Graefe's Archive for Clinical and Experimental Ophthalmology (2023) 261:715–721
https://doi.org/10.1007/s00417-022-05884-3

RETINAL DISORDERS

# Artificial intelligence using deep learning to predict the anatomical outcome of rhegmatogenous retinal detachment surgery: a pilot study

**Timothy H. M. Fung[1] · Neville C. R. A. John[2] · Jean-Yves Guillemaut[2] · David Yorston[3] · David Frohlich[2] · David H. W. Steel[4,5] · Tom H. Williamson[1] · on behalf of the BEAVRS Retinal Detachment Outcomes Group**

Received: 25 July 2022 / Revised: 14 October 2022 / Accepted: 21 October 2022 / Published online: 28 October 2022


## Abstract
**Purpose** To develop and evaluate an automated deep learning model to predict the anatomical outcome of rhegmatogenous retinal detachment (RRD) surgery.
**Methods** Six thousand six hundred and sixty-one digital images of RRD treated by vitrectomy and internal tamponade were collected from the British and Eire Association of Vitreoretinal Surgeons database. Each image was classified as a primary surgical success or a primary surgical failure. The synthetic minority over-sampling technique was used to address class imbalance. We adopted the state-of-the-art deep convolutional neural network architecture Inception v3 to train, validate, and test deep learning models to predict the anatomical outcome of RRD surgery. The area under the curve (AUC), sensitivity, and specificity for predicting the outcome of RRD surgery was calculated for the best predictive deep learning model.
**Results** The deep learning model was able to predict the anatomical outcome of RRD surgery with an AUC of 0.94, with a corresponding sensitivity of 73.3% and a specificity of 96%.
**Conclusion** A deep learning model is capable of accurately predicting the anatomical outcome of RRD surgery. This fully automated model has potential application in surgical care of patients with RRD.

**Keywords** Artificial intelligence · Deep learning · Models · Rhegmatogenous retinal detachment · Retinal detachment surgery

✉ Timothy H. M. Fung
timothyfung@doctors.org.uk

1 Guy's and St Thomas' NHS Foundation Trust, London, UK

2 Centre for Vision, Speech and Signal Processing, Department of Electrical and Electronic Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK

3 Gartnavel Hospital, Glasgow, UK

4 Sunderland Eye Hospital, Sunderland, UK

5 Bioscience Institute, Newcastle University, Newcastle Upon Tyne, UK

**Key messages**

- Deep learning models can accurately predict surgical outcomes.
- In this study, we developed a deep learning model to predict the anatomical outcome of rhegmatogenous retinal detachment (RRD) surgery based on digital RRD images alone.
- Our deep learning model was able to predict the anatomical outcome of RRD surgery with an area under the curve of 0.94 and an accuracy of 93.9%.
- Our deep learning model has potential application in surgical care of patients with RRD.

## Introduction

Rhegmatogenous retinal detachment (RRD) is a major cause of vision loss and its annual incidence has been reported to be between 6.3 and 17.9 cases per 100,000 persons [1]. The treatment of RRD is surgical and the ability to accurately predict the anatomical outcome after RRD surgery is fundamental to providing optimal surgical care. The anatomical outcome of RRD surgery is currently predicted by clinicians using their clinical judgment considering available preoperative clinical data such as the extent of RRD, the presence of inferior retinal breaks, and the presence of proliferative vitreoretinopathy (PVR) [2–10].

Advances in applied modeling using deep learning, a prominent type of artificial intelligence (AI), have provided a novel and more accurate method for predicting surgical outcomes [6]. Deep learning techniques have the advantage over other methods, because deep learning enables a computer model to automatically learn the best and most robust predictive features present in a dataset. Recently, deep convolutional neural networks (CNNs), a special type of deep learning technique, have been applied to produce highly accurate models that predict a range of postoperative outcomes on patients undergoing major surgical procedures within multiple surgical fields, including general, orthopedic, cardiothoracic, otolaryngology, gynecological, urology, neurosurgery, and vascular surgery [11].

Although deep CNNs have the potential to transform our abilities to predict surgical outcomes, this technique remains unexplored for predicting the outcome of RRD surgery. In this study, we aimed to develop and evaluate a CNN-based deep learning model to predict the anatomical outcome of RRD surgery.

## Methods

### Dataset

Using the British and Eire Association of Vitreoretinal Surgeons (BEAVRS) database, we extracted anonymized digital RRD images (Fig. 1) of eyes which underwent pars plana vitrectomy and internal tamponade between June 2008 and September 2019. A total of 6661 digital images of RRD were included in this study. Details about the BEAVRS database, including its inclusion and exclusion

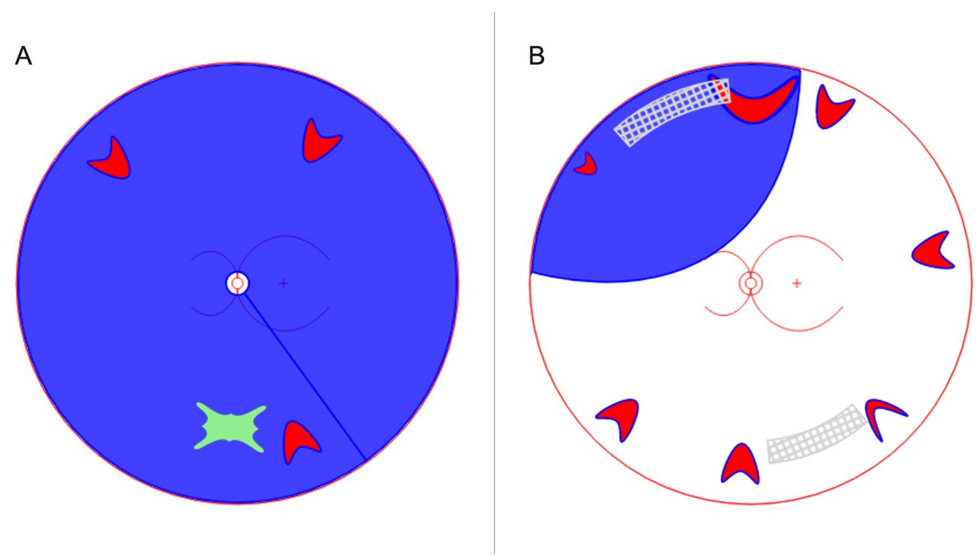


**Fig. 1** Examples of digital rhegmatogenous retinal detachment images from the BEAVRS database. **A** A total (blue area) retinal detachment with the red areas representing the location and type of retinal breaks and the green area representing the presence and extent of PVR. **B** A superonasal (blue area) retinal detachment with the red areas representing the location and type of retinal breaks and the gray areas representing areas of lattice degeneration

**Table 1** Breakdown of training, validation, and testing datasets

| Success | | | Failure | | |
|---|---|---|---|---|---|
| Train | Validation | Test | Train | Validation | Test |
| 5727 | 150 | 150 | 3228 | 150 | 150 |

**Table 2** Performance of the model on the test dataset

| Accuracy (%) | Positive predictive value (%) | Sensitivity (%) | Specificity (%) | ROC AUC (%) |
|---|---|---|---|---|
| 93.9 | 64.7 | 73.3 | 96 | 94 |

criteria, have been reported [12]. Briefly, the BEAVRS database is a web application where anonymized vitreoretinal surgical data are entered prospectively immediately following surgery by multiple clinicians at different sites. A digital drawing tool linked to diagnostic codes is used to record the anatomical details of RRD. This allows the recording of RRD extent, retinal break type and location, and the presence, extent, and severity of PVR at the time of surgery. The BEAVRS dataset also provided demographic data on age, sex, and lens status.

We classified each RRD image from the BEAVRS database as a primary surgical success or a primary surgical failure. The BEAVRS definition of primary surgical success is complete retinal reattachment in the absence of tamponade and without any additional reattachment procedures [12]. Primary surgical failure was defined by BEAVRS as surgeon recorded redetachment, or a record of repeat RRD surgery [12]. The number of images that belonged to primary surgical success was 6027 (90%) whereas 634 (10%) belonged to primary surgical failure. Only cases with a recorded outcome (success or failure) at least 2 months following primary surgery were included.

This study was conducted in accordance with the Declaration of Helsinki and the UK's Data Protection Act. No patient details could be identified with any of the data contained in the BEAVRS database and a unique alphanumeric code is used for internal identification. As the BEAVRS dataset is considered a service evaluation, no IRB approval and/or informed consent were needed according to UK guidelines.

## Model development and training

The imbalance between positive (failure) and negative (success) classes present in the original BEAVRS dataset was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE is an over-sampling approach in which the minority class is over-sampled by creating synthetic examples rather than over-sampling with replacement [13, 14]. Previous studies with varying amounts of imbalance and varying amounts of data have found that SMOTE performs better than other class balancing methods for improving the accuracy of classifiers for a minority class [13, 14]. The balanced BEAVRS dataset was divided into training, validation, and test datasets (Table 1). The training dataset was used for developing and training the state-of-the art CNN architecture Inception v3 available in the Keras application programming interface inside TensorFlow (http://tensorflow.org). Images from the datasets were resized for the CNN architecture Inception v3. Initial layers of the CNN architecture Inception v3 were all pre-trained on the ImageNet database of 1.2 million ontology of images from 1000 output classes [15] and the top layers were replaced and trained with layers that would support the classes from the BEAVRS dataset. This process of transfer learning has been shown to improve classification performance [16]. The validation dataset was used for continuous validation of the model and to prevent overfitting. Training of the network model was achieved by presenting the model with batches of 32 labeled images from the training dataset and 32 labeled images from the validation dataset. The independent test dataset was reserved for model testing after successful model development and training. The model was trained entirely on RRD images with no access to information on patient demographics.

## Model evaluation

The performance of the model for predicting the anatomical outcome of RRD surgery was evaluated using the following performance metrics: sensitivity, specificity, and the area under the curve (AUC) for the receiver operating characteristic (ROC) curve with 95% confidence intervals (CI) (Table 2). All statistical analyses were conducted using Python 2.7.15.

To highlight the visual features of the processed image that contributed the most to the model assignment of the predicted outcome or led to unappropriated misclassification, we generated heatmaps using Gradient Weighted Class Activation Mapping (Grad-CAM) technology. Grad-CAM is a class-discriminative localization technique that uses the average gradient information flowing into the last convolutional layer of any CNN-based network to assign importance values to each neuron for a particular decision of interest [17]. Grad-CAM heatmaps identify discriminative areas of the images that contribute to the decision of the deep learning model in classifying images as a failure or a success using a color-coded scale. A red area represents the area that critically contributed to the model's classification decisions, yellow to green represents medium level, and blue represents the lowest level [17].

## Results

### Demographics

The mean age of patients that underwent pars plana vitrectomy with internal tamponade was 61.5 years (range; 34–96 years). Sixty-four percent of patients were male. At the time of surgery, 69% of eyes were phakic, 30% of eyes were pseudophakic, and 1% of eyes were aphakic.

### Model performance

The AUC of the model was 0.940 (95% CI: 0.93–0.95), with 73.3% (95% CI: 62.5–97.2) sensitivity and 96% (95% CI: 88.3–96.2) specificity (Fig. 2).

### Heatmaps

Examples of heatmaps corresponding to the best model for predicting the outcome of RRD surgery are shown in Fig. 3 A and B. To analyze errors made by our best deep learning model, we checked all misclassified images. Among these images, six were success images misclassified as failure (false positives), and four were failure images misclassified as success (false negatives). Of the six false positive images (Fig. 3C), two showed heatmap visualization to areas of PVR, one showed heatmap visualization to an inferior retinal break, one presented heat map visualization to two temporal retinal breaks, and two showed non-specific heatmap visualization to the posterior pole. Of the four false negative images (Fig. 3D), two showed non-specific heatmap visualization to the posterior pole, one presented non-specific heatmap visualization to the inferior retina, and one showed heatmap visualization to a posterior retinal break.

## Discussion

The application of AI-based learning techniques to retinal pathologies has increased over the last decade, mainly due to larger datasets, electronic medical records, and better application programming interfaces. Artificial neural networks from the 1990s and early 2000s were shown to be capable of performing at a similar level as an expert clinician for detecting normal retinal landmarks and diabetic retinal lesions based on features extracted from color fundus images [18–20]. The subsequent development of deep CNNs has enabled the automated diagnosis and quantification of diseases such as diabetic retinopathy [21–23], age-related macular degeneration [24], and retinopathy of prematurity [25], from retinal images, with comparable accuracy to that of human experts.

In our pilot study, we have shown that the application of a CNN-based deep learning model to digital RRD images alone can be used to accurately predict the anatomical outcome of RRD surgery. Our findings, together with what has been shown in non-RRD AI studies [11, 26], underline the value of deep learning in enhancing surgical outcome predictions through an automated approach.

Before the deep learning era, the anatomical outcome of RRD surgery was predicted based on the clinical characteristics of RRD [2–10]. In a study of 847 eyes, Williamson et al. [4] found on multivariate analysis that the presence of PVR,

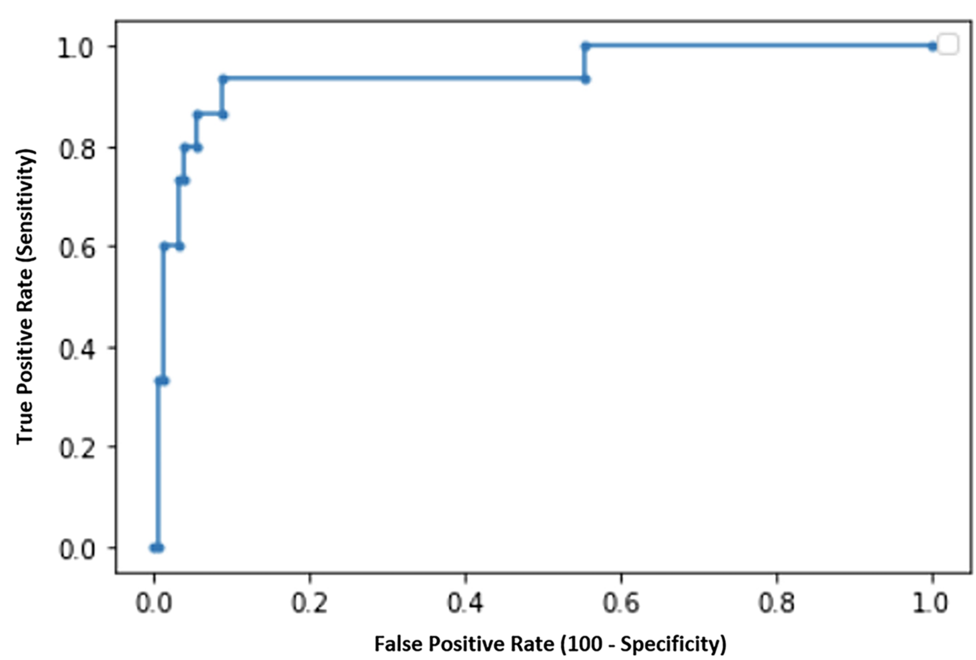


**Fig. 2** Receiver operating characteristic (ROC) curve of the best model for predicting the outcome of rhegmatogenous retinal detachment surgery

inferior positioning of retinal breaks, greater extent of RRD, and increased number of retinal breaks were associated with an increased risk of surgical failure. Using univariate analysis, Wickham et al. [2] identified prior cataract surgery and vitreous hemorrhage as additional risk factors for surgical failure. In the same study, Wickham et al. reported the use of a multivariable logistic regression model based on three risk factors—previous lens extraction, area of retina detached and preoperative grade C PVR—to predict the risk of failure from RRD surgery, showing an AUC of 0.658. Our study showed that a deep learning model can more accurately predict the outcome of RRD surgery with an AUC of 0.940. The accuracy of deep learning in predicting the anatomical outcome of RRD surgery has the potential to equip vitreoretinal surgeons with the tools for optimized patient counseling and decision-making.

Deep learning models are "black boxes" by construction since the features used in multiple layers of the model for prediction are learnt internally and not readily interpretable. In our study, we used Grad-CAM heatmaps to identify areas in the images that the deep learning model might have been using to make its predictions. Although our best deep learning model had high accuracy, misclassification still existed. To analyze errors made by our best deep learning model, we checked all the misclassified images carefully. Grad-CAM heatmap visualization of PVR and inferior retinal breaks appeared to result in false positive predictions in a few images using our deep learning model. In several images, false positive and false negative predictions seemed to be made based on non-specific locations the model used and produced unclear explanations. Due to its gradient averaging step, Grad-CAM can highlight image locations the model did not use to make its predictions and may explain why some of our image heatmap visualizations occurred in non-specific locations [27, 28]. Increasing the training dataset size of RRD images in our model could potentially minimize any false positive and false negative predictions [17]. In future studies, further detailed analysis of heatmaps using a variety of CNN visual explanation methods may reveal new features that are predictive of surgical success or failure, potentially serving as an educational tool for surgeons.

Despite the promising results, our study has several limitations. First, the overall size of the BEAVRS dataset is relatively small for deep learning. However, our work has to be considered a pilot and proof-of-feasibility study, and future research will be needed to validate our

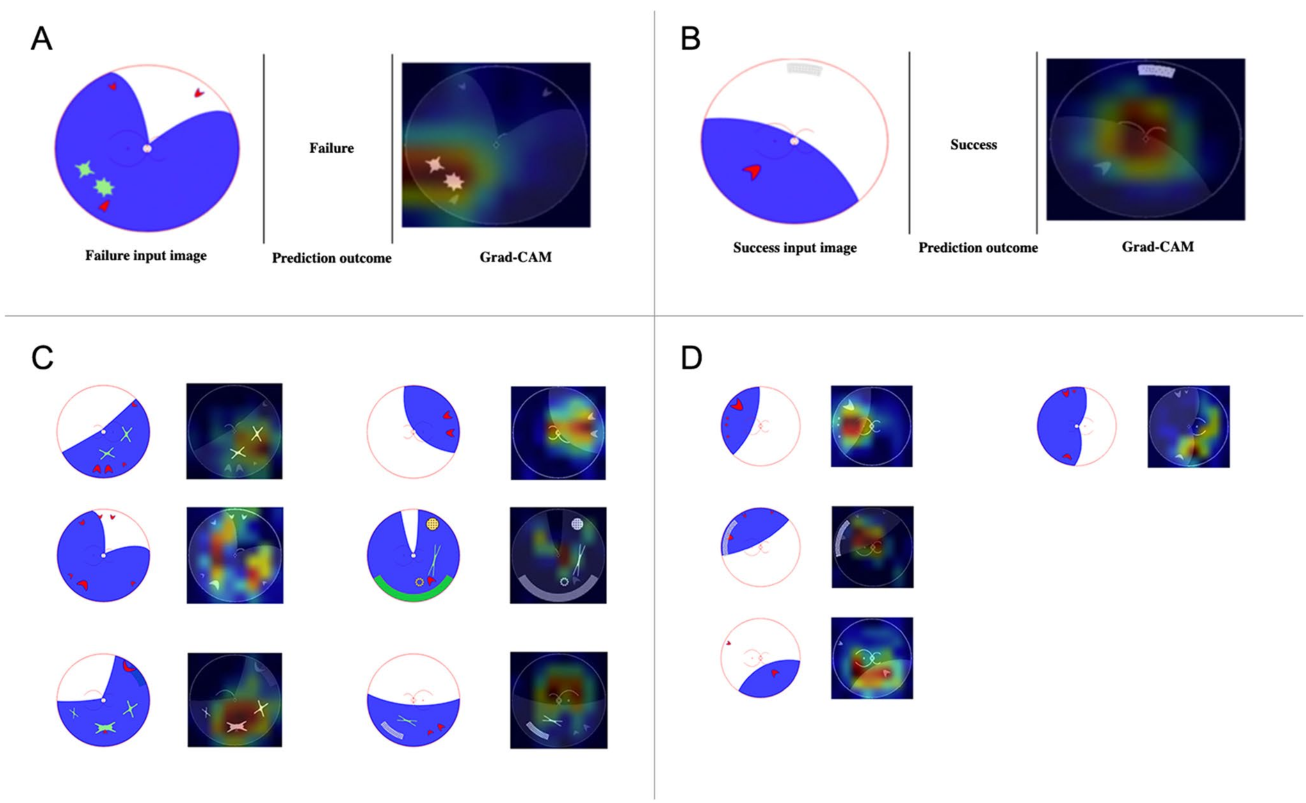


**Fig. 3** Heatmap visualization using gradient-weighted class activation mapping (Grad-CAM). **A** A Grad-CAM heatmap created by the best model that accurately predicted primary rhegmatogenous retinal detachment surgical failure. **B** A Grad-CAM heatmap created by the best model that accurately predicted primary rhegmatogenous retinal detachment surgical success. **C** Grad-CAM heatmap visualization for false positive images. **D** Grad-CAM heatmap visualization for false negative images

deep learning model against significantly larger datasets with more positive (failure) outcomes. Second, only RRD images created from a digital drawing tool were used. These images are reliant on accurate data entry by the surgeons. Any errors in data entry have the potential to affect the reliability of the predictions. However, commercially available digital fundus cameras are currently only successful at detecting the presence of an RRD, with limitations in their optics restricting their ability to accurately detect peripheral retinal pathology including retinal breaks and PVR [29, 30]. Despite this, future work should examine the generalizability of our findings to real color fundus photographs of RRD. Third, all RRD in this study were treated by pars plana vitrectomy and internal tamponade, and thus the use of this deep learning model is only applicable to retinal detachments treated by vitrectomy. Fourth, we included RRD images that had a recorded outcome (success or failure) at least 2 months following primary surgery. It is possible that some failures may have occurred after the follow-up date, and may not have been recorded in the BEAVRS database, leading to an incorrect image classification. Fifth, our deep learning model only predicted the anatomical outcome of RRD surgery. Future deep learning studies should also focus on the functional outcomes of RRD surgery. Lastly, we developed a deep learning model using images alone. The addition of other input variables, such as the patient's age, lens status, the vitrectomy gauge size, the use of cryotherapy or laser for retinopexy, and the type of tamponade is likely to improve the complexity and performance of a deep learning model for predicting the outcome of RRD surgery.

In conclusion, our pilot study shows that a CNN-based deep learning model can predict anatomical outcomes of RRD treated by vitrectomy and internal tamponade. Further deep learning studies are required to validate our deep learning model against other larger datasets for predicting the anatomical outcome of RRD surgery.

**Acknowledgements** Members of the BEAVRS Retinal Detachment Outcomes Group: A.G. Casswell, Andrew H.C. Morris, Assad Jalil, Atiq R. Babar, Craig Goldsmith, David H.W. Steel, Diego Sanchez-Chicharro, E.D. Hughes, E.N. Herbert, Huw Jenkins, Imran J. Khan, John D. Ellis, Jonathan Smith, Kamaljit S. Balaggan, Kurt Spiteri Cornish, Laura Wakeley, Mark Costen, Sonali Tarafdar, Stephen J. Charles, Stephen Winder, Timothy Cochrane, Tsveta Ivanova, Vasileios T. Papastavrou, Vaughan Tanner, David Yorston, D. Alistair Laidlaw, Tom H. Williamson

## Declarations

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**Conflict of interest** The authors declare no competing interests.

## References

1. Mitry D, Charteris DG, Fleck BW, Campbell H, Singh J (2010) The epidemiology of rhegmatogenous retinal detachment: geographical variation and clinical associations. Br J Ophthalmol 94:678–684
2. Wickham L, Bunce C, Wong D, Charteris DG (2011) Retinal detachment repair by vitrectomy: simplified formulae to estimate the risk of failure. Br J Ophthalmol 95:1239–1244
3. Wickham L, Ho-Yen GO, Bunce C, Wong D, Charteris DG (2011) Surgical failure following primary retinal detachment surgery by vitrectomy: risk factors and functional outcomes. Br J Ophthalmol 95:1234–1238
4. Williamson TH, Lee EJK, Shunmugam M (2014) Characteristics of rhegmatogenous retinal detachment and their relationship to success rates of surgery. Retina 34:1421–1427
5. Mitry D, Awan MA, Borooah S et al (2012) Surgical outcome and risk stratification for primary retinal detachment repairs: results from the Scottish Retinal Detachment study. Br J Ophthalmol 96:730–734
6. Thompson JA, Snead MP, Billington BM, Barrie T, Thompson JR, Sparrow JM (2002) National audit of the outcome of primary surgery for rhegmatogenous retinal detachment. II Clinical outcomes. Eye (Lond) 16:771–777
7. Pastor JC, Fernandez I, Rodriguez de la Rua E et al (2008) Surgical outcomes for primary rhegmatogenous retinal detachments in phakic and pseudophakic patients: the Retina 1 project – report 2. Br J Ophthalmol 92:378–382
8. Feltgen N, Heimann H, Hoerauf H, Walter P, Hilgers R-D, Heussen N (2013) Scleral buckling versus primary vitrectomy in rhegmatogenous retinal detachment study (SPR study): risk assessment of anatomical outcome. SPR study report no.7. Acta Ophthalmol 91:282–287
9. Adelman RA, Parnes AJ, Michalewska Z, Ducournau D (2014) Clinical variables associated with failure of retinal detachment repair: the European vitreo-retinal society retinal detachment study report number 4. Ophthalmology 121:1715–1719
10. Haugstad M, Moosmayer S, Bragadottir R (2017) Primary rhegmatogenous retinal detachment – surgical methods and anatomical outcome. Acta Ophthalmol 95:247–251
11. Bonde A, Varadarajan KM, Bonde N et al (2021) Assessing the utility of deep neural networks in predicting postoperative surgical complications: a retrospective study. Lancet Digit Health 3:e471-485
12. Ferrara M, Mehta A, Qureshi H et al (2021) Phenotype and outcomes of phakic versus pseudophakic primary rhegmatogenous retinal detachments: cataract or cataract surgery related? Am J Ophthalmol 222:318–327
13. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357
14. Fernandez A, Garcia S, Herrera F, Chawla NV (2018) SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J Artif Intell Res 61:863–905
15. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. Paper presented at: 2009 Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition; June 20–25, 2009; Miami, FL
16. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Paper presented at: 27th International Conference on Neural Information Processing Systems; December 8–13, 2014: Montreal, Quebec, Canada
17. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Paper presented at: 2017 Institute of

Electrical and Electronics Engineers International Conference on Computer Vision; October 22–29, 2017; Venice, Italy
18. Sinthanayothin C, Boyce JF, Cook HL, Williamson TH (1999) Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. Br J Ophthalmol 83:902–910
19. Gardner GG, Keating D, Williamson TH, Elliott AT (1996) Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. Br J Ophthalmol 80:940–944
20. Usher D, Dumskyjs M, Himaga M, Williamson TH, Nussey S, Boyce J (2004) Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. Diabet Med 21:84–90
21. Gulshan V, Peng L, Coram M et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316:2402–2410
22. Bellermo V, Lim ZW, Lim G et al (2019) Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. Lancet Digit Health 1:e35–e44
23. Ting DSW, Cheung CY-L, Lim G et al (2017) Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 318:2211–2223
24. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM (2017) Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. JAMA Ophthalmol 135:1170–1176
25. Brown JM, Campbell JP, Beers A et al (2018) Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol 136:803–810
26. Hu Y, Xiao Y, Quan W et al (2021) A multi-center study of prediction of macular hole status after vitrectomy and internal limiting membrane peeling by a deep learning model. Ann Transl Med 9:51
27. Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. Paper presented at: 2018 Institute of Electrical and Electronics Engineers Winter Conference on Applications of Computer Vision; March 12–15, 2018; Lake Tahoe, NV
28. Draelos RL, Carin L. Use HiResCAM instead of Grad-CAM for faithful explanations of convolutional neural networks. arXiv:2011.08891 [Pre-print]. November 21, 2021 [cited 2022 Apr 11] https://doi.org/10.48550/arXiv.2011.08891
29. Kornberg DL, Klufas MA, Yannuzzi NA, Orlin A, D'Amico DJ, Kiss S (2016) Clinical utility of ultra-widefield imaging with the optos optomap compared with indirect ophthalmoscopy in the setting of non-traumatic rhegmatogenous retinal detachment. Semin Ophthalmol 31:505–512
30. Khandhadia S, Madhusudhana KC, Kostakou A, Forrester JV, Newsom RSB (2009) Use of optomap for retinal screening within an eye casualty setting. Br J Ophthalmol 93:52–55