PREDICTING FAILURES OF MOLTENO AND BAERVELDT GLAUCOMA DRAINAGE DEVICES USING MACHINE LEARNING MODELS

Paul Morrison[1], Maxwell Dixon[2],
Arsham Sheybani[2], Bahareh Rahmani[1, 3]

[1]Fontbonne University, Mathematics and Computer Science Department, St. Louis, MO
[2]Washington University, Department of Ophthalmology and Visual Sciences, St. Louis, MO
[3]Maryville University, Mathematics and Computer Science Department, St. Louis, MO

ABSTRACT

The purpose of this retrospective study is to measure machine learning models' ability to predict glaucoma drainage device (GDD) failure based on demographic information and preoperative measurements. The medical records of sixty-two patients were used. Potential predictors included the patient's race, age, sex, preoperative intraocular pressure (IOP), preoperative visual acuity, number of IOP-lowering medications, and number and type of previous ophthalmic surgeries. Failure was defined as final IOP greater than 18 mm Hg, reduction in IOP less than 20% from baseline, or need for reoperation unrelated to normal implant maintenance. Five classifiers were compared: logistic regression, artificial neural network, random forest, decision tree, and support vector machine. Recursive feature elimination was used to shrink the number of predictors and grid search was used to choose hyperparameters. To prevent leakage, nested cross-validation was used throughout. Overall, the best classifier was logistic regression.

With a small amount of data, the best classifier was logistic regression, but with more data, the best classifier was the random forest. All five classification methods discussed at this research confirm that race effects on failure glaucoma drainage. Use of topical beta-blockers pre-operatively is related to device failure. In treating glaucoma medically, prostaglandin equivalents are often first-line with beta-blockers used second-line or as a reasonable alternative first-line agent.

1. INTRODUCTION

GDDs are typically utilized in the management of glaucoma refractory to maximal medical therapy or prior failed glaucoma surgery. The devices can be divided into two categories: non-valved (e.g. Molteno and Baerveldt) and valved (e.g. Ahmed). Non-valved GDDs have been shown to be more effective in lowering IOP and have lower rates of reoperation than valved GDDs, but experience more frequent failure leading to dangerously low IOP or reduction of vision to the point of absolute blindness.[1] However, there have been no studies directly comparing the two main types of non-valved GDDs despite their significantly different device profiles and implantation technique.

David C. Wyld et al. (Eds): COMIT, SIPO, AISCA, MLIOB, BDHI - 2020
pp. 109-120, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101610

The accuracy of machine learning models in predicting GDD outcomes based on a minimal feature setprovides a unique strategy to understand differences between these devices. Previous studies have predicted individual outcomes for other ophthalmic surgeries using machine learning and logistic regression. Achiron et al. used extreme gradient boosted decision forests to predict the efficacy (final VA/starting VA) of refractive surgery.[2] Rohm et al. compared five algorithms to predict postoperative VA at 3 and 12 months in patients with neovascular age-related macular degeneration.[3] Valdes-Mas et al. compared an artificial neural network with a decision tree to predict the occurrence of astigmatism and found the neural network superior.[4] Mohammadi et al. used neural networks to predict the occurrence of posterior capsule opacification after phacoemulsification.[5] Gupta et al. used linear regression to determine post-operative visual acuity based on patient demographics and pre-operative predictors.[6] Koprowski et al. compared hundreds of artificial neural network topologies to predict corneal power after corneal refractive surgery.[7] McNabb et al. used OCT (optical coherence tomography) to predict corneal power change after laser refractive surgery.[8] Bowd et al. used Relevance Vector Machines to predict visual field progression in glaucoma patients based on SAP (standard automated perimetry) and CSLO (confocal scanning laser ophthalmoscope) measurements.9 More recently, Lee et al. used random forests and extremely randomized trees to predict glaucoma progression specifically in pediatric patients, also using SAP data.10 Similar to our own study, Baxter et al. used machine learning techniques (random forest, artificial neural network, and logistic regression) to predict surgical intervention for POAG (primary open-angle glaucoma) based on structured EHR data. They identified high blood pressure as a factor increasing the likelihood of surgical intervention, and several categories of ophthalmic and non-ophthalmic factors decreasing the likelihood of surgery.11 In contrast to their study, this one predicts implant failure instead of the need for surgical intervention, and includes more classifiers and types of glaucoma.

When comparing the Molteno and Baerveldt GDDs, demographic predictors included race, sex, and age at surgery. A total of seven clinical predictors were considered including:

Implant Type: Identified by type of implant (Molteno or Baerveldt) and implant plate surface area.
VA (logMAR): "Logarithm of the Minimum Angle of Resolution." A more reproduceable visual acuity measurement often used in research. As Snellen visual acuity is more often collected in the clinic setting, conversion to logMAR allows easier statistical analysis.
IOP: Intraocular pressure. Elevated IOP is the major risk factor for development of glaucoma.
Number of medications: Include usage of beta-blockers, alpha-adrenergicagonists, prostaglandin analogs, or carbonic anhydrase inhibitors. The number of medications was calculated from patient records at each visit.
Number of previous surgeries: Glaucoma drainage implants are typically placed after less-invasive treatments fail but may incidentally be utilized following other ophthalmic surgeries (e.g. phacoemulsification of cataracts or retinal surgeries).
Type of previous surgeries: Include phacoemulsification or extracapsular cataract extraction (ECCE), trabeculectomy, pars planavitrectomy, penetrating keratoplasty, Ex-PRESS shunt, iStent, or diode laser cyclophotocoagulation (dCPC).
Diagnosis: Causes for glaucoma included open-angle, neovascular, uveitic, angle-closure, secondary to trauma, secondary to PKP, pseudoexfoliation, and combined mechanism.

At this study, we describe data in section 2. The methodology and explaining five classification methods come after in section 3. This section covers describing logistic regression, support vector machine, random forest, neural network and decision tree. Section 3 covers the information of train and test datasets as well. Results and discussion and right after conclusion are in section 4 and section 5.

2. DATA DESCRIPTION

Of 62 patients analyzed, 26 (41%) were determined to have device failure. Follow-up time was 573±245.45 days, (range 133-1037 days). Patient samples were balanced between male and female. White race was three times more common than Black race and there was only one Asian patient (Tables 1 and 2). Implant failure was defined as final IOP was greater than or equal to 18, less than 20% reduction from pre-operative levels, or if repeat surgery for glaucoma was required (this did not include in-clinic procedures that did not indicate failure of the device itself). By the last recorded appointment, 35% (22) of patients had a failing IOP and 19% (12) required additional surgery. No patients in this group experienced loss of light perception.

Table 1: Number of participants by race and sex

Race	Male	Female	Total
Asian	1	0	1
Black	8	8	16
White	24	21	45
Total	33	29	62

Table 2: Average Age at Surgery by Race and Sex

Race	Male	Female	Average
Asian	72±0 (n = 1)	-	72±0 (1 total)
Black	61.9±9.7 (n = 8)	68.9±6.53 (8 total)	65.4±8.1 (28 total)
White	65.8±12.2 (n = 24)	69.4±6.53 (21 total)	67.6±9.4 (58 total)
Average	65.0±11.22 (33total)	69.2±6.53 (29 total)	67.1±8.9 (62 total)

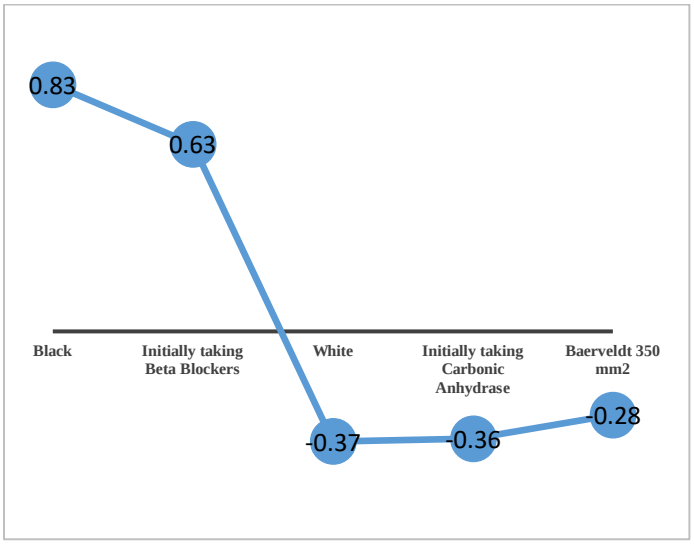
A total of 42 patients received a BaerveldtGDD (67%) and 20 received a Molteno implant (27%). Forty-eight (77%) patients had surgery prior to placement of a GDD. Twelve patients (19%) required repeat surgery after initial placement of a GDD. Open-angle glaucoma was the most common underlying diagnosis (61%, n = 38) with combined mechanism (11%, n = 5) and chronic angle-closure (8%, n = 5) being less common. There were also individual patients with either neovascular, uveitic, traumatic, or pseudo exfoliation glaucoma. A diagnosis of "Other" was given for 8% of the patients, which indicated a singular diagnosis was not able to be determined from chart review.

3. METHODOLOGY

All models in this study were validated using three-fold stratified cross validation, and all but the neural net were developed using recursive feature elimination and grid searches. To prevent data leakage, final validation, grid searching, and feature selection were performed in separate cross validation loops, as recommended by Krstajic et al.[12] In the outer loop, the final model was tested; in the middle loop, the best hyper parameters were chosen; and in the inner loop, the best feature subsets were selected. Within each loop, three-fold stratified cross-validation was used. Scaling and centering for continuous variables was performed as part of the model fitting procedure. The Logistic Regression, SVM, Decision Tree, and Random Forest classifiers were implemented in Python using Scikit-Learn,[13] and the Neural Network classifier was implemented in R using the caret package.[14]

3.1. Logistic Regression

Logistic regression, traditionally used for modeling, determines the class of each input variable by multiplying each feature by a constant, adding a bias term, and applying the logistic function. Any outcome above 0.5 is rounded to 1; any outcome below 0.5 is rounded to 0. The optimal logistic regression classifier used L2 regularization and a C parameter of 1, had an accuracy of 0.66±0.08 and a ROC (Receiver Operating Characteristic) of 0.67±0.08. Based on the coefficients of the logistic regression model listed in Table 3, Black race and initially taking beta-blockers is associated with implant failure.

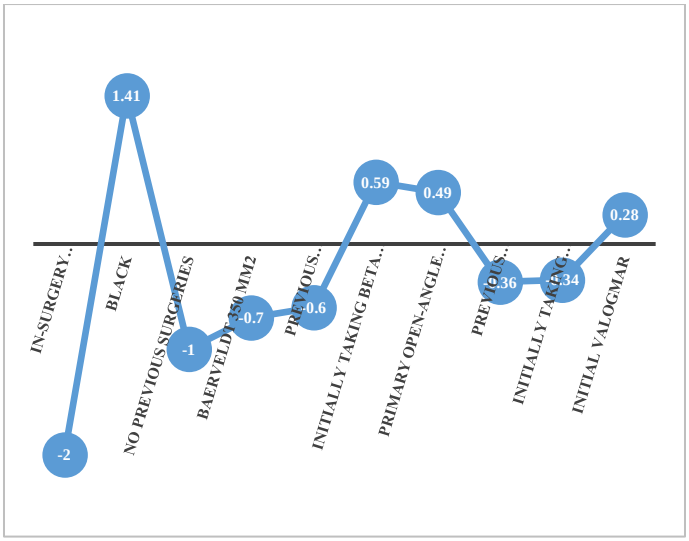


Feature	Sign	Coefficient
Black race	+	0.83
Initially taking Beta Blockers	+	0.63
White	-	0.37
Initially taking Carbonic Anhydrase	-	0.36
Baerveldt 350 mm2	-	0.28

Table 3: Feature Coefficients of Logistic Regression Classifier

3.2. Support Vector Machine (SVM)

A Support Vector Machine uses several data points (support vectors) to find the hyperplanes separating data classes that allow identification of a hyperplane giving the maximum margin. The best classifier had a cost parameter of 0. Table 4 shows the feature coefficients of SVM classifier.[15]



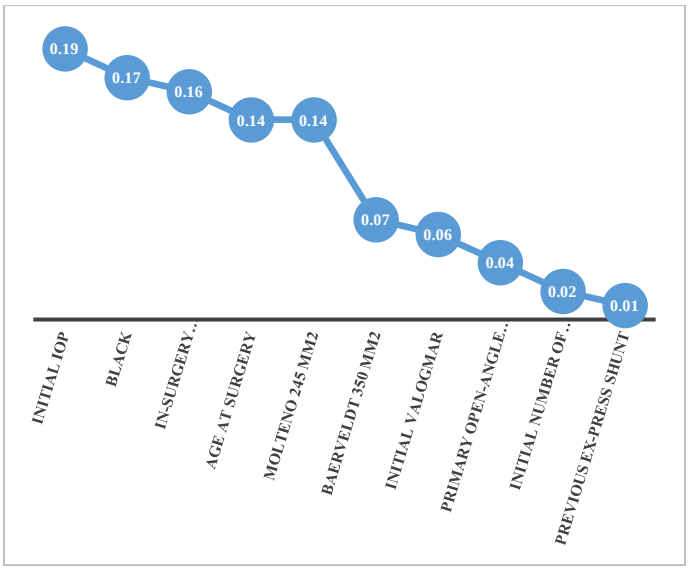
Feature	Sign	Coefficient
Combined tube placement and phacoemulsification	-	2
Black race	+	1.41
No previous surgeries	-	1
Baerveldt 350 mm^2	-	0.7
Previous Phacoemulsification or ECCE	-	0.6
Initially taking beta blockers	+	0.59
Primary Open-Angle Glaucoma	+	0.49
Previous Trabeculectomy	-	0.36
Initially taking carbonic anhydrase inhibitor	-	0.34
Initial VA (logMAR)	+	0.28

Table 4: Feature Coefficients of SVM Classifier

Like regression, race and initially taking beta-blockers have the most weight in causing implant failure. Primary Open-Angle Glaucoma show possibility of implant failure too. The SVM classifier had an accuracy of 0.61%±0.03 and a ROC AUC of 0.62±0.03.

3.3. Decision Tree

A decision tree repeatedly picks a threshold to divide data until it places all data items in groups (mostly) of the same class. First, it finds the threshold for all features dividing data most cleanly. Then it chooses features producing the cleanest split and repeats the process separately for the data on each side of the split. The algorithm stops when the data divide into pure groups or when the number of points in each group is too small to divide further without overfitting[15]. We used a minimum of three data points per leaf nodeand the Gini impurity measure.



Feature	Importance
Initial IOP	0.19
Black	0.17
In-Surgery Phacoemulsification	0.16
Age at Surgery	0.14
Molteno 245 mm2	0.14
Baerveldt 350 mm2	0.07
Initial VA (logMAR)	0.06
Primary Open-Angle Glaucoma	0.04
Initial Number of Medications	0.02
Previous Ex-Press Shunt	0.01

Table 5: Patient Groups Created Using Decision Tree

As described in Tables 5 and 6 and Figure 1, IOP and race are the most important factor in device failure. Combined GDD placement and phacoemulsification, age, and usage of the 245 mm2 Molteno GDD are other factors. Despite regression and SVM, initially taking beta-blockers does not appear in decision tree. The decision tree's overall accuracy was 0.5±0.05, and its ROC AUC was 0.45%±0.04. Low accuracy makes the efficiency of this method less than previous methods.

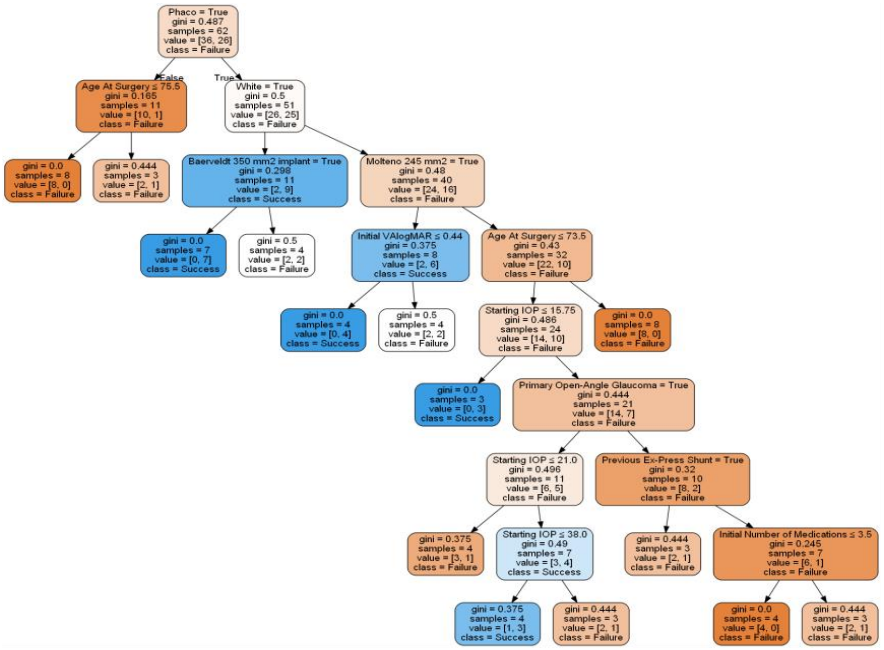


Figure 1: Decision Tree of Glaucoma Data

Table 6: Patient Groups Created Using Decision Tree

	Group Characteristics					Patient Features									
Group Number	Decision	Count	Number of falsuccesses	Number of failures	Gini	In-Surgery Phacoemulsification	Black	Age at Surgery	Molteno 245 mm2	Initial IOP	Primary Open-Angle Glaucoma	Previous Ex-Press Shunt	Initial Number of Medications	Initial VAlogMAR	Baerveldt 350 mm2
0	Success	8	8	0	0	False	Any	<= 75.5	Any	Any	Any	Any	Any	Any	Any
1	Success	3	2	1	0.4	False	Any	> 75.5	Any	Any	Any	Any	Any	Any	Any
2	Failure	7	0	7	0	True	False	Any	Any	Any	Any	Any	Any	Any	False
3	Success	4	2	2	0.5	True	False	Any	Any	Any	Any	Any	Any	Any	True
4	Failure	4	0	4	0	True	True	Any	False	Any	Any	Any	Any	<= 0.44	Any
5	Success	4	2	2	0.5	True	True	Any	False	Any	Any	Any	Any	> 0.44	Any
6	Failure	3	0	3	0	True	True	<= 73.5	True	<= 15.75	Any	Any	Any	Any	Any
7	Success	4	3	1	0.4	True	True	<= 73.5	True	21.0 <= x < 15.75	False	Any	Any	Any	Any
8	Failure	4	1	3	0.4	True	True	<= 73.5	True	38.0 <= x < 21.0	False	Any	Any	Any	Any
9	Success	3	2	1	0.4	True	True	<= 73.5	True	> 38.0	False	Any	Any	Any	Any
10	Success	3	2	1	0.4	True	True	<= 73.5	True	> 15.75	True	False	Any	Any	Any
11	Success	4	4	0	0	True	True	<= 73.5	True	> 15.75	True	True	<= 3.5	Any	Any
12	Success	3	2	1	0.4	True	True	<= 73.5	True	> 15.75	True	True	> 3.5	Any	Any
13	Success	8	8	0	0	True	True	> 73.5	True	Any	Any	Any	Any	Any	Any

3.4. Artificial Neural Network

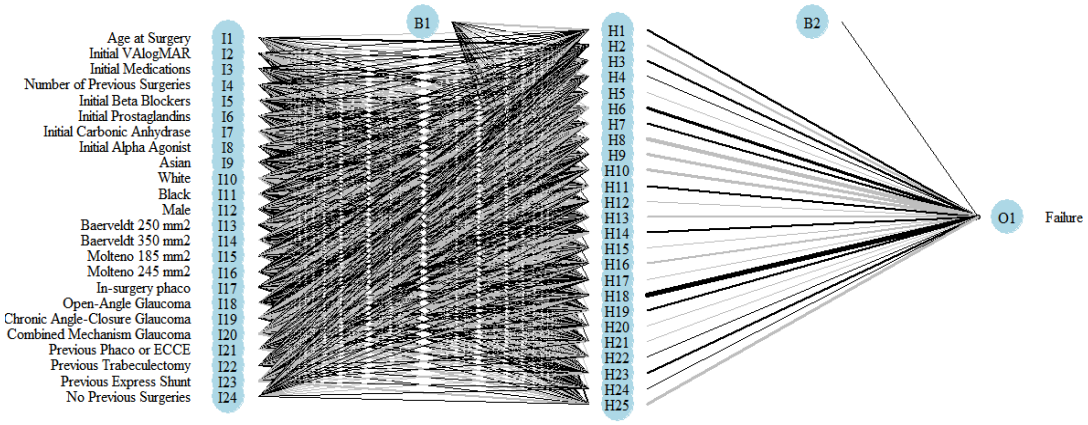


Figure 2: Neural Network Architecture

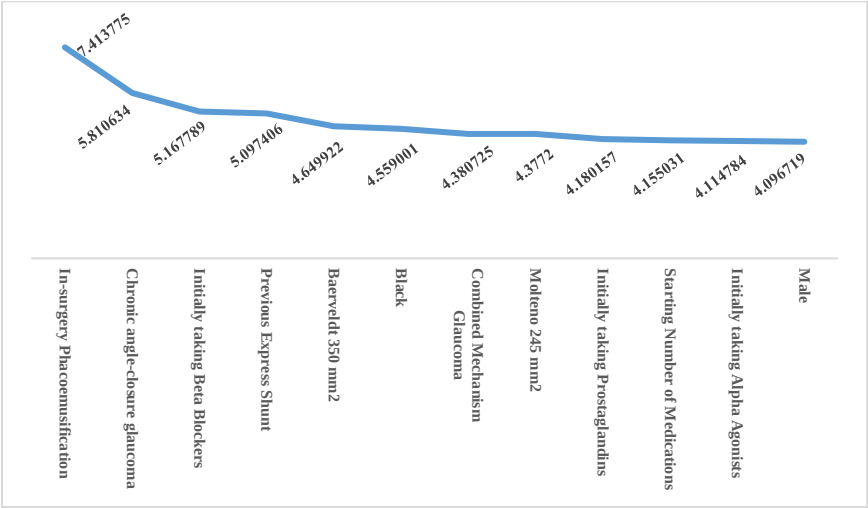


Table 7: Importance of Features in Neural Network Classifier

Feature	Importance	Feature	Importance
In-surgery Phacoemusification	7.413775	No previous surgeries	3.987791
Chronic angle-closure glaucoma	5.810634	Open-Angle Glaucoma	3.899654
Initially taking Beta Blockers	5.167789	Previous Phaco or ECCE	3.894626
Previous Express Shunt	5.097406	Molteno 185 mm2	3.866003
Baerveldt 350 mm2	4.649922	Number of Previous Surgeries	3.782343
Black	4.559001	Age at Surgery	3.535026
Combined Mechanism Glaucoma	4.380725	Starting VAlogMAR	3.53318
Molteno 245 mm2	4.3772	Initially taking Carbonic Anhydrase	3.298634
Initially taking Prostaglandins	4.180157	Previous Trabeculectomy	3.234256
Starting Number of Medications	4.155031	White	2.994674
Initially taking Alpha Agonists	4.114784	Baerveldt 250 mm2	2.98814
Male	4.096719	Initial IOP	2.982531

A feed-forward artificial neural network imitates biological neural tissue using sequential layers of "neurons" that transform the underlying data and pass it on to the next layer. The root of a neural net is a perceptron: two or more inputs connected to a neuron, which then multiplies each input by a weight, adds an intercept, and applies an output function to the result. In a single-layer neural network, many perceptrons extract information from the underlying features, and a final neuron (or more for multiclass classification) combines the output from these nodes[15]. A single-layer network with 25 hidden nodes (Fig. 2) was trained on the data. Combined GDD placement and phacoemulsification was indicated as the most important factor in failure. Chronic angle-closure glaucoma and initially taking beta-blockers were also associated with therapy failure, though to a lesser extent. Race with importance = 4.6 shows high effect on failure. The accuracy was 0.53±0.11 and the ROC AUC was 0.52±0.10.

3.5. Random Forest

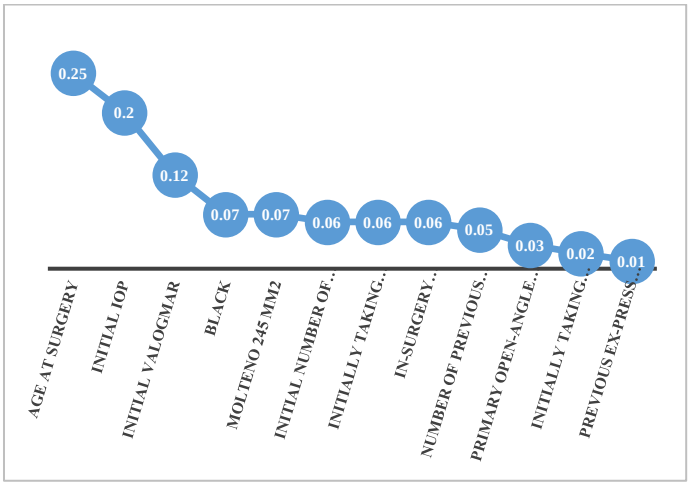


Table 8: Importance of Features in Random Forest Classifier

Feature	Importance
Age at Surgery	0.25
Initial IOP	0.2
Initial VAlogMAR	0.12
Black	0.07
Molteno 245 mm2	0.07
Initial Number of Medications	0.06
On beta-blocker prior to surgery	0.06
Combined device and phacoemulsification	0.06
Number of Previous Surgeries	0.05
Primary Open-Angle Glaucoma	0.03
On CAI prior to surgery	0.02
Previous Ex-Press Shunt	0.01

A random forest averages the predictions of multiple decision trees trained on subsets of the data[15]. Using ten decision trees, the algorithm identified age at surgery, initial IOP, and visual acuity as the most important factors determining device failure. Race and utilization of the larger Molteno device (245 mm^2) were associated with device failure, though to a lesser degree. The overall ROC AUC was 0.58±0.1, and the overall accuracy was 0.58±0.13.

4. CONCLUSION

Table 9: Accuracy and ROC score across all models. Red: high effect, Orange: low effect

Methods	Black race	Beta Blockers	Age	Initial IOP	Molteno 245 mm2	Cataract removal
Logistic Regression						
SVM						
Random Forest						
Neural Network						
Decision Tree						

Comparing results from different models identified Black race as the strongest factor associated with device failure. This finding aligns with existing research in the ophthalmology literature.[16] Such failure rates are believed to be due to genetic differences in wound healing and proliferation of fibrovascular tissue.[17] Use of topical beta-blockers pre-operatively was also associated with device failure. In treating glaucoma medically, prostaglandin analogs are often first-line with beta-blockers used second-line or as a reasonable alternative first-line agent. Alpha-agonists and carbonic anhydrase inhibitors are often added next, though they can cause intolerable allergic reactions and discomfort on instillation, respectively.[18] These side effects can lead to drop intolerance and serve as an impetus for surgery. Therefore, it is perhaps not unsurprising that patients would be on beta-blockers when surgical intervention is needed as they are usually well-tolerated in those without respiratory problems. Nonetheless, beta-blockers association with implant failure in several models may be an area of further investigation. Placement of the larger Molteno GDD was associated with device failure, though this was a weaker association and found in weaker models. Again, this warrants further investigation given the devices function similarly. Lastly, age, increased IOP, and phacoemulsification at the time of GDD implantation were associated with failure in weaker models. Overall, the most accurate model was logistic regression, followed by a support vector machine model with a linear kernel. Our findings suggest machine learning techniques can accurately determine important features leading to failure of GDD implants from a large dataset of common clinical descriptors.

Table 10: Accuracy and ROC score across all models.

Methods	**ROC AUC**	**Accuracy**
Logistic Regression	0.67±0.08	0.66±0.08
SVM	0.62±0.03	0.61±0.03
Random Forest	0.53±0.10	0.58±0.13
Neural Network	0.52±0.10	0.53±0.11
Decision Tree	0.45±0.04	0.50±0.05

Based on this study, we realized that race and Beta blockers are two factors that may cause failure. Considering these two attributes logistic regression and SVM are the most accurate methods to predict the failure.

The restriction of this study is the low number of cases to investigate. Adding more data make the accuracy higher.

AUTHOR CONTRIBUTIONS

PM: Preprocessed data, wrote the code in R and Python, investigated and analyzed the data, and wrote the paper. MD: Prepared the data and wrote the data description. AS: Provided the data and advised the biological and learning approaches. BR: Supervised the computational approaches and conceived the idea of the paper. All authors commented on this paper.

REFERENCES

[1] D. L. Budenz et al., "Five-Year Treatment Outcomes in the Ahmed Baerveldt Comparison Study," Ophthalmology, vol. 122, no. 2, pp. 308–316, Feb. 2015, doi: 10.1016/j.ophtha.2014.08.043.

[2] A. Achiron et al., "Predicting Refractive Surgery Outcome: Machine Learning Approach With Big Data," J Refract Surg, vol. 33, no. 9, pp. 592–597, Sep. 2017, doi: 10.3928/1081597X-20170616-03.

[3] M. Rohm et al., "Predicting Visual Acuity by Using Machine Learning in Patients Treated for Neovascular Age-Related Macular Degeneration," Ophthalmology, vol. 125, no. 7, pp. 1028–1036, Jul. 2018, doi: 10.1016/j.ophtha.2017.12.034.

[4] M. A. Valdes-Mas, J. D. Martin, M. J. Ruperez, C. Peris, and C. Monserrat, "Machine learning for predicting astigmatism in patients with keratoconus after intracorneal ring implantation," in IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, Spain, Jun. 2014, pp. 756–759, doi: 10.1109/BHI.2014.6864474.

[5] S.-F. Mohammadi et al., "Using artificial intelligence to predict the risk for posterior capsule opacification after phacoemulsification:," Journal of Cataract & Refractive Surgery, vol. 38, no. 3, pp. 403–408, Mar. 2012, doi: 10.1016/j.jcrs.2011.09.036.

[6] M. Gupta, P. Gupta, P. K. Vaddavalli, and A. Fatima, "Predicting Post-operative Visual Acuity for LASIK Surgeries," in Advances in Knowledge Discovery and Data Mining, vol. 9651, J. Bailey, L. Khan, T. Washio, G. Dobbie, J. Z. Huang, and R. Wang, Eds. Cham: Springer International Publishing, 2016, pp. 489–501.

[7] R. Koprowski, M. Lanza, and C. Irregolare, "Corneal power evaluation after myopic corneal refractive surgery using artificial neural networks," BioMed EngOnLine, vol. 15, no. 1, p. 121, Dec. 2016, doi: 10.1186/s12938-016-0243-5.

[8] R. P. McNabb, S. Farsiu, S. S. Stinnett, J. A. Izatt, and A. N. Kuo, "Optical Coherence Tomography Accurately Measures Corneal Power Change from Laser Refractive Surgery," Ophthalmology, vol. 122, no. 4, pp. 677–686, Apr. 2015, doi: 10.1016/j.ophtha.2014.10.003.

[9] C. Bowd et al., "Predicting Glaucomatous Progression in Glaucoma Suspect Eyes Using Relevance Vector Machine Classifiers for Combined Structural and Functional Measurements," Invest. Ophthalmol. Vis. Sci., vol. 53, no. 4, p. 2382, Apr. 2012, doi: 10.1167/iovs.11-7951.

[10] J. Lee, Y. K. Kim, J. W. Jeoung, A. Ha, Y. W. Kim, and K. H. Park, "Machine learning classifiers-based prediction of normal-tension glaucoma progression in young myopic patients," Jpn J Ophthalmol, vol. 64, no. 1, pp. 68–76, Jan. 2020, doi: 10.1007/s10384-019-00706-2.

[11] S. L. Baxter, C. Marks, T.-T. Kuo, L. Ohno-Machado, and R. N. Weinreb, "Machine Learning-Based Predictive Modeling of Surgical Intervention in Glaucoma Using Systemic Data From Electronic Health Records," American Journal of Ophthalmology, vol. 208, pp. 30–40, Dec. 2019, doi: 10.1016/j.ajo.2019.07.005.

[12] D. Krstajic, L. J. Buturovic, D. E. Leahy, and S. Thomas, "Cross-validation pitfalls when selecting and assessing regression and classification models," J Cheminform, vol. 6, no. 1, p. 10, Dec. 2014, doi: 10.1186/1758-2946-6-10.

[13] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," arXiv:1201.0490 [cs], Jun. 2018, Accessed: Nov. 22, 2020. [Online]. Available: http://arxiv.org/abs/1201.0490.

[14] M. Kuhn, "Building Predictive Models in R Using the caret Package," J. Stat. Soft., vol. 28, no. 5, 2008, doi: 10.18637/jss.v028.i05.

[15] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to data mining, 1st ed. Boston: Pearson Addison Wesley, 2006.

[16] "The advanced glaucoma intervention study (AGIS)*113. Comparison of treatment outcomes within race: 10-year results," Ophthalmology, vol. 111, no. 4, pp. 651–664, Apr. 2004, doi: 10.1016/j.ophtha.2003.09.025.

[17] D. Broadway, I. Grierson, and R. Hitchings, “Racial differences in the results of glaucoma filtration surgery: are racial differences in the conjunctival cell profile important?,” British Journal of Ophthalmology, vol. 78, no. 6, pp. 466–475, Jun. 1994, doi: 10.1136/bjo.78.6.466.
[18] K. Inoue, “Managing adverse effects of glaucoma medications,” OPTH, p. 903, May 2014, doi: 10.2147/OPTH.S44708.