coefficient r between postoperative anterior chamber depth and preoperative biometry in the training/validation dataset. **B)** Scatter plot of IOL power against the postoperative lens thickness. The dots are 50% transparent. **C)** The distribution of preoperative axial length, corneal power, anterior chamber depth and postoperative anterior chamber depth in male (M) and female (F) patients. One record per patient in the training/validation set was randomly selected to generate the figures (i.e., the same set of records as the "Training/Validation Set" column in **Table 2.1**).

The performance of the three models on the unseen testing dataset is presented in **Table *2.2* and Table *2.3***. The Friedman test for difference in MAE among the methods in **Table *2.2*** was significant ($p < 0.01$). The Base predictors, which included preoperative biometry and

patient sex, achieved an MAE of 0.106 mm. Adding the IOL improved the prediction performance in the test set (MAE = 0.105 mm). Base and Base + IOL significantly outperformed Haigis, Hoffer Q, Holladay I, Olsen, SRK/T, and mean postoperative ACD, based on the post-hoc Wilcoxon signed rank test with Bonferroni correction ($p < 0.01$). When the corneal power was not included (Base − K), which simulates the scenario when the measured corneal power is not reliable, our method maintained good performance, with an MAE = 0.123 mm. The performance of Base − K still significantly outperformed the existing 5 formulas ($p < 0.01$). When we tested the Base − K model on patients with prior refractive surgery (**Table *2.3***), the MAE was 0.129 mm, and this result was not significantly different compared with the performance of Base − K for patients with no history of refractive surgery ($p = 0.13$).

Index	**Method**	**MAE in mm ± SD**	**MedAE in mm (interquartile range)**	R^2
1	**Base = biometry + patient sex**	**0.106 ± 0.098**	**0.082 (0.119)**	**0.777**
2	**Base + IOL**	**0.105 ± 0.091**	**0.080 (0.114)**	**0.781**
3	Haigis	0.680 ± 0.172	0.681 (0.206)	0.681
4	Hoffer Q	1.228 ± 0.251	1.219 (0.318)	0.407
5	Holladay I	0.743 ± 0.283	0.744 (0.403)	0.405
6	Olsen	1.200 ± 0.172	1.199 (0.206)	0.681
7	SRK/T	1.205 ± 0.328	1.183 (0.256)	0.317
8	Average postoperative ACD	0.231 ± 0.195	0.192 (0.000)	/
9	Linear regression	0.116 ± 0.099	0.089 (0.120)	0.746

Table 2.2 Prediction performance on the testing set.

MAE = Mean Absolute Error. MedAE = Median Absolute Error.

Index	Method	Number of Patients	MAE in mm ± SD	MedAE in mm (interquartile range)	R^2
1	Base – K without prior refractive surgery	169	0.123 ± 0.109	0.093 (0.124)	0.711
2	Base – K with prior refractive surgery	78	0.129 ± 0.096	0.110 (0.145)	0.743

Table 2.3 Prediction performance on the testing set without using the corneal power as an input.

The testing set used for "Base – K without prior refractive surgery" was the same as the that in **Table 2.2**. And a separate testing set was used for "Base – K with prior refractive surgery" (see details in Methods). MAE = Mean Absolute Error. MedAE = Median Absolute Error.

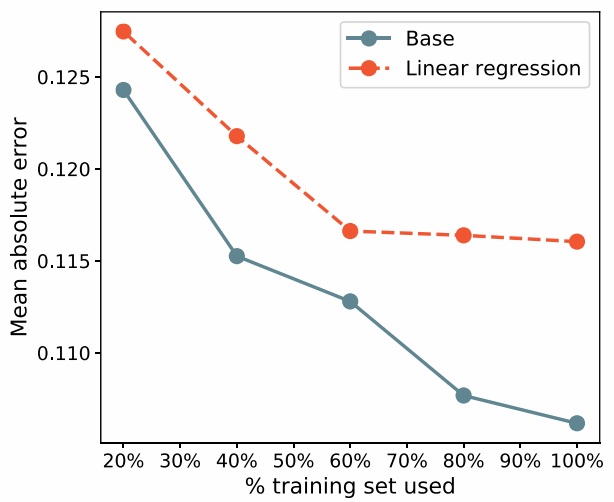

Figure 2.3 Testing set MAE (in mm) of postoperative ACD prediction of the linear regression method and our Base method (Base = biometry + patient sex).

MAE = mean absolute error, ACD = anterior chamber depth.

The performance of our methods and the linear regression method on training datasets of varying size is shown in **Figure *2.3*.** The result demonstrates that the performance attained by our Base method on the testing set continued to improve as the dataset grew to 100% of the available data. On the contrary, the improvement of the linear regression method plateaued at around 60% of the available data. The above result indicates that the large set of paired

preoperative and postoperative biometry provided a significant benefit to our machine learning model and that its performance may continue to improve as more data becomes available.

Feature importance in the Base model and Base model with IOL power is shown in **Figure *2.4***. ACD, LT, and WTW ranked highly in both models. IOL power achieved a high importance score when it was added into the model (**Figure *2.4*B**).

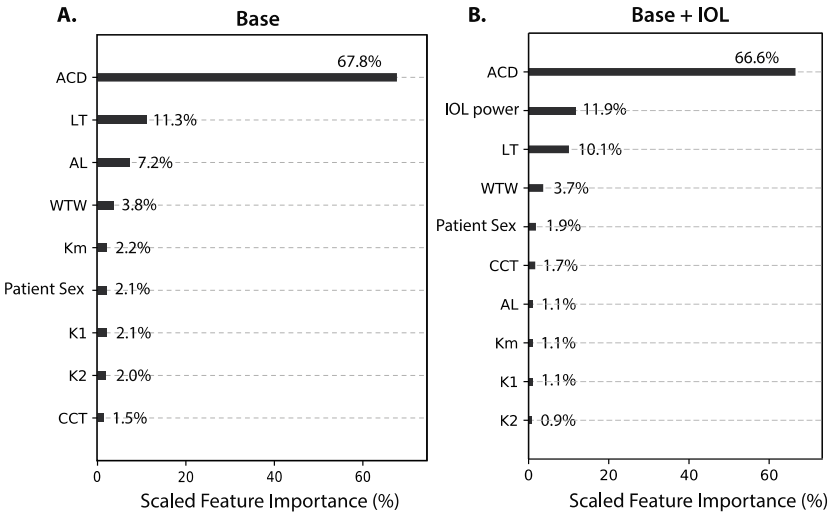

Figure 2.4 Feature importance in the Base model and Base + IOL model, measured by the percentage of total gain.

2.5 Discussion

We have presented here a machine learning approach to predicting postoperative ACD in cataract surgery using standard preoperative optical biometry measurements.

In order to develop this method, we built, to our knowledge, the largest optical biometry dataset involving both preoperative and postoperative measurements reported to date. We found, through sampling of subsets of this dataset of varying size, that the performance of our machine learning method for predicting postoperative ACD improved substantially as the dataset size

grew, while the performance of linear regression plateaued at 60% of the available data. This finding underscores the potential of our machine learning method to continue to improve as the dataset, derived from the SOURCE data repository, continues to grow.

We found that even a linear regression approach to modeling postoperative ACD achieved performance better than that of previously reported methods (including Haigis, Hoffer Q, Holladay I, Olsen, and SRK/T) both in terms of the mean absolute error and the Pearson correlation (r). The performance of linear regression on our biometry dataset also exceeded that of previously reported AS-OCT methods by R^2 value.[109] Since lens constants were optimized individually prior to computing the predictions of each of the aforementioned formulas, the high performance of linear regression relative to existing methods was likely due to the size of the dataset of available, as well as the use of optical biometry to directly measure postoperative IOL position, as opposed to ultrasound biometry or ELP calculations. The existing formulas considered here use a thin lens assumption, wherein the intermediate value referred to as the ACD does not represent the position of either surface of the IOL, but rather the location of the principal plane.[26] Therefore, the estimated ACD terms in these formulas can more accurately be described as providing information about the ELP within the optical models employed by those IOL power calculation formulas. As such, they are not ideal for prediction of the true postoperative anatomy of the eye of a cataract surgery patient.

By employing a gradient boosting machine learning algorithm, we were able to significantly improve ACD prediction performance beyond that of linear regression and existing ACD prediction formulas. Our method also outperformed methods employing more involved measurement techniques such as AS-OCT.[109,110] Evaluation of feature importance demonstrated that preoperative ACD was the most important input parameter, followed by

crystalline lens thickness (LT), axial length (AL), and horizontal white to white (WTW), respectively. Inclusion of patient sex, which is not typically utilized in methods of postoperative ACD prediction, in the model was found to improve performance (**Figure S *2.2***). This finding was consistent with prior studies of patient biometry reporting consistent differences in ocular shape between male and female patients, with female corneal powers measuring greater and axial lengths measuring shorter than those of males on average.[111,112]

In order to enhance performance of our gradient boosting machine learning approach, we included biometry measurements for the same eye taken at multiple preoperative and postoperative time points. The variations in the biometry measurements among different records reflect the margin of error of optical biometry measurements as well as the natural increase in lens thickness and concomitant reduction in preoperative anterior chamber depth over time.

One of the data augmentation methods employed involved modeling IOL thickness based on IOL power to account for potential variations in postoperative ACD due to the thickness of the IOL utilized. This augmentation method utilized the linear relationship between the IOL power and the IOL thickness shown in

Figure *2.2*B. Variations in lens thickness measurements for a given IOL power are also depicted in

Figure *2.2*B, and could be caused by: (1) tilt and decentration of the IOL, (2) variations in the production of the lens, and (3) trace residual or proliferative lens material, which could affect the detection of the surfaces of the IOL. Despite these variations, the data augmentation method presented here resulted in improvements in cross-validation performance (**Figure S *2.2***). The performance enhancement seen with the IOL thickness adjustment method indicates that IOL thickness is indeed relevant to postoperative ACD. It further indicates that customized IOL

thickness modeling should be included in methods for postoperative ACD prediction depending on the model of IOL intended for use in the patient's preoperative plan.

Due to challenges in accurately assessing corneal power in post-refractive surgery patients, we investigated a keratometry-independent (K-independent) approach to the prediction of postoperative ACD as well. Our method outperformed a previously-reported method for K-independent prediction of ACD,[113] and may be applicable in new methods for IOL power calculation in patients with prior refractive surgery. Performance of this K-independent model was not significantly different for patients with or without prior refractive surgery, indicating its applicability to the post-refractive surgery population.

The limitations of our study include the use of a retrospective, rather than prospective, dataset. It was not possible to compare our method for ACD prediction to those of the Barrett Universal II or Holladay 2 formulas, as the ACD predictions of these formulas are not publicly available. An additional limitation of our study is that while a hold-out testing set was used, it was comprised of data obtained at the same institution. Building a separate testing set external to our institution would provide additional validation of our method, and will be a future direction of this work. The presented models were developed using the Alcon SN60WF lens due to the high frequency of its use at our institution, however, the extension of our approach to other IOLs will be a future direction of this work through the expansion of the SOURCE data repository.

The method presented here for the prediction of postoperative ACD has the potential to be integrated into methods for IOL power calculation. Both geometrical optics and ray-tracing methods for IOL power calculation rely on some form of prediction of postoperative IOL position, and could benefit from the accuracy of the approach presented here. Since feature engineering is an important part of optimizing machine learning methods, and postoperative

ACD is known to be a useful predictor in traditional methods of IOL power calculation, it is possible that postoperative ACD may be a useful feature for machine learning approaches to IOL power calculation as well.

In summary, the machine learning method presented here for predicting postoperative ACD in cataract surgery has the potential for integration into novel methods for IOL power calculation, both in standard and post-refractive surgery cases.

2.6 Publication and Acknowledgement

This chapter is a published work [114]: Li, T., Yang, K., Stein, J. D., & Nallasamy, N. (2020). Gradient boosting decision tree algorithm for the prediction of postoperative intraocular lens position in cataract surgery. *Translational vision science & technology*, *9*(13), 38-38.

2.7 Supplementary Materials

2.7.1 Supplementary figures

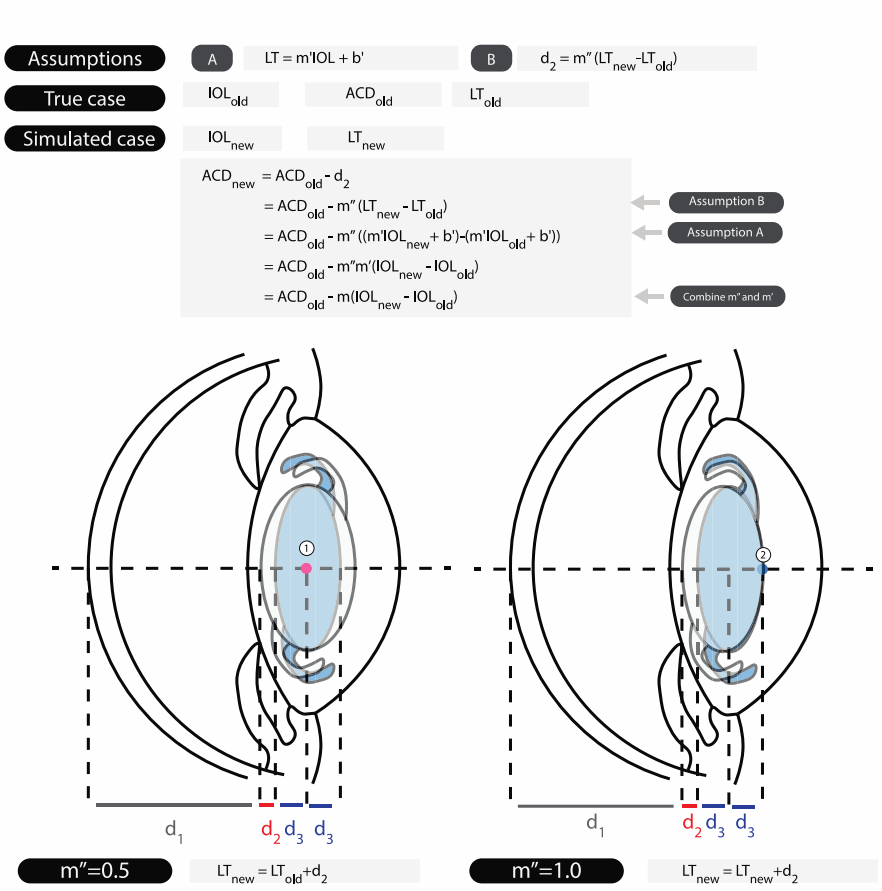

Figure S 2.1 Schematic of the IOL augmentation method.

IOL = the IOL power. LT = the thickness of the intraocular lens. ACD = the postoperative anterior chamber depth. The upper half of the figure elaborates the equation's derivation $ACD_{new} = ACD_{old} - m(IOL_{new} - IOL_{old})$ which we used to augment the IOL power and the postoperative ACD. During the augmentation, simulated cases were created based on real cases in the dataset. The IOL powers for the simulated cases were randomly selected. The equation assumes (1) a simple linear relationship between the IOL power and the thickness of the intraocular lens: $LT_{new} - LT_{old} = m'(IOL_{new} - IOL_{old})$ (LT_{new} and LT_{old} are the thickness of the IOL), and (2) an imaginary anchor point on the visual axis within the IOL that has a fixed location independent of the IOL power. The first assumption bridges the IOL power and the thickness of the IOL. The second assumption is necessary because it provides way to map the change in intraocular lens thickness to the change in the postoperative ACD. The diagrams in the lower half of the figure depict two physiologically plausible assumptions about the location of the IOL: (1) when the anchor point is at $0.5 * LT_{IOL}$ (LT_{IOL} is the thickness of the IOL) and (2) when the anchor point is at $1.0 * LT_{IOL}$. The blue lens stands for the real implanted lens, and the white lens stands for the imaginary lens in a simulated case. Instead of guessing a value for the location of the anchor point, the anchor point location was represented by $m'' * LT_{IOL}$ ($m'' \in [0,1]$) and m'' was combined with m' into one single variable: $m = m'm''$. For a given constant m, we could calculate the corresponding postoperative ACD and evaluate the performance of our method using the augmented data. The m that gave the best performance in cross-validation was chosen.

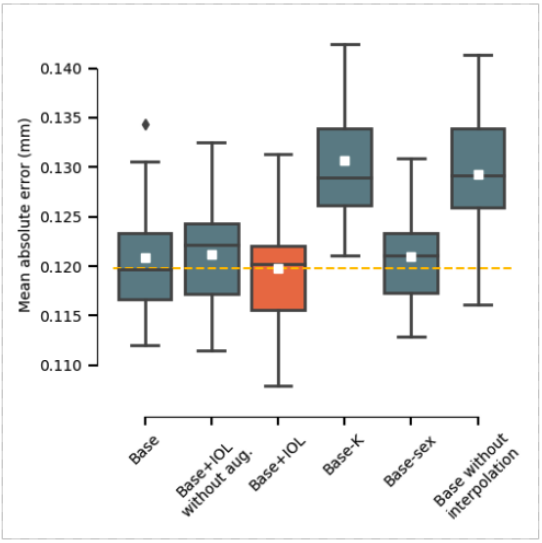

Figure S 2.2 Boxplot of the cross-validation results of six alternative methods.

The means (±SD) of MAE for the methods were: 0.1209 ± 0.00593 for Base, 0.1212 ± 0.00539 for Base + IOL without augmentation, 0.1197 ± 0.00533 for Base + IOL, 0.1306 ± 0.00606 for Base − K, 0.1210 ± 0.00531 for Base – sex, and 0.1293 ± 0.00665 for Base without data interpolation. The mean is represented by a white square inside the boxplot. The yellow dashed line marks the lowest mean (0.1197) among four models. Base + IOL used IOL augmentation and Base + IOL without augmentation did not use IOL augmentation.

2.7.2 Supplementary tables

Formula	**Constant**	**Optimized Constant Value**	**Signed Average Numerical Error in Refraction**
Holladay1	Surgeon factor	1.990	-0.000509
SRK/T	A constant	119.303	-0.000509
Hoffer Q	ACD	5.844	-0.000499
Haigis	a_0	1.655	-0.000405
Olsen	ACDconst	5.155	/

Table S 2.1 Optimized constants and mean error of existing formulas.

The constants for Holladay 1, SRK/T, Hoffer Q and Haigis were selected so that the absolute value of the mean error was minimized. The constants were searched in a space around their default value with a step of 0.001. The constant for the Olsen formula was calculated based on the publication.[58]

Chapter 3 Improve the Accuracy of Existing Lens Formula with Machine Learning-Predicted Lens Position

3.1 Abstract

The aim of this study was to assess whether incorporating a machine learning (ML) method for accurate prediction of postoperative anterior chamber depth (ACD) improves the refraction prediction performance of existing intraocular lens (IOL) calculation formulas. A dataset of 4806 cataract patients were gathered at the Kellogg Eye Center, University of Michigan, and split into a training set (80% of patients, 5761 eyes) and a testing set (20% of patients, 961 eyes). A previously developed ML-based method was used to predict the postoperative ACD based on preoperative biometry. This ML-based postoperative ACD was integrated into new effective lens position (ELP) predictions using regression models to rescale the ML output for each of four existing formulas (Haigis, Hoffer Q, Holladay, and SRK/T). The performance of the formulas with ML-modified ELP was compared using a testing dataset. Performance was measured by the mean absolute error (MAE) in refraction prediction. When the ELP was replaced with a linear combination of the original ELP and the ML-predicted ELP, the MAEs ± SD (in Diopters) in the testing set were: 0.356 ± 0.329 for Haigis, 0.352 ± 0.319 for Hoffer Q, 0.371 ± 0.336 for Holladay, and 0.361 ± 0.331 for SRK/T which were significantly lower (p-value < 0.05) than those of the original formulas: 0.373 ± 0.328 for Haigis, 0.408 ± 0.337 for Hoffer Q, 0.384 ± 0.341 for Holladay, and 0.394 ± 0.351 for SRK/T. In conclusion, using a more accurately predicted postoperative ACD significantly improves the prediction accuracy of four existing IOL power formulas.

3.2 Introduction

As demonstrated in previous chapters, the estimation of postoperative intraocular lens (IOL) position is essential to IOL power calculations for cataract surgery. Methods to improve the accuracy of the prediction of postoperative ACD have been studied for decades. Existing theoretical formulas use an intermediate variable named effective lens position (ELP) to represent the predicted lens position. However, based on our results shown in Chapter 2, the ELPs predicted by the existing formulas are not exact estimates of the physical distance between the cornea and the IOL. Our ML-predicted postoperative ACD was a more accurate representation of the physical location of the lens. In view of the limitations of the ELP in existing formulas, recently, more efforts have been devoted to constructing ELPs that better reflect the true location of the IOL.[110,115–118] New IOL power prediction methods have also been developed based on the new-generation ELP prediction methods, and they have shown that using a more accurately predicted IOL position helps to improve the IOL power prediction accuracy.[115]

It is so far largely unexplored whether inserting a more accurately predicted ELP into existing formulas improves refraction prediction accuracy. This is an important question because: (1) it provides a fast and efficient way to modify and improve on existing IOL formulas whose reliability has been tested extensively. (2) such research can provide supports for translating the continued improvements in accuracy in postoperative ACD prediction into better refraction predictions in published formulas. Several previous studies had modified the ELPs in existing formulas in order to achieve better refraction prediction results in certain cataract cases. Modification of ELP calculation in the Haigis formula for sulcus-implanted IOLs was reported to improve performance.[119] Kim et al. adjusted the ELP estimation in SRK/T formulas with the

corneal height in post-refractive patients and achieved satisfactory accuracy.[120] It remains to be explored whether improvement of ELP estimates for in-the-bag IOL placement can improve IOL power calculations of existing formulas for general cataract patients.

Since most recently published IOL formulas (e.g., Barrett Universal II [121,122], Holladay 2, Olsen formula [123]) are either not disclosed to the public or do not have the option to customize the value of ELP during the prediction of postoperative refraction, here we applied our postoperative ACD prediction methods described in Chapter 2 to a dataset of 4806 cataract surgery patients and replaced the ELP estimates in 4 existing IOL formulas: Haigis, Hoffer Q, Holladay, and SRK/T. We combined our machine learning (ML) prediction of true postoperative ACD with the original ELP estimated by each formula and substituted this updated ELP prediction for each formula. We then compared the refraction prediction performance of each formula using its original and enhanced ELP estimates. The findings reported here demonstrate that existing formulas can benefit from improved methods for predicting true postoperative ACD.

3.3 Materials and Methods

3.3.1 Postoperative ACD prediction machine learning model

In previous work,[124] we developed a machine learning-based postoperative anterior chamber depth (ACD) prediction model, which predicts the postoperative anterior chamber depth (in mm) based on preoperative biometry. Here in the presented study, an ACD prediction machine learning model was trained using the method and dataset (847 patients, 4137 eyes) described in the previous research. The dataset was composed of the preoperative and postoperative biometry measured by the Lenstar LS900 optical biometers (Haag-Streit USA Inc, EyeSuite software version i9.1.0.0) at the University of Michigan's Kellogg Eye Center. The

postoperative ACD was defined as the distance from the front surface of the cornea to the front surface of the intraocular lens (IOL). The postoperative ACD predicted by the machine learning model is referred to as ELP_{ML} in this chapter.

3.3.2 Data collection

In this study, biometry records were collected using the same approach as for the development of the ML postoperative ACD prediction model at University of Michigan's Kellogg Eye Center.[124] Institutional review board approval was obtained for the study, and it was determined that informed consent was not required because of its retrospective nature and the anonymized data used in this study. The study was carried out in accordance with the tenets of the Declaration of Helsinki.

The inclusion criteria were: (1) patients who had cataract surgery (Current Procedural Terminology [CPT] code = 66984 or 66982) but no prior refractive surgery and no additional surgical procedures at the time of cataract surgery. (2) the implanted lens was an Alcon SN60WF single-piece acrylic monofocal lens (Alcon, USA). Each case in the dataset corresponds to one operation of a single eye with preoperative and postoperative information. The preoperative information includes the measurements of the axial length (AL), lens thickness (LT), anterior chamber depth (ACD), flat keratometry (K1), steep keratometry (K2), and the average keratometry which was calculated as $K = \frac{K1+K2}{2}$. The postoperative information includes the postoperative refraction (spherical component SC and cylindrical component CC) where the time when it was recorded was closest to one month (30 days) after surgery. Since the patients were measured in a lane of 10 feet long (3.048 meters), which was shorter than the standard length of 20 feet (6 meters), the SC was adjusted for the vergence distance by adding $\frac{1}{6} -$

$\frac{1}{test\ distance\ in\ meters} = \frac{1}{6} - \frac{1}{3.048} = -0.1614$ according to Simpson and Charman's recommendation.[125] The spherical equivalent (SE) refraction was therefore calculated as $SE\ refraction = (SC - 0.1614) + 0.5CC$. Samples that were used to train the postoperative ACD prediction machine learning model were excluded from the dataset so that the dataset better simulates unseen samples.

The dataset in total consisted of 4806 patients (**Figure *3.1***). The dataset was split into a training dataset used for the development of the methods and a testing dataset used for performance comparison. 80% of the patients were randomly assigned to the training set, and the rest of the patients (20%) were assigned to the testing set. For patients who had more than one associated case in the testing set (i.e., patients who had both eyes operated on), one case was randomly selected to ensure each patient had the same weight when the prediction performance was evaluated. At the end of this process, the training set had 3845 patients (5761 eyes), and the testing set had 961 patients (961 eyes).

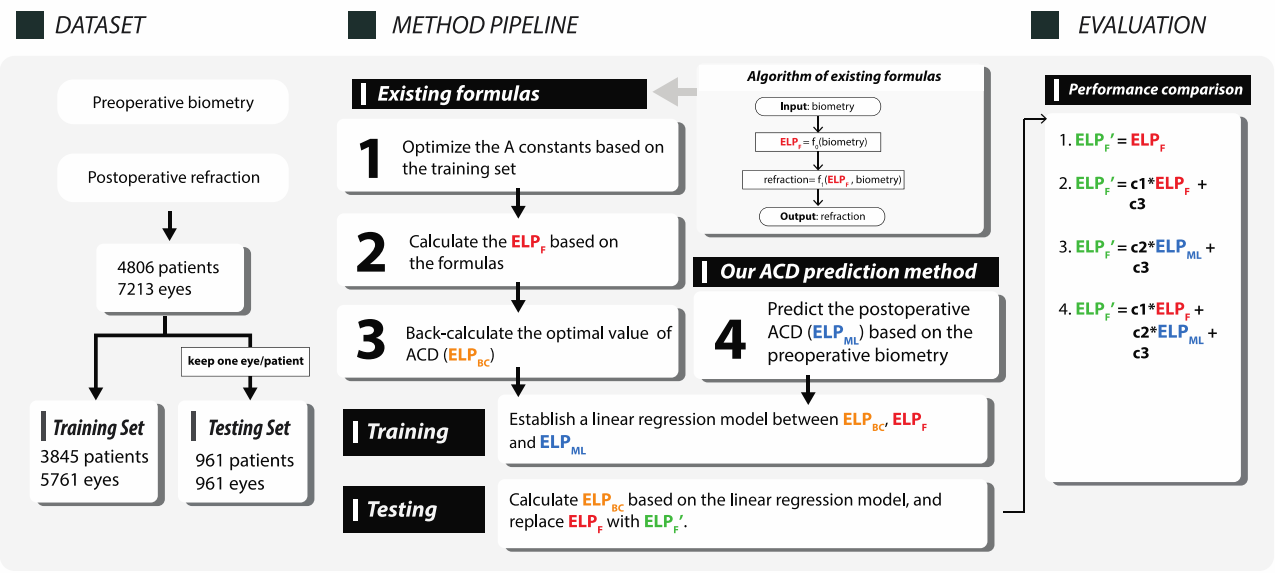

Figure 3.1 The analysis pipeline of the presented study.

ELP_F = the effective lens position (ELP) estimated by the existing formulas. ELP_{ML}= the postoperative anterior chamber depth (ACD) predicted by the machine learning method. ELP_{BC} = the back-calculated ELP (see main text). ELP'_F is a term that refers to a new ELP that is used to replace the ELP_F in the existing formulas.

3.3.3 Linear regression model

We implemented four existing formulas (Haigis, Hoffer Q, Holladay, and SRK/T) in Python based on their publications.[52–55,59,126–128] The existing formulas calculated the effective lens position (ELP_F) as a function of the preoperative biometry (**Figure *3.1***): $ELP_F = f_0(biometry)$. The predicted ELP (ELP_F) was then used to predict the postoperative refraction: $refraction = f_1(ELP_F, biometry)$. Here, the goal was to reduce the refraction prediction error by replacing ELP_F with a different value, ELP'_F. Our approach involves two steps: (1) finding the theoretically most optimal ELP values, (2) modeling the most optimal ELP with ELP_F and the ML-predicted postoperative ACD, denoted ELP_{ML}.

In the first step, the most optimal ELP (denoted ELP_{BC}) was found by the standard method of back-calculating the ELP when the predicted refraction was set to equal the true refraction (i.e., $f_1(ELP_{BC}, biometry) = true\ refraction$). In other words, when $ELP'_F = ELP_{BC}$, the refraction prediction errors of all patients equal zero. More details on the computation of ELP_{BC} can be found in **Supplementary Materials 3.7.3**.

After the computation of ELP_F, ELP_{ML}, and ELP_{BC}, we modeled ELP_{BC} using a linear function of ELP_F and/or ELP_{ML} so as to obtain an approximation of the most optimal ELP using available variables. We compared four different approaches of approximating ELP_{BC}: (1) Original, $ELP'_F = ELP_F$: using the original ELP_F, (2) Formula LR, $ELP'_F = c_1 \cdot ELP_F + c_3$: using linearly adjusted ELP_F, (3) ML LR, $ELP'_F = c_2 \cdot ELP_{ML} + c_3$: using linearly adjusted ELP_{ML}, (4) Formula & ML LR, $ELP'_F = c_1 \cdot ELP_F + c_2 \cdot ELP_{ML} + c_3$: using a linear combination of ELP_F

and ELP_{ML}. Here c_1, c_2, and c_3 are constants. Outliers with large refraction errors (i.e., $error \geq mean\ error + 2 \cdot standard\ deviation$ or $error \leq mean\ error - 2 \cdot standard\ deviation$) were excluded for each formula before establishing the linear regression model, in order to obtain better modeling results. The refraction prediction errors were calculated as $error = predicted\ refraction - true\ refraction$. The linear regression was performed using scikit-learn 0.20.3.

On the testing set, ELP_F' was calculated based on the values of c_1, c_2, and c_3 obtained through linear regression. The predicted refraction was calculated as $refraction = f_1(ELP_F', biometry)$. The mean absolute error (MAE), median absolute error (MedAE) and mean error (ME) were calculated for performance comparison.

3.3.4 A-constant optimization

The A-constants for the formulas were optimized based on the training dataset so that the mean error in refraction prediction was closest to zero. The A-constants were optimized separately for the unmodified formulas and formulas with a modified ELP estimate (see additional details in the "A-Constant Optimization" section and **Figure S *3.1*** in **Supplementary Materials**). The optimized A-constants for the original formulas were: a0 = -0.733, a1 = -0.234, a2 = 0.217 for Haigis, ACD constant = 5.724 for Hoffer Q, surgeon factor = 1.864 for Holladay, and A = 119.089 for SRK/T (**Table S *3.1***).

3.3.5 Statistical analysis

Linear regression analysis was used to assess the significance of the correlation between ELP_F, ELP_{ML}, and ELP_{BC}. To test whether the MAE and ME of different methods were significantly different, a Friedman test followed by a post hoc paired Wilcoxon signed-rank test

with Bonferroni correction was used. Statistical significance was defined as the p-value <0.05. All the above analyses were performed with Python 3.7.3.

3.4 Results

3.4.1 Dataset overview

The cases in the training and testing datasets had a similar distribution according to the summary statistics shown in **Table *3.1***. As elaborated in **Materials and Methods**, we calculated ELP_F, ELP_{ML} , and ELP_{BC} based on the formulas and their optimized A-constants. The mean and standard deviation of the ELPs calculated based on the original formulas were summarized in **Table S *3.2***. ELP_{BC} and ELP_F had similar mean values in contrast to ELP_{ML}.

The Pearson correlation coefficients (R) between ELP_F, ELP_{ML} , and ELP_{BC} were shown in **Table *3.2***. Three ELP-related variables were positively intercorrelated with each other. The correlation coefficients, R, between ELP_{BC} and ELP_{ML} were the weakest among the three pairs of variables across all formulas.

Characteristic	Training set	Testing set
Gender	Male: 2514 eyes (43.6%), Female: 3247 eyes (56.4%)	Male: 425 eyes (44.2%), Female: 536 eyes (55.8%)
Age at surgery (years)	70.99 ± 9.61	70.10 ± 10.24
Preoperative K (D)	43.85 ± 1.64	43.90 ± 1.66
Preoperative AL (mm)	24.19 ± 1.40	24.20 ± 1.41
Preoperative LT (mm)	4.54 ± 0.45	4.53 ± 0.45
Preoperative ACD (mm)	3.24 ± 0.41	3.26 ± 0.41
Postoperative refraction (D)	-0.53 ± 0.96	-0.57 ± 0.90

Table 3.1 The summary statistics for the patient demographics for the training and testing dataset.

For the age at surgery, preoperative biometry, and postoperative refraction, the mean ± SD (standard deviation) is shown in the table. K = keratometry; AL = axial length; LT = lens thickness; ACD = anterior chamber depth; D = Diopter.

Index	Variable Pairs	Haigis	Hoffer Q	Holladay1	SRK/T
1	ELP_F vs. ELP_{ML}	0.751	0.676	0.698	0.636
2	ELP_{BC} vs. ELP_F	0.621	0.730	0.622	0.633
3	ELP_{BC} vs. ELP_{ML}	0.532	0.544	0.534	0.524

Table 3.2 The Pearson correlation coefficients (R) between ELP_F, ELP_{ML} , and ELP_{BC}.
The ELP_{BC} and ELP_F were calculated using the A constants optimized based on the original formulas. P-values of all correlations were < 0.05. All R were rounded to three decimal places.

3.4.2 Linear regression results on the training set

Linear regression models were established based on the training set and the R^2 of alternative linear models were shown in **Table *3.3***. The coefficients of the fitted linear regression line are shown in **Table S *3.3***. The mean and SD of the ELP_F' resulting from different models are shown in **Table S *3.4***. For "Formula LR", the R^2 was larger than that of "ML LR" for all four formulas. For "Formula & ML LR", the R^2 was larger than that when one of ELP_F or ELP_{ML} was excluded from the linear combination for all four formulas.

Index	Methods	Haigis	Hoffer Q	Holladay1	SRK/T
1	Formula LR	0.377	0.541	0.579	0.394
2	ML LR	0.376	0.442	0.426	0.378
3	Formula & ML LR	**0.425**	**0.622**	**0.605**	**0.482**

Table 3.3 The R^2 of alternative least-squares linear regression models in the training set.
The outlier cases were removed before calculating the above values. The largest R^2 among three methods is marked in bold for each formula. P-values of all correlations were < 0.05.

3.4.3 Refraction prediction performance comparison on the testing set

Linear regression analysis was used to assess the significance of the correlation between ELP_F, ELP_{ML}, and ELP_{BC}. To test whether the MAE and ME of different methods were significantly different, a Friedman test followed by a post hoc paired Wilcoxon signed-rank test with Bonferroni correction was used. Statistical significance was defined as the p-value <0.05. All the above analyses were performed with Python 3.7.3.

We tested the performance of four scenarios on the testing set and summarized the MAE and SD in **Table *3.4***. The mean error (ME) and median absolute error (MedAE) were shown in **Table S *3.5*** and **Table S *3.6***. Statistical tests were used to compare the difference in the MAEs of different models (see **Materials and Methods 3.3.5**). Using a linear combination of ELP_F and ELP_{ML}, the refraction prediction results of four existing formulas were significantly improved compared to original ELP_F (statistical test results shown in **Table S *3.7*** and **Table S *3.8***).

We further compared the MAEs of "Original" and "Formula & ML LR" among patients with short, medium, and long axial length (**Table S *3.9***). It was observed that the short and medium axial length groups had a higher percentage decrease in MAE than the long axial length group for Hoffer Q and SRK/T. For Haigis, the medium AL group achieved higher decrease than the other two groups. And for Holladay, the long AL group achieved more decrease in MAE than the other two groups.

Index	Methods	Haigis	Hoffer Q	Holladay1	SRK/T
1	Original	0.373 ± 0.328	0.408 ± 0.337	0.384 ± 0.341	0.394 ± 0.351
2	Formula LR	0.373 ± 0.328 (0.0%)	0.374 ± 0.321 (8.3%)	0.388 ± 0.342 (-1.1%)	0.391 ± 0.345 (0.8%)
3	ML LR	0.391 ± 0.346 (-4.8%)	0.454 ± 0.375 (-21.4%)	0.434 ± 0.364 (-13.0%)	0.397 ± 0.344 (-1.5%)

4	Formula & ML LR	**0.356 ± 0.329 (9.0%)**	**0.352 ± 0.319 (22.5%)**	**0.371 ± 0.336 (3.4%)**	**0.361 ± 0.331 (9.1%)**

Table 3.4 Performance in the testing set.

The mean absolute error (MAE) ± standard deviation (SD) and the percentage reduction in MAE compared to "Original" for alternative linear models in the testing set. All MAE and SD were rounded to three decimal places. The percentage reduction was calculated as $\frac{MAE\ of\ a\ given\ method - MAE\ of\ Original}{MAE\ of\ Original} \cdot 100\%$. All percentage reduction values were rounded to one decimal place. The method with the smallest MAE among four alternative methods is marked in bold for each formula.

3.5 Discussion

In this study, we applied a previously developed machine learning method for postoperative anterior chamber depth (ACD) prediction to an unseen dataset of 4806 cataract surgery patients to assess whether it was possible to improve the performance of existing IOL formulas (Haigis, Hoffer Q, Holladay, and SRK/T) by replacing each formula's ELP estimate.

We computed three ELP-related quantities: the machine learning-predicted postoperative ACD (ELP_{ML}), formula-predicted ELP (ELP_F), and a back-calculated ELP (ELP_{BC}) that minimized the refraction error for each eye in the dataset. They are strongly correlated with each other (**Table *2.2***), which indicates that (1) ELP_F and ELP_{ML} are both predictive of the most optimal ELP ELP_{BC}, (2) ELP_F and ELP_{ML} contain partially overlapping information, which is consistent with our expectation. ELP_{ML} is an estimation of the value of the true postoperative ACD. On the other hand, ELP_F was designed by the originators of each formula to serve a similar purpose but was based on the theoretical assumptions in each formula. Our findings are consistent with observations of previous studies that the ELP estimates made by IOL formulas were numerically different from the true postoperative ACD. [110]

Using a training dataset of 3845 patients, we sought to evaluate whether the machine-predicted postoperative ACD, ELP_{ML}, was able to provide information that could be used to

refine each formula's predicted ELP, ELP_F. We established regression models between the ELP_{ML}, ELP_F, and ELP_{BC} to evaluate whether a linear combination of ELP_{ML} and ELP_F used in place of the original ELP_F could lower the refraction prediction error. Using the modified ELPs, we obtained significantly lower mean absolute errors (MAE) in refraction prediction compared to the formulas with the original ELPs on the unseen testing set (**Table S *3.4***). Notably, the accurately predicted postoperative ACD (ELP_{ML}) alone did not outperform the original ELP (ELP_F) when it was inserted into the formulas (**Table S *3.4***, row 3 compared to row 1). This is likely because the original method of calculating ELP in each formula compensates for its particular model of the eye and its associated assumptions. Our ELP_{ML}, however, does not have any components that compensate for the assumptions and constants in the formulas. On the other hand, ELP_{ML} has information about the true postoperative ACD, which it appears can beneficially alter the original ELP estimate.

In this study, the A-constants were optimized separately when ELP_F was replaced with different ELP_F'. The means of ELP_F', as shown in **Table S *3.4*** were numerically close to those of ELP_F as shown in **Table S *3.2***. However, in our method, the similarity between ELP_F' and ELP_F was not among the restrictions and goals of the optimization. The reason that ELP_F' and the original ELP_F had similar means might be that the other parts of each formula put restrictions on the values of ELP in order to obtain reasonable results. This could also be the reason why ELP_{BC} and ELP_F had similar means as shown in **Table S *3.2***.

Previous studies involving replacement of ELP in existing formulas have focused on special cases, such as sulcus implantation and post-refractive surgery eyes, where ELP estimates of traditional formulas would be expected to be inapplicable.[10,11] However, the method for replacing ELP estimates presented here provides a simple way of improving the refraction

prediction performance of existing formulas for the general cataract surgery population. While it would be ideal to evaluate this method on modern formulas such as Barrett Universal II or Holladay 2, the absence of published equations for these formulas prevents such a study. As such, we studied the application of the machine learning predicted postoperative ACD in four existing formulas whose mathematical equations were published. Although it awaits to be further validated, similar results can likely be transferred to other refraction prediction methods, since many modern IOL power formulas use predicted postoperative ACD as an intermediate step for predicting postoperative refraction. A limitation of the study was the absence of an external validation set, despite the use of a large unseen testing dataset (961 eyes). Accordingly, evaluation of the method at additional institutions and the extension to additional formulas will be future directions of this work.

In summary, the results of this study demonstrate that a machine learning method for postoperative ACD prediction based on postoperative optical biometry can be incorporated into a variety of existing IOL power formulas to improve their accuracy in refraction prediction.

3.6 Publication and Acknowledgement

This chapter is a published work [129]: Li, Tingyang, Joshua Stein, and Nambi Nallasamy. "AI-powered effective lens position prediction improves the accuracy of existing lens formulas." *British Journal of Ophthalmology* (2021).

3.7 Supplementary Materials

3.7.1 Supplementary figures

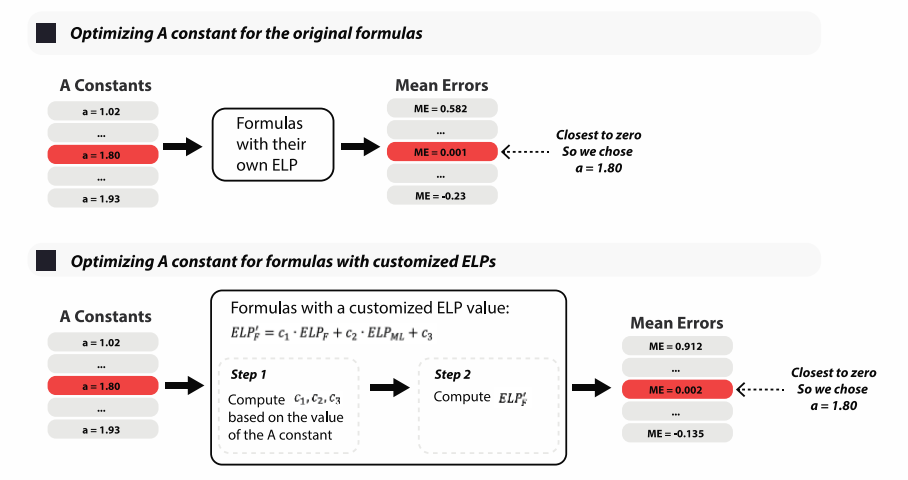

Figure S 3.1 The pipeline of the A-constant optimization procedure.

Numbers in the figure don't represent real data.

3.7.2 Supplementary tables

Formula	Constant	Optimized Constants	Mean Error
Haigis	a0 (a1 = 0.234, a2 = 0.217)	-0.736	0.000319
Hoffer Q	ACD	5.726	0.000157
Holladay1	Surgeon factor	1.867	0.000121
SRK/T	A constant	119.091	-0.000204

Table S 3.1 The optimized A constants and the corresponding mean error for the original formulas in the training dataset.

The A constants were optimized so that the absolute value of the mean error was minimized. The mean errors were calculated after excluding the outliers (see main text). The mean errors in the training set were rounded to three significant figures.

Dataset	Variables	Holladay1	SRK/T	Hoffer Q	Haigis
Training dataset	ELP_F	4.00 ± 0.41	5.82 ± 0.50	5.86 ± 0.38	5.28 ± 0.36
	ELP_{BC}	4.08 ± 0.92	5.85 ± 0.83	5.97 ± 1.00	5.32 ± 0.76

	ELP_{ML}	4.67 ± 0.27			
Testing dataset	ELP_F	4.00 ± 0.39	5.83 ± 0.47	5.87 ± 0.38	5.28 ± 0.36
	ELP_{BC}	4.06 ± 0.72	5.84 ± 0.65	5.95 ± 0.80	5.31 ± 0.61
	ELP_{ML}	4.68 ± 0.26			

Table S 3.2 The mean ± standard deviation (SD) for ELP_F, ELP_{ML}, and ELP_{BC} in the training and testing dataset when $ELP'_F = ELP_F$.

The ELP_{BC} and ELP_F were calculated using the corresponding formula with the optimized A-constants therefore their values vary with different formulas. The values of ELP_{ML} only depend on the values of the preoperative biometry. The outliers were not removed when the above summary statistics were calculated. All values were rounded to two decimal places.

Methods	Holladay1	SRK/T	Hoffer Q	Haigis
Formula LR	c_1 = 1.27 c_3 = -1.01	c_1 = 0.76 c_3 = -1.67	c_1 = 1.44 c_3 = -26.65	c_1 = 1.00 c_3 = -5.88
ML LR	c_2 = 1.65 c_3 = -3.80	c_2 = 1.31 c_3 = -0.33	c_2 = 1.62 c_3 = -1.70	c_2 = 1.24 c_3 = -0.53
Formula & ML LR	c_1 = 0.98 c_2 = 0.61 c_3 = -2.77	c_1 = 0.47 c_2 = 0.79 c_3 = -2.55	c_1 = 1.09 c_2 = 0.65 c_3 = -6.90	c_1 = 0.58 c_2 = 0.68 c_3 = -4.44

Table S 3.3 The coefficients (c_1 and c_2) and the intercept c_3 for the linear regression model established based on the training dataset.

All values were rounded to two decimal places.

Dataset	Method	Holladay1	SRK/T	Hoffer Q	Haigis
Training dataset	Formula LR	4.06 ± 0.52	5.80 ± 0.38	5.90 ± 0.55	5.28 ± 0.36
	ML LR	3.92 ± 0.44	5.78 ± 0.35	5.86 ± 0.43	5.26 ± 0.33
	Formula & ML LR	3.98 ± 0.53	5.80 ± 0.40	5.89 ± 0.55	5.27 ± 0.37
	Formula LR	4.04 ± 0.46	5.80 ± 0.31	5.91 ± 0.53	5.28 ± 0.34

Testing dataset	ML LR	3.90 ± 0.39	5.79 ± 0.32	5.87 ± 0.39	5.27 ± 0.30
	Formula & ML LR	3.97 ± 0.46	5.79 ± 0.34	5.89 ± 0.49	5.28 ± 0.34

Table S 3.4 The mean ± standard deviation (SD) for ELP'_F in the training and testing dataset.

The ELP_{BC} and ELP_F were calculated using the corresponding formula with the optimized A-constants therefore their values vary with different formulas. The values of ELP_{ML} only depend on the values of the preoperative biometry. The outliers were not removed when the above summary statistics were calculated.

Methods	Holladay1	SRK/T	Hoffer Q	Haigis
Original	-0.020 ± 0.513	-0.008 ± 0.528	-0.020 ± 0.529	-0.025 ± 0.496
Formula LR	0.008 ± 0.517	-0.003 ± 0.522	-0.017 ± 0.492	-0.018 ± 0.496
ML LR	0.057 ± 0.563	0.001 ± 0.525	0.007 ± 0.589	-0.001 ± 0.523
Formula & ML LR	0.009 ± 0.500	-0.008 ± 0.490	-0.016 ± 0.475	-0.014 ± 0.484

Table S 3.5 The mean error (ME) ± standard deviation (SD) of alternative linear models in the testing set.

All values were rounded to three decimal places.

Methods	Holladay1	SRK/T	Hoffer Q	Haigis
Original	0.299	0.307	0.330	0.283
Formula LR	0.305	0.310	0.293	0.283
ML LR	0.351	0.308	0.366	0.304
Formula & ML LR	0.290	0.273	0.268	0.263

Table S 3.6 The median absolute error (MedAE) of alternative linear models in the testing set.

All values were rounded to three decimal places.

Statistic	Holladay1	SRK/T	Hoffer Q	Haigis
Friedman chi-square test statistic	37.29	39.13	117.42	37.25
p-value	4.00e-08	1.63e-08	2.78e-25	4.07e-08

Table S 3.7 The Friedman test statistic and the p-values for comparing the testing set results of different methods.

All Friedman statistics were rounded to two decimal places. All p-values were rounded to three significant figures.

Formula	Methods	ML LR	Formula LR	Formula & ML LR
Haigis	Formula LR	**1.7E-01**	/	/
	Formula & ML LR	1.8E-11	2.6E-03	/
	Original	**1.7E-01**	**1.0E+00**	2.7E-03
Hoffer Q	Formula LR	1.5E-10	/	/
	Formula & ML LR	1.7E-26	3.6E-05	/
	Original	4.1E-05	1.5E-10	5.1E-17
Holladay 1	Formula LR	1.5E-04	/	/
	Formula & ML LR	4.4E-12	3.0E-04	/
	Original	1.4E-05	**1.0E+00**	9.9E-03
SRK/T	Formula LR	**1.0E+00**	/	/
	Formula & ML LR	1.7E-12	7.0E-06	/
	Original	**1.0E+00**	**1.0E+00**	1.1E-05

Table S 3.8 The post hoc test results of four existing formulas for comparing the testing set performance of different methods.

The insignificant p-values ($p \geq 0.05$) were highlighted in bold.

Method	Formulas	Short AL ($AL < 22mm$) n=28	Medium AL ($22mm \leq AL \leq 26mm$) n=832	Long AL ($AL > 26mm$) n=100
Original	Haigis	0.321 ± 0.234	0.373 ± 0.332	0.383 ± 0.315
	Hoffer Q	0.524 ± 0.295	0.396 ± 0.335	0.480 ± 0.350
	Holladay1	0.397 ± 0.224	0.364 ± 0.322	0.541 ± 0.464

	SRK/T	0.438 ± 0.236	0.386 ± 0.337	0.452 ± 0.465
Formula & ML LR	Haigis	0.330 ± 0.285 (-2.8%)	0.353 ± 0.331 (5.5%)	0.394 ± 0.319 (-2.8%)
	Hoffer Q	0.338 ± 0.264 (35.5%)	0.344 ± 0.318 (13.1%)	0.420 ± 0.338 (12.5%)
	Holladay1	0.392 ± 0.257 (1.3%)	0.356 ± 0.320 (2.2%)	0.486 ± 0.445 (10.3%)
	SRK/T	0.375 ± 0.284 (14.3%)	0.351 ± 0.324 (9.0%)	0.438 ± 0.391 (3.2%)

Table S 3.9 The mean absolute error (MAE) ± standard deviation in the testing set for patients with short, medium, and long axial length (AL).

All MAE and SD were rounded to three decimal places. For "Formula & ML LR", the percentage reduction in MAE compared to "Original" were shown. The percentage reduction was calculated as $\frac{MAE\ of\ Original - MAE\ of\ Formula\ \&\ ML\ LR}{MAE\ of\ Original} \cdot 100\%$ and was rounded to one decimal place. The letter "n" is the number of cases in each AL group.

3.7.3 Back-calculation of ELP

As described in the main text, the postoperative refraction was predicted using a function of ELP_F and preoperative biometry: $predicted\ refraction = f_1(ELP_F, biometry)$. Here we define ELP_{BC} as follows: when $ELP_F = ELP_{BC}$, $f_1(ELP_{BC}, biometry) - true\ refraction = 0$ holds for all cases. In other words, when the ELP estimation equals ELP_{BC}, the refraction prediction error equals zero for all cases. Based on the above definition, the value of ELP_{BC} can be found by solving for the x in the equation $f_1(x, biometry) - true\ refraction = 0$, where $biometry$and $true\ refraction$ are known. For a given case, there were always no more than two roots for the above function because of the quadratic nature of the formulas. When there were two roots, the smaller root was taken as ELP_{BC} because of two main reasons: (1) the greater root was usually >50, which was not within a physiologically meaningful range for ELP; (2) practically when the larger roots were used as ELP_{BC}, the R^2 in the training set was significantly lower than that obtained with the smaller root (data are not shown). The function

$f_1(x, biometry) - true\ refraction = 0$ was solved programmatically using scipy.optimize.fsolve (scipy 1.2.1) in Python 3.7.3.

3.7.4 A-constant optimization

When ELP_F was not replaced with a modified value ELP'_F (**Figure S *3.1***, upper part), the A-constants of the formulas were optimized in the standard way: first, compute the mean refraction prediction error when the A-constant takes different values, then, the A-constant that gives the smallest absolute mean error is the most optimal A-constant.

When $ELP_F = ELP'_F$ (**Figure S 3.1**, lower part), the A-constants were optimized based on the same concept, the value of ELP'_F changes with the values of the A-constant. The pseudo-code for the A-constant optimization process is shown below. The value of ELP_{ML} does not change with the A-constant. The value of ELP_F and ELP_{BC} changes with the A-constants.

Algorithm: A-constant optimization when $ELP_F = ELP'_F$	
1	ELP_{ML}← compute ELP_{ML} using the machine learning model
2	FOR a IN A-constant search space
3	ELP_F ← compute ELP_F based on the formula with a as the A constant
4	ELP_{BC}← compute ELP_{BC} based on the formula with a as the A constant
5	c_1, c_2, c_3 ← model ELP_{BC} as a linear function of $ELP_{ML}$and/or ELP_F.
6	$ELP'_F \leftarrow c_1 \cdot ELP_F + c_2 \cdot ELP_{ML} + c_3$
7	predicted refraction ← compute the predicted refraction based on a and ELP'_F
8	mean error ← compute ME based on the predicted and true refraction
9	END FOR
10	The most optimal A-constant← the A-constant that minimizes the absolute mean error

Chapter 4 Ray Tracing IOL Calculation Performance Improved by AI-Powered Postoperative Lens Position Prediction

4.1 Abstract

The aim of this study was to assess whether incorporating a machine learning (ML) method for accurate prediction of postoperative anterior chamber depth (ACD) improves cataract surgery refraction prediction performance of a commonly used ray tracing power calculation suite (OKULIX). A dataset of 4357 eyes of 4357 cataract patients was gathered at the Kellogg Eye Center, University of Michigan. A previously developed machine learning-based method was used to predict the postoperative ACD based on preoperative biometry measured with the Lenstar LS900 optical biometer. Refraction predictions were computed with standard OKULIX postoperative ACD predictions and ML-based predictions of postoperative ACD. The performance of the ray tracing approach with and without ML-based ACD prediction was evaluated using mean absolute error (MAE) and median absolute error (MedAE) in refraction prediction as metrics. Replacing the standard OKULIX postoperative ACD with the ML-predicted ACD resulted in statistically significant reductions in both MAE (1.7% after zeroing mean error) and MedAE (2.1% after zeroing mean error). ML-predicted ACD substantially improved performance in eyes with short and long axial lengths ($p < 0.01$). Using an ML-powered postoperative ACD prediction method improves the prediction accuracy of the OKULIX ray tracing suite by a clinically small but statistically significant amount, with the greatest effect seen in long eyes.

4.2 Introduction

Most modern IOL calculation formulas involve computation of postoperative refraction using Gaussian optics, which relies on the assumption that incoming rays are paraxial, in addition to empirically determined adjustment factors. The primary empirical adjustments for these modern formulas (such as Barrett Universal II, Holladay 2, and SRK/T) are made through the use of Effective Lens Position (ELP) as an intermediate quantity to indicate the location of the lens as it relates to a given optical model of the eye.[130] ELP was initially intended to estimate the position of the IOL in the postoperative eye. In practice, however, the postoperative ACD and the optimal location of the principal plane of the IOL in a given formula's optical model of the eye are not numerically equal.[28,131] Achieving an accurate prediction of the optimal ELP represents an ongoing limitation for modern formulas.

Numerical ray tracing represents an alternative to Gaussian optics for the purpose of IOL power calculation. Ray tracing involves the direct calculation of refraction of rays of light at each medium change within the eye using Snell's law. Studies have demonstrated that ray tracing performance is comparable to that of state-of-the-art IOL calculation formulas in normal eyes and may provide improved IOL calculation accuracy in certain populations.[132,133]

The more data available regarding the index of refraction of each medium, the curvature of each surface (including the anterior and posterior surfaces of the cornea and the intraocular lens implant) and position of each of these refracting surfaces relative to one another, the more accurate a ray tracing calculation is. As such, ray tracing calculations are likely to benefit from improved methods for predicting the actual anatomical postoperative IOL position.

In Chapter 2 and Chapter 3, we demonstrated that 1) it is possible to improve on estimates of the true anatomical postoperative IOL position through the use of a gradient-boosting machine learning (ML) approach, and 2) incorporation of this ML-predicted postoperative IOL position can be used to refine ELP estimates in existing IOL formulas and improve their accuracy. [114,129]

In this chapter, we investigate whether our machine learning method for prediction of postoperative IOL position as described in Chapter 2 is able to improve the accuracy of the commonly used OKULIX ray tracing suite for intraocular lens power calculation. The standard approach to postoperative IOL position prediction in the OKULIX suite employs a regression model based on axial length and thickness of the crystalline lens. In the work presented here, we sought to determine whether, holding all else equal, replacement of the standard OKULIX postoperative IOL position prediction method with that of a highly accurate machine learning-based predictor, had the potential to improve the accuracy of ray tracing calculations.

4.3 Materials and Methods

4.3.1 Data collection

Preoperative and postoperative biometry records of cataract patients were exported from the Lenstar LS900 optical biometers (Haag-Streit USA Inc, EyeSuite software version i9.1.0.0) at the University of Michigan's Kellogg Eye Center. Patients who had cataract surgery but no prior corneal surgery and no additional surgical procedures at the time of cataract surgery were included. Only surgery cases involving the implantation of Alcon SN60WF single-piece acrylic monofocal lenses (Alcon Inc., USA) were included in the study, because it is the most commonly implanted lens at the Kellogg Eye Center. Cases that were used to train the postoperative ACD

prediction machine learning model were excluded from the dataset so that the dataset involved only unseen samples. One eye was selected at random for each patient who had undergone surgery in both eyes, so that all cases in the final dataset were independent of each other. The preoperative information gathered included the measurements of the axial length (AL), crystalline lens thickness (LT), anterior chamber depth (ACD), the radius of curvature in the flat meridian (R1), the radius of curvature in the steep meridian (R2), patient gender, and selected IOL power. As defined by the Lenstar LS900 optical biometer, the postoperative ACD represents the distance from the anterior surface of the cornea to the anterior surface of the IOL. The postoperative refraction (including the spherical component [SC] and cylindrical component [CC]) records were obtained. The spherical equivalent (SE) refraction was calculated as $SE\ refraction = (SC - 0.1614) + 0.5CC$. The constant 0.1614 was used to account for the length (10 feet, 3.048 meters) of the examination lane according to Simpson and Charman's recommendation.[134] For each patient, the postoperative record that was generated closest to one month (i.e., 30 days) after surgery was included. Details about the collection and processing of the dataset can be found in our previous publications. [114,129]

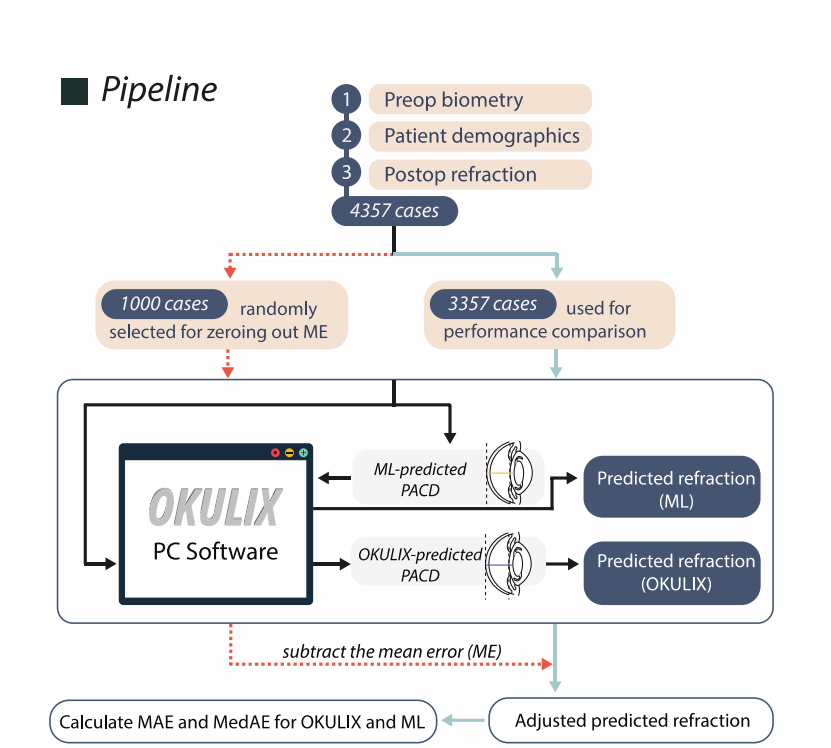

Figure 4.1 The analysis pipeline of the presented study.

PACD = postoperative anterior chamber depth. MAE = mean absolute error. MedAE = median absolute error. The "OKULIX-predicted PACD" was not a direct output from OKULIX, but calculated as $PACD = PAD - CCT$, where PAD is the OKULIX-predicted postoperative aqueous depth (AD).

4.3.2 Performance comparison between OKULIX and ML-based approach

The dataset in total consisted of the aforementioned preoperative and postoperative data for 4357 eyes of 4357 patients (**Figure *4.1***). Our postoperative ACD prediction machine learning model (referred to as the "Base" model in our prior work) was used to compute predictions of postoperative ACD (in mm) for each eye in the dataset based on the preoperative data (AL, CCT, preoperative ACD, LT, R1, R2, and WTW) and patient gender. Details about the ML method can be found in our previous publication.[114] The OKULIX standalone PC software suite (OKULIX v9.20, Panopsis GmbH, Mainz, Germany) was utilized to compute refraction

predictions based on the available preoperative biometry (AL, LT, R1, R2, and preoperative ACD) and laterality of the case. In addition to the postoperative refraction, OKULIX also predicts the postoperative aqueous depth (AD) as an intermediate value which in downstream pipelines serves as one of the input variables for the prediction of the refraction. By design, the OKULIX software allows the users to replace its predicted postoperative aqueous depth (AD) with a custom value. We therefore compared the refraction prediction accuracy of OKULIX when different PACD were used: (1) its standard internal prediction of postoperative ACD and (2) the postoperative ACD prediction from our ML model. OKULIX by default outputs a predicted aqueous depth (AD) defined as the distance from the posterior surface of the cornea to the anterior surface of the IOL, instead of a predicted PACD (defined as distance from the front surface of the cornea to the anterior surface of the IOL). In order to transform the ML-predicted postoperative ACD to a predicted postoperative AD, the preoperative central corneal thickness (CCT) was subtracted from the ML-predicted postoperative ACD. Similarly, to compare the OKULIX and ML-predicted PACD, we transformed the OKULIX-predicted postoperative AD to postoperative ACD prediction by adding the preoperative CCT. The equations used to transform predicted postoperative ACD to AD and AD to ACD are shown as follows, where the OKULIX predicted postoperative AD is the direct output from OKULIX, and the ML predicted postoperative ACD is the direct output from the ML method.

$$OKULIX\ predicted\ postop.\ ACD = OKULIX\ predicted\ postop.AD + preop.CCT$$

$$ML\ predicted\ postop.\ AD = ML\ predicted\ postop.ACD - preop.CCT$$

The dataset was split into an optimization dataset and a performance comparison dataset, where the former was utilized for zeroing out the mean error and the latter was used for prediction performance comparison. A total of 1000 cases were randomly set aside as the

optimization dataset. Increasing the size of the optimization dataset did not significantly change the results (results not shown). The remaining 3357 cases were utilized to assess performance of the ray tracing calculations with and without the ML-predicted PACD. Mean absolute error (MAE), median absolute error (MedAE), and mean error (ME) of refraction predictions were computed for performance comparison. The prediction error was defined as follows. A negative error corresponds to a more myopic prediction, and a positive error corresponds to a more hyperopic prediction.

$$prediction\ error = predicted\ refraction - actual\ refraction$$

We further bootstrapped the performance comparison dataset to obtain the confidence intervals for the MAE, MedAE, and differences in MAE and MedAE. Specifically, a random sampling of 3357 cases with replacement was applied 10,000 times to the performance comparison dataset, and four metrics were calculated for those 10,000 bootstrap samples as follows. Let p_i be the absolute error of the OKULIX-based approach for the i-th case and q_i be the absolute error of the ML-based approach for the i-th case in the bootstrap dataset. For each bootstrap dataset we calculated (1) the mean absolute error as $mean\{p_1, p_2, \ldots, p_{3357}\}$ and $mean\{q_1, q_2, \ldots, q_{3357}\}$, (2) the median absolute error as $median\{p_1, p_2, \ldots, p_{3357}\}$ and $median\{q_1, q_2, \ldots, q_{3357}\}$, (3) the mean of the differences between the absolute error as $mean\{p_1 - q_1, p_2 - q_2, \ldots, p_{3357} - q_{3357}\}$, (4) the median of the differences between the absolute error as $median\{p_1 - q_1, p_2 - q_2, \ldots, p_{3357} - q_{3357}\}$.

4.3.3 Zeroing of mean error

To account for possible systematic differences in our patient population, we zeroed out the mean errors of the OKULIX predictions individually for the OKULIX and ML approaches. This was done by computing the mean error on the optimization dataset and subtracting it from

the predictions on the performance comparison dataset (**Figure *4.1***). The aforementioned scoring metrics (e.g. MAE) were computed after the zeroing of mean error.[135]

4.3.4 Statistical analysis

A paired Wilcoxon test was performed to evaluate the significance of the difference between the OKULIX and ML-predicted PACD. The same test (Wilcoxon test) was performed to test whether the prediction errors of the OKULIX and ML-based approach were significantly different. Statistical significance was defined as the p-value <0.05. All the above analyses were performed with Python 3.7.3.

4.4 Results

4.4.1 Dataset overview

The dataset in total consisted of 4357 cataract surgery cases of 4357 patients. Among those patients, 1919 (44.04%) were males and 2438 (55.96%) were females. A summary of the dataset is shown in **Table *4.1***. The OKULIX-predicted postoperative ACD had a mean of 5.13 mm and was significantly longer than the ML-predicted postoperative ACD (mean=4.68 mm) (Wilcoxon-test p-value<0.01).

Characteristic	**Mean ± SD**	**Median**	**Range**	
			Min	**Max**
Age at surgery (years)	70.66 ± 9.53	71.24	13.23	89.45
Preoperative K (D)	43.85 ± 1.61	43.84	33.42	50.39
Preoperative AL (mm)	24.21 ± 1.38	24.01	20.44	31.22
Preoperative LT (mm)	4.52 ± 0.45	4.51	2.50	5.99
Preoperative ACD (mm)	3.26 ± 0.41	3.27	2.08	5.15

OKULIX-predicted postoperative ACD (mm)	5.13 ± 0.29	5.12	4.25	6.24
ML-predicted postoperative ACD (mm)	4.68 ± 0.26	4.69	3.85	5.59
Postoperative refraction (D)	-0.56 ± 0.96	-0.41	-12.16	3.34

Table 4.1 Baseline characteristics of the dataset.

The highly myopic postoperative refraction value ("-12.16" in the last row in the table) was from a patient with high myopia and unilateral cataract who chose a highly myopic target refraction to avoid anisometropia. D: diopter; SD: standard deviation.

4.4.2 Refraction prediction performance comparison

The OKULIX and ML-based approaches were compared using the performance comparison dataset, and the results are shown in (**Table *4.2***. The refraction prediction errors in the optimization dataset were significantly different for the OKULIX and ML-based approaches (**Table *4.2***, column 1) (Wilcoxon-test p-value < 0.01). The unadjusted predictions in the optimization dataset from OKULIX tended to be more myopic compared to the true refraction (ME = -0.329 D), while the unadjusted predictions from the ML-based method tended to be more hyperopic (ME = 0.211 D) (though to a lesser extent). After subtracting the mean errors from the optimization dataset, the mean errors of the OKULIX and ML-based approaches became much closer to zero in the performance comparison dataset (**Table *4.2***, column 2). The mean absolute errors (MAEs) in refraction prediction were significantly lower for the ML-based approach compared to the OKULIX-based approach (**Table *4.2***, column 3) (Wilcoxon-test p-value < 0.01).

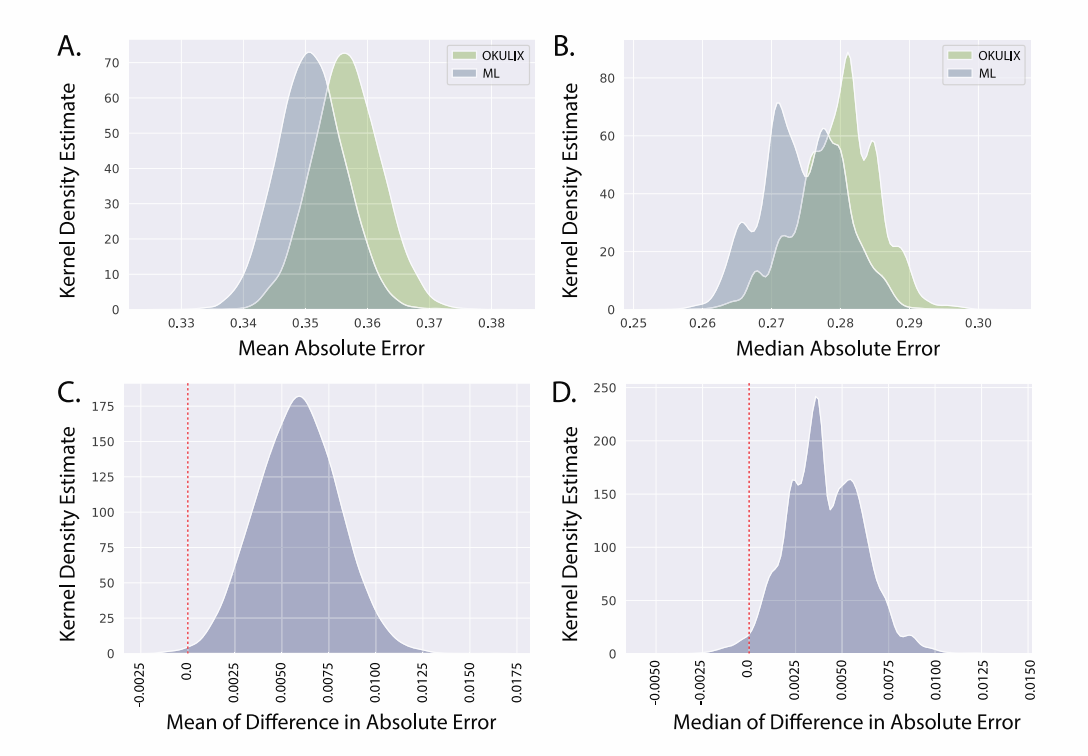

Figure 4.2 Bootstrap results in the performance comparison dataset.

A. The distribution of the mean absolute error. **B**. The distribution of the median absolute error. **C**. The distribution of the means of the difference in the absolute errors of OKULIX and the ML method. **D**. The distribution of the medians of the difference in the absolute errors of OKULIX and the ML method. The red dashed line marks the location of 0.0.

The bootstrap distributions of MAE, MedAE and the mean/median difference in absolute errors are depicted in **Figure *4.2***. According to the bootstrap results, the 95% confidence interval (CI) for MAEs was [0.3459, 0.3677] for OKULIX and [0.3404, 0.3615] for ML. The 95% CI for MedAE was [0.2679, 0.2901] for OKULIX and [0.2638, 0.2858] for ML. The 95% CI for the mean of the differences in absolute errors was [0.0017, 0.0101]. The 95% CI for the median of differences in absolute errors was [0.0003, 0.0083].

	Unadjusted ME in optimization dataset ± SD	ME in performance comparison dataset ± SD	MAE in performance comparison dataset ± SD	MedAE in performance comparison dataset
OKULIX	-0.329 ± 0.471	0.018 ± 0.478	0.357 ± 0.318	0.280
ML	0.211 ± 0.470	0.019 ± 0.470	0.351 ± 0.313	0.274
% improvement	/	/	1.7%	2.1%

Table 4.2 Performance comparison set results.

The first column shows the ME ± SD in the optimization dataset, where the ME was calculated from the unadjusted predictions from OKULIX and ML. The second column shows the ME ± SD in the performance comparison dataset, after subtracting the corresponding ME in the optimization dataset from the original prediction from OKULIX and ML. The percent improvement was calculated as $\frac{OKULIX\ performance - ML\ performance}{OKULIX\ performance}$. ME: Mean Error; MAE: Mean Absolute Error; MedAE: Median Absolute Error.

4.4.3 Refraction prediction performance comparison in different axial length groups

We further summarized the performance of the OKULIX and ML-based approaches in different axial length (AL) groups in **Table *4.3***. There were significant differences in the short and long axial length groups and insignificant difference in the medium length group.

AL Group	Number of cases (%)	Wilcoxon test p-value	Method	MAE ± SD	MedAE
Short (AL<=23mm)	589 (17.5%)	<0.01	**OKULIX**	0.394 ± 0.333	0.313
			ML	0.373 ± 0.328	0.290
			% improvement	5.3 %	7.3%
Medium (23mm<AL<=26mm)	2429 (72.4%)	0.22	**OKULIX**	0.340 ± 0.306	0.265
			ML	0.343 ± 0.307	0.270
			% improvement	-0.9%	-0.5%

Long (AL>26mm)	339 (10.1%)	<0.01	**OKULIX**	0.409 ± 0.361	0.322
			ML	0.365 ± 0.328	0.271
			% improvement	10.8%	15.8%

Table 4.3 Results comparison in different axial length groups.

This table shows the performance of OKULIX and ML-based approaches for patients in the short, medium and long axial length group. The percent improvement was calculated as $\frac{OKULIX\ performance - ML\ performance}{OKULIX\ performance}$. The Wilcoxon test was performed to compare the difference in the absolute errors between OKULIX and ML-based approaches in three different axial length groups. MAE: Mean Absolute Error; MedAE: Median Absolute Error.

4.5 Discussion

In this study, we sought to determine whether utilizing a machine-learning based approach to prediction of postoperative IOL position could improve the performance of a ray tracing approach to IOL power calculations.

Ray tracing offers several advantages over traditional geometrical optics approaches to IOL power calculation. In particular, through the calculation of refraction of light rays at each refracting surface in the eye, factors such as irregular corneal curvature and varying pupil sizes can be accounted for with a ray tracing approach. The incorporation of more detailed surface information, including corneal tomography with Scheimpflug imaging or optical coherence tomography, offers the potential to maintain accuracy for a broader range of eyes than traditional methods. Since ray tracing accuracy depends primarily on the accuracy of anatomical measurements, rather than on theoretical quantities such as ELP, ray tracing may have greater long-term potential in comparison to traditional IOL calculation methods as techniques for measuring the size, shape, and anatomical location of components of the eye's optical system continue to evolve.

The physical location of the IOL (postoperative ACD) is one component of the post-cataract surgery optical system that is not directly measurable preoperatively. However, our group has previously developed and validated a machine learning method for accurate postoperative ACD prediction using preoperative optical biometry, gender, and intended IOL power.

In previous work, our group demonstrated that incorporating this ML method for postoperative ACD prediction into the ELP calculations of traditional IOL calculation formulas significantly improved the refraction prediction performance of these IOL formulas.[129] Since ray tracing methods in theory require only the true postoperative ACD, rather than an ELP, it would logically follow that utilization of our validated ML method for postoperative ACD prediction could improve the accuracy of ray tracing IOL power calculations. In the study presented here, we evaluated whether replacing the standard model for postoperative ACD in the commonly used OKULIX ray tracing suite with our ML-based postoperative ACD prediction could improve the refraction prediction performance of the OKULIX suite.

We found that replacing the standard OKULIX postoperative ACD model (based on a linear regression of axial length and lens thickness) with our ML-based postoperative ACD model resulted in statistically significant improvements in the MAE and MedAE of ray tracing refraction predictions. It appears likely that the internal ACD predictions of OKULIX were sufficiently accurate in normal length eyes (as would be expected for a linear regression approach) such that no significant difference in refraction prediction was seen in the normal length eye population. However, utilization of our ML approach significantly improved refraction prediction MAE and MedAE in patients with short eyes (5.3% reduction in MAE, 7.3% reduction in MedAE) and those with long eyes (10.8% reduction in MAE, 15.8% reduction

in MedAE). Although the overall improvement and the improvement in the short eyes were clinically small, we believe the results are of clinical significance for the long eyes.

The results highlighted the impact of the accuracy of postoperative ACD prediction on the accuracy of the refraction prediction for ray tracing methods. In relation to this study, our previous research has shown that substituting the ELP with a more accurately predicted postoperative ACD improved the accuracy of existing IOL formulas.[129] Since more and more studies on ELP (or postoperative ACD) prediction have been published, it is important to synchronize the progress in ELP prediction with efforts on refraction prediction. Our research provided a support for the vital role of postoperative ACD prediction in refraction prediction, suggesting that new generation ELP prediction methods should be considered when developing new IOL formulas and should be considered for incorporation with existing IOL formulas as an easily achievable refinement.

Since our ML model for postoperative ACD prediction was demonstrated to outperform linear regression in its prior validation study, it would be expected that incorporation of this ML model into the OKULIX suite would improve refraction prediction performance. While the means of the ACD predictions and refraction predictions were different between the standard and ML-based approaches, subtracting off the mean error in refraction prediction on an optimization subset represents a straightforward (and previously described) method for addressing these systematic differences. The postoperative ACD prediction of OKULIX appeared to be longer than the ones predicted by the ML method (**Table *4.1***), and naturally the unadjusted refraction predictions of OKULIX were more myopic (**Table *4.2***). We corrected this systematic error by subtracting the mean error for the performance comparison dataset.

While our study was intentionally limited to eyes with normal corneas (in order to test the standard version of the ML model described in our group's prior work), a clear future direction is to apply the keratometry-independent version of our ML model for predicting postoperative ACD to the ray tracing analysis of eyes with abnormal corneas, such as those with ectasia, prior refractive surgery, or prior keratoplasty. This subset of patients with abnormal corneas is one group for which ray tracing has been demonstrated to have clear advantages over traditional methods for IOL power calculation. However, the accurate prediction of postoperative ACD in this population is more challenging due to the absence of reliable keratometry data.

In addition, our study was limited to a retrospective sample and further investigation in a prospective manner would be of value.

In summary, this study demonstrates that incorporation of a validated machine learning method for postoperative ACD prediction can significantly improve ray tracing IOL calculation performance. Further investigation into the efficacy of this approach in eyes with ectatic and post-refractive surgery corneas is warranted.

4.6 Publication

This chapter is a published work [136]: Li, Tingyang, Aparna Reddy, Joshua D. Stein, and Nambi Nallasamy. "Ray tracing intraocular lens calculation performance improved by AI-powered postoperative lens position prediction." *British Journal of Ophthalmology* (2021).

Chapter 5 Postoperative Refraction Prediction with Ensemble Machine Learning

5.1 Abstract

The aim of this study is to develop a new intraocular lens (IOL) power selection method with improved accuracy for general cataract patients receiving Alcon SN60WF lenses. 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 to that of Barrett Universal II, EVO, Haigis, Hoffer Q, Holladay 1, PearlDGS, and SRK/T. Mean absolute error (MAE) of the Nallasamy formula in the testing dataset was 0.312 Diopters 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-value < 0.05). In conclusion, 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.

5.2 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 IOL power based on accurate prediction of postoperative refraction is necessary for achieving a favorable 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.[137,138]

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 and keratometry, the accuracy of IOL power calculation has been substantially improved. Among existing formulas, the Barrett Universal II formula [57] is widely used and several publications have demonstrated that Barrett Universal II has greater accuracy than other traditional formulas.[139,140] 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 [68] which is a theoretical thick lens formula, and the PearlDGS formula [69,141], which is a machine learning-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. Machine learning (ML) and artificial intelligence have proven

to be successful in many medical applications, including ophthalmology.[78,142] 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,[84] (2) failure to achieve statistically significant improvement over current generation formulas,[86] and (3) small datasets that leave the robustness and generalizability of methods in question.[85]

From Chapter 2 to Chapter 4, we presented our machine learning-based postoperative ACD prediction method, which provides a more accurate estimation of the postoperative lens position compared to existing IOL formulas, and we have shown that our ML-predicted ACD was able to improve the postoperative refraction prediction accuracy of existing Gaussian optics-based IOL formulas as well as ray tracing-based formulas. With a goal of advancing the understanding of IOL power selection for general cataract patients and improving refraction prediction accuracy, in the current chapter, we present a novel machine learning-based IOL power calculation method, the Nallasamy formula, based on a large dataset of 5016 cataract patients. In this model, we employed ensemble machine learning methods and novel data augmentation methods. The performance of our method was compared to that of Barrett Universal II, EVO, Haigis, Hoffer Q, Holladay 1, PearlDGS, and SRK/T on an unseen testing dataset of 1003 patients.

5.3 Materials and Methods

5.3.1 Data collection and preprocessing

The presented study focused on a subset of patients receiving care at the University of Michigan between August 25, 2015 and June 27, 2019. The preoperative biometry records were obtained from Lenstar LS 900 optical biometers (Haag-Streit USA Inc, EyeSuite software

version 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 (EHR) 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 (ICD) 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.[114,129,143–145] 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.

Institutional review board approval was obtained for the presented study. All subjects were fully anonymized and, therefore an informed consent was not required for this retrospective study. The study was carried out in accordance with the tenets of the Declaration of Helsinki.

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. (6) Data was complete and was not out of bounds for any of the formulas with which performance was compared.

5.3.2 Stacking ensemble machine learning 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 lens thickness (LT), anterior chamber depth (ACD), aqueous depth (AD), astigmatism, white-to-white (WTW), central corneal thickness (CCT), and keratometry (K1 and K2, $K = \frac{K1+K2}{2}$). 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 feet, 3.048 meters):
spherical equivalent (SE) $refraction = (\mathrm{SC} - 0.1614) + 0.5\mathrm{CC}$ according to Simpson and Charman's recommendation.[125]

The prediction task was framed as a regression problem where the goal was to build a machine learning 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 *5.1***) and the inputs that are used to make the predictions are referred to as features or predictors (represented as X in **Figure *5.1***). 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 *5.1***). 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 machine learning 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.[146] Stacking (or stacked generalization) 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.[147] The reason for using an ensemble ML model in this study is to take advantage of different classes of ML 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 *5.1***).

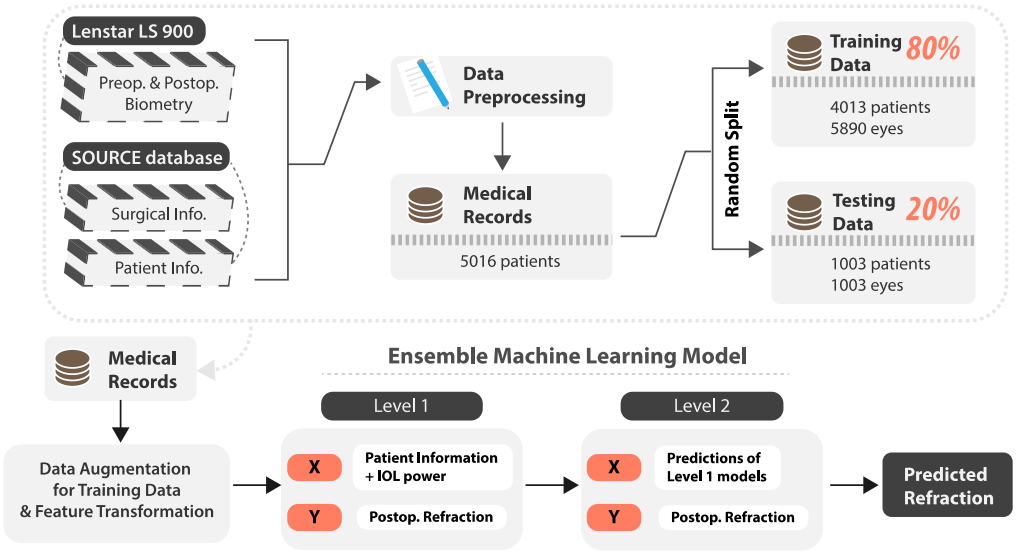

Figure 5.1 The overall method pipeline.

5.3.3 Lens constant optimization of existing IOL formulas

The existing formulas Haigis, Hoffer Q, Holladay 1, SRK/T were implemented in Python based on their specific equations.[52–55,59,126–128] The results obtained were validated against printouts from Haag-Streit USA Inc, EyeSuite software version i9.1.0.0. The prediction results of Barrett Universal II [57], EVO (v2.0) [68] and PearlDGS [69,141] were obtained through their online calculators. The constants of the corresponding formulas were optimized based on the cases in the training dataset (4013 patients). The most optimal constant was selected by zeroing the mean prediction error. The optimized constants are listed in **Table *5.1***.

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	Personalized ACD	5.727
Holladay 1	Surgeon factor	1.860
PearlDGS	IOL A constant	119.1
SRK/T	A constant	119.082

Table 5.1 The optimized lens constants.

5.3.4 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 optimization of hyperparameters of the machine learning models, the combination of the level-1 models, and the selection of the level-2

model were performed by minimizing the averaged mean absolute error (MAE) based on the cross-validation results.

5.3.5 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), mean absolute error (MAE), median absolute error (MedAE) of the postoperative refraction predictions and the standard deviation (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 et al[89] 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-value < 0.05. All statistical analyses were scripted with Python 3.9.5.

$$\text{error} = \text{true postoperative refraction} - \text{predicted postoperative refraction}$$

5.4 Results

You'll Out of 5016 patients, 4013 patients (5890 eyes) were assigned to the training/validation dataset, and 1003 cases were isolated 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 *5.2***. A total of 49 surgeons performed the surgeries included in the dataset. The distribution of data is shown in **Figure S *5.1***.

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

Table 5.2 Summary of patient demographics.

SD: standard deviation; K: keratometry; AL: axial length; LT: lens thickness; ACD: anterior chamber depth; D: diopter.

The performance of our method and existing methods is shown in **Table *5.3***. 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 **Table S *5.1***. Our method also achieved the highest Formula Performance Index (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 *5.2***. Our method resulted in a larger percentage of patients in the absolute error ≤ 0.5 D group (80.2%) compared to 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 to 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 **Table S *5.2***) on this metric.

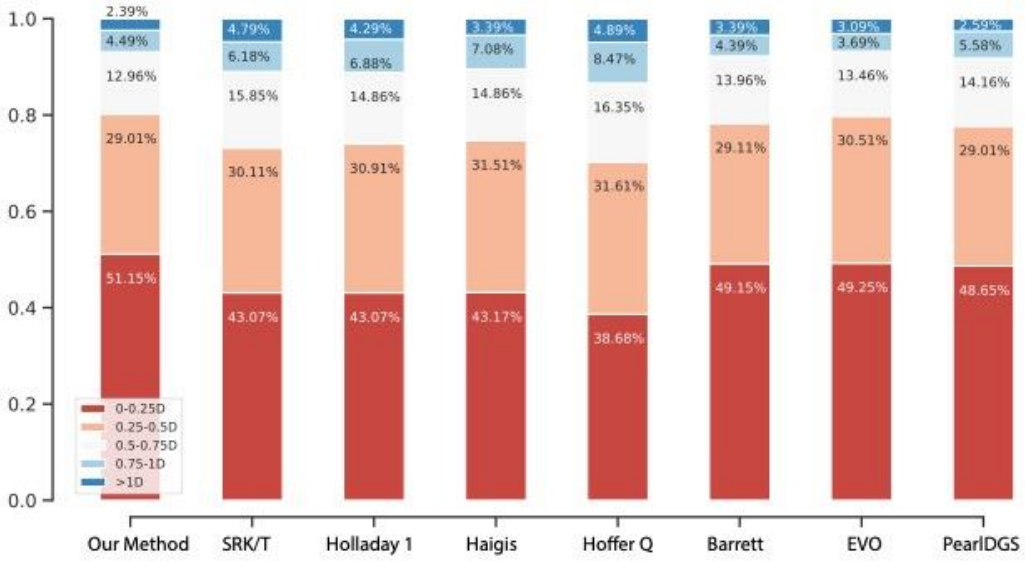

Figure 5.2 The percentage of patients in each error category for each formula, calculated based on the results in the testing dataset.

	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	$p < 0.05$
EVO	0.322	0.251	0.043	0.427	0.128	79.8%	0.312	$p < 0.05$
Haigis	0.363	0.289	0.024	0.469	0.226	74.7%	0.230	$p < 0.05$
Hoffer Q	0.404	0.331	0.009	0.518	0.951	70.3%	0.085	$p < 0.05$
Holladay 1	0.371	0.298	0.021	0.487	0.773	74.0%	0.101	$p < 0.05$
PearlDGS	0.329	0.258	0.044	0.438	0.408	77.7%	0.165	$p < 0.05$
SRK/T	0.376	0.300	0.014	0.485	0.486	73.2%	0.143	$p < 0.05$
Our Method	0.312	0.242	0.015	0.418	-0.033	80.2%	0.447	/

Table 5.3 Performance summary in the testing set.

MAE: mean absolute error; MedAE: median absolute error; ME: mean error, SD: standard deviation of the 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 absolute error (AE) less than or equal to 0.5 D; FPI: Formula Performance Index. The unit for the errors is Diopter (D). Wilcoxon test p-value < 0.05 indicates the statistical significance of the difference between the performance of our method and an existing method.

We compared the performance of the tested formulas among patients with different axial lengths in **Table *5.4***. Numerically, our method achieved the lowest MAEs and SDs among all 8 formulas in all 3 AL groups. The relationship between the prediction errors and the ALs is shown in **Figure *5.3***. The errors of our method remained close to zero across the whole span of ALs.

Method	**Short AL (< 22.0 mm, n = 32)**				**Medium AL (≥ 22.0 and ≤ 26.0 mm, n = 878)**				**Long AL (> 26.0 mm, n = 93)**			
	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**

Table 5.4 The postoperative refraction prediction performance of existing formulas and our method in short/medium/long AL groups in the testing set.

AL: axial length; n: number of eyes in each group; MAE: mean absolute error; MedAE: median absolute error; ME: mean error; SD: standard deviation of the prediction error.

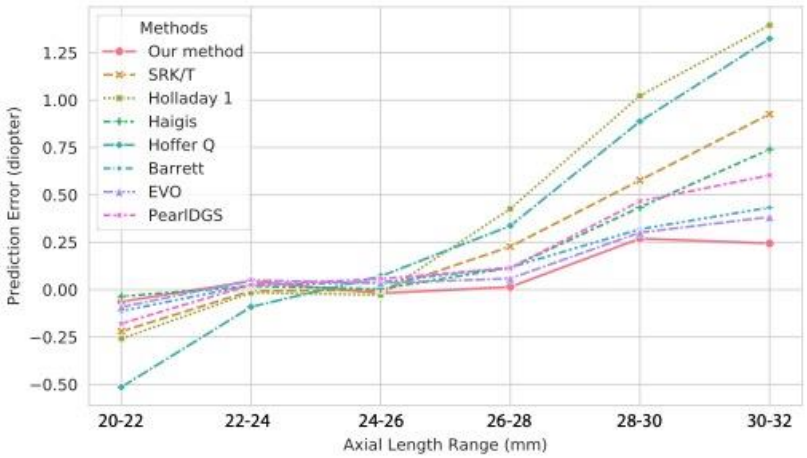

Figure 5.3 The mean prediction errors in the testing set grouped based on axial lengths.

Each dot represents the mean prediction error of eyes with an axial length between a specific range.

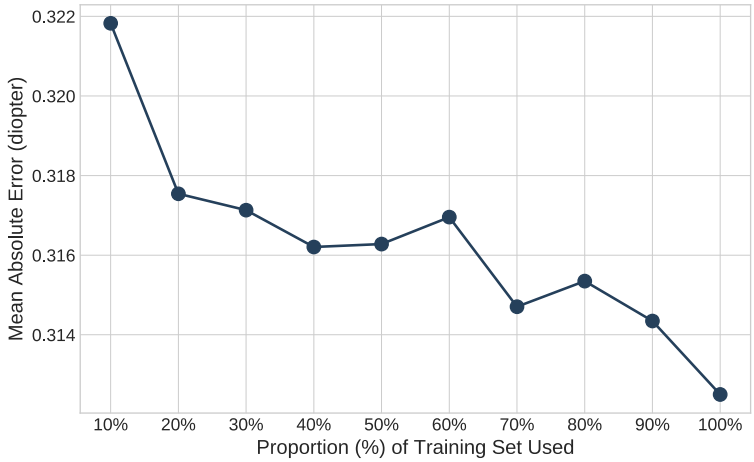

Figure 5.4 The change of the mean absolute prediction error in the testing set when the machine learning method uses 10%, 20%, …, 100% of the training data.

When the model was trained with different proportions of the training data (**Figure *5.4***), the corresponding performance on the testing set displayed a trend toward improving performance (decreased MAE) with increasing training set sizes.

5.5 Discussion

We have presented here a new machine learning-based IOL power calculation method which performs statistically significantly better than Barrett Universal II, EVO (v2.0) and PearlDGS on a large unseen testing dataset. We chose an ensemble machine learning 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 **Figure S *5.1***. For example, the axial length has more instances in the medium group (between 22 mm and 26 mm) compared to 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 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 *5.4***), 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 mean absolute error 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 *5.4*** 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 mean absolute error 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 *5.3*** 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 Formula Performance Index (FPI) as a means of evaluating and ranking the performance of IOL power calculation methods.[89] 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 *5.3***). 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 summarized in **Table *5.3***. 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 characterize the performance of an IOL power calculation formula.”[148]

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 *5.3***. Machine learning-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 (e.g., -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.[97,139,140] An additional limitation is that at present, our method has been customized 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 effective lens position (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 anterior chamber depth (ACD) estimation.[114,129] However, the method presented here does not rely upon 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, machine learning offers improved performance for large populations through the identification of latent patterns in historical data that can go unrecognized by existing methods. To that end, we have reported here the successful development and testing of a machine learning-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/.

5.6 Publication

This chapter is a published work[149]: Li, Tingyang, Joshua Stein, and Nambi Nallasamy. "Evaluation of the Nallasamy formula: a stacking ensemble machine learning method for refraction prediction in cataract surgery." *British Journal of Ophthalmology* (2022).

5.7 Supplementary Materials

5.7.1 Supplementary figures

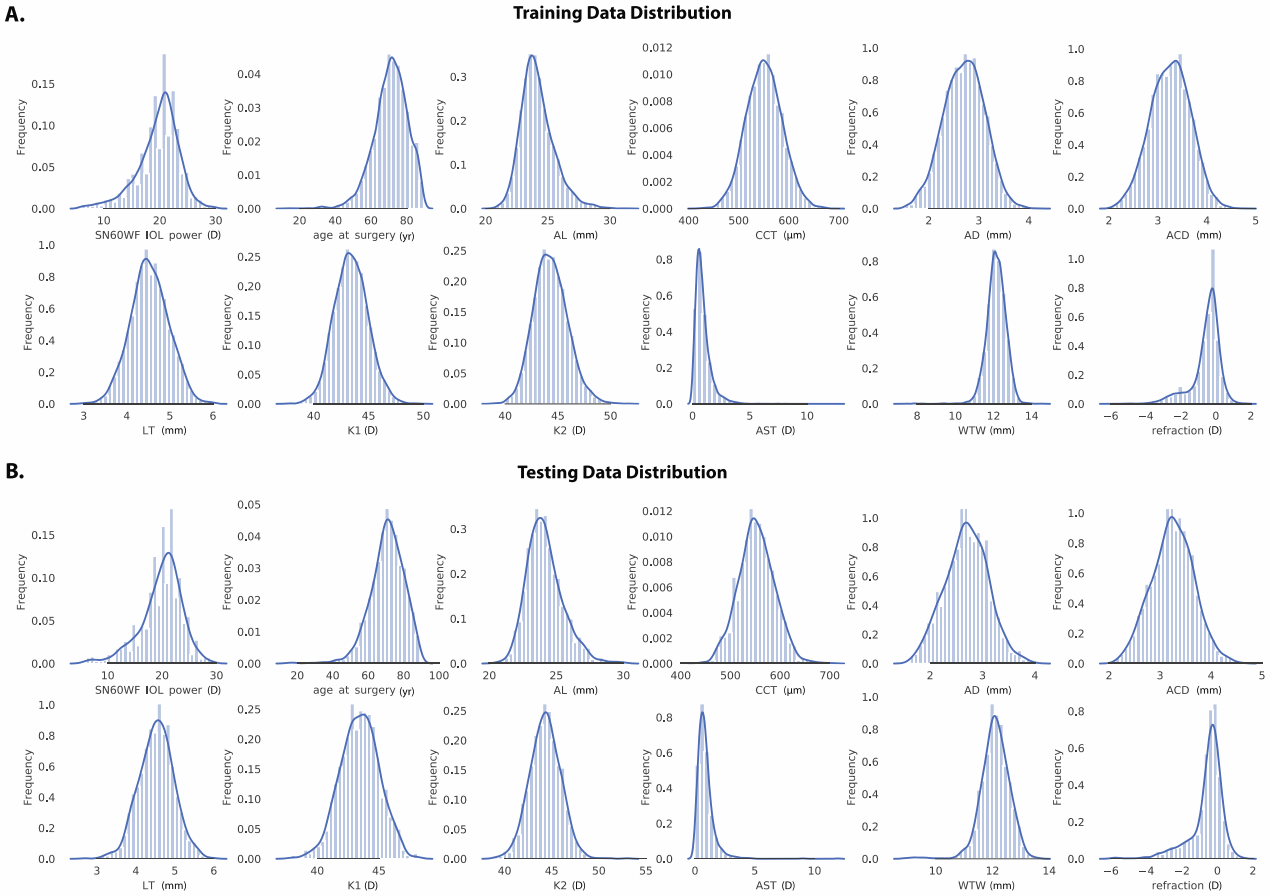

Figure S 5.1 Distribution of data in the training and testing datasets.

A. Training data distribution. **B.** Testing data distribution. The number of bins for the bar plots is 30. The curve in each plot is a gaussian kernel density estimate of the distribution. AL: axial length; CCT: central corneal thickness; AD: aqueous depth; ACD: anterior chamber depth; LT: lens thickness; K1: flat keratometry; K2: steep keratometry, AST: astigmatism; WTW: white-to-white; D: diopter

5.7.2 Supplementary tables

Formula	Our Method	SRK/T	Holladay 1	Haigis	Hoffer Q	Barrett	EVO
SRK/T	2.71E-18	/	/	/	/	/	/
Holladay 1	9.31E-14	1.00E+00	/	/	/	/	/

Haigis	9.51E-15	1.00E+00	1.00E+00	/	/	/	/
Hoffer Q	1.26E-23	7.74E-02	1.24E-08	2.60E-09	/	/	/
Barrett	1.08E-03	7.02E-14	3.00E-09	1.11E-08	1.99E-22	/	/
EVO	2.41E-02	1.27E-14	5.04E-10	1.57E-09	7.28E-20	1.00E+00	/
PearlDGS	1.20E-02	1.37E-07	5.71E-07	3.29E-07	3.70E-22	1.00E+00	1.00E+00

Table S 5.1 The p-values from the post-hoc paired Wilcoxon tests, following the Friedman test, for the comparison of the testing set performance between methods.

The p-values were adjusted with Bonferroni correction. The p-values between our method and the conventional methods were shown in the first column. All significant p-values (p-value < 0.05) were underscored. The Friedman test statistic was 235.65, and the associated p-value was 3.13e-47, which was statistically significant.

Comparison	**Within ±0.25 D**	**Within ±0.50 D**	**Within ±0.75 D**	**Within ±1.00 D**
Vs. Barrett	1.18e-1	4.16e-2	8.33e-2	1.24e-2
Vs. EVO	1.07e-1	6.66e-1	8.41e-1	5.22e-2
Vs. Haigis	2.06e-6	9.98e-6	4.83e-6	1.84e-2
Vs. Hoffer Q	3.70e-11	3.37e-11	2.25e-10	3.96e-5
Vs. Holladay 1	4.83e-7	8.12e-7	9.63e-8	2.56e-4
Vs. PearlDGS	5.30e-2	1.47e-2	4.74e-2	5.93e-1
Vs. SRK/T	1.89e-6	9.43e-8	7.98e-7	5.74e-6

Table S 5.2 The p-values from Cochran's Q test for the comparison of number cases within 0.25 D, 0.50 D, 0.75 D, and 1.00 D between our method and existing methods in the testing set.

All significant p-values (p-value < 0.05) were underscored. D: diopter.

Chapter 6 MAEPI and CIR: New Metrics for Robust Evaluation of the Prediction Performance of AI-Based IOL Formulas

6.1 Abstract

The purpose of this study is to develop a class of new metrics for evaluating the performance of intraocular lens power calculation formulas robust to issues that can arise with AI-based methods. The dataset consists surgical information and biometry measurements of 6893 eyes of 5016 cataract patients who received Alcon SN60WF lenses at University of Michigan's Kellogg Eye Center. We designed two types of new metrics: the MAEPI (Mean Absolute Error in Prediction of IOL) and the CIR (Correct IOL Rate) and compared the new metrics with traditional metrics including the mean absolute error (MAE), median absolute error (MedAE), and standard deviation (SD). We evaluated the new metrics with simulation analysis, machine learning (ML) methods as well as existing optics-based IOL formula (Barrett Universal II, Haigis, Hoffer Q, Holladay 1, and SRK/T). Results of traditional metrics did not accurately reflect the performance of overfitted ML formulas. By contrast, MAEPI and CIR discriminated between accurate and inaccurate formulas. The standard IOL formulas received low MAEPI and high CIR, which were consistent with the results of the traditional metrics. In conclusion, MAEPI and CIR provide a more accurate reflection of the real-life performance of AI-based IOL formula than traditional metrics. They should be computed in conjunction with conventional metrics when evaluating the performance of new and existing IOL formulas. The proposed new metrics would help cataract patients avoid the risks caused by inaccurate AI-based formulas, whose true performance cannot be determined by traditional metrics.

6.2 Introduction

The prediction performance of intraocular lens (IOL) formulas for cataract patients is usually evaluated with the following metrics: the mean prediction error (ME), the mean absolute error (MAE), median absolute error (MedAE) and standard deviation (SD) of the prediction error (PE), as recommended in multiple publications.[88–90] These are standard evaluation metrics commonly used for regression problems in which the target value is a scalar. The MAE summarizes the average distance between the prediction and the true value. The MedAE evaluates the median deviation and is less sensitive to outliers and extreme values. The standard deviation (SD) measures the extent of scattering of the PE. Aside from these standard metrics, ophthalmologists also calculate the percentage of PEs within a certain range (e.g., $\pm\, 0.25\ D$, $\pm\, 0.5\ D$), and the performance in different axial length (AL) groups (short, medium, and long). The former is a convenient way of investigating the distribution of PEs. The latter aids in determining whether a formula has consistent performance among myopic, hyperopic, and regular eyes. Recently, Hoffer et al[89] demonstrated a new evaluation metric, the IOL Formula Performance Index, which combines multiple metrics into one: (1) the SD (2) the MedAE (3) the AL bias, and (4) the percentage of eyes with PE within $\pm\, 0.5\ D$. Holladay et al reviewed IOL calculation evaluation metrics and recommended the standard deviation (SD) as the single best measurement because SD allows the use of heteroscedastic statistical methods and SD predicts the percentage of cases within a given interval, the mean absolute error, and the median of the absolute errors.[91] However, this conclusion was drawn based on the results of 11 optics-based IOL formulas (Barrett, Olsen, Haigis, Haigis WK, Holladay 1, Holladay 1 WK, Holladay 2, SRK/T, SRK/T WK, Hoffer Q, and Hoffer Q WK), which have been validated extensively with real-world datasets. For machine learning (ML) based formulas, the algorithm is oftentimes a

black box, of which the exact behavior is not known a priori. When evaluating or developing novel ML-based IOL formulas, it is important that the evaluation metric is appropriately selected and robust enough so that the trained model can be generalized to unseen data.

A special characteristic of cataract patient datasets is that the data is highly imbalanced, because the IOL powers were manually selected, rather than randomly drawn, and some powers were selected more often than the others. The postoperative refractions therefore represent a highly biased view of the expected outcome, assuming the IOL power is not specified. When trained on such imbalanced datasets, ML predictions are likely to be dominated by the over-represented domains, or in other words, the algorithm will tend to always predict the most common numbers (for regression) and labels (for classification).

For imbalanced regression problems, standard evaluation metrics (such as the MAE) are known to provide frequently misleading conclusions,[150,151] because they measure the average behavior of the most frequent cases, while rare cases may be of greater interest. In contrast to numerous publications focused on imbalanced classifications, little research has been conducted on the metrics for imbalanced regressions. As summarized in **Table *6.1***, previously proposed metrics for imbalanced regression problems include weighted errors, asymmetric loss functions,[152] precision-recall evaluation framework-based metrics,[153,154] Receiver Operating Characteristic (ROC) curves for regression,[155–157] and ranking-based evaluation.[158] Unfortunately, none of these metrics are targeting the underlying optimization goal within the context of IOL power prediction.

Choosing the correct scoring metric is a critical first step for developing or evaluating IOL formulas. Through our previous work developing machine learning-based IOL formulas (as described in Chapter 5) [159], we identified multiple weaknesses of conventional metrics that, to

our knowledge, have not been discussed in the literature. Traditional metrics can generate misleading information for formulas developed solely based on historical data, as is commonly the case with machine learning-based methods. In this work, we demonstrate a series of new IOL formula accuracy evaluation metrics, which should be used alongside traditional metrics when evaluating the performance of IOL formulas.

Method Type	Examples/Explanations
Weighted error	$err = \frac{\sum_{i=1}^{n} w_i \cdot L(y_i, \hat{y}_i)}{\sum_{i=1}^{n} w_i}$ where w_i is the weight of the i^{th} case, L is a loss function
Asymmetric loss functions	LINLIN asymmetric linear loss function[152]: $L(y_i, \hat{y}_i) = \begin{cases} a\lvert y_i - \hat{y}_i \rvert & if\ \hat{y}_i \geq y_i; \\ b\lvert y_i - \hat{y}_i \rvert & if\ \hat{y}_i < y_i; \end{cases}$ where a and b are constants, L means the loss function
Precision/recall evaluation framework, based on the concept of utility-based regression	The utility of a prediction model is defined as: $U = TB - TC$ where costs (TC) and benefits (TB) are defined based on a loss function and a relevance function.[153,154]
Adapt ROC curves to regression	Regression ROC space[157], Regression Error Characteristic (REC) curve[155], Regression Error Characteristic Surfaces (RECS)[156]
Ranking-based evaluation	As opposed to residual-based additive measures of errors (such as the MAE), ranking-based measures evaluate the model's performance in sorting y values and maintaining the ranking.[158]

Table 6.1 Metrics for imbalanced regression.
Notations: y_i is the true value of the i^{th} case; $\hat{y}_i$ is the predicted value of the i^{th} case; n is the total number of cases.

6.3 Methods

6.3.1 Data collection

We collected medical records of patients receiving care at the University of Michigan between August 25, 2015 and June 27, 2019. A total of 49 surgeons performed the surgeries included in the dataset. The patients were all measured preoperatively and postoperatively with

Lenstar LS 900 optical biometers (Haag-Streit USA Inc, EyeSuite software version i9.1.0.0) at University of Michigan's Kellogg Eye Center. Patient demographics (including patient age, gender, and ethnicity), the implanted IOL powers and the postoperative refraction were obtained from the Sight Outcomes Research Collaborative (SOURCE) Ophthalmology Data Repository. The data in SOURCE were described and used in various studies.[114,129,143–145] We included the manifest refractions measured by trained technicians employed by University of Michigan's Kellogg Eye Center at or closest to one month after the surgery. The postoperative refraction was computed with the following equation using an adjustment with regard to the lane length at Kellogg Eye Center (10 feet, 3.048 meters): $\text{spherical equivalent (SE) refraction} = (\text{spherical component} - 0.1614) + 0.5 \times \text{cylindrical component}$. The adjustment factor was determined according to Simpson and Charman's recommendation.[125]

Patients who received uneventful phacoemulsification cataract surgery (Current Procedural Terminology code = 66982 or 66984) and implantation of Alcon SN60WF one-piece acrylic monofocal lenses were included in this study. We excluded (1) patients who received previous refractive surgery or additional procedures during cataract surgery; (2) patients with postoperative visual acuity worse than 20/40; (3) records that were incomplete or out of bounds for any of the five IOL formulas analyzed in this study (Barrett Universal II, Haigis, Hoffer Q, Holladay 1 and SRK/T).

This research was conducted in compliance with the Institutional Review Board (IRB) at the University of Michigan. Informed consent was not applicable because this is a retrospective study, and all cases were fully anonymized. The study was carried out in accordance with the tenets of the Declaration of Helsinki.

6.3.2 Conventional metrics

Among a total of n eyes, for the i^{th} eye, we shall denote the actual postoperative refraction as y_i, and the predicted postoperative refraction with a given prediction method for the implanted IOL as $\hat{y}_i$. Further we define e_i as the refraction prediction error (PE) of the i^{th} eye, and the PE equals the actual refraction minus the predicted refraction: $e_i = y_i - \hat{y}_i$. The mean absolute error can then be calculated as follows:

$$refraction\ MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$$

The standard deviation of the PE can be calculated as:

$$SD = \sqrt{\frac{\sum_{i=1}^{n} |e_i - \bar{e}|^2}{n-1}},\ \bar{e} = \frac{\sum_{i=1}^{n} e_i}{n}$$

The Formula Performance Index (FPI) is calculated as demonstrated by Hoffer et al.[89]

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

Four key elements are involved in FPI: (1) the SD (2) the MedAE (3) the axial length (AL) bias m, calculated as the slope of the correlation between the prediction error and the AL. (4) the percentage of patients with predictions errors within $\pm$ 0.5 D, represented as n.

6.3.3 The MAEPI and CIR

In this section, we demonstrate the specific definitions of the MAEPI (Mean Absolute Error of the Prediction of the IOL) and CIR (Correct IOL Rate). In addition to the previously defined notations, for the i^{th} eye, we define its implanted IOL power as p_i, and the predicted IOL power as $\hat{p}_i$. For IOL power p (for example, $6\ D \leq p \leq 30\ D, step = 0.5\ D$), we defined the corresponding predicted postoperative refraction as $\hat{y}_i^p$. The predicted IOL power $\hat{p}_i$ is found by minimizing the absolute difference between the actual postoperative refraction y_i and the

predicted postoperative refraction $\hat{y}_i^p$ while altering the value of p. Therefore, the relationship between $\hat{p}_i$, and $\hat{y}_i^p$ can be defined with the following equation (**Table *6.2***).

$$\hat{p}_i = \underset{6 \le p \le 30, step=0.5}{\operatorname{argmin}} |y_i - \hat{y}_i^p|$$

As a comparison, the predicted postoperative refraction $\hat{y}_i$ is defined as:

$$\hat{y}_i = \hat{y}_i^{p=p_i}$$

Similar to the definition of refraction PE, we define the IOL power prediction error as follows:

$$IOL\ power\ PE = actual\ IOL\ power - predicted\ IOL\ power$$

The MAEPI is therefore defined as:

$$MAEPI = \frac{\sum_{i=1}^{n} |p_i - \hat{p}_i|}{n}$$

In addition to MAEPI, we define the CIR (Correct IOL Rate) as the proportion of predicted IOL powers $\hat{p}_i$ having a deviation within 0.0 D, $\pm$ 0.5 D, $\pm$ 1.0 D from the implanted IOL power p_i. Specific examples for calculating the MAEPI and CIR are shown in **Supplementary Material A.** Lower MAEPI and higher CIR mean better prediction performance.

$$CIR(0) = \frac{\sum_{i=1}^{n} I(|p_i - \hat{p}_i| = 0.0)}{n} \times 100\%$$

$$CIR(0.5) = \frac{\sum_{i=1}^{n} I(|p_i - \hat{p}_i| \le 0.5)}{n} \times 100\%$$

$$CIR(1) = \frac{\sum_{i=1}^{n} I(|p_i - \hat{p}_i| \le 1.0)}{n} \times 100\%$$

The above definitions of MAEPI and CIR assume that the formula predicts the postoperative refraction as the response variable and uses the given IOL power as an explanatory variable. If a predictive model is instead designed to predict the IOL power as the response

variable according to designated target refractions, $\hat{p}_i$ is simply the directly output of the model. The MAEPI and CIR can be calculated with the same equations as shown above.

Variable	True Value	Predicted Value
Postoperative refraction	y_i	$\hat{y}_i = \hat{y}_i^{p=p_i}$
Postoperative refraction when IOL power = 6 D, 6.5D, …., 30 D	unknown	$\hat{y}_i^{p=6}, \hat{y}_i^{p=6.5}, \ldots, \hat{y}_i^{p=30}$
Implanted IOL power	p_i	$\hat{p}_i = \underset{6 \le p \le 30, step=0.5}{\operatorname{argmin}} \lvert y_i - \hat{y}_i^p \rvert$

Table 6.2 A summary of the variables used in this study.

Variable i is an index of the cases, referring to the i^{th} eye among all cases.

6.3.4 Simulation analysis: MAEPI vs. refraction MAE

As shown in **Figure *6.1***, for a given eye and a given IOL power prediction formula, we use e ($e \in \mathbb{R}$) to represent the refraction PE of the implanted IOL. We denote by d the increment of the prediction value when the IOL power is increased by 0.5 D. Since the spherical equivalent refraction of the eye should decrease with increasing power of the IOL, we assume here that d is always negative.

In order to characterize the behavior of the conventional metrics and the new metrics, we performed data simulations under different conditions and restrictions. Scenario (1): The IOL power and predicted refraction have a linear relationship, meaning that $d < 0$ is a constant for all patients. Scenario (2): The IOL power and predicted refraction have a non-linear relationship, meaning that $d < 0$ is not a constant. Scenario (3): predictions are random.

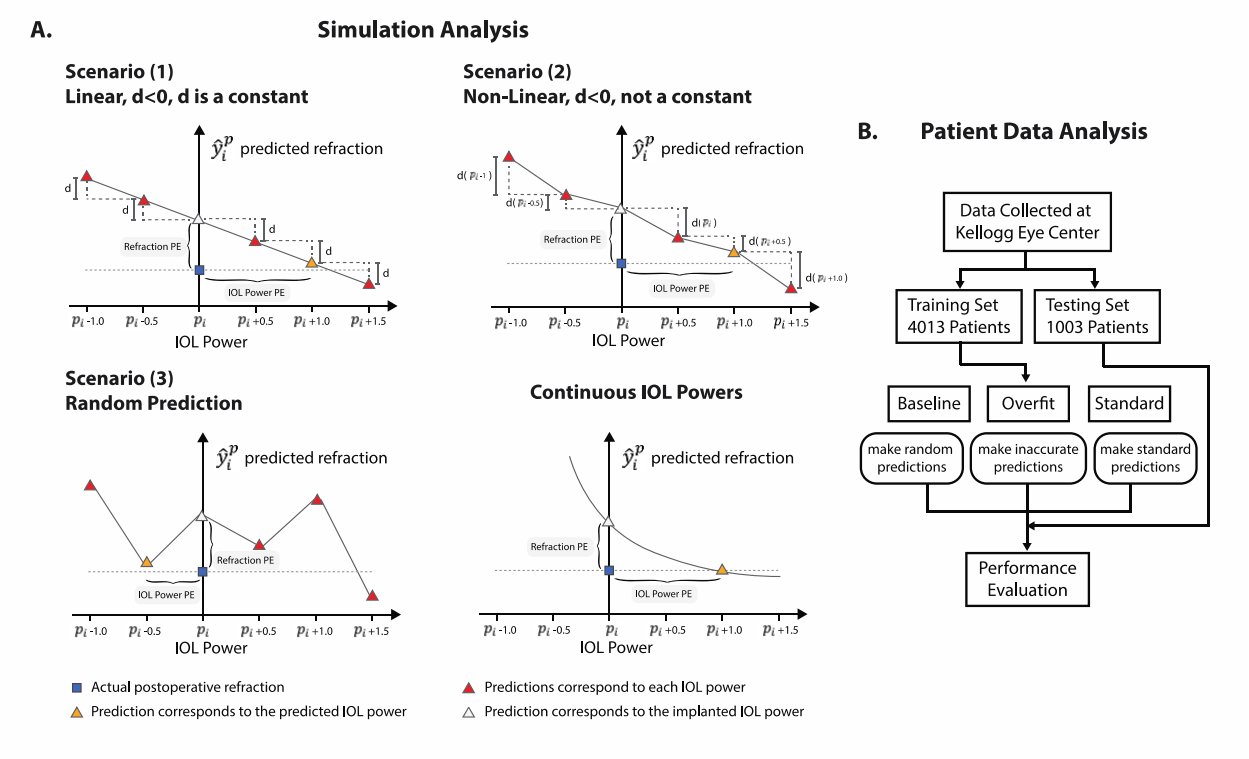

Figure 6.1 Schematics of three simulation scenarios and when the IOL powers are continuous numbers. PE: prediction error. Variable p_i is the implanted IOL power for the i^{th} eye among all cases. Variable $\hat{y}_i^p$ is the predicted postoperative refraction corresponds different IOL powers (p) for the i^{th} eye. Variable d is the increment of the predicted refraction.

In scenario (1), when d is a negative constant, it can be proven that the MAEPI and refraction MAE are always consistent, meaning that they always have a non-negative correlation. Consider a fictitious case A, wherein the PE of the implanted IOL is e_a ($e_a \in R$). We can represent the predicted IOL power as $(implanted\ IOL\ power + step\ of\ the\ IOL\ power * n_a)$, where $n_a \in \mathbb{Z}$. The difference between the actual postoperative refraction and the predicted refraction for the predicted IOL power and is then $e_{pa} = e_a - n_a d$. Similarly, for case B, the PE of the implanted IOL is e_b ($e_b \in R$), and the difference between the actual refraction and the

predicted refraction for the predicted IOL power is $e_{pb} = e_b - n_b d$, where $n_b \in \mathbb{Z}$. Based on the above relationships, we have

$e_a = e_{pa} + n_a d$, $e_b = e_{pb} + n_b d$

Accordingly,

$$|e_a| = |e_{pa} + n_a d|, |e_b| = |e_{pb} + n_b d|$$

Thus, if $|e_a| > |e_b|$, it follows that:

$$\left|\frac{e_a}{d}\right| > \left|\frac{e_b}{d}\right|$$

$$\left|\frac{e_{pa} + n_a d}{d}\right| > \left|\frac{e_{pb} + n_b d}{d}\right|$$

$$\left|\frac{e_{pa}}{d} + n_a\right| > \left|\frac{e_{pb}}{d} + n_b\right|$$

By a simple feature of absolute values, we have:

$$\left|\frac{e_{pa}}{d}\right| + |n_a| \geq \left|\frac{e_{pa}}{d} + n_a\right| > \left|\frac{e_{pb}}{d} + n_b\right| \geq |n_b| - \left|\frac{e_{pb}}{d}\right|$$

Since e_{pa} is, by definition, the error associated with optimal IOL power by the prediction method under consideration, and *d* is the difference in refraction prediction for each step in IOL power, it follows that

$$|e_{pa}| \leq \left|\frac{d}{2}\right|$$

and

$$\left|\frac{e_{pa}}{d}\right| < \frac{\left(\frac{d}{2}\right)}{d} = \frac{1}{2}$$

Similarly, the minimum value of $|n_b| - \left|\frac{e_{pb}}{d}\right|$ is $|n_b| - \frac{1}{2}$. Therefore, we have

$$\frac{1}{2} + |n_a| \geq \left|\frac{e_{pa}}{d}\right| + |n_a| \geq \left|\frac{e_{pa}}{d} + n_a\right| > \left|\frac{e_{pb}}{d} + n_b\right| \geq |n_b| - \left|\frac{e_{pb}}{d}\right| \geq |n_b| - 1/2$$

and thus

$$|n_a| > |n_b| - 1$$

Since $n_a, n_b \in \mathbb{Z}$, this implies

$$|n_a| \geq |n_b|$$

Hence, we see that $|e_a| > |e_b|$ implies $|n_a| \geq |n_b|$ for arbitrary eyes A and B and thus MAEPI and refraction MAE have a non-negative correlation under the aforementioned conditions.

In scenario (2), the MAEPI and refraction MAE are not always consistent. The corresponding counterexamples are shown in the **Supplementary Material B**. In conventional vergence formulas such as the Haigis, SRK/T, Holladay 1, and Hoffer Q formula, the postoperative refraction is represented as a reciprocal function of the IOL power:

$$refraction = A + \frac{B}{C + D * IOL\ power}$$

where A, B,C and D are functions of preoperative biometry measurements. The first derivative of this function is

$$\frac{d(refraction)}{d(IOL\ power)} = -\frac{DB}{(C + D * IOL\ power)^2}$$

which is not a constant because it changes with the IOL power. In the case of IOL power calculation, the first derivative is always negative, because the predicted postoperative refraction should always decrease with increasing IOL power. This above-described scenario fits the assumptions of scenario (2): $d < 0$ and d is not a constant.

In scenario (3), the predictions did not depend on the input data, but were completely random. We generated the predictions for each case and each IOL power randomly based on a uniform distribution, without assuming the values of e and d. This scenario helps to demonstrate

the general behavior of the MAEPI and refraction prediction MAE, with no assumptions on the characteristics of the IOL formula.

In the above analysis we assumed a step of 0.5 D for the IOL powers. It is obvious that the general behavior of the IOL metrics will not be affected when the IOL power has a different increment step (see also **Figure *6.1*A**). To facilitate comprehension, we have provided examples in **Supplementary Material C** for situations when the new metrics are consistent/inconsistent with the refraction metrics assuming continuous IOL powers.

Scenario	Number of Cases	Range of e	Value/Range of d	Range of Refraction Prediction
(1)	500	[-1.0, 1.0]	-0.5	/
(2)	500	[-1.0, 1.0]	[-1, -0.1]	/
(3)	500	/	/	[-5,5]

Table 6.3 The simulation parameters.

Variable e is the prediction error (PE) of the implanted IOL. Variable d is the signed increment of the refraction prediction when the IOL power increases by 0.5 D.

For the above-described three scenarios, we simulated the refraction predictions for different patients and IOL powers. The parameters used for the simulation are shown in **Table *6.3***. For scenario (1) and (2), the value of e was randomly generated for each simulated patient. A fixed value of d was used across all cases and all IOL powers for scenario (1). A random value of d was selected for each simulated patient and each IOL power for scenario (2). Schematics of each scenario are illustrated in **Figure *6.1*A**, and specific examples of simulated patients are shown in the **Supplementary Material D**.

6.3.5 Patient data analysis

In addition to the simulation analysis, we investigated the relationship between MAEPI and refraction MAE using real patient data collected at University of Michigan's Kellogg Eye

Center. We analyzed the performance of the following models on the dataset: (1) the baseline formula (2) the overfitted formulas (3) the standard formulas.

To build these models, we randomly separated the patients in the dataset into a training set (80%, 4013 patients, 5890 eyes) and a testing set (20%, 1003 patients, 1003 eyes) (**Figure *6.1*B**). One random eye was kept for patients with both eyes available in the testing set.

The baseline formula randomly samples from a normal distribution centered around the training dataset's mean postoperative refraction. This is done to simulate a method with poor prediction accuracy, but generates plausible results given the tendency to target postoperative refractions within a narrow range. Specifically, we generated the predictions from a normal distribution where the mean equaled the mean of the postoperative refraction in the training dataset, and the standard deviation was 0.01 (to simulate a tight grouping around the mean).

Overfitting is a term in machine learning which describes the situation when the algorithms are memorizing not only the underlying patterns but also the noise and biases present in the training dataset. With the overfitted formulas, we simulated prediction models which have appealing prediction accuracies for the implanted IOLs in the historical dataset, but fail to make accurate predictions for new unseen patients due to the lack of ability of making predictions for IOL powers that are not included in the historical dataset. Based on our experience, formulas that were developed solely or mostly based on historical datasets are especially vulnerable to overfitting. We then trained support vector machines (SVM) (implemented by scikit-learn 0.24.2[75]) and XGBoost[76] (a machine learning [ML] framework for gradient boosting tree-based algorithms) with the training set. SVM and XGBoost are both state-of-the-art machine learning frameworks, commonly used for a wide variety of applications, including ophthalmology. The preoperative biometry measurements and patient information used as

features are listed as follows: the SN60WF IOL power (D), patient gender, patient age at surgery (years), eye laterality, axial length (mm), central corneal thickness (μm), aqueous depth (mm), anterior chamber depth (mm), lens thickness (mm), flat and steep keratometry (D), astigmatism (D) and white-to-white (mm). The hyperparameters for the machine learning models were optimized based on five-fold cross-validation. The models were optimized by minimizing the cross-validation MAE. The values of the hyperparameters are shown in **Supplementary Material E**.

Formula	**Constant**	**Value**
Barrett	Lens factor	1.94
Haigis	a0, a1, a2	-0.739, 0.234, 0.217
Hoffer Q	Personalized ACD	5.727
Holladay 1	Surgeon factor	1.860
SRK/T	A constant	119.082

Table 6.4 The optimized formula constants.

For the standard formulas, we computed the refraction MAE, MedAE, SD, FPI, as well as the MAEPI and CIRs for five well-established IOL formulas: Barrett Universal II, Haigis, Hoffer Q, Holladay 1 and SRK/T. Friedman test followed by a post-hoc Wilcoxon signed-rank test with Bonferroni correction was used to compare the difference in the IOL power MAE (MAEPI) and refraction MAE. The predictions of Barrett Universal II were retrieved from the online calculator.[57] The Haigis, Hoffer Q, Holladay 1 and SRK/T formula were implemented in Python based on their publications.[52–56,59,127,128] The formula constants were optimized by zeroing out the mean prediction error in the training data (**Table *6.4***). We plotted and calculated the correlation between the IOL power prediction errors and the refraction prediction error for the above-mentioned formulas.

We generated partial dependence plots (PDP) for the above formulas to visualize the effect of the IOL powers on the predicted refractions. The PDP calculates the predicted refraction while altering the IOL powers and keeping all the other features unchanged. Barrett Universal II was removed from this analysis because of technical difficulty in the implementation using a web-based calculator.

In order to demonstrate that the step of the IOL powers will not influence the conclusions, we computed the IOL metrics assuming the IOL power step = 1.0 D and 0.01 D. The former represents a step larger than the common 0.5 D step, and the latter resembles a continuous variable. The Barrett Universal II calculator does not allow altering the IOL power step, therefore the Barrett formula was excluded from this analysis.

All statistical analyses were performed with Python 3.9.5. The criterion for statistical significance was p-value < 0.05.

6.4 Results

6.4.1 Simulation analysis results

We simulated three main situations to compare the properties of the IOL prediction errors and the refraction prediction errors. The corresponding results are shown in **Figure *6.2***. Consistent with the theoretical derivation in the Methods section above, **Figure *6.2*A** shows no overlap between the horizontal error intervals, which infers that when a case A has a lower absolute IOL power PE than case B, the refraction absolute error of case A can never be higher than that of case B, and vice versa. On the contrary, **Figure *6.2*B** and **Figure *6.2*C** show instances where a case A had a lower IOL absolute error than case B, but the refraction absolute error of case A was higher than that of case B, meaning the refraction accuracy and IOL accuracy suggest contradictory conclusions about case A and B. Increasing the number of

simulated cases beyond 500 or altering the value/range of e, d, or the refraction prediction in **Table *6.3*** did not change the general characteristics of the simulation results (results not shown).

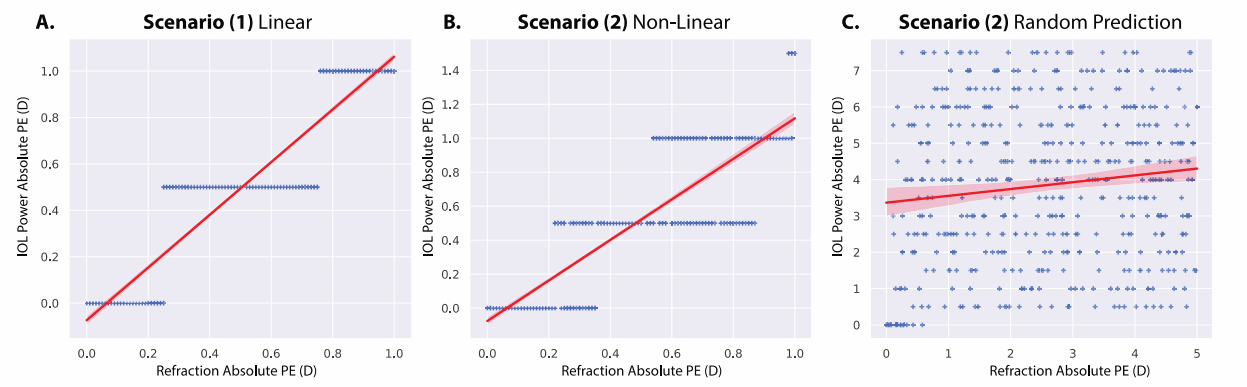

Figure 6.2 Scatter plots of simulation results under three conditions.

A. Refraction absolute PE vs. the scatter plot of IOL power absolute PE when the d < 0 and d is a constant. **B**. Refraction absolute PE vs. the scatter plot of IOL power absolute PE when the d < 0 and d is not a constant. **C.** Refraction absolute PE vs. the scatter plot of IOL power absolute PE when the predictions are random. Each dot in the scatterplot represents a simulated case. Variable d is the signed increment of the refraction prediction when the IOL power increases by 0.5 D. PE: prediction error.

6.4.2 Patient data analysis results

To investigate the behavior of the MAEPI and CIR in clinical data, we utilized the aforementioned dataset of 5016 patients (6893 eyes) at the Kellogg Eye Center and compared the performance of different methods using a testing subset of 1003 patients (1003 eyes). A summary of the dataset is shown in **Table *6.5***. The distribution of IOL powers and postoperative refractions in the training and testing dataset are shown in **Figure *6.3***. Distribution of other measurements are shown in **Figure S *6.1***. The prediction results of different methods are shown in **Table *6.6***.

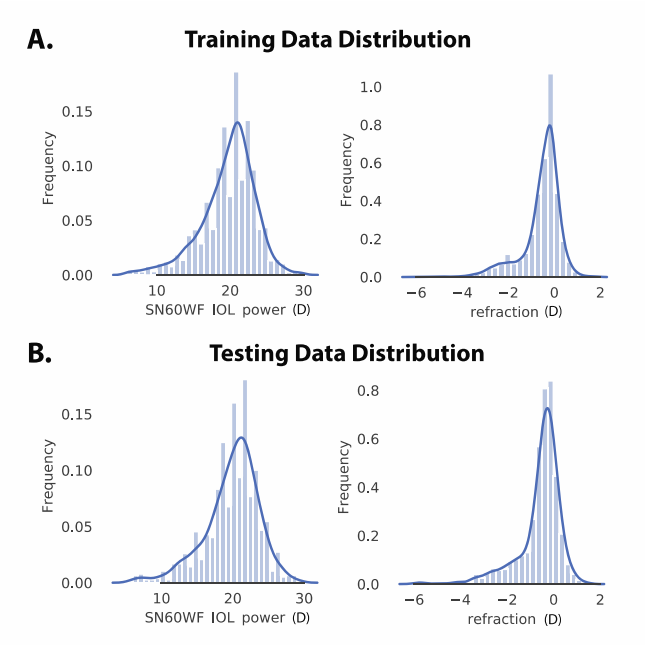

Figure 6.3 Distribution of the IOL power and postoperative refraction in the training and testing dataset. The number of bins for the bar plots was set to 30. The curve in each plot represents a gaussian kernel density estimate of the distribution. D: diopter.

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

Table 6.5 The distribution of the dataset.

AL: axial length; LT: lens thickness, ACD: anterior chamber depth; SD: standard deviation; D: diopter.

Index	Method	MAEPI (D)	CIR(0)	CIR(0.5)	CIR(1)	Ref. MAE (D)	Ref. MedAE (D)	Ref. ME (D)	Ref. SD (D)	AL bias	±0.5 D	FPI
(1)	**Random**	6.836	1.9%	6.3%	10.0%	0.628	0.393	-0.036	0.934	-0.302	55.7%	0.163
(2)	**SVM**	0.546	32.4%	76.5%	92.9%	0.329	0.245	0.023	0.455	-0.076	78.4%	0.366
	XGBoost	1.451	15.7%	43.9%	64.8%	0.368	0.286	0.016	0.488	-0.037	73.3%	0.398
(3)	**Barrett**	0.450	35.8%	81.3%	95.5%	0.328	0.256	0.0376	0.437	0.307	78.3%	0.198
	Haigis	0.497	31.3%	77.1%	94.1%	0.363	0.289	0.0237	0.469	0.226	74.7%	0.230
	Hoffer Q	0.591	24.6%	69.8%	90.7%	0.404	0.331	0.0091	0.517	0.951	70.3%	0.085
	Holladay	0.529	30.2%	74.4%	92.5%	0.371	0.298	0.0207	0.486	0.773	74.0%	0.101
	SRK/T	0.540	29.1%	73.5%	92.5%	0.376	0.300	0.0144	0.485	0.486	73.2%	0.143

Table 6.6 Performance of individual methods in the testing set.

MAEPI: mean absolute error in prediction of the IOL. CIR: correct IOL rate; MAE: mean absolute error; MedAE: median absolute error; ME: mean of prediction error; Ref: refraction; SD: standard deviation of the prediction error; AL: axial length; AL bias was calculated based on PE; FPI: formula performance index; D: diopter.

The formula that provided predictions by sampling from the aforementioned normal distribution centered on the training dataset mean yielded a high MAEPI (6.836 D) and extremely low IOL accuracies. On the contrary the refraction metrics were closer to normal range: the MAE was 0.628 D; the MedAE (0.393 D) was less than 0.5 D; the SD (0.934 D) was less than 1.0 D; the percentage of errors within ± 0.5 D was 55.7%; and the FPI (0.163) was higher than those of Hoffer Q, Holladay 1 and SRK/T.

The overfitted formulas (SVM and XGBoost) resulted in appealing refraction prediction performance, and poor IOL power prediction performance. The refraction MAE (0.329 D) and SD (0.455 D) of the SVM method were numerically close to those of the Barrett Universal II formula (0.328 D and 0.437 D). On the contrary, the MAEPI (0.546 D) of the SVM method was worse than that of SRK/T (0.540 D).

The Pearson correlation scores between the IOL prediction errors and the refraction prediction errors are shown in **Table *6.7***. The scatter plots of the refraction absolute errors of each eye and the IOL absolute errors of each eye for each method are shown in **Figure *6.4***. "Random" formula and the "Overfitted" formulas had lower correlation coefficients compared to those of the existing formulas.

In order to demonstrate that the same conclusions still hold when the step of IOL powers does not equal 0.5 D, we calculated the IOL power metrics assuming the step of IOL powers was 1 D and 0.01 D (**Table S *6.1***).

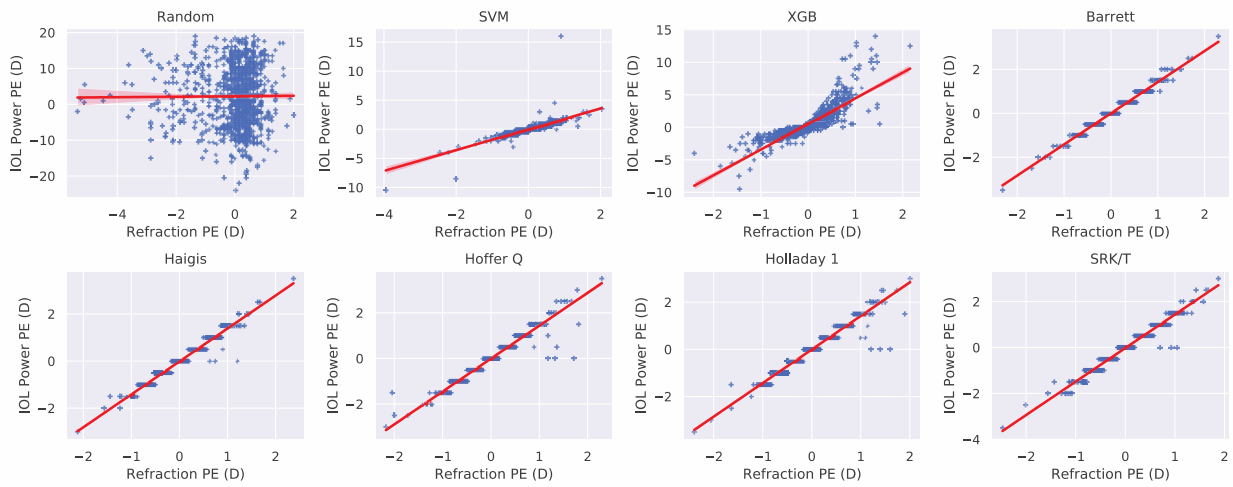

Figure 6.4 The scatter plots of the IOL power PE and the refraction PE for each method.

PE: prediction error.

Index	Method	Correlation between IOL power PE and Refraction PE	P-value
(1)	**Random**	-0.019	>0.05
(2)	**SVM**	0.827	<0.05
	XGBoost	0.846	<0.05
(3)	**Barrett**	0.973	<0.05
	Haigis	0.970	<0.05
	Hoffer Q	0.963	<0.05
	Holladay	0.962	<0.05
	SRK/T	0.972	<0.05

Table 6.7 The Pearson correlation coefficient and p-value between the IOL power prediction error and the refraction prediction PE.

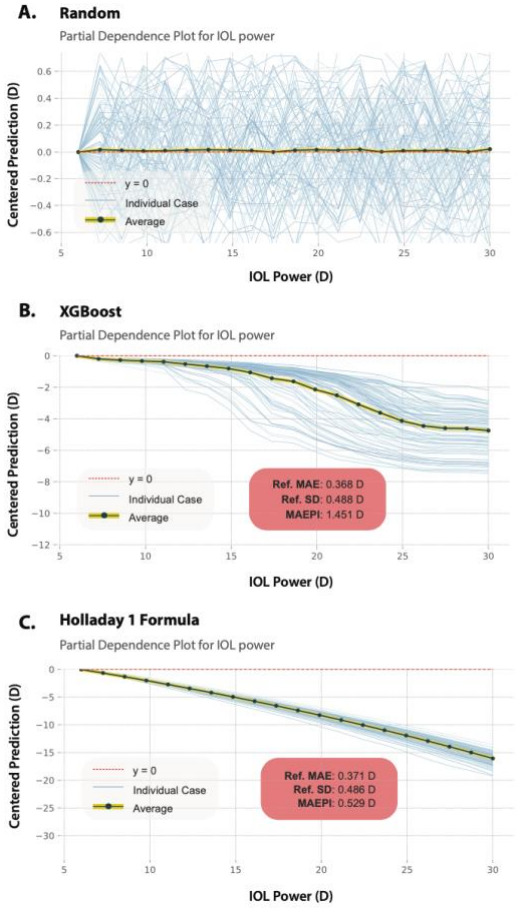

Figure 6.5 The partial dependence plots (PDP) for IOL power.

IOL powers range from 6 D to 30 D. The blue lines depict how the predicted refractions change with the IOL powers for individual patients. To facilitate comparisons and visualization, the heads of blue lines are centered at zero. The thick black line with yellow highlight represents the average of all blue lines. The refraction prediction MAE ("Ref. MAE"), the standard deviation of the refraction prediction error ("Ref. SD") and the IOL power prediction MAE (MAEPI) of the corresponding formulas are marked on the plots for the convenience of comparison (they are also listed in **Table 6.6**).

6.4.3 Analysis of overfitted formulas

The partial dependence plots of the "Random" formula, the XGBoost method, and the Holladay formula are shown in **Figure *6.5***. Partial dependence plots of other methods are shown in **Figure S *6.2***. We have selected the Holladay 1 formula for the comparison with the XGBoost method because the refraction ME, MAE, and SD of Holladay were similar to those of the XGBoost method, and yet the IOL power MAE (MAEPI) was substantially different (see **Table *6.6*** and **Figure *6.5***). The Wilcoxon test p-value for the refraction MAEs of Holladay and XGBoost was not significant ($p > 0.05$) (complete results shown in **Table S *6.2***). The Wilcoxon test p-value for the IOL power MAEs (MAEPIs) was significant ($p = 2.34e\text{-}16$) (complete results shown in **Table S *6.3***).

Prediction results of XGBoost and Holladay for a specific patient in the testing dataset are shown in **Table *6.8***. The actual implanted IOL power of this patient was 20.5 D, and the actual postoperative refraction was -0.0364 D. The refraction PE for XGBoost was $(-0.0346) - (-0.312) = 0.2774\ D$, and for Holladay 1 it was $(-0.0346) - (-0.347) = 0.3124\ D$. On the other hand, the IOL power PE for XGBoost can be calculated as $actual\ IOL\ power - predicted\ IOL\ power = 20.5 - 19 = 1.5\ D$, and for Holladay 1 it was $20.5 - 20 = 0.5\ D$.

Index	IOL Power (D)	Refraction Predicted by XGBoost (D)	Refraction Predicted by Holladay 1 (D)
1	18.5	0.047	0.945
2	19.0	-0.062	0.627
3	19.5	-0.124	0.306
4	20.0	-0.214	-0.019
5	20.5	-0.312	-0.347

Table 6.8 Prediction results of one patient from the testing set.

D: diopter.

6.5 Discussion

In this study, we identified potential problems with the traditional metrics for IOL formulas. These metrics focus on the deviation of a formula's predicted refraction for the implanted IOL power from the true postoperative refraction. We have presented here two new metrics, the Mean Absolute Error in Prediction of the IOL (MAEPI) and Correct IOL Rate (CIR) that are not susceptible to biases related to the clustering of real-world refraction targets in historical data.

Based on the simulation analysis results, only under special circumstances will the IOL metrics and the refraction metrics always be consistent and as such, IOL metrics (MAEPI and CIR) should be considered essential for the assessment and optimization of the performance of IOL formulas. As shown in the patient data analysis results, the conventional metrics can generate misleading information with ML prediction results. We believe this issue is the result of the following factors:

(1) The basic mechanism of ML models is to learn patterns from historical data, which sometimes makes them vulnerable to noise and biases in the training data, thus overfitting to the historical data if not properly trained. This property is in opposition to the theoretical optics-based methods. We trained machine learning models with the SVM algorithm and the XGBoost framework, and showed that both can be overfitted. It is possible to overcome overfitting through machine learning techniques such as data augmentation, resampling or by integrating theoretical components to the ML model. While of potential interest, a review of these methods is beyond the scope of this work.

(2) The IOL powers in the dataset were chosen through clinical decision making, and the postoperative refractions were influenced by the IOL powers. Both quantities have a unimodal

distribution (**Figure *6.3***), meaning that the distribution has one clear peak. Because of this reason, a calculator that simply predicts random values around the mean of historical data resulted in a reasonable MAE and SD of refraction prediction errors (**Table *6.6***, "Random" method). This method of fooling the refraction metrics can be easily "learned" by machine learning algorithms. As mentioned before, the dataset can overfit the model with meaningless information or information that is specific to the dataset. In the case of IOL prediction, the postoperative refractions are not randomly drawn from a distribution but are influenced by the IOL power that is manually selected by a surgeon based on discussions of refractive target with a patient (typically between -3.0 and 0.0 D). The fact that the postoperative refraction in historical datasets follows a certain distribution can be misleading for ML models. Practically, there is a tendency for ML algorithms to take advantage of such properties during training and inadvertently develop models that are not representative of the true system underlying the observations presented. This is a generalization issue caused by the mismatch between historical datasets and unseen patients.

(3) When evaluating formulas using a historical dataset, the conventional metrics assume known implanted IOL powers correspond to the postoperative refraction, which is not practical in real clinical settings. In contrast to that, the IOL metrics make no assumption about which IOL power was implanted. Suppose that the patient in **Table *6.8*** were an incoming patient participating in a clinical trial testing the performance of XGBoost and Holladay 1, with the target refraction = -0.0364 D, the surgeon using XGBoost would end up picking a 19.0 D lens, and the surgeon using Holladay 1 would pick a 20.0 D lens. It is likely that clinical trial results would agree with the MAEPI rather than the refraction MAE calculated from **Table *6.8***.

Therefore, the IOL metrics are more intuitive and are better representations of real-life performance of IOL formulas using modern empirical methods such as ML.

As mentioned in the Introduction, previous research has shown that standard metrics are not adequate for ML with imbalanced data, both in the case of regression and classification.[150,151,160] However, the existing evaluation frameworks for imbalanced regression does not provide an immediate solution for evaluating IOL formulas' performance. The Regression Error Characteristics (REC) curve, which plots the absolute deviation tolerance versus accuracy, can be viewed as an expanded version of proportions of patients with errors within different intervals.[155] The REC curve provides a description of the cumulative distribution function of the errors, however its limitation is also recognized in previous research.[156] The standard deviation of the prediction error was recently put forth as the "single best parameter" to characterize the performance of an IOL formula.[91] However, the paper investigated only geometric optics-based formulas that have been tested extensively over the years. From the point of view of a surgeon, a critical question is whether reported evaluation metrics are able to predict the real-life performance of a new formula. In this study, we have demonstrated that the SD, as well as the other traditional metrics such as MAE, MedAE and the percentage of patients in different error intervals, can be easily fooled by ML methods (**Table *6.6***). The Random, SVM and XGBoost model achieved abnormal high FPI (**Table *6.6***), which implies that the FPI is not a reliable evaluation metric.

Well-performing IOL formulas should have low MAEPI as well as low refraction MAE and SD. When comparing the performance of existing formulas, it is normal that a formula outperforms another in multiple metrics. However, this will not eliminate the necessity of taking each of these metrics into consideration, because such results are not guaranteed to be

generalizable to all relevant metrics. We believe the traditional metrics and the new IOL power metrics both describe important aspects of the performance of a formula, and therefore we recommend using both at the same time.

In this study, we assumed that the IOL formula predicts the postoperative refraction as a function of the IOL power. On the other hand, if the IOL formula is formulated as predicting the appropriate IOL power based on a specific refraction, then the MAEPI can simply be calculated as the mean absolute prediction error based on the standard equation. However, there is a persistent imbalance in the data, so the ML model will tend to predict common IOL powers (e.g., near 20 D) more frequently, regardless of how the refraction is specified. As an example, a model that outputs random IOL powers in a narrow interval around 20 D is capable of achieving a low MAEPI, which is similar to the results we obtained from the Random model in **Table *6.6***. Therefore, we recommend that in this scenario, both IOL power-based metrics and refraction-based metrics still need to be calculated to obtain a more accurate assessment of the formula's performance.

In our dataset the spherical component and cylindrical component had a step of 0.25 D, and the IOL powers had a step of 0.5 D. The presented analysis did not assume a fixed step for the measured refraction, therefore if the refraction had a different step or became continuous the results and conclusion would still be the same. We have also demonstrated in the Methods and Results sections that the conclusions drawn do not depend on the step of the IOL powers.

Despite the benefits of using the IOL accuracy metrics, we have identified the following limitations: (1) Calculating the IOL accuracy metrics may require programming knowledge and may consume more computing resources compared to the traditional metrics, because the predictions of multiple IOL powers have to be computed for each patient to determine the

predicted IOL power. This issue might be especially prominent for machine learning models since the scoring metrics may need to be computed repetitively during model selection. (2) The refraction MAE can be utilized as the loss function for optimization algorithms in machine learning. On the contrary, the MAEPI cannot be used directly as a loss function.

We are aware of potential confusion regarding the definition of the predicted IOL power. In the clinical setting, surgeons often choose an IOL power for which the predicted refraction is closest to the target refraction while remaining negative and/or more myopic than the target refraction. However, when calculating the MAEPI, the predicted IOL power should simply be the one of which the predicted refraction is closest to the true refraction. This is because the IOL power metrics are calculated according to historical datasets and the aim is to evaluate the accuracy of a formula. For instance, for a patient with a postoperative refraction of 2 D, the IOL power metrics should evaluate whether the IOL formula is able to pick the correct IOL power when aiming at a refraction of 2 D, instead of a refraction of -0.5 D.

In sum, we have demonstrated the potential pitfalls when using traditional metrics for performance evaluation of IOL formulas, including refraction MAE and SD. We recommend using the newly proposed IOL metrics, MAEPI and CIR, in addition to the conventional metrics for evaluating the performance of an IOL formula, especially for pure or largely ML-based formulas.

6.6 Publication

This chapter is under revision for publication at *Translational Vision Science & Technology*: Li, Tingyang, Joshua Stein, and Nambi Nallasamy. "MAEPI and CIR: new metrics for robust evaluation of the prediction performance of AI-based IOL formulas.".

6.7 Supplementary Materials

Supplementary Material A

Example calculation for MAEPI and CIR

Given the following dataset of two patients:

Patient ID	Implanted IOL Power (D)	Postop. Refraction (D)
1	20.0	-0.45
2	15.5	-1.25

For each patient, compute IOL predictions of all IOL powers (6 D to 30 D).

Patient ID	IOL Power (D)	Predicted Refraction (D)
1	6.0	5.28
1	6.5	4.89
1	…	…
1	19.0	-0.11
1	19.5	-0.40
1	20.0	-0.78
1	20.5	-1.21
1	…	…
1	30.0	-3.23

Patient ID	IOL Power (D)	Predicted Refraction (D)
2	6.0	5.28
2	6.5	4.89
2	…	…
2	13.5	-0.82
2	14.0	-1.20
2	14.5	-1.52
2	15.0	-1.89

2	15.5	-2.01
2	…	…
2	30.0	-3.23

Based on the above results for patient 1, when IOL power = 19.5 D, the predicted refraction (-0.40 D) is closest to the actual postoperative refraction (-0.45 D). Therefore, the predicted IOL power for patient 1 is 19.5 D. Similarly, the predicted IOL power for patient 2 is 14.0 D.

The refraction prediction error of patient 1 =-0.45-(-0.78) = 0.33 D

The refraction prediction error of patient 2 =-1.25-(-2.01) = 0.76 D

The refraction prediction MAE = (0.33+0.76)/2 = 0.545 D

The IOL power prediction error of patient 1 = 20.0 - 19.5 = 0.5D

The IOL power prediction error of patient 2 = 15.5 - 14.0 = 1.5D

The MAEPI = (0.5+1.5)/2 = 1.0

The CIR(0) = 0%

The CIR(0.5) = 50%

The CIR(1) = 50%

In practice it is not necessary to compute predictions for all 6 D to 30 D IOL powers. A more efficient algorithm can be designed to determine the IOL power that has a predicted refraction closest to the actual postoperative refraction.

Supplementary Material B

Case A

Implanted IOL power for case 1 = 20 D, postoperative refraction = 0 D

IOL Power (D)	19.5	20	20.5	21.0
Predicted Refraction (D)	0.6	0.2	-0.1	-0.3

For this case, the predicted IOL power = 20.5 D, absolute refraction PE = 0.2 D, absolute IOL power PE = 0.5 D

Case B: Example wherein the two metrics are not consistent when d < 0, d is not a constant

Implanted IOL power for case 2 = 20 D, postoperative refraction = 0 D.

IOL Power (D)	19.5	20	20.5	21.0
Predicted Refraction (D)	0.6	-0.3	-0.5	-0.8

For this case, the predicted IOL power = 20.0 D, absolute refraction PE = 0.3 D, absolute IOL power PE = 0.0 D.

The refraction absolute PE was higher than case A, but IOL absolute PE was lower than case A.

Case C: Example wherein the two metrics are consistent when d < 0, d is not a constant

Implanted IOL power for case 2 = 20 D, postoperative refraction = 0 D.

IOL Power (D)	19.5	20	20.5	21.0
Predicted Refraction (D)	0.5	0.0	-0.4	-0.7

For this case, absolute refraction PE = 0.0 D, absolute IOL power PE = 0.0 D.

The refraction absolute PE was lower than case A, the IOL absolute PE was also lower than case A.

Supplementary Material C

Case A (IOL power is continuous)

Implanted IOL power for case 1 = 20 D, postoperative refraction = 0 D

IOL Power (D)	18.32	20.00
Predicted Refraction (D)	0.0	-0.2

For this case, the predicted IOL power = 18.32 D, absolute refraction PE = 0.2 D, absolute IOL power PE = 1.68 D

Case B: Example wherein the two metrics are not consistent when d < 0, d is not a constant, and the IOL power is continuous

Implanted IOL power for case 2 = 20 D, postoperative refraction = 0 D.

IOL Power (D)	19.51	20
Predicted Refraction (D)	0.0	-0.3

For this case, the predicted IOL power = 19.51 D, absolute refraction PE = 0.3 D, absolute IOL power PE = 0.49 D.

The refraction absolute PE was higher than case A, but IOL absolute PE was lower than case A.

Case C: Example wherein the two metrics are consistent when d < 0, d is not a constant, and the IOL power is continuous

Implanted IOL power for case 2 = 20 D, postoperative refraction = 0 D

IOL Power (D)	19.51	20
Predicted Refraction (D)	0	-0.1

For this case, the predicted IOL power = 19.51 D, absolute refraction PE = 0.1 D, absolute IOL power PE = 0.49 D.

The refraction absolute PE was lower than case A, the IOL absolute PE was also lower than case A.

Supplementary Material D

Examples of simulated samples:

(1) Linear: d = -0.5, range of e is [-1,1]

Patient 1: The postoperative refraction = -0.5 D, implanted IOL power = 21 D, e = 0.8, d = -0.5

IOL Power (D)	…	20.0	20.5	21.0	21.5	…
Predicted Refraction (D)	…	-0.3	-0.8	-1.3	-1.8	…

Patient 2: The postoperative refraction = +0.7 D, implanted IOL power = 15 D, e = -0.3, d = -0.5

IOL Power (D)	…	14.0	14.5	15	15.5	…
Predicted Refraction (D)	…	2.0	1.5	1.0	0.5	…

(**2) None-linear**: range of d is [-1, -0.1], range of e is [-1,1]

Patient 1: The postop. refraction = -0.5 D, implanted IOL power = 21 D, e = 0.8, range of d is [-1, -0.1]

IOL Power (D)	…	20.0	20.5	21.0	21.5	…
Predicted Refraction (D)	…	-0.5	-1.0	-1.3	-1.5	…

Patient 2: The postop. refraction = +0.7 D, implanted IOL power = 15 D, e = -0.3, range of d is [-1, -0.1]

IOL Power (D)	…	14.0	14.5	15	15.5	…
Predicted Refraction (D)	…	2.1	1.7	1.0	0.6	…

(3) Random: range of refraction predictions is [-5,5]

Patient 1: The postop. refraction = -0.5 D, implanted IOL power = 21 D, range of refraction predictions is [-5,5]

IOL Power (D)	…	20.0	20.5	21.0	21.5	…
Predicted Refraction (D)	…	-1.5	1.0	-0.3	2.7	…

Patient 2: The postop. refraction = +0.7 D, implanted IOL power = 15 D, range of refraction predictions is [-5,5]

IOL Power (D)	…	14.0	14.5	15	15.5	…
Predicted Refraction (D)	…	4.3	-0.8	3.5	1.6	…

Supplementary Material E

Hyperparameters of the machine learning models.

Hyperparameters for SVM:

sklearn.svm.SVR(kernel='rbf',epsilon=0.4,C=100, gamma=0.0005)

Hyperparameters for XGBoost:

learning_rate = 0.05, n_estimators = 400, max_depth= 6, min_child_weight = 5, gamma = 0.5, subsample = 0.5, colsample_bytree = 0.8, scale_pos_weight = 0.5, reg_alpha = 1e-5, objective = 'reg:squarederror'

Index	Method	MAEPI (D)	CIR(0)	CIR(0.5)	CIR(1)	MAEPI (D)	CIR(0)	CIR(0.5)	CIR(1)
		IOL power step = 0.01 D				IOL power step = 1.0 D			
(1)	Random	6.845	0.0%	4.1%	8.1%	7.092	1.8%	6.7%	11.2%
(2)	SVM	0.559	0.0%	60.5%	86.7%	0.575	30.6%	74.7%	91.4%
	XGBoost	1.518	0.0%	32.2%	54.9%	1.454	16.6%	44.5%	64.2%
(3)	Haigis	0.514	0.0%	57.5%	87.3%	0.544	28.6%	72.9%	92.3%
	Hoffer Q	0.602	0.0%	50.1%	82.5%	0.624	25.0%	66.0%	88.3%
	Holladay	0.546	0.0%	55.2%	86.0%	0.574	27.2%	70.6%	91.1%
	SRK/T	0.557	0.0%	54.3%	85.7%	0.583	26.4%	69.6%	90.9%

Table S 6.1 The MAEPI and CIRs calculated assuming the step of the IOL powers equals 0.01 D and 1.0 D. Because the step of the IOL power was 0.5 D in the dataset, it is not technically meaningful to evaluate the percentage of predicted IOL powers that are identical to the ones in the dataset. Here we included CIR(0) only for the sake of comparison. MAEPI: mean absolute error in prediction of IOL; CIR: correct IOL rate; D: diopter.

	Barrett	Haigis	Hoffer Q	Holladay 1	SRK/T	Random	SVM
Haigis	1.00E+00	/	/	/	/	/	/
Hoffer Q	5.65E-01	1.00E+00	/	/	/	/	/

Holladay 1	**6.59E-04**	1.00E+00	**4.62E-02**	/	/	/	/
SRK/T	3.97E-01	1.00E+00	1.00E+00	1.00E+00	/	/	/
Random	**1.03E-13**	**3.39E-11**	**4.49E-10**	**3.10E-12**	**9.44E-14**	/	/
SVM	1.00E+00	1.00E+00	2.12E-01	**3.46E-02**	1.00E+00	**3.43E-11**	/
XGBoost	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	**3.70E-11**	1.00E+00

Table S 6.2 The post-hoc Wilcoxon test p-values (with Bonferroni correction) for the refraction absolute prediction error.

The Friedman chi-square test statistic, correcting for ties: 402.99. The associated p-value was 5.45e-83. Significant p-values are marked in bold with underscore.

	Barrett	**Haigis**	**Hoffer Q**	**Holladay 1**	**SRK/T**	**Random**	**SVM**
Haigis	5.32E-02	/	/	/	/	/	/
Hoffer Q	**2.21E-02**	1.00E+00	/	/	/	/	/
Holladay 1	1.64E-01	1.00E+00	1.00E+00	/	/	/	/
SRK/T	**6.02E-05**	1.00E+00	1.00E+00	2.17E-01	/	/	/
Random	**3.54E-14**	**2.23E-14**	**2.91E-14**	**2.58E-14**	**8.04E-15**	/	/
SVM	1.00E+00	1.00E+00	1.00E+00	1.00E+00	3.50E-01	**1.74E-14**	/
XGBoost	**1.35E-13**	**7.71E-16**	**3.06E-16**	**2.34E-16**	**4.04E-18**	**8.23E-09**	**1.84E-18**

Table S 6.3 The post-hoc Wilcoxon test p-values (with Bonferroni correction) for the IOL power absolute prediction error.

The Friedman chi-square test statistic, correcting for ties: 146.27. The associated p-value was 2.476e-28. Significant p-values are marked in bold with underscore.

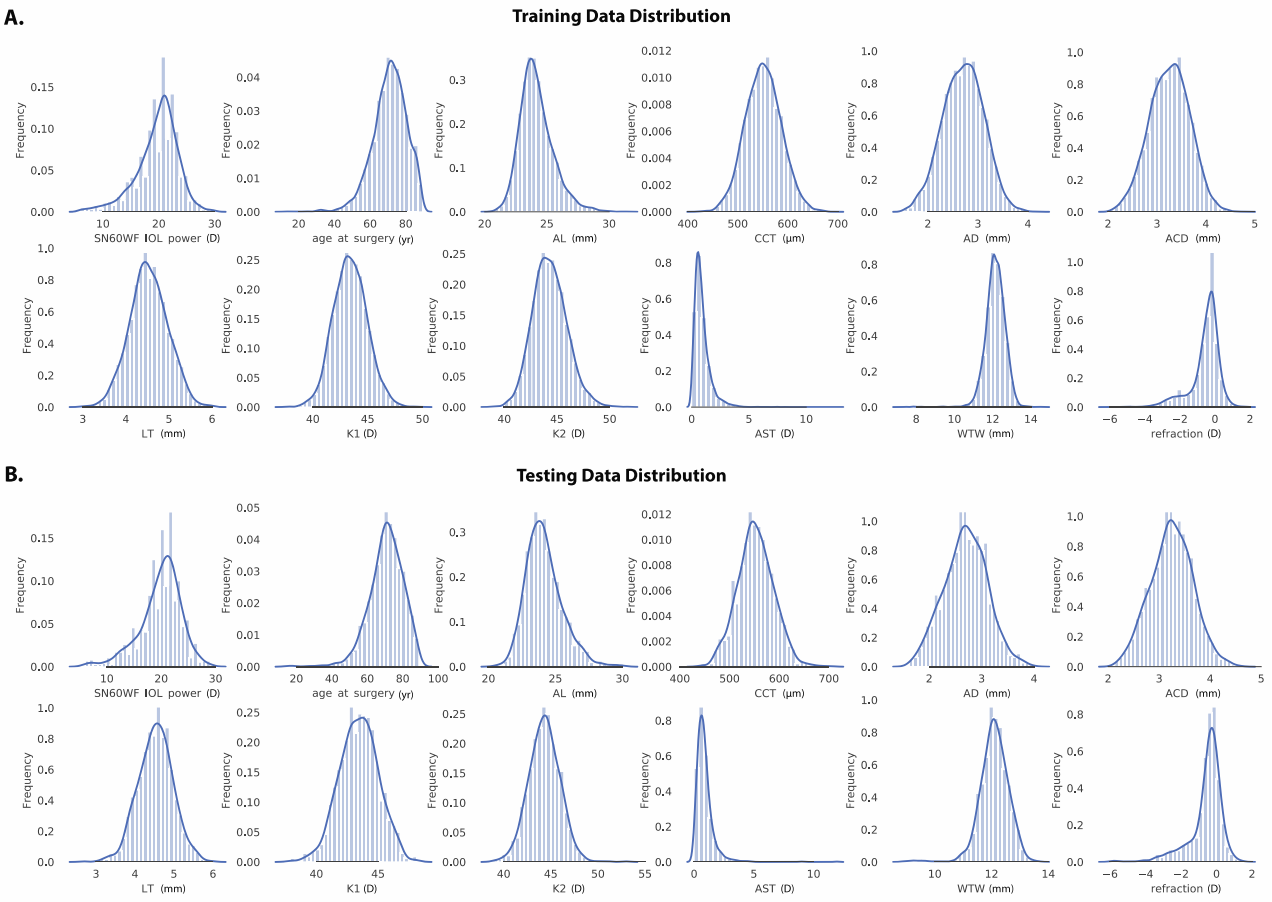

Figure S 6.1 Distribution of data in the training and testing datasets.

A. Training data distribution. **B.** Testing data distribution. The number of bins for the bar plots was set to 30. The curve in each plot represents a gaussian kernel density estimate of the distribution. AL: axial length; CCT: central corneal thickness; AD: aqueous depth; ACD: anterior chamber depth; LT: lens thickness; K1: flat keratometry; K2: steep keratometry, AST: astigmatism; WTW: white-to-white; D: diopter

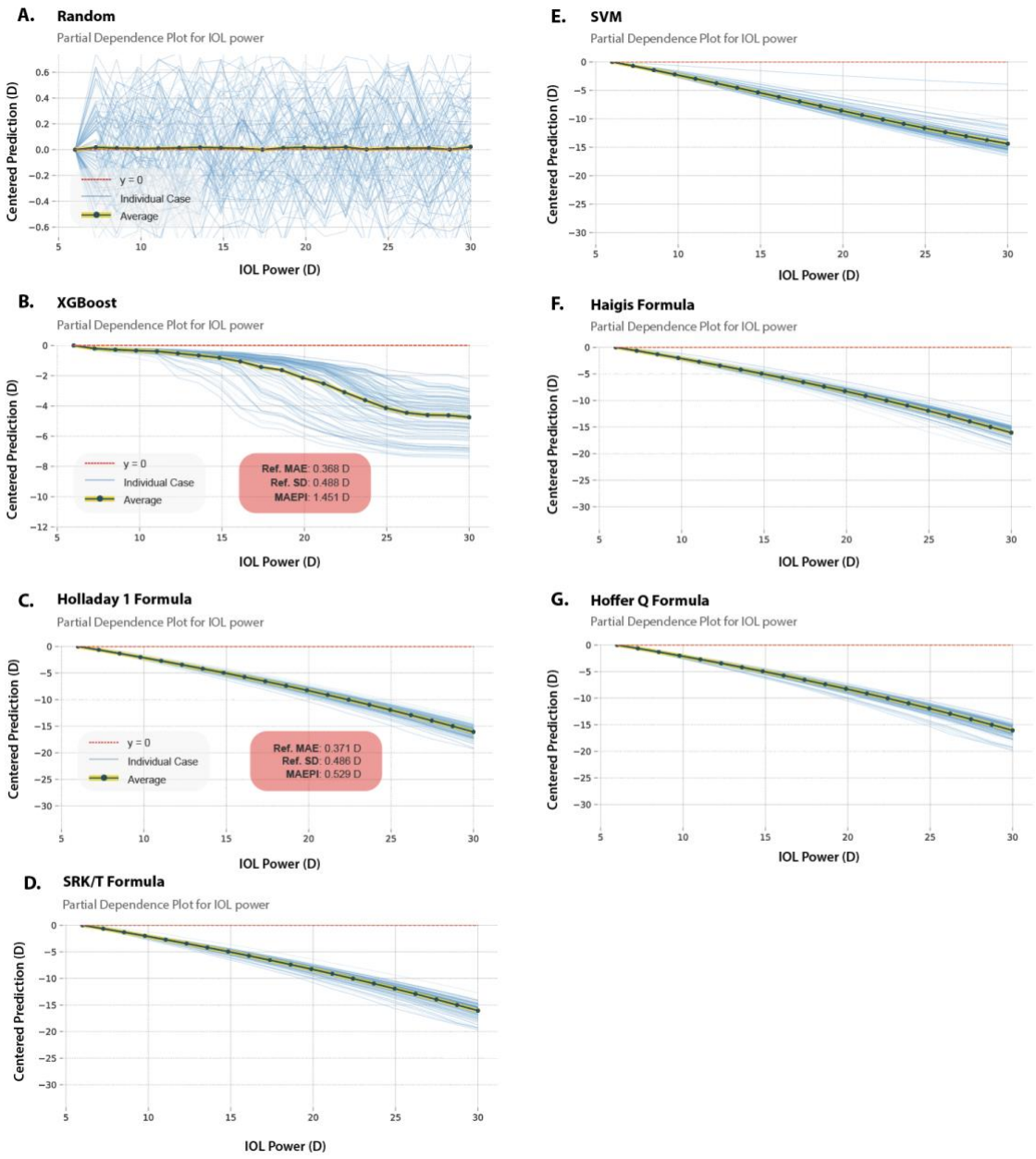

Figure S 6.2 The partial dependence plots (PDP) for IOL power.

IOL powers range from 6 D to 30 D. The blue lines depict how the predicted refractions change with the IOL powers for individual patients. The heads of blue lines are centered at zero for the sake of comparison and visualization. The thick black line with yellow highlight represents the average of all blue lines. The refraction prediction MAE ("Ref. MAE"), the standard deviation of the refraction prediction error ("Ref. SD") and the IOL power prediction MAE (MAEPI) of the corresponding formulas are marked on the plots for the convenience of comparison (they are also listed in **Table 6.6**).

Chapter 7 Conclusions and Future Directions

7.1 Summary of Findings

Beyond the surgeon's technique, accurate prediction of postoperative refraction lies at the heart of any successful cataract surgery. Although many IOL formulas have been developed in the past decades, the accuracy of IOL formulas still needs to be enhanced to address the ever-increasing expectations of cataract patients. In light of the large number of cataract surgeries performed every year, improving the accuracy of IOL power selection has become a crucial endeavor for both academic researchers and ophthalmologists. However, a complicated interplay of interdependent factors involved in IOL power selection made modeling the vision of the eye a rather challenging task.

Theoretical Gaussian optics-based methods have demonstrated good prediction abilities, despite their limitations in terms of model complexity. In contrast, supervised machine learning algorithms are able to circumvent the limitations of theoretical methods and learn complex patterns without relying on experts' knowledge, making it a viable option for future IOL formulas. The objective of this dissertation is describing a series of retrospective studies where we developed machine learning-based methods using cataract patients' data attained at the University of Michigan's Kellogg Eye Center to aid the decision-making process for cataract surgery.

In Chapter 2, using the preoperative biometry and postoperative ACD of 847 patients, we built a machine learning based method for postoperative lens position (ACD) prediction. Previous methods estimate the ELP as a function of AL, K, ACD and LT. In order to improve the accuracy of prediction, we utilized a gradient boosted decision tree-based algorithm in addition to adding more features (such as WTW and patient sex) to the algorithm. The prediction performance was significantly better than that of existing formulas Haigis, Hoffer Q, Holladay 1, Olsen, and SRK/T. The MAE and MedAE of the prediction slightly decreased when IOL power was added as an additional feature. Additionally, our algorithm maintained good performance when K (corneal power) was not available.

In Chapter 3, in order to investigate whether the ML-predicted postoperative ACD can improve the accuracy of existing optics-based IOL formulas for predicting postoperative refraction, we combined the ML model discussed in Chapter 2 with four existing formulas: Haigis, Hoffer Q, Holladay, and SRK/T. The performance of all formulas was significantly improved when formula-predicted ELPs were replaced with a linear combination of ML and formula-predicted ELPs.

In Chapter 4, we investigated into methodologies of combining ML-predicted ACD with ray tracing based IOL prediction formula OKULIX. OKULIX computes the postoperative refraction with exact calculations based on Snell's law but the postoperative lens position is not an exact calculation but a prediction.[161] Accordingly, we hypothesized that OKULIX should produce a more accurate prediction of postoperative refraction when the lens position is more accurately predicted. The results that we obtained corroborated our hypothesis: when combined with ML-predicted postoperative ACD, OKULIX achieved significantly better prediction

performance. Together with Chapter 3, we have shown that more accurate predictions of the postoperative lens position can significantly improve the performance of existing prediction methods.

In Chapter 5, we compiled a large dataset of 5016 cataract patients' preoperative biometry and postoperative manifest refraction, and utilized this dataset for developing the AI-based Nallasamy formula. The Nallasamy formula was developed using a stacking ensemble framework whereby the first-level models are combined using the second-level meta-algorithm. Furthermore, we devised novel feature engineering and data augmentation methods to integrate expert knowledge into the ML model for further improvement of prediction performance. Based on our results on a testing set of 1003 patients (1003 eyes), the Nallasamy formula significantly outperformed existing formulas including Barrett Universal II, EVO, Haigis, Hoffer Q, Holladay 1, PearlDGS, and SRK/T. The Nallasamy formula is freely available online at https://lenscalc.com/.

In Chapter 6, we proposed two metrics for robust evaluation of machine learning based IOL formulas. It is long-established that for imbalanced dataset, standard metrics (such as accuracy for classification tasks) may lead to misleading conclusions. However, the cataract patient dataset exhibits an extremely unbalanced distribution where the majority of IOL powers and the majority of postoperative refraction values are approximately centered around their average values. Unlike supervised classification tasks for which many metrics for imbalanced dataset have been proposed (such as the F-score), metrics for imbalanced regression and specifically for IOL formulas are underexplored. The MAEPI and CIR that we proposed in Chapter 6 calculate the accuracy in selecting the most appropriate IOL power based on the

predicted refraction of all IOL powers. Our results showed that MAEPI and CIR were capable of distinguishing overfitted ML models and providing unbiased evaluation scores.

7.2 Future Directions

7.2.1 Machine learning in cataract surgery

Our method and current machine learning-based formulas have demonstrated excellent performance and a promising future for AI-aided cataract surgery planning. In light of our experience and the opinions expressed in other publications, we summarize the advantages and disadvantages of machine learning methods as follows.

Machine learning has the following merits that make it a great candidate for building new IOL power selection methods: (1) The traditional vergence formulas only use a limited number of parameters (i.e., coefficients and constants). This is because these formulas were derived manually and with unavoidable simplifications. Comparatively, ML models present greater flexibility and scalability. Algorithms such as neural networks do not theoretically place a limit on the number of parameters that can be included in the prediction model. In this way, ML models are able to capture intervariable relationships with a higher level of complexity. (2) ML provides a more flexible framework that facilitates easy combination of different types of information (e.g., text and images), as well as different types of prediction models (e.g., integrating theoretical methods into machine learning-based methods). Such flexibility is especially important as more and more medical devices with different capabilities become available to patients. There will be difficulties leveraging data of other formats using theoretical optics-based methods. (3) There is a large degree of reliance on expert knowledge in theoretical

formulas. However, because of the eye's complicated structure, many aspects of the eye and the changes caused by cataract surgery cannot be readily represented mathematically. By contrast, ML can identify patterns purely based on the dataset, thus allowing it to uncover rules and patterns that were previously unknown to experts. (4) The ELP must be specifically predicted in theoretical formulas in order to be able to predict the postoperative refraction. Meanwhile, ML models can directly predict postoperative refraction using preoperative biometry. This feature might help ML-based methods avoid errors caused by ELP prediction.

In addition to the above advantages, machine learning has limitations compared to theoretical methods. We also suggest potential solutions to these limitations. (1) The performance of ML models is highly dependent on the size and quality of data. This issue is especially relevant for atypical eyes (such as eyes with extreme ALs). Many commonly-known ML-based IOL formulas were developed using large datasets: the RBF calculator project (for developing the Hill-RBF formula) involved 44 study sites in 20 countries; the Kane formula was built with ~30000 cases; the PearlDGS formula used more than 4000 cases; the Ladas Super Formula (LSF) AI claims that more than 4000 eyes were used to improve the original LSF. Establishing cross-institutional databases such as the SOURCE database may facilitate the gathering of data in future studies. Further, the use of data augmentation techniques such as interpolation may help to increase the dataset size and improve prediction generalizability. (2) Overfitting is an issue that can occur with any ML model. When overfitting happens, the ML model has high prediction accuracy with known cases, but poor prediction accuracy with unseen cases. ML learning models are particularly vulnerable to overfitting when the datasets are small, as is the case for many medical datasets. Since overcoming overfitting is a long-standing

endeavor, there have been many sophisticated techniques that have been developed and are widely used in modern ML model development.[162] (3) Difficulty in interpretation is another common issue with ML algorithms. Essentially this means that it is difficult to explain in human language how a machine learning algorithm makes a decision or predicts a certain number (with exceptions of algorithms such as decision trees). Although there are indirect ways to examine the underlying logic of a model, such as assessing feature importance and feature dependence, currently it is still impossible to accurately interpret ML models. Interdisciplinary researchers and computer science researchers are both actively working on these topics, with promising results constantly emerging. (4) Unlike theoretical methods, ML models do not have an optimizable lens-specific constant, therefore ML models are not directly generalizable to completely unseen lens types. A possible way to resolve this issue is by devising methods or equations that can be used to adjust predictions of IOL formulas for different lenses. A second solution is to retrain the ML-based formula with different datasets. This idea has been successfully implemented via transfer learning with the Nallasamy formula using the Aravind Eye Care System's dataset collected in southern India where patients received the Aravind-designed Auroflex IOLs. This study was presented at the 2022 American Society of Cataract and Refractive Surgery (ASCRS) annual meeting.

According to the above description, despite inherent limitations of current machine learning algorithms, we can anticipate that artificial intelligence will play a key role in the development of next-generation IOL formulas. In addition to general cataract patients, artificial intelligence also has a great potential for difficult cataract surgery cases where theoretical methods may not be the best choice. This may include patients with combined cataract surgery

and endothelial keratoplasty (EK) surgery, such as Descemet's Stripping Endothelial Keratoplasty (DSEK) or Descemet Membrane Endothelial Keratoplasty (DMEK). It may also include patients with corneal ectasia or prior penetrating keratoplasty. Future ML applications that are most relevant to this dissertation will be discussed in the following sections.

7.2.2 Machine learning predicted lens position combined with intraoperative aberrometry

Intraoperative aberrometry (IA) measures the power of the aphakic or pseudophakic eye during surgery. The Optiwave Refractive Analysis (ORA) system (Alcon, Fort Worth, TX, USA) uses a proprietary algorithm and IA measurements to predict postoperative refractions. Recent publications found mixed performance results when comparing ORA to Barrett True K No History (Barrett TKNH) among post-refractive surgery patients [163–165]. For regular cataract patients (with no refractive surgery history), IA displays no significant advantage compared to standard IOL formulas. [166] In practice, IA is typically used to validate the IOL power selection. A prominent source of error in the IA-based formula is the error in the estimation of the postoperative lens position. A study in this direction may investigate methodologies to combine IA-based methods with more accurate lens position prediction methods such as our ML-based method described in Chapter 2.

7.2.3 Evaluation metrics for IOL formulas

Previous discussions on the choice of evaluation metrics for IOL formulas (SD [91] and FPI [89]) have not considered the imbalances in the dataset. It was our publication that first discussed the limitations of standard regression metrics (such as MAE, MedAE and SD) resulting in false conclusions when ML-based formulas are overfitted. We have demonstrated in

Chapter 6 that MAEPI and CIR effectively distinguished overfitted, poorly performing ML models built with state-of-the-art ML algorithms. We anticipate more experimentation and validation of these two metrics MAEPI and CIR in the future.

In addition, we noticed that the data imbalance issues do not arise only in cataract patients, but rather occur in a wide range of medical datasets as well. For example, for diagnostic classification, it is expected that the dataset will have fewer positive cases than negative cases. As an example of imbalanced regression, consider an AI-powered grading system for medical education. If not enough poor grades are present in the training dataset, the ML algorithm may tend to give high grades when it encounters uncertain data. Standard evaluation metrics will not accurately reflect such errors because poor grades are rare in the dataset, and thus the value of the evaluation metric is mostly driven by the algorithm's performance in cases with high grades. Problems of imbalanced regression in other fields of medical research need to be addressed with the same rigor. Our recommendation is that the selection of evaluation metrics should depend on the application's goals. For example, a grading system may employ a ranking-based evaluation metric which evaluates the ability of the ML system to maintain the predicted value rank rather than calculating the absolute difference.

7.2.4 IOL power prediction in short and long eyes

When compared with existing formulas, the Nallasamy formula achieved a smaller variation in errors between eyes with different ALs. This result is in line with the performance of other new IOL formulas such as Kane and EVO, which were shown to have relatively small variations of errors among different ALs compared to older formulas (such as Hoffer Q and Holladay 1).[80,99] However, despite the trend of increasing accuracy, methods to reduce

prediction errors specifically for extreme ALs have been sparsely explored. Based on our experience and our review of relevant publications, we identified the following factors that may be the cause of errors in prediction for short and long axial lengths.

Inaccurate ELP (postoperative ACD) predictions could be one source of error, particularly in short eyes which require high IOL powers and thick IOLs. A high-power IOL is more susceptible to the ELP prediction error, while on the other extreme, the impact of ELP diminishes as the IOL power approaches zero. [52,167,168] The issue may not arise in ML-based formulas that do not predict ELP as an intermediate variable.

A second source of error is the keratometry measurements. Short eyes may have steeper corneas, and long eyes may have flatter corneas, both of which could affect the accuracy of the keratometry measurements of biometer.[167] Similar to the issues with post-refractive surgery eyes, the assumptions made by the biometer about the corneal index and the relationship between the anterior and posterior corneal surfaces may not be accurate for eyes of extreme AL. This issue could be mitigated through the use of devices that are able to measure the total corneal power based on both the anterior and posterior corneal surface. Many publications have proven that total K measurements (e.g., K values considering both the anterior and posterior corneal surface) are significantly different from standard K measurements (e.g., K values considering only the anterior corneal surface), and IOL power calculations resulted in statistically significant differences. [23,169–171] By analyzing the formula prediction errors using total K and standard K respectively, the influence of K measurements on the prediction accuracy of extreme eyes can be determined.

Another source of error is the scarcity of cases of extreme ALs as compared to cases with medium ALs. Previous sections have discussed the effect of data size on ML algorithms. A lack of data may also affect theoretical formulas, since the relationship between variables may be different for extreme ALs, in comparison with average ALs. Without enough data, it is difficult to accurately approximate such a complex nonlinear relationship for postoperative refraction or ELP prediction.

Most studies have focused on the prediction error variations related to ALs, but it has also been demonstrated that varied ACDs and LTs also contribute to error variations.[172] In reality, many biometric variables are correlated. For example, eyes with shorter ALs tend to have shorter ACD and thicker LT, and vice versa. The inclusion of patient demographic information may be one way to take into consideration such differences between patients.

7.2.5 IOL power prediction in post-refractive surgery eyes

Postoperative refraction prediction has always been difficult for patients who have previous refractive surgery such as laser-assisted in situ keratomileusis (LASIK), photorefractive keratectomy (PRK), or radial keratotomy (RK). These procedures alter the shape of the cornea, making the K measurement inaccurate. Formulas have been developed for these post-refractive eyes, such as the Barrett True-K formula and Haigis-L. Previous studies have compared the performance of standard formulas + total K (measured based on the anterior and posterior corneal surfaces) with that of formulas for post-refractive surgery eyes + standard K (measured only based on the anterior corneal surface). They found that the results differed based on the lens type and subgroup of patients.[173–175]

There is still room for improvement in the accuracy of the IOL power prediction for post-refractive surgery eyes, as more accurate measurement devices and more advanced algorithms become available. In future studies, ML-based methods include information of both anterior and posterior surface corneal measurements, as well as to devise IOL formulas using standard K for post-refractive surgery patients.

7.2.6 IOL power prediction of the second eye using data of the first eye

There has been considerable effort put into determining how to correct the second eye's IOL prediction based on the first operated eye's results in bilateral sequential cataract surgery. There are two main methods of adjusting the prediction of the second eye based on the errors of predicted postoperative refraction of the first operated eye [176]. One method is the formula-specific adjustment, where the second eye predictions are adjusted by formula-specific regression coefficients established based on the prediction errors of the first and second eye:

$$Rx_2^{New} = Rx_2^{Old} + \beta PE_1$$

$\mathrm{Rx}_2^{\mathrm{New}}$ is the corrected refraction prediction of the second eye; $\mathrm{Rx}_2^{\mathrm{Old}}$ is the uncorrected refraction prediction of the second eye; PE_1 is the refraction prediction error of the first eye; β is a formula-specific coefficient between 0 to 1, which is estimated based on a retrospective correlation analysis of prediction errors of the first eye and the second eye. Sometimes a constant value of 0.5 (50%) is used instead of β.

Another method is the patient-specific IOL constant adjustment, where the second eye predictions are adjusted by regression coefficients established based on the patients-specific IOL constants optimized based on the first eye and second eye.

The first method has been widely adopted in previous publications and was shown to be effective in reducing the prediction error of the second eye.[177–179]

Previous methods mostly investigated linear correlations between the results of the first and second eye. The methodology could be further improved by introducing non-linear modeling and sophisticated pattern recognition with ML algorithms.

7.3 Conclusions

Our ML-based postoperative lens position prediction method has achieved outstanding performance and we have further shown that the ML-predicted postoperative ACD can be used to improve the prediction performance of existing IOL formulas. The ML-based Nallasamy formula has also shown superior prediction performance compared to existing formulas based on our results. In the Future Directions section, we reviewed the advantages and limitations of machine learning applied to IOL power prediction and outlined potential future directions for further research.

In conclusion, as more data become available and advanced medical devices and technologies become available, the future holds great potential for AI-powered applications in cataract surgery and ophthalmology research.

Appendices

Appendix A: Derivation of the Thin Lens Formula

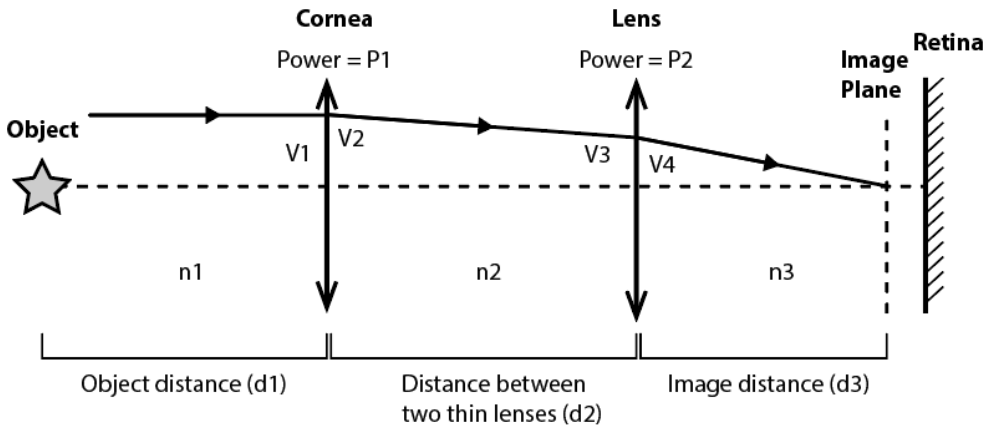

Figure A.1 Diagram of the thin lens assumption.

The thin lens schematic is shown in **Figure A.1** The cornea and the intraocular lens (IOL) are assumed to be thin lenses with zero thickness. The light ray only bends once when it passes through the thin lens, whereas in reality the lens has two refracting surfaces (the anterior surface and the posterior surface). The refractive index of air is $n1$ (which is usually set to $n1 = 1$). The refractive index of the aqueous humor is $n2$. The refractive index of the vitreous humor is $n3$. The power of the cornea is $P1$. The power of the IOL is $P2$. The distance between the object and the cornea is $d1$. The distance between the cornea and the IOL is $d2$. The distance between the IOL and the image of the object is $d3$. All quantities given are in SI units. The vergences can be then calculated with the following equations. $V1$ is the vergence of light just before entering the

cornea. $V2$ is the vergence of the light right after leaving the cornea. $V3$ is the vergence of light just before entering the IOL. $V4$ is the vergence of the light right after leaving the IOL.

$$V1 = \frac{n1}{d1}$$

$$V2 = V1 + P1 = \frac{n2}{d'}$$

$$V3 = \frac{n2}{d' - d2} = \frac{n2}{\frac{n2}{\frac{n1}{d1} + P1} - d2}$$

$$V4 = V3 + P2 = \frac{n2}{\frac{n2}{\frac{n1}{d1} + P1} - d2} + P2 = \frac{n3}{d3}$$

$$d3 = \frac{n3}{V4} = \frac{n3}{\frac{n2}{\frac{n2}{\frac{n1}{d1} + P1} - d2} + P2}$$

Here d' is an intermediate variable referring to a distance where the light coming out of the cornea will converge (media refractive index = $n2$).

To obtain emmetropia, the image distance $d3$ should equal the lens-to-retina distance $(AL - ACD)$ so that the image of the object falls precisely on the retina. Now rewrite the expression of $d3$ while letting $P1 = K$ (the keratometry measurement), $\frac{n1}{d1} \rightarrow 0$ (because $\frac{n1}{d1}$ is much smaller than $P1$), and $d2$ = postoperative ACD.

$$d3 = \frac{n3}{\frac{n2}{\frac{n2}{0 + K} - ACD} + P2} = AL - ACD$$

Solving the above equation gives the expression of $P2$, which is the power of the IOL to obtain emmetropia. Here $n2$, n3 are the refractive index of the aqueous and vitreous humor, respectively.

$$P2 = \frac{n3}{AL - ACD} - \frac{n2}{\frac{n2}{K} - ACD}$$

Here AL, K and ACD are postoperative AL, postoperative K and postoperative ACD respectively. The above expression represents a generic form for early generation thin lens formulas (e.g., the Thijssen formula). The vergence formulas approximate the postoperative AL and K with preoperative AL and K. The postoperative ACD is an estimated value. The aqueous humor refractive index $n2$ and vitreous humor refractive index $n3$ are usually set to $n2 = n3 = 1.336$.

Appendix B: Derivation of the Equations for Keratometry Measurements

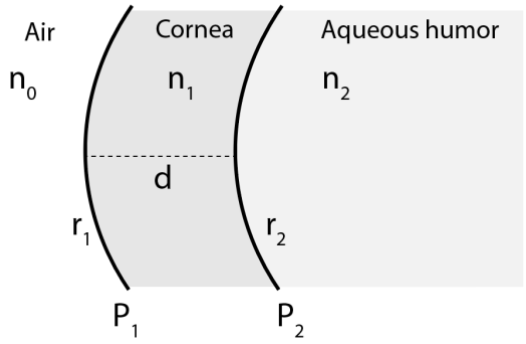

Figure B.1 Diagram of the anterior segment of the eye

The following derivation is based on the publication of Olsen.[40] Based on the schematic shown above in **Figure B.1**, the total power of the cornea (including the power of the anterior and posterior surface) can be given by the following equations. Here, n_0 is the refractive index of air; n_1 is the refractive index of the cornea; n_2 is the refractive index of the aqueous humor; d is the central corneal thickness; r_1 is the radius of curvature of the anterior corneal surface; r_2 is the radius of curvature of the posterior corneal surface; P_1 is the refractive power of the anterior corneal surface; P_2 is the refractive power of the posterior corneal surface. The unit of P_1, P_2, and P_{total} is diopter. The unit of r_1, r_2 and d is meter.

$$P_1 = \frac{n_1 - n_0}{r_1}$$

$$P_2 = \frac{n_2 - n_1}{r_2}$$

$$P_{total} = P_1 + P_2 - \frac{d}{n_1} P_1 P_2$$

Now set $n_0 = 1$, AP ratio $a = \frac{r_1}{r_2}$, $b = \frac{n_1}{n_2}$.

$$P_{total} = \frac{\left(\frac{a}{b} - a + 1\right) n_1 - 1}{r_1} - \frac{(\frac{1}{b} - 1)(n_1 - 1) a d n_1}{{r_1}^2}$$

Now make the following assumptions: $n_1 = 1.376$, $n_2 = 1.336$, AP ratio $a = 7.7/6.8$, $d = 0.5\ mm = 0.0005\ m$, $\frac{(\frac{1}{b}-1)(n_1-1)adn_1}{r_1} = -0.0008$ (let $r_1 = 7.7\ mm = 0.0077\ m$ for this part), then:

$$P_{total} = \frac{1.3315 - 1}{r_1}$$

Bibliography

1 Shichi H. Cataract formation and prevention. *Expert Opin Investig Drugs* 2004;**13**:691–701.

2 Ang MJ, Afshari NA. Cataract and systemic disease: A review. *Clin Experiment Ophthalmol* 2021;**49**:118–27.

3 Asbell PA, Dualan I, Mindel J, *et al.* Age-related cataract. *Lancet* 2005;**365**:599–609.

4 Cataract Tables | National Eye Institute. https://www.nei.nih.gov/learn-about-eye-health/outreach-campaigns-and-resources/eye-health-data-and-statistics/cataract-data-and-statistics/cataract-tables (accessed 5 Apr 2022).

5 Hashemi H, Pakzad R, Yekta A, *et al.* Global and regional prevalence of age-related cataract: a comprehensive systematic review and meta-analysis. *Eye* 2020;**34**:1357–70.

6 Ascaso FJ, Huerva V. *The history of cataract surgery*. chapter 2013.

7 Ascaso FJ, Lizana J, Cristóbal JA. Cataract surgery in ancient Egypt. *J Cataract Refract Surg* 2009;**35**.https://journals.lww.com/jcrs/Fulltext/2009/03000/Cataract_surgery_in_ancient_Egypt.43.aspx

8 Schuster AK, Schlichtenbrede FC, Harder BC, *et al.* Target Refraction for Best Uncorrected Distance and near Vision in Cataract Surgery. *Eur J Ophthalmol* 2013;**24**:509–15. doi:10.5301/ejo.5000414

9 Leffler CT, Klebanov A, Samara WA, *et al.* The history of cataract surgery: from couching to phacoemulsification. *Ann Transl Med* 2020;**8**:1551. doi:10.21037/atm-2019-rcs-04

10 Bilak S, Simsek A, Capkin M, *et al.* Biometric and intraocular pressure change after cataract surgery. *Optom Vis Sci Off Publ Am Acad Optom* 2015;**92**:464–70. doi:10.1097/OPX.0000000000000553

11 Khan AM, Waldner DM, Luong M, *et al.* Stabilization of refractive error and associated factors following small incision phacoemulsification cataract surgery. *BMC Ophthalmol*

2022;**22**:1–10.

12 Charlesworth E, Alderson AJ, de Juan V, *et al.* When is refraction stable following routine cataract surgery? A systematic review and meta-analysis. *Ophthalmic Physiol Opt* 2020;**40**:531–9.

13 Brunin G, Sajjad A, Kim EJ, *et al.* Secondary intraocular lens implantation: Complication rates, visual acuity, and refractive outcomes. *J Cataract Refract Surg* 2017;**43**:369–76. doi:https://doi.org/10.1016/j.jcrs.2016.12.024

14 Cruysberg LPJ, Doors M, Verbakel F, *et al.* Evaluation of the Lenstar LS 900 non-contact biometer. *Br J Ophthalmol* 2010;**94**:106–10. doi:10.1136/bjo.2009.161729

15 Garza-Leon M, Fuentes-de la Fuente HA, García-Treviño A V. Repeatability of ocular biometry with IOLMaster 700 in subjects with clear lens. *Int Ophthalmol* 2017;**37**:1133–8.

16 Bullimore MA, Slade S, Yoo P, *et al.* An Evaluation of the IOLMaster 700. *Eye Contact Lens* 2019;**45**:117–23. doi:10.1097/ICL.0000000000000552

17 MANUAL I. Lenstar LS 900®.

18 Song JS, Yoon DY, Hyon JY, *et al.* Comparison of Ocular Biometry and Refractive Outcomes Using IOL Master 500, IOL Master 700, and Lenstar LS900. *kjo* 2020;**34**:126–32. doi:10.3341/kjo.2019.0102

19 Ortiz A, Galvis V, Tello A, *et al.* Comparison of three optical biometers: IOLMaster 500, Lenstar LS 900 and Aladdin. *Int Ophthalmol* 2019;**39**:1809–18. doi:10.1007/s10792-018-1006-z

20 Kunert KS, Peter M, Blum M, *et al.* Repeatability and agreement in optical biometry of a new swept-source optical coherence tomography–based biometer versus partial coherence interferometry and optical low-coherence reflectometry. *J Cataract Refract Surg* 2016;**42**:76–83. doi:https://doi.org/10.1016/j.jcrs.2015.07.039

21 Shetty N, Kaweri L, Koshy A, *et al.* Repeatability of biometry measured by IOLMaster 700, Lenstar LS 900 and Anterion, and its impact on predicted intraocular lens power. *J Cataract Refract Surg* 2020.

22 Dong J, Tang M, Zhang Y, *et al.* Comparison of anterior segment biometric measurements between Pentacam HR and IOLMaster in normal and high myopic eyes. *PLoS One* 2015;**10**:e0143110.

23 Pereira JMM, Neves A, Alfaiate P, *et al.* Lenstar® LS 900 vs Pentacam®-AXL: Comparative study of ocular biometric measurements and intraocular lens power calculation. *Eur J Ophthalmol* 2018;**28**:645–51. doi:10.1177/1120672118771844

24 HELMY T, ALI H, SAMEH M, *et al.* Comparative Study between Pentacam and IOL Master in Measuring Anterior Segment Parameters in Eyes of Egyptian Individuals. *Med J Cairo Univ* 2019;**87**:4087–92.

25 Olsen T. Theoretical approach to intraocular lens calculation using Gaussian optics. *J Cataract Refract Surg* 1987;**13**:141–5. doi:10.1016/s0886-3350(87)80128-1

26 Norrby S, Bergman R, Hirnschall N, *et al.* Prediction of the true IOL position. *Br J Ophthalmol* 2017;**101**:1440–6. doi:10.1136/bjophthalmol-2016-309543

27 Schröder S, Langenbucher A. Relationship between effective lens position and axial position of a thick intraocular lens. *PLoS One* 2018;**13**:e0198824.

28 Norrby S. Sources of error in intraocular lens power calculation. *J Cataract Refract Surg* 2008;**34**:368–76. doi:10.1016/j.jcrs.2007.10.031

29 Sella R, Chou L, Schuster AK, *et al.* Accuracy of IOL power calculations in the very elderly. *Eye* 2020;**34**:1848–55. doi:10.1038/s41433-019-0752-0

30 Hayashi K, Ogawa S, Yoshida M, *et al.* Influence of Patient Age on Intraocular Lens Power Prediction Error. *Am J Ophthalmol* 2016;**170**:232–7. doi:10.1016/j.ajo.2016.08.016

31 Nuzzi G, Cantu C, De Giovanni MA. Older age as risk factor for deviation from emmetropia in pseudophakia. *Eur J Ophthalmol* 2001;**11**:133–8.

32 Reitblat O, Gali HE, Chou L, *et al.* Intraocular lens power calculation in the elderly population using the Kane formula in comparison with existing methods. *J Cataract Refract Surg* 2020;**46**:1501–7.

33 Shrestha S, Kaini KR, Basnet B. Gender differences in ocular biometry among cataract patients of western nepal. *Am J Public Heal Res* 2015;**3**:31–4.

34 Yoon JJ, Misra SL, McGhee CNJ, *et al.* Demographics and ocular biometric characteristics of patients undergoing cataract surgery in Auckland, New Zealand. *Clin Experiment Ophthalmol* 2016;**44**:106–13.

35 Nizamani NB, Surhio SA, Memon S, *et al.* Axial length variability in cataract surgery. *J Coll Physicians Surg Pak* 2014;**24**:918–21.

36 Wang D, Amoozgar B, Porco T, *et al.* Ethnic differences in lens parameters measured by

ocular biometry in a cataract surgery population. *PLoS One* 2017;**12**:e0179836.

37 Haigis W. Intraocular lens calculation after refractive surgery. *Journal-Intraocular Lens Calc After Refract Surg* 2012.

38 Hamilton DR, Hardten DR. Cataract surgery in patients with prior refractive surgery. *Curr Opin Ophthalmol* 2003;**14**:44–53.

39 Fam H-B, Lim K-L. Validity of the keratometric index: large population-based study. *J Cataract Refract Surg* 2007;**33**:686–91.

40 Olsen T. On the calculation of power from curvature of the cornea. *Br J Ophthalmol* 1986;**70**:152–4.

41 Patel R, Karp CL, Yoo SH, *et al.* Cataract surgery after refractive surgery. *Int Ophthalmol Clin* 2016;**56**:171.

42 Haigis W. Intraocular lens calculation after refractive surgery for myopia: Haigis-L formula. *J Cataract Refract Surg* 2008;**34**:1658–63. doi:https://doi.org/10.1016/j.jcrs.2008.06.029

43 Lawless M, Jiang JY, Hodge C, *et al.* Total keratometry in intraocular lens power calculations in eyes with previous laser refractive surgery. *Clin Experiment Ophthalmol* 2020;**48**:749–56.

44 Wang L, Koch DD. Intraocular Lens Power Calculations in Eyes with Previous Corneal Refractive Surgery: Review and Expert Opinion. *Ophthalmology* 2021;**128**:e121–31. doi:https://doi.org/10.1016/j.ophtha.2020.06.054

45 Binkhorst RD. The optical design of intraocular lens implants. *Ophthalmic Surg* 1975;**6**:17–31.

46 Colenbrander MC. Calculation of the power of an iris clip lens for distant vision. *Br J Ophthalmol* 1973;**57**:735.

47 Fyodorov SN, Galin MA, Linksz A. Calculation of the optical power of intraocular lenses. *Invest Ophthalmol Vis Sci* 1975;**14**:625–8.

48 Thijssen JM. The emmetropic and the iseikonic implant lens: computer calculation of the refractive power and its accuracy. *Ophthalmologica* 1975;**171**:467–86.

49 Sanders DR, Kraff MC. Improvement of intraocular lens power calculation using empirical data. *J Am Intraocul Implant Soc* 1980;**6**:263–7. doi:10.1016/s0146-2776(80)80075-9

50 Retzlaff J. Posterior chamber implant power calculation: regression formulas. *J Am Intraocul Implant Soc* 1980;**6**:268–70. doi:10.1016/s0146-2776(80)80076-0

51 Sanders DR, Retzlaff J, Kraff MC. Comparison of the SRK II formula and other second generation formulas. *J Cataract Refract Surg* 1988;**14**:136–41. doi:10.1016/s0886-3350(88)80087-7

52 Hoffer KJ. The Hoffer Q formula: A comparison of theoretic and regression formulas. *J Cataract Refract Surg* 1993;**19**:700–12. doi:https://doi.org/10.1016/S0886-3350(13)80338-0

53 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. doi:https://doi.org/10.1016/S0886-3350(13)80705-5

54 Erratum. *J Cataract Refract Surg* 1990;**16**:528. doi:https://doi.org/10.1016/S0886-3350(13)80820-6

55 Correction. *J Cataract Refract Surg* 1994;**20**:677. doi:https://doi.org/10.1016/S0886-3350(13)80677-3

56 Holladay JT, Musgrove KH, Prager TC, *et al.* A three-part system for refining intraocular lens power calculations. *J Cataract Refract Surg* 1988;**14**:17–24. doi:10.1016/S0886-3350(88)80059-2

57 Barrett Universal II Formula V1.05. https://calc.apacrs.org/barrett_universal2105/ (accessed 31 Aug 2021).

58 Olsen T, Corydon L, Gimbel H. Intraocular lens power calculation with an improved anterior chamber depth prediction algorithm. *J Cataract Refract Surg* 1995;**21**.https://journals.lww.com/jcrs/Fulltext/1995/05000/Intraocular_lens_power_calculation_with_an.22.aspx

59 Haigis W, Lege B, Miller N, *et al.* Comparison of immersion ultrasound biometry and partial coherence interferometry for intraocular lens calculation according to Haigis. *Graefe's Arch Clin Exp Ophthalmol* 2000;**238**:765–73. doi:10.1007/s004170000188

60 Holladay JT. Holladay IOL consultant computer program. 1996.

61 Sheard RM, Smith GT, Cooke DL. Improving the prediction accuracy of the SRK/T formula: the T2 formula. *J Cataract Refract Surg* 2010;**36**:1829–34. doi:10.1016/j.jcrs.2010.05.031

62 Ladas J, Ladas D, Lin SR, *et al.* Improvement of multiple generations of intraocular lens calculation formulae with a novel approach using artificial intelligence. *Transl Vis Sci Technol* 2021;**10**:7.

63 Ladas JG, Siddiqui AA, Devgan U, *et al.* A 3-D "super surface" combining modern intraocular lens formulas to generate a "super formula" and maximize accuracy. *JAMA Ophthalmol* 2015;**133**:1431–6.

64 IOLcalc - Ladas Super Formula. https://www.iolcalc.com/ (accessed 6 Apr 2022).

65 Preussner P-R, Wahl J, Lahdo H, *et al.* Ray tracing for intraocular lens calculation. *J Cataract Refract Surg* 2002;**28**:1412–9. doi:10.1016/s0886-3350(01)01346-3

66 Kane Formula. https://www.iolformula.com/ (accessed 31 Aug 2021).

67 IOL Power Calculator for Cataract Surgery | Hill-RBF Calculator. https://rbfcalculator.com/ (accessed 31 Aug 2021).

68 EVO Formula. https://www.evoiolcalculator.com/ (accessed 20 Jan 2022).

69 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. doi:10.1016/j.ajo.2021.05.004

70 Hipólito-Fernandes D, Luís ME, Gil P, *et al.* VRF-G, a New Intraocular Lens Power Calculation Formula: A 13-Formulas Comparison Study. *Clin Ophthalmol (Auckland, NZ)* 2020;**14**:4395.

71 CookeFormula.com. https://cookeformula.com/ (accessed 15 Apr 2022).

72 Chollet F, others. Keras. 2015.

73 Abadi M, Agarwal A, Barham P, *et al.* Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv Prepr arXiv160304467* 2016.

74 Paszke A, Gross S, Massa F, *et al.* Pytorch: An imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 2019;**32**.

75 Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;**12**:2825–30.

76 Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: : Association for Computing Machinery 2016. 785–794. doi:10.1145/2939672.2939785

77 Ke G, Meng Q, Finley T, *et al.* Lightgbm: A highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 2017;**30**.

78 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. doi:10.1097/IIO.0000000000000298

79 Balyen L, Peto T. Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology. *Asia-Pacific J Ophthalmol* 2019;**8**:264–72.

80 Pereira A, Popovic MM, Ahmed Y, *et al.* A comparative analysis of 12 intraocular lens power formulas. *Int Ophthalmol* Published Online First: July 2021. doi:10.1007/s10792-021-01966-z

81 Langenbucher A, Szentmáry N, Cayless A, *et al.* Prediction of the axial lens position after cataract surgery using deep learning algorithms and multilinear regression. *Acta Ophthalmol* 2022.

82 Wu Y, Zhang S, Zhong Y, *et al.* Prediction of effective Lens position using anterior segment optical coherence tomography in Chinese subjects with angle closure. *BMC Ophthalmol* 2021;**21**:454. doi:10.1186/s12886-021-02213-w

83 Sramka M, Slovak M, Tuckova J, *et al.* Improving clinical refractive results of cataract surgery by machine learning. *PeerJ* 2019;**7**:e7202–e7202. doi:10.7717/peerj.7202

84 Clarke GP, Kapelner A. The Bayesian Additive Regression Trees Formula for Safe Machine Learning-Based Intraocular Lens Predictions. *Front Big Data* 2020;**3**:46. doi:10.3389/fdata.2020.572134

85 Carmona González D, Palomino Bautista C. Accuracy of a new intraocular lens power calculation method based on artificial intelligence. *Eye (Lond)* 2021;**35**:517–22. doi:10.1038/s41433-020-0883-3

86 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

87 Moutari S, Moore JE. An ensemble-based approach for estimating personalized intraocular lens power. *Sci Rep* 2021;**11**:22961. doi:10.1038/s41598-021-02288-x

88 Hoffer KJ, Aramberri J, Haigis W, *et al.* Protocols for studies of intraocular lens formula accuracy. Am. J. Ophthalmol. 2015;**160**:403-405.e1. doi:10.1016/j.ajo.2015.05.029

89 Hoffer KJ, Savini G. Update on Intraocular Lens Power Calculation Study Protocols: The Better Way to Design and Report Clinical Trials. *Ophthalmology* Published Online First: July 2020. doi:10.1016/j.ophtha.2020.07.005

90 Wang L, Koch DD, Hill W, *et al.* Pursuing perfection in intraocular lens calculations: III. Criteria for analyzing outcomes. *J Cataract Refract Surg* 2017;**43**:999–1002. doi:10.1016/j.jcrs.2017.08.003

91 Holladay JT, Wilcox RR, Koch DD, *et al.* Review and recommendations for univariate statistical analysis of spherical equivalent prediction error for IOL power calculations. *J Cataract Refract Surg* 2021;**47**:65–77. doi:10.1097/j.jcrs.0000000000000370

92 Xia T, Martinez CE, Tsai LM. Update on intraocular lens formulas and calculations. *Asia-Pacific J Ophthalmol (Philadelphia, Pa)* 2020;**9**:186.

93 Sheard R. Optimising biometry for best outcomes in cataract surgery. *Eye* 2014;**28**:118–25.

94 ULIB. http://ocusoft.de/ulib/ (accessed 6 Apr 2022).

95 Melles RB, Holladay JT, Chang WJ. Accuracy of Intraocular Lens Calculation Formulas. *Ophthalmology* 2018;**125**:169–78. doi:10.1016/J.OPHTHA.2017.08.027

96 Connell BJ, Kane JX. Comparison of the Kane formula with existing formulas for intraocular lens power selection. *BMJ open Ophthalmol* 2019;**4**:e000251. doi:10.1136/bmjophth-2018-000251

97 Kane JX, Van Heerden A, Atik A, *et al.* Intraocular lens power formula accuracy: Comparison of 7 formulas. *J Cataract Refract Surg* 2016;**42**.https://journals.lww.com/jcrs/Fulltext/2016/10000/Intraocular_lens_power_formula_accuracy_.15.aspx

98 Darcy K, Gunn D, Tavassoli S, *et al.* Assessment of the accuracy of new and updated intraocular lens power calculation formulas in 10 930 eyes from the UK National Health Service. *J Cataract Refract Surg* 2020;**46**:2–7. doi:10.1016/j.jcrs.2019.08.014

99 Carmona-González D, Castillo-Gómez A, Palomino-Bautista C, *et al.* Comparison of the accuracy of 11 intraocular lens power calculation formulas. *Eur J Ophthalmol* 2020;**31**:2370–6. doi:10.1177/1120672120962030

100 Cheng H, Kane JX, Liu L, *et al.* Refractive Predictability Using the IOLMaster 700 and Artificial Intelligence–Based IOL Power Formulas Compared to Standard Formulas. *J*

Refract Surg 2020;**36**:466–72.

101 Rocha-de-Lossada C, Colmenero-Reina E, Flikier D, *et al.* Intraocular lens power calculation formula accuracy: Comparison of 12 formulas for a trifocal hydrophilic intraocular lens. *Eur J Ophthalmol* 2020;**31**:2981–8. doi:10.1177/1120672120980690

102 Savini G, Hoffer KJ, Balducci N, *et al.* Comparison of formula accuracy for intraocular lens power calculation based on measurements by a swept-source optical coherence tomography optical biometer. *J Cataract Refract Surg* 2020;**46**.https://journals.lww.com/jcrs/Fulltext/2020/01000/Comparison_of_formula_accuracy_for_intraocular.6.aspx

103 Nemeth G, Kemeny-Beke A, Modis L. Comparison of accuracy of different intraocular lens power calculation methods using artificial intelligence. *Eur J Ophthalmol* 2021;**32**:235–41. doi:10.1177/1120672121994720

104 Ryu S, Jun I, Kim T-I, *et al.* Accuracy of the Kane Formula for Intraocular Lens Power Calculation in Comparison with Existing Formulas: A Retrospective Review. *Yonsei Med J* 2021;**62**:1117–24. doi:10.3349/ymj.2021.62.12.1117

105 Savini G, Di Maita M, Hoffer KJ, *et al.* Comparison of 13 formulas for IOL power calculation with measurements from partial coherence interferometry. *Br J Ophthalmol* 2021;**105**:484–9. doi:10.1136/bjophthalmol-2020-316193

106 Holladay JT. Refractive power calculations for intraocular lenses in the phakic eye. *Am J Ophthalmol* 1993;**116**:63–6. doi:10.1016/s0002-9394(14)71745-3

107 Kriechbaum K, Findl O, Preussner PR, *et al.* Determining postoperative anterior chamber depth. *J Cataract Refract Surg* 2003;**29**:2122–6. doi:10.1016/s0886-3350(03)00414-0

108 Plat J, Hoa D, Mura F, *et al.* Clinical and biometric determinants of actual lens position after cataract surgery. *J Cataract Refract Surg* 2017;**43**:195–200. doi:10.1016/j.jcrs.2016.11.043

109 Goto S, Maeda N, Koh S, *et al.* Prediction of Postoperative Intraocular Lens Position with Angle-to-Angle Depth Using Anterior Segment Optical Coherence Tomography. *Ophthalmology* 2016;**123**:2474–80. doi:10.1016/j.ophtha.2016.09.005

110 Tamaoki A, Kojima T, Tanaka Y, *et al.* Prediction of effective lens position using multiobjective evolutionary algorithm. *Transl Vis Sci Technol* 2019;**8**. doi:10.1167/tvst.8.3.64

111 Behndig A, Montan P, Lundström M, *et al.* Gender differences in biometry prediction error and intra-ocular lens power calculation formula. *Acta Ophthalmol* 2014;**92**:759–63. doi:10.1111/aos.12475

112 Hoffer KJ, Savini G. Effect of Gender and Race on Ocular Biometry. *Int Ophthalmol Clin* 2017;**57**:137–42. doi:10.1097/IIO.0000000000000180

113 Dooley I, Charalampidou S, Nolan J, *et al.* Estimation of effective lens position using a method independent of preoperative keratometry readings. *J Cataract Refract Surg* 2011;**37**:506–12. doi:10.1016/j.jcrs.2010.09.027

114 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. doi:10.1167/tvst.9.13.38

115 Martinez-Enriquez E, Pérez-Merino P, Durán-Poveda S, *et al.* Estimation of intraocular lens position from full crystalline lens geometry: towards a new generation of intraocular lens power calculation formulas. *Sci Rep* 2018;**8**:9829. doi:10.1038/s41598-018-28272-6

116 Hirnschall N, Amir-Asgari S, Maedel S, *et al.* Predicting the postoperative intraocular lens position using continuous intraoperative optical coherence tomography measurements. *Invest Ophthalmol Vis Sci* 2013;**54**:5196–203. doi:10.1167/iovs.13-11991

117 Chang Y-C, Cabot F, Williams S, *et al.* Pre-operative Prediction of Post-cataract Surgery IOL Position Using Anterior Chamber Depth and Lens Thickness Determined with Extended-depth OCT. *Invest Ophthalmol Vis Sci* 2017;**58**:2717.

118 Satou T, Shimizu K, Tsunehiro S, *et al.* Development of a new intraocular lens power calculation method based on lens position estimated with optical coherence tomography. *Sci Rep* 2020;**10**:6501. doi:10.1038/s41598-020-63546-y

119 Eom Y, Song JS, Kim HM. Modified Haigis Formula Effective Lens Position Equation for Ciliary Sulcus-Implanted Intraocular Lenses. *Am J Ophthalmol* 2016;**161**:142. doi:10.1016/j.ajo.2015.09.040

120 Kim DH, Kim MK, Wee WR. Estimation of intraocular lens power calculation after myopic corneal refractive surgery: using corneal height in anterior segment optical coherence tomography. *Korean J Ophthalmol* 2015;**29**:195–202. doi:10.3341/kjo.2015.29.3.195

121 Barrett GD. Intraocular lens calculation formulas for new intraocular lens implants. *J*

Cataract Refract Surg 1987;**13**:389–96. doi:10.1016/s0886-3350(87)80037-8

122 Barrett GD. An improved universal theoretical formula for intraocular lens power prediction. *J Cataract Refract Surg* 1993;**19**:713–20. doi:10.1016/s0886-3350(13)80339-2

123 Olsen T. The Olsen formula. In: *Intraocular Lens Power Calculations*. 2004. 27–40.

124 Li T, Yang K, Stein J, *et al.* Gradient Boosting Decision Tree Algorithm for the Prediction of Postoperative Intraocular Lens Position in Cataract Surgery. *medRxiv* Published Online First: 2020. doi:10.1101/2020.08.26.20181156

125 Simpson MJ, Charman WN. The effect of testing distance on intraocular lens power calculation. J. Refract. Surg. 2014;**30**:726. doi:10.3928/1081597X-20141021-01

126 Holladay JT, Musgrove KH, Prager TC, *et al.* A three-part system for refining intraocular lens power calculations. *J Cataract Refract Surg* 1988;**14**:17–24. doi:https://doi.org/10.1016/S0886-3350(88)80059-2

127 Hoffer KJ. Reply: Errata in printed Hoffer Q formula. *J Cataract Refract Surg* 2007;**33**.https://journals.lww.com/jcrs/Fulltext/2007/01000/Reply__Errata_in_printed_Hoffer_Q_formula.3.aspx

128 Zuberbuhler B, Morrell AJ. Errata in printed Hoffer Q formula. *J Cataract Refract Surg* 2007;**33**:2.

129 Li T, Stein J, Nallasamy N. AI-powered effective lens position prediction improves the accuracy of existing lens formulas. *Br J Ophthalmol* Published Online First: 2021. doi:10.1136/bjophthalmol-2020-318321

130 Holladay JT. Refractive Power Calculations for Intraocular Lenses in the Phakic Eye. *Am J Ophthalmol* 1993;**116**:63–6. doi:10.1016/S0002-9394(14)71745-3

131 Norrby S, Bergman R, Hirnschall N, *et al.* Prediction of the true IOL position. *Br J Ophthalmol* 2017;**101**:1440–6. doi:10.1136/bjophthalmol-2016-309543

132 Einighammer J, Oltrup T, Bende T, *et al.* Calculating Intraocular Lens Geometry by Real Ray Tracing. *J Refract Surg* 2007;**23**:393–404. doi:10.3928/1081-597X-20070401-12

133 Hoffmann PC, Lindemann CR. Intraocular lens calculation for aspheric intraocular lenses. *J Cataract Refract Surg* 2013;**39**:867–72. doi:10.1016/j.jcrs.2012.12.037

134 Simpson MJ, Charman WN. The effect of testing distance on intraocular lens power calculation. J. Refract. Surg. 2014;**31**:726. doi:10.3928/1081597X-20141021-01

135 Preußner P-R, Hoffmann P, Petermeier K. Vergleich zwischen Raytracing und IOL-

Formeln der 3. Generation. *Klin Monbl Augenheilkd* 2009;**226**:83–9. doi:10.1055/s-2008-1027966

136 Li T, Reddy A, Stein JD, *et al.* Ray tracing intraocular lens calculation performance improved by AI-powered postoperative lens position prediction. *Br J Ophthalmol* Published Online First: 2021. doi:10.1136/bjophthalmol-2021-320283

137 Jones JJ, Jones YJ, Jin GJC. Indications and outcomes of intraocular lens exchange during a recent 5-year period. *Am J Ophthalmol* 2014;**157**:154-162.e1. doi:10.1016/j.ajo.2013.08.019

138 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. doi:10.1097/j.jcrs.0000000000000349

139 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. doi:10.1016/j.jcrs.2016.12.021

140 Melles RB, Holladay JT, Chang WJ. Accuracy of Intraocular Lens Calculation Formulas. *Ophthalmology* 2018;**125**:169–78. doi:10.1016/j.ophtha.2017.08.027

141 Solver. https://iolsolver.com/ (accessed 29 Jan 2022).

142 Akkara J, Kuriakose A. Role of artificial intelligence and machine learning in ophthalmology. *Kerala J Ophthalmol* 2019;**31**:150–60.

143 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. doi:10.1001/jamaophthalmol.2018.7051

144 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. doi:10.1001/jamaophthalmol.2020.2974

145 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. doi:10.1186/s12886-021-01950-2

146 Dietterich TG. Ensemble Methods in Machine Learning. In: *Multiple Classifier Systems*. Berlin, Heidelberg: : Springer Berlin Heidelberg 2000. 1–15.

147 Wolpert D. Stacked Generalization. *Neural Networks* 1992;**5**:241–59. doi:10.1016/S0893-

6080(05)80023-1

148 Holladay JT, Wilcox RR, Koch DD, *et al.* Re: Hoffer et al.: Update on intraocular lens power calculation study protocols: the better way to design and report clinical trials (Ophthalmology. 2020; Jul 9 [Epub ahead of print]). Ophthalmology. 2021;**128**:e20. doi:10.1016/j.ophtha.2020.10.039

149 Li T, Stein J, Nallasamy N. Evaluation of the Nallasamy formula: a stacking ensemble machine learning method for refraction prediction in cataract surgery. *Br J Ophthalmol* 2022.

150 Japkowicz N. Assessment metrics for imbalanced learning. *Imbalanced Learn Found algorithms, Appl* 2013;:187–206.

151 Moniz N, Branco P, Torgo L. Evaluation of ensemble methods in imbalanced regression tasks. In: *First International Workshop on Learning with Imbalanced Domains: Theory and Applications*. PMLR 2017. 129–40.

152 Christoffersen PF, Diebold FX. Optimal prediction under asymmetric loss. *Econom theory* 1997;**13**:808–17.

153 Torgo L, Ribeiro R. Utility-based regression. In: *European conference on principles of data mining and knowledge discovery*. Springer 2007. 597–604.

154 Ribeiro RP. Utility-based regression. *Ph D Diss* 2011.

155 Bi J, Bennett KP. Regression error characteristic curves. In: *Proceedings of the 20th international conference on machine learning (ICML-03)*. 2003. 43–50.

156 Torgo L. Regression error characteristic surfaces. In: *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. 2005. 697–702.

157 Hernández-Orallo J. ROC curves for regression. *Pattern Recognit* 2013;**46**:3395–411.

158 Rosset S, Perlich C, Zadrozny B. Ranking-based evaluation of regression models. *Knowl Inf Syst* 2007;**12**:331–53. doi:10.1007/s10115-006-0037-3

159 Li T, Stein JD, Nallasamy N. Evaluation of the Nallasamy Formula: A Stacking Ensemble Machine Learning Method for Refraction Prediction in Cataract Surgery. *medRxiv* Published Online First: 2021. doi:10.1101/2021.10.25.21265489

160 Lemnaru C, Potolea R. Imbalanced classification problems: systematic study, issues and best practices. In: *International Conference on Enterprise Information Systems*. Springer

2011. 35–50.

161 Gjerdrum B, Gundersen KG, Lundmark PO, *et al.* Refractive Precision of Ray Tracing IOL Calculations Based on OCT Data versus Traditional IOL Calculation Formulas Based on Reflectometry in Patients with a History of Laser Vision Correction for Myopia. *Clin Ophthalmol* 2021;**15**:845–57. doi:10.2147/OPTH.S298007

162 Ying X. An overview of overfitting and its solutions. In: *Journal of Physics: Conference Series*. IOP Publishing 2019. 22022.

163 Gouvea L, Sioufi K, Brown CE, *et al.* Refractive Accuracy of Barrett True-K vs Intraoperative Aberrometry for IOL Power Calculation in Post-Corneal Refractive Surgery Eyes. *Clin Ophthalmol (Auckland, NZ)* 2021;**15**:4305.

164 Dawson VJ, Patnaik JL, Ifantides C, *et al.* Comparison of refractive prediction for intraoperative aberrometry and Barrett True K no history formula in cataract surgery patients with prior radial keratotomy. *Acta Ophthalmol* 2021;**99**:e844–51.

165 Wei L, Meng J, Qi J, *et al.* Comparisons of intraocular lens power calculation methods for eyes with previous myopic laser refractive surgery: Bayesian network meta-analysis. *J Cataract Refract Surg* 2021;**47**:1011–8.

166 Davison JA, Potvin R. Preoperative measurement vs intraoperative aberrometry for the selection of intraocular lens sphere power in normal eyes. *Clin Ophthalmol (Auckland, NZ)* 2017;**11**:923.

167 Hoffer KJ, Savini G. IOL Power Calculation in Short and Long Eyes. *Asia-Pacific J Ophthalmol* 2017;**6**.https://journals.lww.com/apjoo/Fulltext/2017/07000/IOL_Power_Calculation_in_Short_and_Long_Eyes.6.aspx

168 Wang Q, Jiang W, Lin T, *et al.* Meta-analysis of accuracy of intraocular lens power calculation formulas in short eyes. *Clin Experiment Ophthalmol* 2018;**46**:356–63.

169 Yang Y-H. Comparison of Pentacam and IOL Master for measurement of anterior segment parameters in eyes with cataracts. *Int Eye Sci* 2019;:796–800.

170 Uçakhan OÖ, Akbel V, Bıyıklı Z, *et al.* Comparison of corneal curvature and anterior chamber depth measurements using the manual keratometer, Lenstar LS 900 and the Pentacam. *Middle East Afr J Ophthalmol* 2013;**20**:201–6. doi:10.4103/0974-9233.114791

171 Hashemi H, Asgari S, Miraftab M, *et al.* Agreement study of keratometric values

measured by Biograph/LENSTAR, auto-kerato-refractometer and Pentacam: Decision for IOL calculation. *Clin Exp Optom* 2014;**97**:450–5.

172 Hipólito-Fernandes D, Luís ME, Serras-Pereira R, *et al.* Anterior chamber depth, lens thickness and intraocular lens calculation formula accuracy: nine formulas comparison. *Br J Ophthalmol* 2022;**106**:349–55.

173 Yeo TK, Heng WJ, Pek D, *et al.* Accuracy of intraocular lens formulas using total keratometry in eyes with previous myopic laser refractive surgery. *Eye* 2021;**35**:1705–11. doi:10.1038/s41433-020-01159-5

174 Lee H, Chung JL, Kim YJ, *et al.* Prediction accuracy of standard and total keratometry by swept-source optical biometer for multifocal intraocular lens power calculation. *Sci Rep* 2021;**11**:4794. doi:10.1038/s41598-021-84238-1

175 Wang L, Spektor T, de Souza RG, *et al.* Evaluation of total keratometry and its accuracy for intraocular lens power calculation in eyes after corneal refractive surgery. *J Cataract Refract Surg* 2019;**45**:1416–21. doi:https://doi.org/10.1016/j.jcrs.2019.05.020

176 Turnbull AMJ, Barrett GD. Using the first-eye prediction error in cataract surgery to refine the refractive outcome of the second eye. *J Cataract Refract Surg* 2019;**45**:1239–45. doi:https://doi.org/10.1016/j.jcrs.2019.04.008

177 Olsen T. Use of Fellow Eye Data in the Calculation of Intraocular Lens Power for the Second Eye. *Ophthalmology* 2011;**118**:1710–5. doi:https://doi.org/10.1016/j.ophtha.2011.04.030

178 Covert DJ, Henry CR, Koenig SB. Intraocular lens power selection in the second eye of patients undergoing bilateral, sequential cataract extraction. *Ophthalmology* 2010;**117**:49–54.

179 Zhang J, Ning X-N, Yan H. Adjustment of IOL power for the second eye based on refractive error of the first-operated eye. *Int J Ophthalmol* 2019;**12**:1348–50. doi:10.18240/ijo.2019.08.18