

# Development and Validation of Machine Learning Models: Electronic Health Record Data To Predict Visual Acuity After Cataract Surgery

Stacey E Alexeeff, PhD<sup>1</sup>; Stephen Uong, MPH<sup>1</sup>; Liyan Liu, MSc<sup>1</sup>; Neal H Shorstein, MD<sup>2</sup>; James Carolan, MD<sup>3</sup>; Laura B Amsden, MSW, MPH<sup>1</sup>; Lisa J Herrinton, PhD<sup>1</sup>

Perm J 2020;25:20.188

E-pub: 12/23/2020

<https://doi.org/10.7812/TPP/20.188>

## ABSTRACT

**Background:** To develop predictive models of final corrected distance visual acuity (CDVA) following cataract surgery using machine learning algorithms and electronic health record data.

**Methods:** In this predictive modeling study we used decision tree, random forest, and gradient boosting. We included the first surgical eye of 64,768 members of Kaiser Permanente Northern California who underwent cataract surgery from June 1, 2010 through May 31, 2015. We measured discrimination and calibration of machine learning models for predicting postoperative CDVA 20/50 or worse vs 20/40 or better.

**Results:** The training set included 51,712 patients, and the validation set included 13,056 patients. We compared 3 machine learning models and found that the gradient boosting model provided the best discrimination ability for CDVA. The most important variables for predicting final CDVA 20/50 or worse were preoperative CDVA, age, and age-related macular degeneration, which together accounted for 41% of the gain in optimization of the gradient boosting model. Other important variables in the model included dispensed glaucoma medication, epiretinal membrane, cornea disorder, cataract surgery operating time, surgeon experience, and census block neighborhood characteristics (household income, family income, family poverty, college education, and home residence by owner).

**Conclusion:** For predicting CDVA after cataract surgery, gradient boosting had the best ability to discriminate patients with postoperative CDVA 20/50 or worse from patients with postoperative CDVA 20/40 or better. Machine learning has the potential to improve prognosis and can improve patient information when making decisions to undergo cataract surgery.

## INTRODUCTION

Machine learning is the ability of a statistical method to learn without explicit programming, thereby semiautomating data analysis and increasing analytic output.<sup>1</sup> Machine learning has been used to automate the interpretation of images from retinal photography or optical coherence tomography for the diagnosis of diabetic retinopathy, age-related macular degeneration (AMD), glaucoma, and other eye diseases.<sup>2-4</sup> Machine learning has also been used to predict visual acuity among patients with AMD after treatment with antivascular endothelial growth factor (anti-VEGF).<sup>5</sup> In addition, machine learning can be used in research to explore data at relatively low cost.<sup>6</sup> For these reasons, machine learning has the potential to transform

healthcare delivery, and understanding its application and implications is of broad interest.

Cataract is a leading cause of blindness and vision impairment in the US and the world.<sup>7,8</sup> By 2020, it is expected that 57.1 million people affected by cataract will have moderate or severe vision impairment.<sup>8</sup> Cataract surgery is a common and effective procedure to treat vision loss due to cataract, with more than 3 million routine cataract surgeries performed in Medicare beneficiaries each year.<sup>9</sup> With so many people affected by cataract and treated by cataract surgery, it is important to evaluate whether postoperative vision can be accurately predicted in patients who receive cataract surgery.

We used machine learning and data from an electronic health record to predict final corrected distance visual acuity (CDVA) following phacoemulsification cataract surgery. We considered 3 machine learning models based on classification trees, each with an increasingly complex statistical algorithm. We compared the models' discrimination to determine the model that best differentiated patients with postoperative CDVA 20/50 or worse from those with CDVA 20/40 or better. We also evaluated the calibration of the best performing model by comparing the predicted vs observed probabilities of the outcome.

## METHODS

The Declaration of Helsinki was adhered to, and Institutional Review Board/Ethics Committee approval was obtained. Prediction models were developed and validated following the TRIPOD guidelines.<sup>10</sup>

### Population and Data

The dataset used for the present study was developed for a retrospective cohort study that is described in detail in an earlier publication.<sup>11</sup> Briefly, the cohort included 65,370 members of Kaiser Permanente Northern California who

### Author Affiliations

<sup>1</sup>Division of Research, Kaiser Permanente Northern California, Oakland, CA

<sup>2</sup>Departments of Ophthalmology and Quality, Kaiser Permanente, Walnut Creek, CA

<sup>3</sup>Department of Ophthalmology, Kaiser Permanente, San Rafael, CA

### Corresponding Author

Lisa J Herrinton, PhD ([lisa.herrinton@kp.org](mailto:lisa.herrinton@kp.org))

Keywords: cataract surgery, electronic health record, machine learning, visual acuity

underwent phacoemulsification cataract surgery during June 1, 2010 through May 31, 2015. For this machine learning study, we excluded 602 patients with preoperative CDVA 20/20, leaving a cohort of 64,768 patients.

Patient data were extracted from the EPIC-based (Verona, WI) electronic health record via a research data warehouse maintained on an Oracle platform. The data warehouse aggregates the many data sources that researchers in our organization use most often to access key data elements. Database structures are optimized for research query and retrieval. At the core of the data warehouse are a series of standardized file definitions. Content areas and data elements are documented in data dictionaries. The data warehouse leverages the capabilities of the longitudinal clinical information documented in the electronic health record and facilitates the extraction of hundreds of variables for use in this modeling study.

The outcome, postoperative CDVA in the first surgical eye, was obtained for the period 21-365 days after cataract surgery. We used SAS PERL regular expressions to extract CDVA and laterality from free text and excluded patients whose preoperative CDVA was 20/20.<sup>12</sup> We dichotomized postoperative CDVA at 20/50 or worse vs 20/40 or better because 20/40 or better is required by many states' Department of Motor Vehicle Services for driving.<sup>13</sup> To increase readability, we converted Snellen to the logarithm of the minimum angle of resolution, performed calculations, and then converted the logarithm of the minimum angle of resolution back to Snellen.

Predictors included patient demographic and insurance information, systemic and ocular comorbidities based on diagnostic codes, care utilization data, and surgeon experience. Neighborhood socioeconomic data were obtained at the block-group level from 2010 census data based on geocoding of each patient's address; socioeconomic data included median household income (all households), median family income (households with at least 2 residents), proportion of family households with below-poverty income, proportion of population with a Bachelor's degree (college education), and proportion of homes with residence by owner. AMD was defined as ICD-9 diagnostic codes 362.5X but not 362.53 or 362.56; this definition includes both wet and dry forms, with 10% being wet. Epiretinal membrane (ERM) was identified from ICD-9 diagnostic code 362.56, cornea disorders from diagnostic code 371.XX, and glaucoma from diagnostic code 365.XX. Procedure codes were extracted from inpatient and outpatient encounter data. Retinal procedures were identified from ICD-9 procedure code 14.XX except 14.7X. We also considered "other intraocular therapeutic procedures" and other broad categories of procedures as defined by the Healthcare Cost and Utilization Project.<sup>14</sup> Medication data

were extracted from pharmacy dispensing records, reflecting medication that was prescribed and then dispensed to the patient. Glaucoma medications included prostaglandin analogues,  $\beta$ -adrenergic antagonists,  $\alpha$ -adrenergic agonists, carbonic anhydrase inhibitors, and muscarinic agents. Care utilization variables included inpatient visits, outpatient visits, and emergency department visits; we quantified utilization in 2 key time periods before surgery: the preoperative period (0-30 days before surgery) and the year before surgery other than the preoperative period (31-365 days before surgery). Surgeon experience was defined as the total number of cataract surgeries performed by the surgeon during all their years of employment with the health plan.

### Predictive Modeling

We evaluated 3 tree-based machine learning models for predicting postoperative CDVA: decision tree, random forest, and gradient boosting. We used a split-sample design to randomize and partition the study population into training (80%) and validation (20%) sets. The training set was used to build the machine learning models. The validation set was held out during model building and used to evaluate model performance. The initial list of 473 variables was reduced using forward stepwise logistic regression with a 0.2 significance level, resulting in 203 candidate variables to be used in the machine learning modeling. The 3 machine learning methods are described below.

### Decision Tree

Decision tree models use recursive binary partitioning algorithms that divide multidimensional predictor space into regions through successive splits that optimize the classification model by minimizing the variance. The result of this partitioning process can be visualized as a tree.<sup>15</sup> At each step, 1 branch of the tree is split into 2 branches based on selecting a predictor variable and choosing a cutoff value that reduces the model error. The leaves of the tree are the terminal nodes of the branches, where each leaf represents a subregion of the predictor space based on all the cumulative splits of the branches leading to that terminal node. Predictions for new patients are made by finding what leaf that patient falls into, such that the predicted outcome for each leaf is the observed proportion of subjects in that leaf who have the outcome in the training data. To prevent overfitting, a method called "cost-complexity pruning" is used with cross-validation to find the subtree that minimizes the prediction error. An advantage of the decision tree is that the final model can be easily interpreted and visualized. A disadvantage is relatively high variance, so that predictions may not be as accurate as other methods.

### Random Forest

Random forest models use an ensemble of decision trees (a forest) to generate predictions. When considering each potential branch split, the random forest algorithm randomly selects a subset of variables from the full set of predictors and only considers splitting that branch using the variables in the selected subset.<sup>15</sup> The predictions from each decision tree in the forest are then averaged to generate the final predicted outcome for each patient. Creating random subsets of the predictors allows the trees in the forest to be less correlated with one another because they rely on different combinations of predictors. An advantage of this model is that the reduced correlation leads to reduced variance in the final predictions and more accurate predictions. A disadvantage is that interpretation can be difficult because the model uses hundreds or thousands of trees, so the result cannot be synthesized into a single diagram. However, the importance of each predictor across all trees in the forest can be summarized by the average reduction in model error or the increase in node purity that is attained by splits on that predictor.

### Gradient Boosting

Gradient boosting models also generate predictions by using an ensemble of decision trees. These decision trees are grown sequentially by fitting each new tree to the model residuals (unexplained variance) to slowly explain more and more of the variance as more trees are added to the model.<sup>16</sup> A key difference from the random forest algorithm is that gradient boosting focuses on the sequentially updated model residuals so that new trees account for the trees that have already been grown. Each tree uses a small number of splits to prevent overfitting. A shrinkage parameter is used to average across all the decision trees to generate the final predicted outcome for each patient. An advantage of the gradient boosting model is improved prediction by using a large number of trees and a shrinkage parameter that lowers variance and allows the model to learn slowly. A disadvantage of the interpretation of the final gradient boosting model cannot be directly visualized. Instead, each predictor's contribution to the model optimization is used to determine the importance that predictor.

### Model Evaluation

We assessed discrimination ability using the concordance statistic (C statistic) (ie, the probability that a patient with the CDVA 20/50 or worse has a higher predicted risk than a patient with 20/40 or better).<sup>17</sup> The C statistic ranges from 0 (complete discordance) to 1 (perfect concordance), with 0.5 indicating random chance, and is numerically equal to the area under the receiver operating characteristic curve.<sup>18</sup> We assessed calibration by graphically comparing the models'

predicted probabilities to the observed event rates. We formally tested calibration using the Hosmer-Lemeshow statistic.<sup>19</sup> In subgroup analyses, we evaluated discrimination and calibration after stratifying on comorbid AMD, diabetes, and glaucoma and in those with preoperative CDVA 20/100 or worse. Modeling was carried out using SAS 9.4 and R version 3.4.4.

### RESULTS

Characteristics of the study population of 64,768 cataract surgery patients are provided in [Table 1](#). Because patients were randomly selected into the training and validation sets, the characteristics were nearly identical. In the training set, average postoperative CDVA was 20/28 (standard deviation = 20/23), whereas in the validation set, it was 20/29 (standard deviation = 20/25).

[Figure 1](#) illustrates the decision tree for predicting CDVA 20/50 or worse vs 20/40 or better. The variables selected by the decision tree model were preoperative CDVA, AMD, dispensed glaucoma medication, ERM, and proliferative diabetic retinopathy. Preoperative CDVA was treated as a continuous variable, and the decision tree algorithm identified the cutpoints of 20/75 and 20/225 as optimal for discrimination. The model resulted in 7 leaves, each defined by a combination of covariate values determined by following the branch splits. The boxes at the bottom show the leaves of the tree with the probability of postoperative CDVA 20/50 or worse and the number of patients in each group. The leaf at the far left includes patients who have a preoperative CDVA worse than 20/75 but do not have AMD, dispensed glaucoma medications, ERM, or proliferative diabetic retinopathy. Patients who meet these criteria ( $n = 12,905$  in the training dataset) have an 8% probability of CDVA 20/50 or worse. The leaf at the far right includes patients who have preoperative CDVA 20/75 or better. Patients who meet this criterion ( $n = 34,070$  in the training dataset) have a 4% probability of CDVA 20/50 or worse. No other predictors mattered to an important extent. The highest risk of CDVA 20/50 or worse (50%) is predicted for 373 patients in the training dataset who have AMD and have a preoperative CDVA worse than 20/225. Although it is possible to directly visualize the decision tree model using [Figure 1](#), it is not possible to directly visualize models using random forest or gradient boosting because they are ensembles of decision trees, as described in the Methods.

[Table 2](#) shows the discrimination ability of each machine learning model for predicting postoperative CDVA. The gradient boosting model had the best discrimination ability overall and in patient subgroups as indicated by its relatively high C statistics. Discrimination ability using the gradient boosting model was slightly lower among patients with

Table 1. Characteristics of the patients in the training and validation datasets		
	Sample (N = 64,768)	
Characteristic	Training set (n = 51,712)	Validation set (n = 13,056)
Postoperative CDVA, Snellen (SD)	20/28 (20/23)	20/29 (20/26)
Preoperative CDVA, Snellen (SD)	20/91 (20/85)	20/91 (20/84)
Age, mean (SD)	73.1 (9.1) y	73.2 (8.9) y
Female, %	58.9	58.4
Race, %		
White	64.4	64.9
Hispanic	9.3	9.2
Asian	14.0	13.6
African American	5.0	5.4
Other	7.3	7.0
Body mass index, mean (SD)	27.6 (5.7) kg/m <sup>2</sup>	27.7 (5.8) kg/m <sup>2</sup>
Block-group characteristics		
Household income, mean (SD)	\$66,778 (\$26,141)	\$66,603 (\$25,697)
Family income, mean (SD)	\$73,923 (\$28,263)	\$73,811 (\$28,108)
Family poverty, mean (SD)	6.9% (6.0%)	6.3% (6.2%)
College education, mean (SD)	21.8% (10.0%)	21.8% (10.0%)
Home residence by owner, median (IQR)	73.3% (28.4%)	73.6% (28.7%)
Charlson comorbidity score, mean (SD)	0.9 (1.3)	0.9 (1.3)
Systemic comorbidity, %		
Hypertension	76.7	76.0
Chronic obstructive pulmonary disease	20.7	21.2
Cardiovascular disease	14.7	14.7
Renal disease	21.3	21.4
Diabetes mellitus	24.4	23.9
Ocular comorbidity, medications, procedures, %		
Mild diabetic retinopathy	2.7	2.6
Moderate/severe diabetic retinopathy	0.8	0.9
Nonproliferative diabetic retinopathy, NOS	1.7	1.7
Proliferative diabetic retinopathy	1.2	1.3
Epiretinal membrane	5.0	5.1
Retinal procedure	1.3	1.2
Focal laser coagulation	0.3	0.2
Age-related macular degeneration	11.5	11.8
Anti-VEGF therapy	0.6	0.6
Glaucoma diagnosis	19.8	19.6
Glaucoma medication dispensing	10.5	10.1
Corneal disorder	4.1	4.0
Preoperative ocular medications, 0-30 d, %		
Prednisolone only	40.2	40.6
NSAID ± prednisolone	55.7	55.1
Preoperative utilization, 0-30 d, mean (SD)		
Outpatient visits	2.4 (1.9)	2.4 (1.9)
Drug classes	4.6 (2.6)	4.6 (2.6)

(continued on following page)

**Table 1. Characteristics of the patients in the training and validation datasets (continued)**

Characteristic	Sample (N = 64,768)	
	Training set (n = 51,712)	Validation set (n = 13,056)
Preoperative utilization, 31-365 d, mean (SD)		
Inpatient visits	0.1 (0.5)	0.1 (0.5)
Outpatient visits	12.3 (11.6)	12.3 (11.4)
Emergency dept visits	0.4 (1.0)	0.4 (1.0)
Drug classes	7.9 (5.1)	7.9 (5.1)
No. of procedures	16.1 (17.1)	15.9 (15.8)

CDVA = corrected distance visual acuity; IQR = interquartile range; NOS = not otherwise specified; NSAID = nonsteroidal antiinflammatory drug; SD = standard deviation; VEGF = vascular endothelial growth factor.

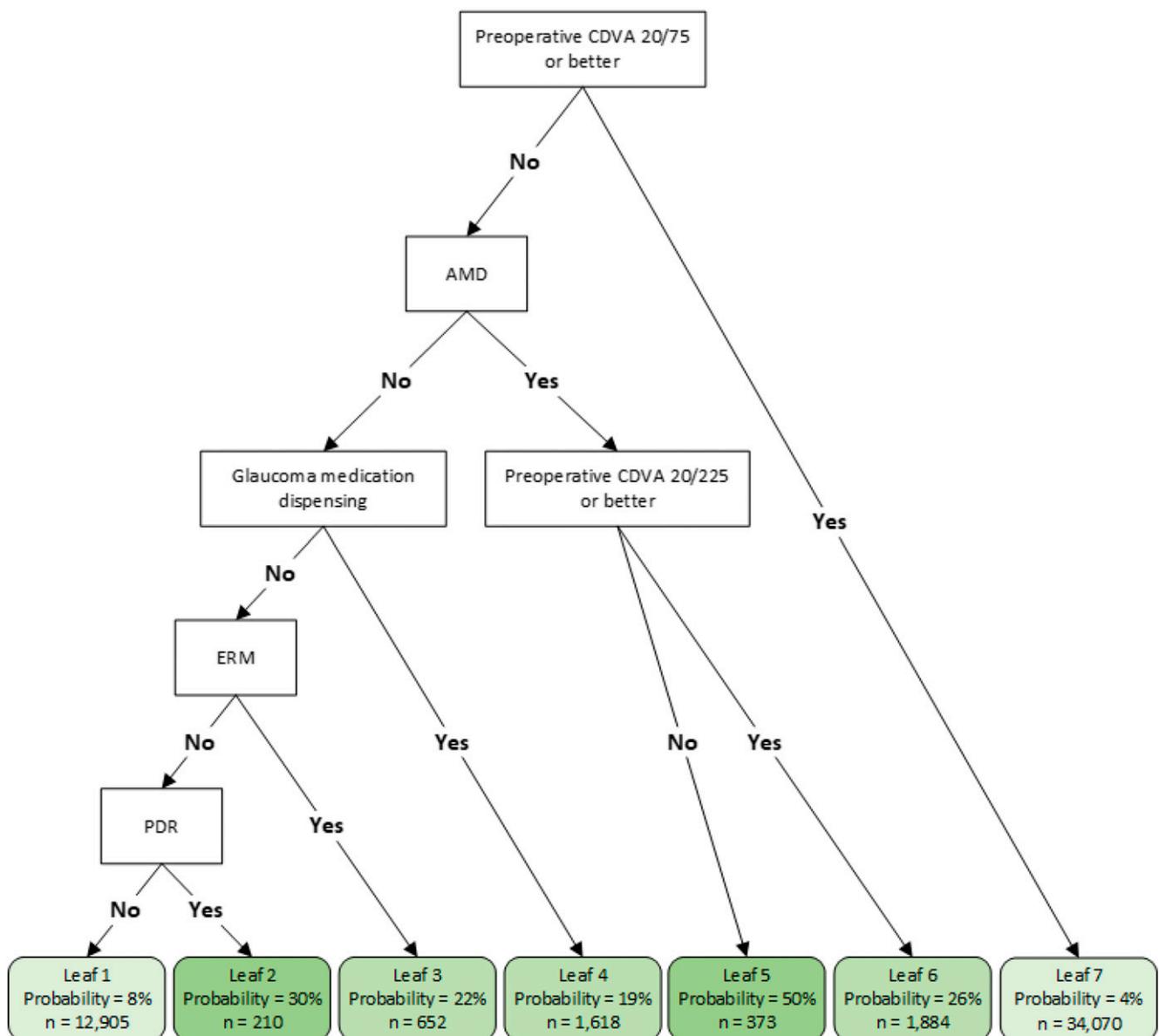


Figure 1. Decision tree for predicting postoperative CDVA 20/50 or worse.

**Table 2. Discrimination ability (C statistic and 95% confidence interval) of each machine learning model to predict postoperative corrected distance visual acuity 20/50 or worse vs 20/40 or better in the overall population and in subgroups**

Model	Overall		Subgroup			
	N	C (95% CI)	AMD (n = 7493)	Diabetes (n = 15,717)	Glaucoma (n = 12,805)	Preoperative CDVA 20/100 or worse (n = 17,569)
Decision tree	64,768	0.678 (0.674-0.682)	0.716 (0.706-0.726)	0.653 (0.646-0.660)	0.713 (0.705-0.721)	0.674 (0.667-0.681)
Random forest		0.716 (0.713-0.719)	0.749 (0.739-0.759)	0.686 (0.679-0.693)	0.729 (0.721-0.737)	0.707 (0.700-0.714)
Gradient boosting		0.756 (0.753-0.759)	0.773 (0.764-0.782)	0.733 (0.726-0.740)	0.770 (0.763-0.777)	0.733 (0.726-0.740)

CDVA = corrected distance visual acuity.

diabetes or with preoperative CDVA 20/100 or worse, indicating that the postoperative CDVA 20/50 or worse was more difficult to predict in these subgroups. In contrast, we found C statistics were higher among patients with AMD or glaucoma. Predictions of postoperative CDVA using the gradient boosting model for hypothetical patients with various preoperative characteristics are shown in Table 3. A hypothetical patient aged 74 years with preoperative CDVA of 20/60 and no ocular comorbidities has 3% probability of postoperative vision 20/50 or worse. In contrast, the probability is 42% for a 90-year-old patient with AMD who is starting with preoperative CDVA of 20/200.

Out of 203 candidate variables, 168 were selected into the gradient boosting model. Because the gradient boosting model uses hundreds or thousands of decision trees simultaneously to generate predictions, the model itself cannot easily be visualized. Instead, the percent gain in model optimization is used as a metric to understand the impact of predictive variables in the model. This is shown in Figure 2 for the 15 most important predictive variables. Together, the top 15 predictors account for 72% of the model optimization. Preoperative CDVA, age, and AMD were the most important variables in the model, contributing 22.3%, 9.4%, and 8.9% of the gain in model optimization, respectively. History of a dispensed glaucoma medication identifies patients with more severe glaucoma and was more predictive than a glaucoma diagnosis alone. Calibration of the gradient boosting model is shown in Figure 3. The figure shows good agreement between the predicted and observed probability. Our formal assessment of model calibration via the Hosmer-Lemeshow test found good calibration ( $p = 0.43$ , where  $p > 0.05$  indicates no evidence of poor fit).

**DISCUSSION**

We compared the accuracy of 3 machine learning models using electronic health record data for 64,768 cataract surgery patients to predict CDVA following cataract surgery. Gradient boosting, the most complex method, demonstrated the best ability to discriminate patients with postoperative CDVA 20/50 or worse from patients with postoperative CDVA 20/40 or better. Preoperative CDVA, age, and AMD were the most important variables in the final model, accounting for 41% of the gain in model optimization. The remaining 59% came from the other 165 variables selected into the gradient boosting model, including ocular comorbidities; previous ocular and nonocular procedures; medication use such as glaucoma medication, anti-VEGF, and ophthalmic prednisolone and nonsteroidal antiinflammatory agents, among others; surgery characteristics; and neighborhood characteristics. Although each

**Table 3. Model predictions for 4 example patients undergoing cataract surgery**

	Base patient	Patient 2	Patient 3	Patient 4
<b>Predictor variables</b>				
Preoperative CDVA	20/60	20/200	20/60	20/60
Age, y	74	90	74	74
Age-related macular degeneration	No	Yes	No	No
Surgeon experience, cases	1200	1200	500	1200
Cornea disorder	No	No	Yes	No
Epiretinal membrane	No	No	Yes	No
Block-group family income, \$	65,000	65,000	65,000	30,000
Block-group college education, %	20	20	20	2
Block-group home residence by owner, %	75	75	75	20
<b>Predicted outcome</b>				
Probability of post-operative CDVA 20/50 or worse, %	3	42	10	4

CDVA = corrected distance visual acuity.

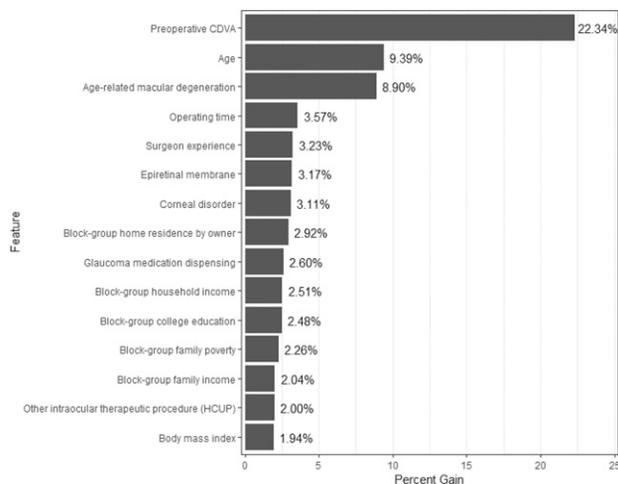


Figure 2. Predictive variables ranked by importance in gradient boosting model, showing the percent gain in model optimization for each predictor, showing the 15 most important predictive variables, for the outcome of CDVA 20/50 or worse vs 20/40 or better.

variable alone contributed only a small amount of information, the gradient boosting model used all this information together to obtain highly accurate predictions. In general, sophisticated machine learning methods can leverage small bits of information across many predictor variables to a surprising degree to provide accurate predictions.

We observed that discrimination ability using the gradient boosting model was slightly lower among patients with diabetes. About 80% of diabetes patients have no retinopathy, and their risk of visual acuity mirrors nondiabetics.<sup>11</sup> The other 20% have varying degrees of retinopathy, with increasing severity of retinopathy being associated with worsening chances of a good visual outcome.

To our knowledge, this is the first report describing the use of machine learning methods with electronic health

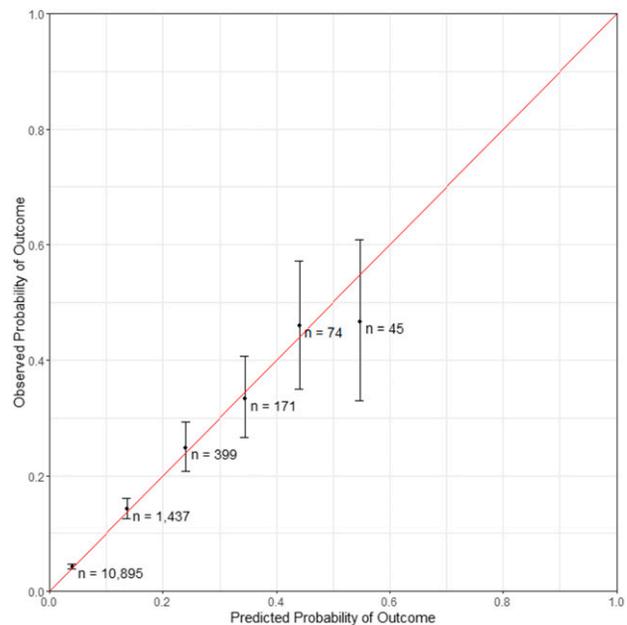


Figure 3. Calibration plot for the gradient boosting model showing predicted versus observed probability of CDVA 20/50 or worse vs 20/40 or better in the validation dataset.

record data to predict postoperative CDVA in a community-based population following cataract surgery. A similar study used machine learning methods with electronic health record data from 653 neovascular AMD patients to predict CDVA at 3 and 12 months after treatment with anti-VEGF injections.<sup>5</sup> That study compared 5 different machine learning models, including gradient boosting, and found that all models had similar accuracy. The study was similar to ours in using a data warehouse, although the setting was an ophthalmology clinic and eye hospital, with minimal information available on nonophthalmic diseases and utilization.

This study used electronic health record data that had been carefully curated into a research data warehouse, with the final dataset undergoing further cleaning, structuring, and curation by an experienced data scientist. A limitation is that electronic health record data are not available in all health care systems, and, without the infrastructure of a data warehouse, it may be difficult for other health care systems to create machine learning models that are tuned to their systems and patient populations. However, our primary goal was to familiarize ophthalmologists with machine learning methods because we believe they will be used increasingly to deliver health care. This study demonstrates the potential use of this technology as it is more widely adopted and as data become more accessible. Another limitation is that the gradient boosting model cannot be easily described the way a single decision tree can be illustrated, and the influence of each predictive variable in the model is harder to understand, so implementing this model in practice may appear to clinicians and patients to be a “black box.”

The study included variables collected during the cataract surgery (eg, cataract surgery operating time). Longer surgical times generally indicate more difficult conditions and intraoperative complications. In some cases, patients are known to have characteristics that will make the surgery longer, and the operating time is a proxy for patient complexity. Surgical time tends to be longer for patients with brunescant cataract, phacodonesis, pseudoexfoliation, previous vitrectomy, and advanced age, among other factors.<sup>20</sup> In other cases, the surgeon may encounter an unexpected complication, with the opportunity to update the predicted postoperative CDVA after the surgery has been completed. By including variables collected during cataract surgery, our model was set up to predict CDVA after surgery, predicting forward from the time that surgery was completed to several weeks or months later when the new CDVA was obtained. Depending on how the model will be used in practice, inclusion of surgery variables may or may not be appropriate.

The gradient boosting and decision tree models selected predictors using different computational approaches. Nonetheless, both identified preoperative CDVA as an important predictor of postoperative CDVA, consistent with past reports.<sup>21</sup> In addition, both models identified AMD, glaucoma, and ERM as important predictors. Thus, the key predictor variables were similar for both models. Notably, the gradient boosting algorithm allows for predictors with smaller contributions to still have influence in the model, and there were other predictors that were selected in the gradient boosting that were overlooked in the decision tree model. Specifically, the gradient boosting model also identified age, operative time and surgeon experience, and several block-group measure of socioeconomic status as meaningful predictors. The inclusion of these

additional variables in the gradient boosting model resulted in better predictive discrimination, as measured by the C statistics.

Each of these machine learning models could be implemented in clinical practice to obtain reliable predictions of CDVA. However, the gradient boosting model would need to be implemented via additional software to calculate the predictions using the predictor variables and the fitted model information. These types of tools can now be embedded within a medical record systems that can extract the predictor variables needed from the medical record and plug into model formulas to generate predictions. In contrast, the decision tree model could be implemented in clinical practice by an ophthalmologist using readily available information without the need to implement additional software.

Machine learning offers promise for enabling rapid analysis of complex observational health data toward the goals of improving patient care, rapidly responding to sudden increases in adverse events, identifying targets for quality improvement, and reducing the cost of research by enabling lower-cost data exploration.<sup>6</sup> The potential benefits of machine learning for analyzing imaging for diagnosis have been demonstrated in diabetic retinopathy and other eye diseases.<sup>2-4</sup> Based on the methodological literature and our own experience with comparative effectiveness research, defined as the use of observational data to make causal inferences about the effectiveness of a treatment, factors that affect patient selection for surgical or medical treatment can be strongly related to outcomes and cause bias.<sup>22</sup> Thus, using ordinary machine learning methods to make prognoses must take into consideration that the patients in the dataset used to generate the predictions were fit enough to undergo treatment. We believe that these methods are best for predicting short-term outcomes among patients who have committed to the treatment and not patients who could opt-out or who have complex treatment trajectories. The development of machine learning methods that can be used with sequential decision-making points to predict longer-term outcomes is a topic of active research.<sup>23</sup>

Our study showed that machine learning combined with well-characterized data could be used to predict CDVA after cataract surgery with good accuracy. There is great potential for improvement in these models as new machine learning methods continue to be developed, data stored in electronic medical records continue to grow, and text mining and natural language processing are able to extract additional information from the medical record. As these predictive models continue to improve through the addition of new predictive variables and new machine learning methodologies, a highly predictive model for uncorrected visual acuity for near and far distances is potentially achievable. This has the potential to increase the functional level of

patients for activities of daily living for many tasks post-operatively. As the demand increases for more exacting visual and refractive outcomes, machine learning may be useful for satisfying patient expectations. In conclusion, the prediction of CDVA by using machine learning with electronic health record data is a valuable first step toward demonstrating the potential for individual-level patient predictive modeling. ❖

#### Disclosure Statement

The author(s) have no conflicts of interest to disclose.

#### Authors' Contributions

Stacey Alexeeff, PhD, participated in study design, statistical analysis, and drafting and critical review of the final manuscript. Stephen Uong, MPH, participated in data acquisition and drafting and critical review of the final manuscript. Liyan Liu, MSc, participated in study design, data acquisition, statistical analysis, and drafting and critical review of the manuscript. Neal Shorstein, MD, participated in critical review and drafting and submission of the final manuscript. James Carolan, MD, participated in critical review and in drafting and submission of the final manuscript. Laura B. Amsden, MSW, MPH, participated in critical review and drafting and submission of the final manuscript. Lisa Herrinton, PhD, participated in study design, critical review, and drafting and submission of the final manuscript. All authors have given final approval of the manuscript.

#### Financial Support

This project was funded by the National Eye Institute R01 EY027329. The project also used computer programs developed under earlier research grants provided by NEI R21 EY022989, Kaiser Permanente's Community Benefit program, and the Garfield Memorial Fund, Kaiser Permanente. These sponsors had no role in the design or conduct of this research.

#### How to Cite this Article

Alexeeff SE, Uong S, Liu L, et al. Development and validation of machine learning models: Electronic health record data to predict visual acuity after cataract surgery. *Perm J* 2020;25:20.188. DOI: <https://doi.org/10.7812/TPP/20.188>

#### References

- Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev* 1959 Jul;44:206-26.
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016 Dec;316(22):2402-10. DOI: <https://doi.org/10.1001/jama.2016.17216>, PMID:27898976.
- Du XL, Li WB, Hu BJ. Application of artificial intelligence in ophthalmology. *Int J Ophthalmol* 2018 Sep;11(9):1555-61. DOI: <https://doi.org/10.18240/ijo.2018.09.21>, PMID:30225234.
- Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019 Feb;103(2):167-75. DOI: <https://doi.org/10.1136/bjophthalmol-2018-313173>, PMID:30361278.
- Rohm M, Tresp V, Müller M, et al. Predicting visual acuity by using machine learning in patients treated for neovascular age-related macular degeneration. *Ophthalmology* 2018 Jul;125(7):1028-36. DOI: <https://doi.org/10.1016/j.ophtha.2017.12.034>, PMID:29454659.
- Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: A review. *Clin Exp Ophthalmol* 2019 Jan;47(1):128-39. DOI: <https://doi.org/10.1111/ceo.13381>, PMID:30155978.
- Congdon N, O'Colmain B, Klaver CC, et al. Causes and prevalence of visual impairment among adults in the United States. *Arch Ophthalmol* 2004 Apr;122(4):477-85. DOI: <https://doi.org/10.1001/archophth.122.4.477>, PMID:15078664.
- Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990-2020: A systematic review and meta-analysis. *Lancet Glob Health* 2017 Dec;5(12):e1221-34. DOI: [https://doi.org/10.1016/S2214-109X\(17\)30393-5](https://doi.org/10.1016/S2214-109X(17)30393-5), PMID:29032195.
- Moshirfar M, Milner D, Patel BC. Cataract surgery. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2020.
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Ann Intern Med* 2015 May;162(10):735-6. DOI: <https://doi.org/10.7326/L15-5093-2>, PMID:25984857
- Liu L, Herrinton LJ, Alexeeff S, et al. Visual outcomes after cataract surgery in patients with type 2 diabetes. *J Cataract Refract Surg* 2019 Apr;45(4):404-13. DOI: <https://doi.org/10.1016/j.jcrs.2018.11.002>, PMID:30638823.
- Liu L, Shorstein NH, Amsden LB, Herrinton LJ. Natural language processing to ascertain two key variables from operative reports in ophthalmology. *Pharmacoepidemiol Drug Saf* 2017 Apr;26(4):378-85. DOI: <https://doi.org/10.1002/pds.4149>.
- State vision screening and standards for license to drive. 2003. Accessed August 21, 2019. <https://lowvision.preventblindness.org/2003/06/06/state-vision-screening-and-standards-for-license-to-drive/>.
- Agency for Healthcare Research and Quality HcaUP. Clinical classifications software for Services and procedures. 2018. Accessed April 3, 2019. <https://www.hcup-us.ahrq.gov/toolsoftware/procedure/procedure.jsp>.
- James G, Witten D, Hastie T, Tibshirani R. Tree-based methods. In: An introduction to statistical learning with applications in R. 1<sup>st</sup> ed. New York, NY: Springer; 2013; p 303-36.
- Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann. Statist* 2001 Oct;29(5):1189-232. DOI: <https://doi.org/10.1214/aos/1013203451>.
- Pencina MJ, D'Agostino RB. Evaluating discrimination of risk prediction models: The c statistic. *JAMA* 2015 Sep;314(10):1063-4. DOI: <https://doi.org/10.1001/jama.2015.11082>, PMID:26348755.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982 Apr;143(1):29-36. DOI: <https://doi.org/10.1148/radiology.143.1.7063747>, PMID:7063747.
- Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3<sup>rd</sup> ed. New York, NY: Wiley; 2013.
- Muhtaseb M, Kalthoro A, Ionides A. A system for preoperative stratification of cataract patients according to risk of intraoperative complications: A prospective analysis of 1441 cases. *Br J Ophthalmol* 2004 Oct;88(10):1242-6. DOI: <https://doi.org/10.1136/bjo.2004.046003>, PMID:15377542.
- Modjtahedi BS, Hull MM, Adams JL, Munz SJ, Luong TQ, Fong DS. Preoperative vision and surgeon volume as predictors of visual outcomes after cataract surgery. *Ophthalmology* 2019 Mar;126(3):355-61. DOI: <https://doi.org/10.1016/j.ophtha.2018.10.030>, PMID:30808486
- Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: Challenges and potential approaches. *Med Care* 2010 Jun;48(6 Suppl):S114-20. DOI: <https://doi.org/10.1097/MLR.0b013e3181d8be3>, PMID:20473199.
- Zhao YQ, Zeng D, Laber EB, Kosorok MR. New statistical learning methods for estimating optimal dynamic treatment regimes. *J Am Stat Assoc* 2015 Jul;110(510):583-98. DOI: <https://doi.org/10.1080/01621459.2014.937488>, PMID:26236062.