Machine Learning for the Prediction of Refractive Surprise after Cataract Surgery

Subject Area: Artificial Intelligence & Machine Learning, Data Science, Ophthalmology

Student: Boas Meier

Advisor: Dr. sc. ETH Andreas Streich

Expert: Dr. Rémi Janner

Principal: Univ.-Prof. Dr. Achim Langenbucher

Keywords: Cataract, Intraocular Lens Power Calculation, Machine Learning, Ophthalmology, Refractive Surprise

1. Problem Statement

A cataract is the clouding of the eye's focusing lens that results in blurry vision and, if left untreated, eventually leads to vision loss. Today, cataract surgery is the most common procedure performed around the world and in all of medicine. With an overall success rate of approximately 97 percent when performed in appropriate settings, it is as well the most effective procedure (Feldman et al., 2022). Due to the high success rate, patient expectations are today at an all-time high and so is the dissatisfaction in cases where vision is not restored as expected. One of the major reasons for dissatisfaction after cataract surgery is refractive surprise. (Peck et al., 2022) Refractive surprise refers to cases, where the intended post-operative refractive target is missed. This can lead to follow-up interventions up to and including replacement of the lens. This thesis aims to preoperatively predict whether a case at hand may yield a refractive surprise, using machine learning. By predicting refractive surprises, the surgeon could take preventive measures, such as choosing a lens type that is less sensitive to refractive error. Avoiding complications would increase patient satisfaction and save additional post-operative treatments, thus also saving costs.

2. Concept

The refractive prediction error dataset for this task consists of 2626 eyes. To assess how the final algorithm performs on previously unseen data, the dataset was split into a training set with 2363 eyes and a testing set consisting of 263 eyes. The splits were stratified based on the prediction error (PE) and the different studies the data comes from which ensures a similar distribution of the target in both datasets. Additionally, the training set was further divided with ratio 0.2 into a training and validation set, using the same stratification.

Refractive surprises were predicted both through classification and continuously through regression. Different models with increasing complexity were trained iteratively and incrementally on the training set, including support vector machines, decision trees, random forests, gradient boosted trees, and neural networks. Hyperparameters for the models were tuned on the validation set, and the final model was selected based on its validation performance.

In addition, an intraocular lens (IOL) power calculation formula based on a stacking ensemble machine learning algorithm was implemented. The ensemble consists of two levels of predictors. The first level consists of the Castrop IOL power calculation formula (Langenbucher et al., 2021) and a neural network regressor that predicts the PE of the Castrop formula. The second-level predictor is another neural network regressor, which is trained on the output of the first-level models and predicts the post-operative spherical equivalent (SEQ) (see Figure

1). The constants of the existing IOL formulas were optimized for each study and dataset separately using the Levenberg-Marquardt algorithm with a mean squared error (MSE) loss function.

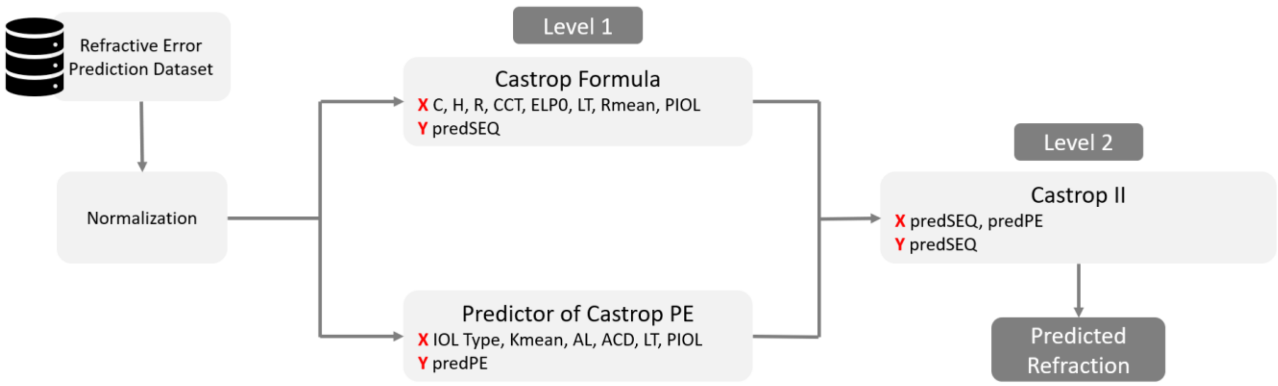

Figure 1: A stacking ensemble machine learning algorithm consisting of two levels. The first level consists of the Castrop formula and a model that predicts the PE of the Castrop formula. The second level is a model which is trained on the output of the fist level models and predicts the SEQ.

3. Challenges

One of the main challenges was that refractive surprises were heavily overlapping with normal cases. Additionally, the class distribution was moderately imbalanced, resulting in models that almost completely ignored refractive surprises by default. These two factors made it a particularly challenging task for machine learning and made it difficult to develop a model that can properly distinguish refractive surprises. The problem of class imbalance could be partially addressed through re-sampling, class-weight-based adjustments of the loss-function, and post-prediction threshold tuning. However, due to the overlap between classes, better recall came at the expense of worse precision.

4. Results

The regression of the Castrop PE is done using a neural network, which achieves a mean absolute error (MAE) of 0.334 +/- 0.422 and a median absolute error (MedAE) of 0.283. Neural networks are also used for the classification of the Castrop PE. When predicting whether the absolute PE will be greater than 0.5 diopters (D), the neural network achieved a precision of 0.3 and a recall of 0.69, resulting in an F1 score of 0.42. When predicting whether the absolute PE will be greater than 0.25 D, the neural network achieved a precision of 0.58 and a recall of 0.52, resulting in an F1 score of 0.55.

The performance of the ensemble was compared to the Castrop and the SRKT formulas (Retzlaff et al., 1990) based on the number of absolute PEs within the limits of 0.25 D, 0.5 D, 0.75 D, and 1.0 D. The ensemble performed slightly worse within 0.25 D but slightly better within 0.75 D and 1.0 D (see **Error! Reference source not found.**). However, McNemar tests did not report significant differences. Additionally, the MAE, MedAE, and SD as well as the formula performance index (FPI), were also compared. The MAE, MedAE, and SD were smaller for the ensemble for both the validation and testing sets (see **Error! Reference source not found.**). However, T-tests did not detect significance. The ensemble shows similar performance to other publications of IOL formulas based on machine learning (Li et al., 2022).

5. Outlook

In the future, the performance of both PE prediction and IOL power calculation could certainly be improved with more data and features. A desk study showed that preoperative visual acuity, sex, and age, which were not present in the data for this task, are all somehow correlated with refractive surprise. It would be interesting to see the influence of these additional features on the performance of the developed models. Furthermore, a gold standard benchmark for PE prediction and IOL power calculation could be created, making comparisons between such studies more reliable. The problem of little and imbalanced data could be addressed through data augmentation. Data augmentation experiments conducted during this thesis did not significantly affect performance. However, there have been recent advancements in data synthesizing for tabular data using transformer-based architectures, which might bring better results when applied. Finally, it may also be worth experimenting with some recent deep learning architectures for tabular data, which were not covered in this thesis. Although deep learning has not yet fully overtaken tree-based models for tabular data, these architectures could potentially further increase performance for this particular task.

Bibliography

Feldman, B. H., Heersink, S., & Alpa S., P. (2022, July 7). *Cataract—EyeWiki*. https://eyewiki.aao.org/Cataract

Langenbucher, A., Szentmáry, N., Cayless, A., Weisensee, J., Fabian, E., Wendelstein, J., & Hoffmann, P. (2021). Considerations on the Castrop formula for calculation of intraocular lens power. *PLOS ONE*, *16*(6), e0252102. https://doi.org/10.1371/journal.pone.0252102

Li, T., Stein, J., & Nallasamy, N. (2022). Evaluation of the Nallasamy formula: A stacking ensemble machine learning method for refraction prediction in cataract surgery. *British Journal of Ophthalmology*. https://doi.org/10.1136/bjophthalmol-2021-320599

Peck, T., Feldman, B. H., & Patel, A. S. (2022, May 8). *Refractive Error After Cataract Surgery—EyeWiki*. https://eyewiki.aao.org/Refractive_Error_After_Cataract_Surgery

Retzlaff, J. A., Sanders, D. R., & Kraff, M. C. (1990). Development of the SRK/T intraocular lens implant power calculation formula. *Journal of Cataract & Refractive Surgery*, *16*(3), 333–340. https://doi.org/10.1016/S0886-3350(13)80705-5