Received: 5 December 2022 | Accepted: 31 January 2023

DOI: 10.1111/aos.15648

Acta Ophthalmologica

ORIGINAL ARTICLE

Development of machine learning models to predict posterior capsule rupture based on the EUREQUO registry

Ron J. M. A. Triepels[1] | Maartje H. M. Segers[2] | Paul Rosen[3] | Rudy M. M. A. Nuijts[2] | Frank J. H. M. van den Biggelaar[2] | Ype P. Henry[4] | Ulf Stenevi[5] | Marie-José Tassignon[6] | David Young[7] | Anders Behndig[8] | Mats Lundström[9] | Mor M. Dickman[2]

[1]Department of Data Analytics and Digitalisation, Maastricht University, Maastricht, the Netherlands

[2]University Eye Clinic, Maastricht University Medical Center+, Maastricht, the Netherlands

[3]Department of Ophthalmology, Oxford Eye Hospital, Oxford, UK

[4]Department of Ophthalmology, Amsterdam UMC, Amsterdam, the Netherlands

[5]Department of Ophthalmology, Sahlgrenska University Hospital, Göteborg, Sweden

[6]Department of Ophthalmology, Antwerp University Hospital, Edegem, Belgium

[7]Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK

[8]Department of Clinical Sciences, Ophthalmology, Umeå University, Umeå, Sweden

[9]Department of Clinical Sciences, Ophthalmology, Lund University, Lund, Sweden

Correspondence
Ron J.M.A. Triepels, Department of Data Analytics and Digitalisation, Maastricht University, Tongersestraat 53, 6211 LM, Maastricht, the Netherlands.
Email: r.triepels@maastrichtuniversity.nl

Abstract

Purpose: To evaluate the performance of different probabilistic classifiers to predict posterior capsule rupture (PCR) prior to cataract surgery.

Methods: Three probabilistic classifiers were constructed to estimate the probability of PCR: a Bayesian network (BN), logistic regression (LR) model, and multi-layer perceptron (MLP) network. The classifiers were trained on a sample of 2 853 376 surgeries reported to the European Registry of Quality Outcomes for Cataract and Refractive Surgery (EUREQUO) between 2008 and 2018. The performance of the classifiers was evaluated based on the area under the precision-recall curve (AUPRC) and compared to existing scoring models in the literature. Furthermore, direct risk factors for PCR were identified by analysing the independence structure of the BN.

Results: The MLP network predicted PCR overall the best (AUPRC 13.1 ± 0.41%), followed by the BN (AUPRC 8.05 ± 0.39%) and the LR model (AUPRC 7.31 ± 0.15%). Direct risk factors for PCR include preoperative best-corrected visual acuity (BCVA), year of surgery, operation type, anaesthesia, target refraction, other ocular comorbidities, white cataract, and corneal opacities.

Conclusions: Our results suggest that the MLP network performs better than existing scoring models in the literature, despite a relatively low precision at high recall. Consequently, implementing the MLP network in clinical practice can potentially decrease the PCR rate.

KEYWORDS
artificial intelligence, Bayesian network, cataract surgery, logistic regression, machine learning, multi-layer perceptron, posterior capsule rupture

1 | INTRODUCTION

Cataract remains the leading cause of blindness, affecting an estimated 95 million people worldwide in 2020 (Steinmetz et al., 2021). The standard treatment of cataract is to surgically remove the crystalline lens and replace it with an artificial intraocular lens (IOL). Over the past decade, this procedure has improved with more advanced surgical techniques allowing cataract surgery to become a minimally invasive surgery with fast visual recovery, good visual outcomes, and few complications. Nonetheless, it has been reported that 0.2%–1.8% of cataract surgeries are complicated by a posterior capsule rupture (PCR) (Terveen et al., 2022; Ti et al., 2014), an intraoperative complication in which a breach occurs in the posterior capsule of the crystalline lens. This complication is feared because of its severe and potential sight-threatening consequences, such as failure to implant an IOL, endophthalmitis, and cystoid macular oedema.

To mitigate the risks of PCR, surgeons typically assess the probability of PCR prior to surgery. Risk assessment is performed by an experienced surgeon, possibly guided

Ron J.M.A. Triepels is a member of the Dutch Ophthalmological Societies.

by a scoring system, which determines the probability of PCR based on a subjective weighting of known risk factors by severity (Han et al., 2019; Muhtaseb et al., 2004). The outcome of the risk assessment can contribute to a better allocation of patients to junior or experienced surgeons and better communication of risks to patients. Although risk assessments have been shown to reduce the occurrence of PCR (Han et al., 2019), clinical judgement is subjective and dependent on the experience of the surgeon and designer of the scoring system.

The application of machine learning may prove helpful in estimating the probability of PCR more reliably and objectively. Instead of manually weighting the severity of known risk factors, a probabilistic classifier can be constructed based on a large data set of cataract surgeries to predict PCR. In this process, the classifier identifies risk factors for PCR and automatically determines their optimal weighting. When a cataract procedure is planned, the classifier can estimate the probability of PCR and, depending on whether the probability is higher than or equal to a given probability threshold, classify the procedure as high risk for PCR or low risk for PCR otherwise. Machine learning has already been applied to aid in other assessments in ophthalmology, such as the detection of edema on optical coherence tomography images (Potapenko et al., 2022) and visual improvement after macular hole surgery (Lachance et al., 2022). However, the application of machine learning for the prediction of PCR has, to the best of our knowledge, not been studied before.

In this study, we develop various probabilistic classifiers to predict PCR prior to cataract surgery. This includes a Bayesian network (BN), logistic regression (LR) model, and multi-layer perceptron (MLP). These classifiers are constructed based on a large data set derived from the European Registry of Quality Outcomes for Cataract and Refractive Surgery (EUREQUO) (Lundstrom et al., 2014). The performance of the classifiers is evaluated and compared. Furthermore, risk factors for PCR identified by the classifiers are analysed.

2 | METHODS

2.1 | Data source

A large data set of cataract surgeries derived from the EUREQUO was used to construct and evaluate the classifiers. The EUREQUO is a large international data repository that collects, stores, and processes data of cataract surgeries carried out at clinics located mainly in Europe. The data set consists of 19 features that describe surgeries, including patient demographics (age, gender, right eye, and preoperative best-corrected visual acuity (BCVA)), operation type (including phacoemulsification, extracapsular cataract extraction, and laser-assisted cataract surgery), operation characteristics (year of surgery, anaesthesia, and target refraction), ocular comorbidities (amblyopia, macular degeneration, glaucoma, diabetic retinopathy, other ocular comorbidities), and complicating ocular comorbidities (small pupil, white cataract, corneal opacities, pseudoexfoliation, previous vitrectomy, previous corneal refractive surgery, and other complicating ocular comorbidities). Moreover, the data set reports whether PCR occurred at each surgery. In our study, PCR is defined as an intraoperative tear in the posterior capsule with or without zonular dialysis and vitreous loss.

2.2 | Bayesian network

A BN is a probabilistic graphical model where the joint probability distribution over a set of random variables factorizes according to a directed and acyclic independence graph (Koller & Friedman, 2009). Nodes in the independence graph represent random variables, whereas edges between the nodes represent conditional dependencies. A convenient property of a BN is that it allows modelling the joint probability distribution over the variables efficiently by the estimation of less and more reliable (conditional) probabilities. In addition, a BN can deal with missing data during model construction. This latter property is particularly useful since about 40% of all surgeries in the data set have at least one missing feature. Features with the most missing values were anaesthesia (39.7%), target refraction (37.4%), and right eye (5.5%). Possible explanations for this missing data are that not all clinics report outcome data to the EUREQUO and some features are optional.

We used R package bnlearn (Scutari, 2009) to construct the independence graph of the BN from the data set and estimate its model parameters. A limitation of the package is that it cannot model conditional dependencies of discrete nodes having one or more numerical parents at the time of writing. To avoid unnecessary restrictions on the independence graph of the BN, we transformed all numerical features to discrete features based on bins used in previous studies. The preoperative BCVA was discretized according to Lundström et al. (2017), and the target refraction was discretized according to Segers et al. (2022).

The BN was constructed by structural expectation–maximization (EM) (Friedman, 1997). The structural EM algorithm started with the empty independence graph and estimated the parameters of the network based on the complete cases in the data set. In the expectation step, missing features in the data set were imputed by performing inference based on the parent configurations of the nodes in the current independence graph. In the maximization step, a network structure search was performed, and the parameters corresponding to the obtained structure were estimated from the data set. Network structure search was performed by minimizing the Bayesian information criteria (Schwarz, 1978) using hill climbing with the current independence graph as the initial graph. Parameter estimation was done by Bayesian parameter estimation with an imaginary sample size of 1. The structural EM algorithm kept alternating between the expectation and maximization step until the independence graph of the network converged.

Once the BN was constructed, we applied Bayesian inference based on likelihood weighting (Fung & Chang, 1990) to estimate the probability of PCR. During the inference procedure, 500 cases of PCR were sampled from the network while taking into account the observed values of the remaining nodes in the network. Accordingly, the probability of PCR was estimated by normalizing the likelihoods of the sampled cases.

2.3 | Discriminative models

A BN is a generative model that can be employed to model the joint probability distribution of the features and the occurrence of PCR. Bayesian inference needs to be performed in turn on the network to estimate the posterior probability of PCR given an observation of the features. Alternatively, the same probability can be estimated directly by a discriminative model, for example, an LR model. It has been shown that discriminative models tend to generate more accurate predictions than generative models when constructed on a data set of sufficient size (Ng & Jordan, 2002). However, a limitation of discriminative models is that they typically cannot handle incomplete data, and a large portion of the data set would have to be discarded for model construction. To avoid this limitation, we constructed discriminative models based on the original data set with missing data imputed by the BN.

An LR model was constructed using Julia package Flux.jl (Innes, 2018). The weights of the model were initialized by the heuristic provided by Glorot and Bengio (2010), and the bias term was initially set to zero. Accordingly, the model was optimized by minimizing the binary cross-entropy between the probability of PCR estimated by the model and the corresponding point estimate derived from the data set. This optimization was carried out by ADAM (Kingma & Ba, 2014) using minibatches of 128 data points for a maximum of 100 epochs. After each epoch, the convergence of the optimization was checked by determining whether the relative improvement in the cross-entropy was smaller than 1e-8.

Besides the LR model, we also constructed a multilayer perceptron (MLP) network in a similar way. The MLP network has an architecture similar to the LR model, except it processes the features through an intermediate hidden layer consisting of a set of neurons with rectified linear activations (Glorot et al., 2011). This hidden layer enables the MLP to learn non-linear relationships between the features and the occurrence of PCR. The weights of the hidden layer were initialized by the heuristic provided by He et al. (2015), and the bias terms were initially set to zero. Dropout (Srivastava et al., 2014) was applied on the activations of the hidden layer to prevent overfitting.

2.4 | Performance evaluation

The performance of the classifiers was evaluated by a precision-recall (PR) curve. A PR curve can be used to visualize the trade-off of a probabilistic classifier between precision and recall. Precision is, in our application, the probability of PCR given that a classifier predicted PCR at a given threshold. Likewise, recall is the probability of a classifier predicting PCR at a given threshold given that PCR occurred. We estimated the area under the PR curve (AUPRC) for each classifier. The AUPRC can be interpreted as the precision of a classifier averaged over all possible thresholds, which provides a single measure independent of the threshold to compare the performance of a set of competing classifiers.

A PR curve is preferable over a more widely used receiver operating characteristic (ROC) curve in this application since PCR occurs only rarely (in about 1.1% of the surgeries in the data set). It is well known that a ROC curve for a severely imbalanced classification task is biased towards the over-represented class. This means that, in our application, an ROC curve is biased towards the correct prediction of the negative class (i.e., surgeries without PCR), while the correct prediction of the positive class (i.e., surgeries with PCR) is more relevant from a clinical perspective. A PR curve more accurately reflects the performance of the classifiers in predicting PCR by taking into account the rarity of PCR.

The evaluation of the classifiers was carried out by nested cross-validation. In the outer loop of the validation procedure, the data set was repeatedly partitioned in a training and test set based on fivefold cross-validation. The training set was used for the construction of the classifiers, and subsequently, the AUPRC of each classifier was estimated on the test set. In the inner loop of the validation procedure, a grid search was performed in combination with holdout sampling to estimate the hyper-parameters of the MLP network. Approximately 20% of the training set was randomly removed and put in a validation set for the estimation of the hyperparameters. Several MLP networks were constructed on the remaining training set while varying the number of neurons (4, 8, 12, 16, and 20) and dropout rate (5%, 10%, 15%, and 20%) of the hidden layer. The AUPRC of the networks was evaluated on the validation set, and the best configuration of neurons and dropout rate was selected accordingly.

2.5 | Analysis of risk factors

To identify risk factors for PCR, we analysed the independence graph of the BN. A useful property of the independence graph is that when no path connects two nodes, the nodes are marginally independent and do not influence each other's state. We can use this property to define risk factors for PCR. If there is no path connecting a node to PCR, then the node cannot influence the probability of PCR estimated by the BN. Likewise, if there is a path connecting a node to PCR, then the node can influence the probability of PCR estimated by the BN. All nodes that satisfy this latter requirement are risk factors for PCR.

Furthermore, we can distinguish between direct and indirect risk factors for PCR by inspecting the Markov blanket of PCR. The Markov blanket of a node consists of the node's parents, its children, and the parents of its children in the independence graph (Pearl, 1988). It can be shown that a node, say A, is conditionally independent of any other node not in the Markov blanket of A given the Markov blanket of A. This means that when data of all nodes is available, only those nodes in the Markov blanket of PCR can influence the probability of PCR estimated by the BN. These nodes are direct risk factors for PCR. Nodes for which a path connects them with PCR but are not in the Markov blanket of PCR are indirect risk factors for PCR. These nodes influence the probability of PCR estimated by the BN only when data of one or more direct risk factors is missing and cannot be used for prediction.

3 | RESULTS

The data set includes all 2 853 376 surgeries recorded between January 2008 and December 2018 in the EUREQUO. A summary of descriptive statistics of the data set can be found in Table 1. The mean age of patients is 73.9±9.7 and 58.7% (n = 1 674 242) of all patients are female. Common ocular comorbidities are macular degeneration (11.3%, n = 323 495) and glaucoma (7.1%, n = 202 926). Complicating comorbidities that occur most frequently include small pupil (2.9%, n = 81 827) and white cataract (2.5%, n = 72 484). Of all surgeries, 1.1% (n = 31 749) were complicated by PCR.

The PR curves of the classifiers are depicted in Figure 1. The PR curves display a trade-off between precision and recall depending on the threshold. If we set the threshold such that the recall is 10%, the corresponding precision is about 50%. This implies that the network correctly predicts 10% of the surgeries that are subject to PCR and, if the network generates a positive prediction, the network is correct in about 50% of the cases. If we decrease the threshold such that the recall is 20%, the corresponding precision drops to about 20%. Clearly, we can set the threshold to either maximize precision or recall, but not both at the same time. In practice, an appropriate threshold must be determined by a panel of experienced surgeons by carefully weighting the clinical importance of precision and recall.

TABLE 1 Demographics of patients included in the study.

Feature	Frequency/mean
Age (years)	73.9±9.7
Female gender	1 674 242 (58.7)
Right eye	1 362 404 (50.5)
Preoperative best-corrected visual acuity (Snellen)	0.45±0.23
Ocular comorbidities	
Amblyopia	33 443 (1.2)
Macular degeneration	323 495 (11.3)
Diabetic retinopathy	97 365 (3.4)
Glaucoma	202 926 (7.1)
Other co-existing eye disease	225 296 (7.9)
Complicating ocular comorbidities	
White cataract	72 484 (2.5)
Small pupil	81 827 (2.9)
Pseudoexfoliation	12 696 (0.4)
Corneal opacities	30 803 (1.1)
Previous vitrectomy	19 553 (0.7)
Previous corneal refractive surgery	4179 (0.1)
Other complicating comorbidity	148 876 (5.2)
Posterior capsule rupture	31 749 (1.1)

Note: Data are N (%) or mean±standard deviation.

The AUPRC of each PR curve is given in Table 2. Close examination of the AUPRC of each classifier reveals that the MLP network performs the best overall, followed by the BN and the LR model.

Each BN constructed during the validation procedure has the same independence graph depicted in Figure 2. The independence graph reveals that all nodes are risk factors of PCR. Direct risk factors for PCR are highlighted in grey and include preoperative BCVA, year of surgery, operation type, anaesthesia, target refraction, other ocular comorbidities, white cataract, and corneal opacities. Indirect risk factors for PCR are highlighted in white and include age, gender, right eye, amblyopia, macular degeneration, glaucoma, diabetic retinopathy, small pupil, pseudoexfoliation, previous vitrectomy, previous corneal refractive surgery, and other complicating ocular comorbidities.

4 | DISCUSSION

When evaluating the performance of classifiers, it is important to consider the PCR rate. A classifier that randomly predicts PCR without having any knowledge about the patient or procedure has an expected AUPRC equal to the PCR rate, which is about 1.1% in our large European cohort. Taking this into account, it can be concluded that the classifiers perform relatively well. The MLP network achieved an AUPRC that is more than 12 times higher than the AUPRC of a random classifier.

Our results also indicate that an MLP network can predict PCR much better than a BN. The MLP network has an AUPRC that is on average more than 1.5 times higher than the AUPRC of the BN and is much more precise at a low recall. It should be noted that this performance improvement cannot be entirely attributed to the MLP network. The MLP network likely performed worse if missing features in the sample would not have been imputed by the BN, and the network had to be constructed based on significantly fewer surgeries. A hybrid approach, whereby a BN is constructed to impute missing features, and subsequently, an MLP network is constructed to predict PCR, appears to work well in this application.

It is challenging to compare the performance of the classifiers to existing scoring systems in the literature, as these systems have not been evaluated based on a PR curve. Nevertheless, the precision and recall of the scoring system proposed by Muhtaseb et al. (2004) can be calculated for different criteria. The scoring system classifies surgeries into four groups with increasing levels of risk (i.e., group 1 is no added risk and group 4 is high risk). Because the PCR rate in their study is comparable to the PCR rate in our study, precision and recall can be safely compared across studies. If surgeries in groups 3 and 4

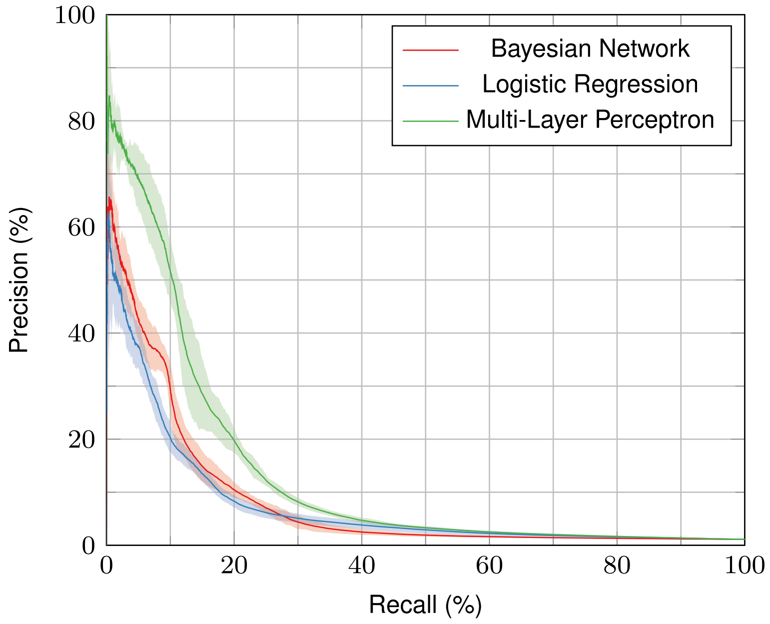

FIGURE 1 The PR curves of the classifiers. The shaded areas highlight the two standard deviation error bands around the mean precision.

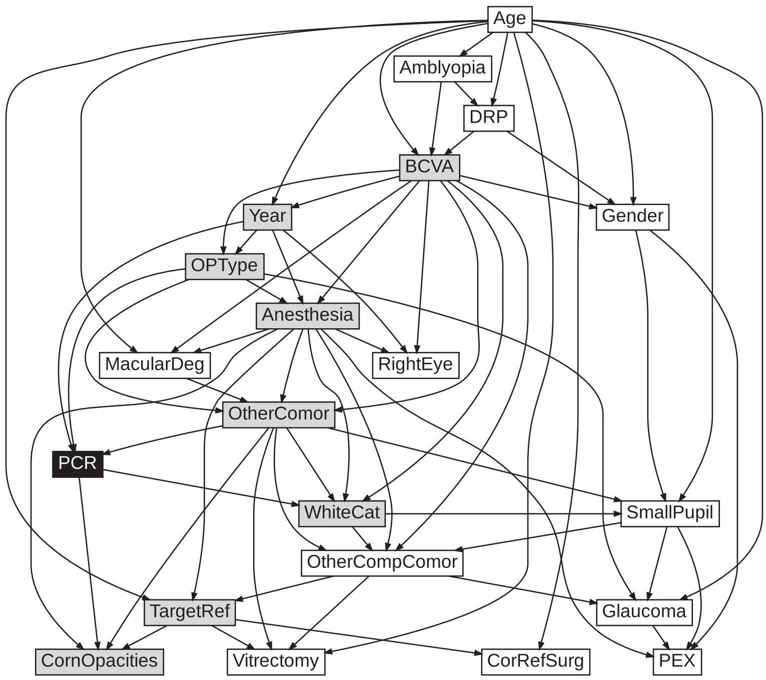

FIGURE 2 The independence graph of a BN learned from data of the EUREQUO. Nodes in the Markov blanket of PCR are highlighted in grey. These constitute direct risk factors of PCR. Indirect risk factors of PCR are highlighted in white. BCVA, preoperative best corrected visual acuity; CornOpacities, corneal opacities; CorRefSurg, previous corneal refractive surgery; DRP, diabetic retinopathy; MacularDeg, macular degeneration; OPType, operation type; OtherComor, other ocular comorbidity; OtherCompComor, other complicating ocular comorbidity; PCR, posterior capsule rupture; PEX, pseudoexfoliation; TargetRef, target refraction; WhiteCat, white cataract.

TABLE 2 The AUPRC of the classifiers.

Classifier	AUPRC
Bayesian network	8.05 ±0.39
Logistic regression	7.31 ±0.15
Multi-layer perceptron	13.10 ±0.41

Note: Data are mean±standard deviation.

are considered high risk, and those in groups 1 and 2 are considered low risk, the scoring system has a precision of 5% at a recall of 41%, which is similar to the precision of the MLP network in our study (see Figure 1). If only surgeries in group 4 are considered high risk, and those in the remaining groups are considered low risk, the scoring system has a precision of 8% at a recall of 9%. At this level of recall, the precision of the MLP network in our study is about seven times higher (see Figure 1). These statistics do not prove that the MLP network performs better than the scoring system of Muhtaseb et al. (2004) since we only compare performance based on two fixed points on the PR curve. Nonetheless, they strongly suggest that the MLP network has a much higher precision at a lower recall.

Implementation of a probabilistic classifier such as the MLP network in clinical practice might decrease the PCR rate. Preoperatively, all available data about the patient and surgery can be processed through the network to estimate the probability of PCR. When the classifier predicts a high risk for PCR, risk mitigation measures can be taken to minimize the risk and its potential consequences. These measures include, for example, ensuring an experienced surgeon carries out the surgery and equipment, such as a dispersive ophthalmic viscoelastic device, is already available in the operating room. Research has shown that implementing a scoring system to guide such risk mitigation measures decreased the PCR rate from 2.6% to 0.6% (Han et al., 2019). It can be envisaged that a further decrease in PCR rate can be achieved by implementing the MLP network which can predict PCR even better. Further research is needed to assess this in a clinical setting.

Previous studies have investigated risk factors for PCR, such as age, diabetic retinopathy, gender, small pupil, glaucoma, and pseudoexfoliation (Chancellor et al., 2021; Salowi et al., 2017; Segers et al., 2022; Segerstad, 2021; Theodoropoulou et al., 2019; Zetterberg et al., 2021). A limitation of these studies is that they are based on a traditional regression analysis, which can only measure the effect of individual risk factors on the probability of PCR, while interactions between risk factors are ignored. The BN and MLP network do not suffer from this limitation. In the case of the BN, interactions between risk factors are modelled through conditional independencies in the independence graph. Close examination of the independence graph in Figure 2 reveals that the aforementioned risk factors are indirect risk factors that do not provide any new information about the occurrence of PCR when data of all direct risk factors is available. Therefore, these risk factors could be removed to reduce model complexity when risk assessment is always performed based on complete data. This observation stresses the importance of modelling interactions between risk factors of PCR.

A few risk factors identified by the BN stand out. It seems counter-intuitive that 'right eye' is an indirect risk factor for PCR, while there are no anatomical reasons why PCR would be more likely to occur in a left or right eye. We suspect 'right eye' to be an indirect risk factor for PCR because the right eye is often operated before the left eye. If a PCR occurs when operating the right eye, it is likely that additional measures are taken to prevent this from happening in the left eye. Moreover, 'year of surgery' was included in the model to account for the time dimension. The PCR rate has been decreasing over the years due to innovation and technological development. By including 'year of surgery' in the model, we account for these changes and their impact on PCR.

A limitation of this study is that the data is self-reported by surgeons and clinics and, therefore, might be subject to under-reporting. The EUREQUO has implemented several measures to prevent under-reporting. For example, surgeons and clinics are encouraged to report consecutive cases and can only consult their own data. Given these measures, we deem it unlikely that under-reporting affects our results. Moreover, we were limited to the features that are recorded in the EUREQUO. Other features that might be relevant for the prediction of PCR likely exist, such as the axial length (Day et al., 2015), intake of a1-adrenergic receptor antagonists and other medication that could lead to intraoperative floppy iris syndrome (Christou et al., 2022), or whether the patient received previous intravitreal injections (Bjerager et al., 2022). Future research could study whether the addition of such features improves performance.

In conclusion, we have studied how different probabilistic classifiers can predict PCR prior to cataract surgery. Our results indicate that the MLP network predicts PCR the best. Although the precision is relatively low at high recall, the network appears to perform better than existing scoring models in the literature. Implementation of the MLP network in clinical practice can potentially decrease the PCR rate even further than existing scoring models.

ORCID

Ron J. M. A. Triepels https://orcid.org/0000-0003-1962-8239
Maartje H. M. Segers https://orcid.org/0000-0002-4565-5559
Anders Behndig https://orcid.org/0000-0001-6652-7436
Mats Lundström https://orcid.org/0000-0002-4489-7118
Mor M. Dickman https://orcid.org/0000-0001-9545-343X

REFERENCES

Bjerager, J., van Dijk, E.H.C., Holm, L.M., Singh, A. & Subhi, Y. (2022) Previous intravitreal injection as a risk factor of posterior capsule rupture in cataract surgery: a systematic review and meta-analysis. *Acta Ophthalmologica*, 100(6), 614–623. Available from: https://doi.org/10.1111/aos.15089

Chancellor, J., Soliman, M.K., Shoults, C.C., Faramawi, M.F., Al-Hindi, H., Kirkland, K. et al. (2021) Intraoperative complications and visual outcomes of cataract surgery in diabetes mellitus: a multicenter database study. *American Journal of Ophthalmology*, 225, 47–56. Available from: https://doi.org/10.1016/j.ajo.2020.12.027

Christou, C.D., Esagian, S.M., Ziakas, N., Prousali, E. & Tzamalis, A. (2022) Factors predisposing to intraoperative floppy-iris syndrome: an up-to-date meta-analysis. *Journal of Cataract and Refractive Surgery*, 48(11), 1335–1341. Available from: https://doi.org/10.1097/j.jcrs.0000000000001017

Day, A.C., Donachie, P.H., Sparrow, J.M., Johnston, R.L. & Royal College of Ophthalmologists' National Ophthalmology Database. (2015) The Royal College of Ophthalmologists' National Ophthalmology Database Study of cataract surgery: report 2, relationships of axial length with ocular copathology, preoperative visual acuity, and posterior capsule rupture. *Eye (London, England)*, 29(12), 1528–1537. Available from: https://doi.org/10.1038/eye.2015.198

Friedman, N. (1997) *Learning belief networks in the presence of missing values and hidden variables.* San Francisco, CA: Morgan Kaufmann.

Fung, R. & Chang, K.-C. (1990) Weighing and integrating evidence for stochastic simulation in Bayesian networks. In: Henrion, M., Shachter, R.D., Kanal, L.N. & Lemmer, J.F. (Eds.) *Machine intelligence and pattern recognition*, vol. 10. North-Holland: Elsevier, pp. 209–219.

Glorot, X. & Bengio, Y. (2010) Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the thirteenth international conference on artificial intelligence and statistics.*

Glorot, X., Bordes, A. & Bengio, Y. (2011) Deep sparse rectifier neural networks. *Proceedings of the fourteenth international conference on artificial intelligence and statistics.*

Han, J.V., Patel, D.V., Wallace, H.B., Kim, B.Z., Sherwin, T. & McGhee, C.N.J. (2019) Auckland cataract study III: refining preoperative assessment with cataract risk stratification to reduce intraoperative complications. *American Journal of Ophthalmology*, 200, 253–254. Available from: https://doi.org/10.1016/j.ajo.2019.01.013

He, K., Zhang, X., Ren, S. & Sun, J. (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. *Proceedings of the IEEE international conference on computer vision*.

Innes, M. (2018) Flux: elegant machine learning with Julia. *Journal of Open Source Software*, 3(25), 602.

Kingma, D.P. & Ba, J. (2014) Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koller, D. & Friedman, N. (2009) *Probabilistic graphical models: principles and techniques*. Massachusetts, MA: MIT press.

Lachance, A., Godbout, M., Antaki, F., Hebert, M., Bourgault, S., Caissie, M. et al. (2022) Predicting visual improvement after macular hole surgery: a combined model using deep learning and clinical features. *Translational Vision Science & Technology*, 11(4), 6. Available from: https://doi.org/10.1167/tvst.11.4.6

Lundstrom, M., Barry, P., Brocato, L., Fitzpatrick, C., Henry, Y., Rosen, P. et al. (2014) European registry for quality improvement in cataract surgery. *International Journal of Health Care Quality Assurance*, 27(2), 140–151. Available from: https://doi.org/10.1108/IJHCQA-10-2012-0101

Lundstrom, M., Dickman, M., Henry, Y., Manning, S., Rosen, P., Tassignon, M.J. et al. (2017) Femtosecond laser-assisted cataract surgeries reported to the European registry of quality outcomes for cataract and refractive surgery: baseline characteristics, surgical procedure, and outcomes. *Journal of Cataract and Refractive Surgery*, 43(12), 1549–1556. Available from: https://doi.org/10.1016/j.jcrs.2017.09.029

Muhtaseb, M., Kalhoro, A. & Ionides, A. (2004) A system for preoperative stratification of cataract patients according to risk of intraoperative complications: a prospective analysis of 1441 cases. *The British Journal of Ophthalmology*, 88(10), 1242–1246. Available from: https://doi.org/10.1136/bjo.2004.046003

Ng, A.Y. & Jordan, M.I. (2002) On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. *Advances in Neural Information Processing Systems*, 14, 841–848.

Pearl, J. (1988) *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Francisco, CA: Morgan Kaufmann.

Potapenko, I., Kristensen, M., Thiesson, B., Ilginis, T., Lykke Sorensen, T., Nouri Hajari, J. et al. (2022) Detection of oedema on optical coherence tomography images using deep learning model trained on noisy clinical data. *Acta Ophthalmologica*, 100(1), 103–110. Available from: https://doi.org/10.1111/aos.14895

Salowi, M.A., Chew, F.L.M., Adnan, T.H., King, C., Ismail, M. & Goh, P.P. (2017) The Malaysian cataract surgery registry: risk indicators for posterior capsular rupture. *The British Journal of Ophthalmology*, 101(11), 1466–1470. Available from: https://doi.org/10.1136/bjophthalmol-2016-309902

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.

Scutari, M. (2009) Learning Bayesian networks with the bnlearn R package. *arXiv preprint arXiv:0908.3817*.

Segers, M.H.M., Behndig, A., van den Biggelaar, F., Brocato, L., Henry, Y.P., Nuijts, R. et al. (2022) Risk factors for posterior capsule rupture in cataract surgery as reflected in the European registry of quality outcomes for cataract and refractive surgery. *Journal of Cataract and Refractive Surgery*, 48(1), 51–55. Available from: https://doi.org/10.1097/j.jcrs.0000000000000708

Segerstad, P.H. (2021) Risk model for intraoperative complication during cataract surgery based on data from 900 000 eyes: previous intravitreal injection is a risk factor. *British Journal of Ophthalmology*, 106(10), 1373–1379.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014) Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.

Steinmetz, J.D., Bourne, R.R., Briant, P.S., Flaxman, S.R., Taylor, H.R., Jonas, J.B. et al. (2021) Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the right to sight: an analysis for the global burden of disease study. *The Lancet Global Health*, 9(2), e144–e160.

Terveen, D., Berdahl, J., Dhariwal, M. & Meng, Q. (2022) Real-world cataract surgery complications and secondary interventions incidence rates: an analysis of US Medicare claims database. *Journal of Ophthalmology*, 2022, 8653476. Available from: https://doi.org/10.1155/2022/8653476

Theodoropoulou, S., Grzeda, M.T., Donachie, P.H.J., Johnston, R.L., Sparrow, J.M. & Tole, D.M. (2019) The royal college of ophthalmologists' national ophthalmology database study of cataract surgery. Report 5: clinical outcome and risk factors for posterior capsule rupture and visual acuity loss following cataract surgery in patients aged 90 years and older. *Eye (London, England)*, 33(7), 1161–1170. Available from: https://doi.org/10.1038/s41433-019-0389-z

Ti, S.E., Yang, Y.N., Lang, S.S. & Chee, S.P. (2014) A 5-year audit of cataract surgery outcomes after posterior capsule rupture and risk factors affecting visual acuity. *American Journal of Ophthalmology*, 157(1), 180–185. Available from: https://doi.org/10.1016/j.ajo.2013.08.022

Zetterberg, M., Kugelberg, M., Nilsson, I., Lundstrom, M., Behndig, A. & Montan, P. (2021) A composite risk score for capsule complications based on data from the Swedish National Cataract Register: relation to surgery volumes. *Ophthalmology*, 128(3), 364–371. Available from: https://doi.org/10.1016/j.ophtha.2020.07.033

How to cite this article: Triepels, R.J.M.A., Segers, M.H.M., Rosen, P., Nuijts, R.M.M.A., van den Biggelaar, F.J.H.M., Henry, Y.P. et al. (2023) Development of machine learning models to predict posterior capsule rupture based on the EUREQUO registry. *Acta Ophthalmologica*, 00, 1–7. Available from: https://doi.org/10.1111/aos.15648