Prediction of Reactivation after Anti-VEGF Monotherapy for Retinopathy of Prematurity Using Multimodal Machine Learning models

Rong Wu, Yu Zhang, Peijie Huang, Yiling Xie, Jianxun Wang, Shuangyong Wang, Qiuxia Lin, Yichen Bai, Songfu Feng, Nian Cai, Xiaohe Lu

Submitted to: Journal of Medical Internet Research
on: May 11, 2024

Table of Contents

Prediction of Reactivation after Anti-VEGF Monotherapy for Retinopathy of Prematurity Using Multimodal Machine Learning models

Rong Wu[1]; Yu Zhang[2]; Peijie Huang[3]; Yiling Xie[3]; Jianxun Wang[4]; Shuangyong Wang[5]; Qiuxia Lin[2]; Yichen Bai[2]; Songfu Feng[2]; Nian Cai[3]; Xiaohe Lu[1]

[1]Department of Ophthalmology, Zhujiang Hospital, Southern Medical University Guangzhou CN
[2]Department of Ophthalmology, Zhujiang Hospital, Southern Medical University, Guangzhou CN
[3]School of Information Engineering, Guangdong University of Technology Guangzhou CN
[4]Department of Pediatric Ophthalmology, Guangzhou Children's Hospital and Guangzhou Women and Children's Medical Center, Guangzhou Medical Uni-versity Guangzhou CN
[5]Department of ophthalmology, Guangdong Provincial Key Laboratory of Major Obstetric Diseases, Guangdong Provincial Clinical Research Center for Obstet-rics and Gynecology, The Third Affiliated Hospital of Guangzhou Medical Uni-versity Guangzhou CN

Corresponding Author:
Xiaohe Lu
Department of Ophthalmology, Zhujiang Hospital, Southern Medical University
No.253 Gongyedadao Middle Road
Guangzhou
CN

Abstract

Background: Retinopathy pf prematurity is the leading preventable cause of childhood blind-ness, timely intravitreal injection of anti-vascular endothelial growth factor is re-quired to prevent retinal detachment with consequent vision impairment and loss. However, anti- vascular endothelial growth factor has been reported to be associated with ROP reactivation. Therefore, prediction of reactivation after treatment is urgent need.

Objective: To develop and validate prediction models for reactivation after anti-vascular endothelial growth factor intravitreal injection in infants with retinopathy of prema-turity using multimodal machine learning algorithms.

Methods: Infants with ROP undergoing anti-vascular endothelial growth factor treatment were recruited from three hospitals, conventional machine learning, deep learning and fusion models were constructed. The areas under the curve, accurancy, sensitivity and specificity were used to show the performances of the prediction models.

Results: 239 cases with anti-vascular endothelial growth factor treatment were recruit-ed, including 90 with reactivation and 149 non-reactivation cases. The area under the curve for the conventional machine learning model was 0.806 and 0.805 in the inter-nal and external validation groups, respectively. The average area under the curve, sensitivity, and specificity in the external validation for the deep learning model were 0.787, 0.800 and 0.570, respectively. The specificity, area under the curve, and sensitivity for the fusion model were 0.686, 0.822, and 0.800 in external validation, separately.

Conclusions: We constructed three prediction models for the reactivation of retinopathy of prematurity, fusion model achieved the best performance. Using this prediction model, we may optimize strategies for treating retinopathy of prematurity infants and developing better screening plans after treatment.

(JMIR Preprints 11/05/2024:60367)
DOI: https://doi.org/10.2196/preprints.60367

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Prediction of Reactivation after Anti-VEGF Monotherapy for Retinopathy of Prematurity Using Multimodal Machine Learning models

Rong Wu[1#], MD; Yu Zhang [1#], MD; Peijie Huang[2#], MD; Yiling Xie[2], MD; Jianxun Wang[4], MD; Shuangyong Wang[3], MD; Qiuxia Lin[1], MD; Yichen Bai[1], MD; Songfu Feng[1*], MD; Nian Cai[2*], MD; Xiaohe Lu[1*], MD.

[1] Department of Ophthalmology, Zhujiang Hospital, Southern Medical University, Guangzhou, China.
[2] School of Information Engineering, Guangdong University of Technology, Guangzhou, China.
[3] Department of ophthalmology, Guangdong Provincial Key Laboratory of Major Obstetric Diseases, Guangdong Provincial Clinical Research Center for Obstetrics and Gynecology, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China.
[4] Department of Pediatric Ophthalmology, Guangzhou Children's Hospital and Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China
[#] These authors contributed equally to this work
[*] Xiaohe Lu, Nian Cai and Songfu Feng should be co-corresponding authors.

Abstract

Background: Retinopathy pf prematurity is the leading preventable cause of childhood blindness, timely intravitreal injection of anti-vascular endothelial growth factor is required to prevent retinal detachment with consequent vision impairment and loss. However, anti- vascular endothelial growth factor has been reported to be associated with ROP reactivation. Therefore, prediction of reactivation after treatment is urgent need.
Aim: To develop and validate prediction models for reactivation after anti-vascular endothelial growth factor intravitreal injection in infants with retinopathy of prematurity using multimodal machine learning algorithms.
Methods: Infants with ROP undergoing anti-vascular endothelial growth factor treatment were recruited from three hospitals, conventional machine learning, deep learning and fusion models were constructed. The areas under the curve, accurancy, sensitivity and specificity were used to show the performances of the prediction models.
Results: 239 cases with anti-vascular endothelial growth factor treatment were recruited, including 90 with reactivation and 149 non-reactivation cases. The area under the curve for the conventional machine learning model was 0.806 and 0.805 in the internal validation and test groups, respectively. The average area under the curve, sensitivity, and specificity in the test for the deep learning model were 0.787, 0.800 and 0.570, respectively. The specificity, area under the curve, and sensitivity for the fusion model were 0.686, 0.822, and 0.800 in test, separately.
Conclusion: We constructed three prediction models for the reactivation of retinopathy of prematurity, fusion model achieved the best performance. Using this prediction model, we may optimize strategies for treating retinopathy of prematurity infants and developing better screening plans after treatment.

Keywords: retinopathy of prematurity; reactivation; prediction; machine learning; deep learning

Introduction

Retinopathy of prematurity (ROP) is characterized by retinal ischemia-hypoxia in preterm infants. Worldwide, it is a leading cause of vision loss and blindness in childhood [1]. More than 20,000 infants suffer from blindness and an estimated 12,300 infants have different levels of visual impairment due to ROP [2]. Dysregulation of vascular endothelial growth factor (VEGF) was proven to play an important role in the development of ROP [3]. Laser photocoagulation and anti-VEGF agents are the mainstay treatments for treatment-requiring ROP. The clinical application of intravitreal injection (IVI) of anti-VEGF agents has recently increased due to their fewer side effects and more advantages, including a lower risk of future myopia, better peripheral vision, and faster regression of acute-phase ROP [4-7]. However, ROP reactivation after anti-VEGF therapy is still a

concern. Given the short half-life of these agents, the beneficial effects of anti-VEGF might be transient, potentially increasing the risk of ROP reactivation. Previous studies evaluating the clinical outcomes of anti-VEGF agents for ROP have reported reactivation rates of 6.8%-64% after treatment [5, 6, 8-10].

Early detection and timely treatment of reactivation is critical for ROP infants who undergo anti-VEGF therapy. Previous studies on ROP reactivation after anti-VEGF therapy have mainly focused on the risk factors of reactivation. Factors related to ocular conditions include zone I ROP, severe retinal neovascularization, preretinal hemorrhage, and aggressive ROP [9, 11]. Neonatal factors include low gestational age, low birth weight, early postmenstrual age at initial treatment, and low Apgar scores [9]. Maternal factors include multiple births [9]. Other risk factors related to neonatal interventions include oxygen requirement before or after treatment and longer duration of hospitalization [6]. Prediction models that can identify infants with a high risk of ROP reactivation are needed in clinical practice.

Artificial intelligence has recently optimized medical practice [12-14]. Artificial intelligence has been mainly applied to ROP diagnosis and prediction based on imaging [15, 16]. Machine learning is a subset of artificial intelligence and includes conventional machine learning and deep learning [17-19]. In the present study, we developed prediction models for reactivation after anti-VEGF treatment in infants with ROP using machine learning algorithms based on clinical risk factors and retinal images before treatment.

Methods

This study was approved by the hospitals' institutional ethics committee and adhered to the tenets of the Declaration of Helsinki. All parents or guardians of the recruited infants provided written informed consent prior to participation. Data were anonymized and de-identified before analysis.

Study Population

We colloected data retrospectively on infants who received anti-VEGF monotherapy for treatment-requiring ROP between April 2016 and November 2022 at Hospital I (Zhujiang Hospital, Southern Medical University, Guangzhou, China), Hospital Ⅱ (The Third Affiliated Hospital, Guangzhou Medical University, Guangzhou, China) and Hospital ⅢI Guangzhou Children's Hospital and Guangzhou Women and Children's Medical Center, Guangzhou, China), respectively. Infants with incomplete data or any other ocular diseases besides ROP were excluded. Also, infants who underwent anti-VEGF therapy as adjunctive treatment before planned vitrectomy or received follow-up examinations for less than 12 months were excluded.

Ocular Examinations

During each ROP screening examination, retinal photographs were captured using RetCam Ⅲ digital fundus camera (Natus, San Carlo, CA, USA). The diagnosis of ROP and indication of treatment for ROP followed the International Classification of ROP Revisited and the Early Treatment for ROP Study, respectively, treatment-requiring ROP include threshold disease, stage 4 or 5 ROP, and type 1 or aggressive ROP. Ocular examinations were conducted before and 1, 7, 14, and 28 days after anti-VEGF therapy, and biweekly or monthly, depending on the ocular findings and systemic conditions. Reactivation of ROP was defined as the recurrence of acute phase features including a range of signs from a new demarcation line to reactivated stage 3 with plus disease, vascular dilation, tortuosity, or new/recurrent neovascularization that required further treatment.[10]

Intravitreal Injections of Anti-vascular Endothelial Growth Factor

All parents or guardians of the infants were fully informed of the efficacy and possible complications prior to intravitreal injection of conbercept, and provided written informed consent. The anti-VEGF treatment was performed as monotherapy for treatment-naive patients. Anti-VEGF

agents were injected intravitreally at 1.5 mm posterior to the limbus with a 30-gauge needle under topical anaesthesia. Topical tobramycin dexamethasone was administered for 3 days after the injection. All operations were done by a trained pediatric ophthalmologist (SF).

Clinical Risk Factors

According to previous studies and clinical experience, the potential clinical risk factors of ROP reactivation extracted from electronic medical records were related to maternal factors, neonatal factors, ocular conditions, laboratory factors, and neonatal interventions. Specifically, maternal factors of interest included maternal age, gestational hypertension, gestational diabetes mellitus, premature rupture of membranes, cesarean delivery, and in vitro fertilization-embryo transfer. Neonatal factors included gestational age, birth weight, postmenstrual age(PMA) at initial ROP treatment, fetal distress, sex, small for gestational age, Apgar scores (1 and 5 min), multiple births, asphyxia, sepsis, respiratory distress syndrome, bronchopulmonary dysplasia, pneumonia, intraventricular hemorrhage, necrotizing enterocolitis, hypoxic-ischemic encephalopathy, atrial septal defect, Patent foramen ovale, patent ductus arteriosus, and hyperbilirubinemia. Ocular conditions included zone I ROP, preretinal hemorrhage, and aggressive ROP. The hemoglobin concentration before treatment was included in laboratory factor. Finally, neonatal interventions included mechanical ventilation and oxygen therapy (before-treatment). [6, 8, 9]

Image Collection and Case Labeling

All retinal photographs were captured using a commercially camera (Retcam; Natus Medical Inc). Retinal photographs of poor photographic quality were excluded by two experienced ophthalmologists (S.F. and R). Since the prediction for ROP reactivation would be performed on each infant rather than each retinal photograph independently, retinal photographs of both eyes from the same infant were labeled as one case. All cases were labeled independently by two clinical ophthalmologists (S.F. and R). According to the ocular findings after anti-VEGF therapy, each case was annotated reactivation or non-reactivation. If ROP reactivation occurred in one of the eyes, the case would be labeled as reactivation.

The κ was 0.81 for annotation of ROP reactivation, suggesting good agreement of the two ophthalmologists in labeling. Moreover, the labels were further confirmed by a senior retinal specialist (X.H.) to generate a final annotation. These annotations were used as the ground-truth labels in the development and validation of the prediction model of ROP reactivation.

Development and Validation of Prediction Models

Conventional Machine Learning Model

An illustration of the conventional machine learning prediction model is presented in **Figure 1**. The importance ranking of the clinical risk factors of ROP reactivation was assessed, and prediction models based on clinical risk factors were developed using conventional machine learning algorithms. The process was implemented with Python (version 3.9).

During the data preprocessing stage, we employ two filling strategies. For discrete data, we employ mean filling followed by rounding; for continuous data, mean filling is applied. Subsequently, continuous data undergoes standardization, transforming it into a standard normal distribution with a mean of 0 and a standard deviation of 1, ensuring uniform scales across different features. This aids in mitigating potential model biases arising from scale differences, enhancing overall model robustness and performance.

In this study, we employed a comprehensive approach to assess the feature importance in predictive models, utilizing five different algorithms: RandomForest (RF), AdaBoost (AB), XGBoost (XGB), CatBoost (CB), and Logistic Regression (LR) (see **eTable 1 in the supplementary** for parameter details). The method involves ranking the feature importance or weights for each algorithm, where for Logistic Regression, feature importance is determined by the absolute values of the weights. Subsequently, these rankings are visualized using scatter plots, with each feature represented along the x-axis and its importance rank along the y-axis. This comprehensive visualization offers a comparative analysis of feature contributions across different algorithms, providing insights into the robustness and consistency of feature importance.

In this study, we conducted an in-depth investigation into the impact of varying the number of selected features (N) on the predictive model performance, employing five distinct algorithms. We systematically varied the number of selected features, choosing from the set (5, 10, 15, 20, 25, 30),

and assessed the resulting model performance by calculating the Area Under the Curve (AUC) for each model across different feature subsets.

We conducted a comprehensive evaluation of five different predictive models, with a specific focus on their Receiver Operating Characteristic (ROC) curves. The models were trained and evaluated using all features on a standardized dataset. The emphasis of this research lies in presenting the performance of different models through ROC curves, offering crucial insights into their predictive capabilities. Furthermore, we conducted ten iterations of training for each of the five distinct predictive models and computed their average Area Under the Curve (AUC) values on the corresponding validation sets, along with their respective standard deviations.

Deep Learning Model

An illustration of the deep learning prediction model is presented in **Figure 1**. Retinal photographs captured before anti-VEGF therapy of ROP infants were ob-tained to develop a prediction model using deep learning algorithms. ResNet-50, a deep convolutional neural network classifier, was applied to fine-tune with our da-taset in order to predict the ROP reactivation in these cases. The process was imple-mented with Python (version 3.9) and PyTorch (Version 1.11.0).

In this study, we designed a series of image preprocessing methods. Initially, we enhance the features of fundus images and suppress noise based on median filtering technology. This approach is due to the excellent filtering effect of median filtering on isolated noise pixels while maintaining the clarity of important features such as blood vessel edges. Subsequently, the images are converted into grayscale for processing to reduce computational complexity.

For the training set, we employed a series of preprocessing steps. These included resizing the images to 224x224 pixels, introducing RandAugment for random data augmentation, and applying both random horizontal and vertical flips. These measures aimed to increase the diversity of the data, enhancing the model's robustness. For the internal validation and test sets, we consistently resized the images to 224x224 pixels and transformed them into tensors. This ensured uniformity in the model's input during the validation process. This diversified preprocessing approach contributes to adapting to the distinct characteristics of each dataset, ultimately improving the model's generalization performance.

We employed a pre-trained ResNet-50 model and iteratively trained it on the image training set, incorporating image preprocessing and data augmentation techniques to enhance the model's generalization performance. We adopted the cross-entropy loss function, whereby the model learns to adjust its parameters by minimizing the cross-entropy loss, to more accurately predict the actual categories of the samples. During the training process, we utilized the Adam optimizer and a cosine annealing learning rate scheduler, conducting a total of 100 epochs. At the end of each epoch, we evaluated the model's performance on both an internal validation set and an external test set. Throughout the training, we also applied the Exponential Moving Average (EMA) technique to stabilize model parameters and improve overall model generalization.

We constructed a dataset in which each case consists of multiple medical images. When performing case-level predictions, we treated the model's output as the probability score for the entire case. By averaging the probability scores from multiple images within each case, we computed the final case-level prediction results.

To enhance the visualization and interpretability of ROP reactivation prediction, we applied Grad-CAM technique to generate heatmaps of key regions. This technique highlights areas crucial to ROP reactivation prediction by analyzing feature weights in specific network layers, such as layer4 of ResNet50. This approach allows us to intuitively identify retinal features that significantly contribute to the prediction outcome.

Fusion Model

An illustration of the prediction model combining the conventional machine learning model and the deep learning model is presented in **Figure 1**. Among the five conventional machine learning prediction models, the model based on the CatBoost algorithm exhibited the best predictive performance. The optimal conventional machine learning model were concatenated with the final fully connected layer set before the output layer in the deep learning prediction model. The process was implemented with Python (version 3.9) and PyTorch (Version 1.11.0).

First, we modeled the basic clinical features of patients using a risk factor model, considering traditional risk factors such as maternal age and infant weight. Subsequently, we employed a deep learning model (ResNet-50) to process retinal images of newborns, capturing more comprehensive information. The deep learning model, by learning features from the images, can capture complex patterns that traditional models find challenging.

To better leverage the strengths of both models, we adopted a fusion strategy, performing a weighted average of the outputs from the risk factor model and the deep learning model. We obtained the deep learning probability scores for each sample through independent training on the training set, while simultaneously predicting risk factor probabilities using the risk factor model. Finally, we linearly combined the probability scores from both models to derive the ultimate fusion result.

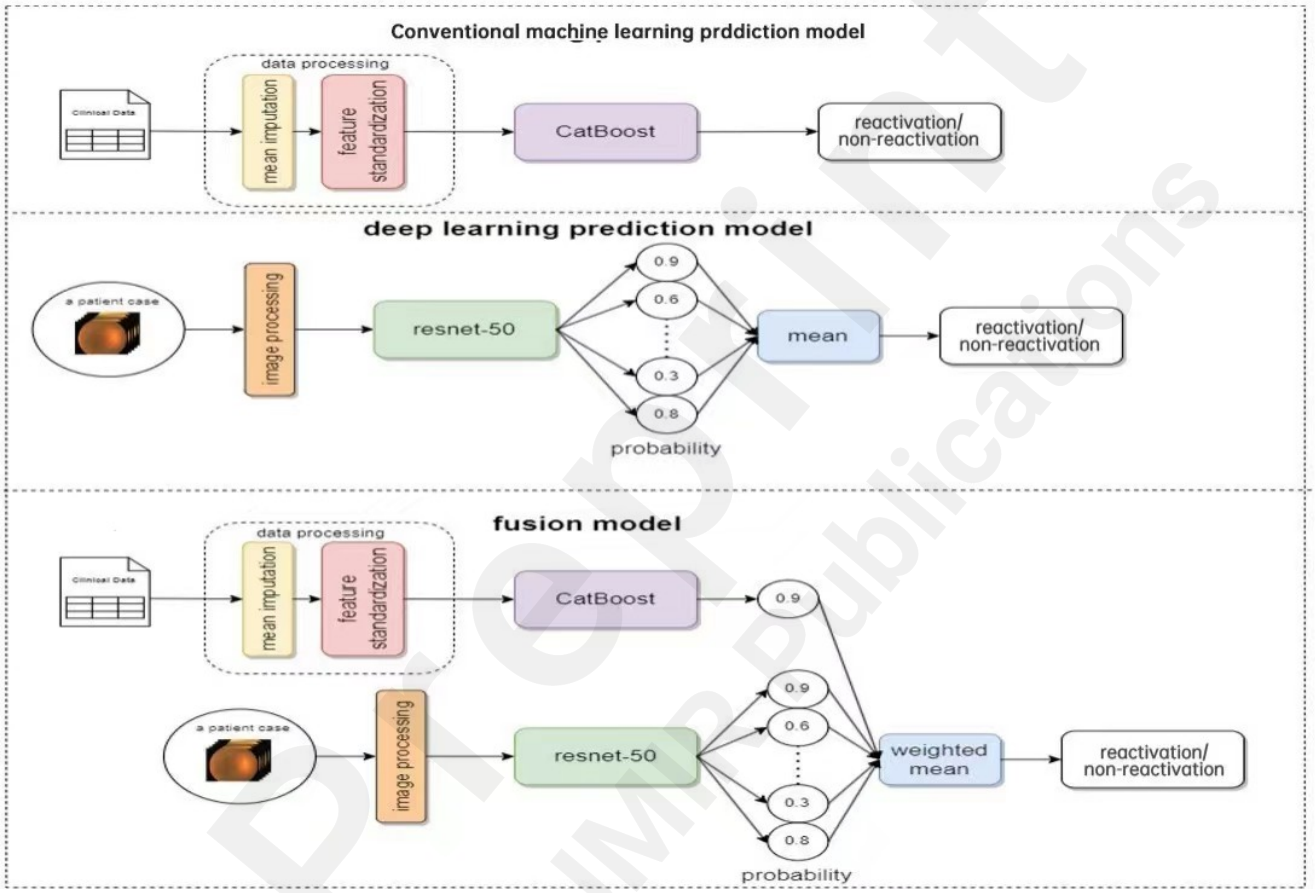


Figure 1. Illustration of the Proposed Algorithmic Pipeline

Statistical Analysis

Statistical analyses were performed using R Language (3.0.2; R Core Team, 2013, Vienna, Austria). Categorical variables were expressed as numbers and percentages and analyzed using chi-squared tests, if any cell number was less than five, Fisher,s exact test was applied. Continuous variables conforming to normal distribution were expressed as means and standard deviations (SD) and compared using the independent two-sample *t*-test. If a normal distribution was not confirmed, the median and interquartile ranges were used, and the Mann-Whitney U test was performed. Breakdowns of the predictive labels with reference to the ground-truth labels were depicted as confusion matrices, which were used to calculate the accuracy, sensitivity, and specificity of the prediction models with the 3 training schemes (Python version 3.7.0 [Python Software Foundation]). P □ 0.05 was considered statistically significant.

Results

Demographics of Study Groups

The study group finally included 239 infants. The infants recruitment flow was presented in **Figure 2**. The detailed demographic characteristics of the enrolled infants were summarized in **Table 1**. Among the 179 included infants of Hospital Ⅰ, the mean (SD) gestational age was 28.87 (2.46) weeks, and the mean (SD) BW was 1.18(0.35) kg. Among 60 included infants of Hospital Ⅱ and Hospital III, the mean (SD) gestational age was 28.30 (1.90) weeks, and the mean (SD) birth weight was 1.09(0.28) kg. There was no significant difference between the demographic characteristic of the two cohorts (all P>0.05).

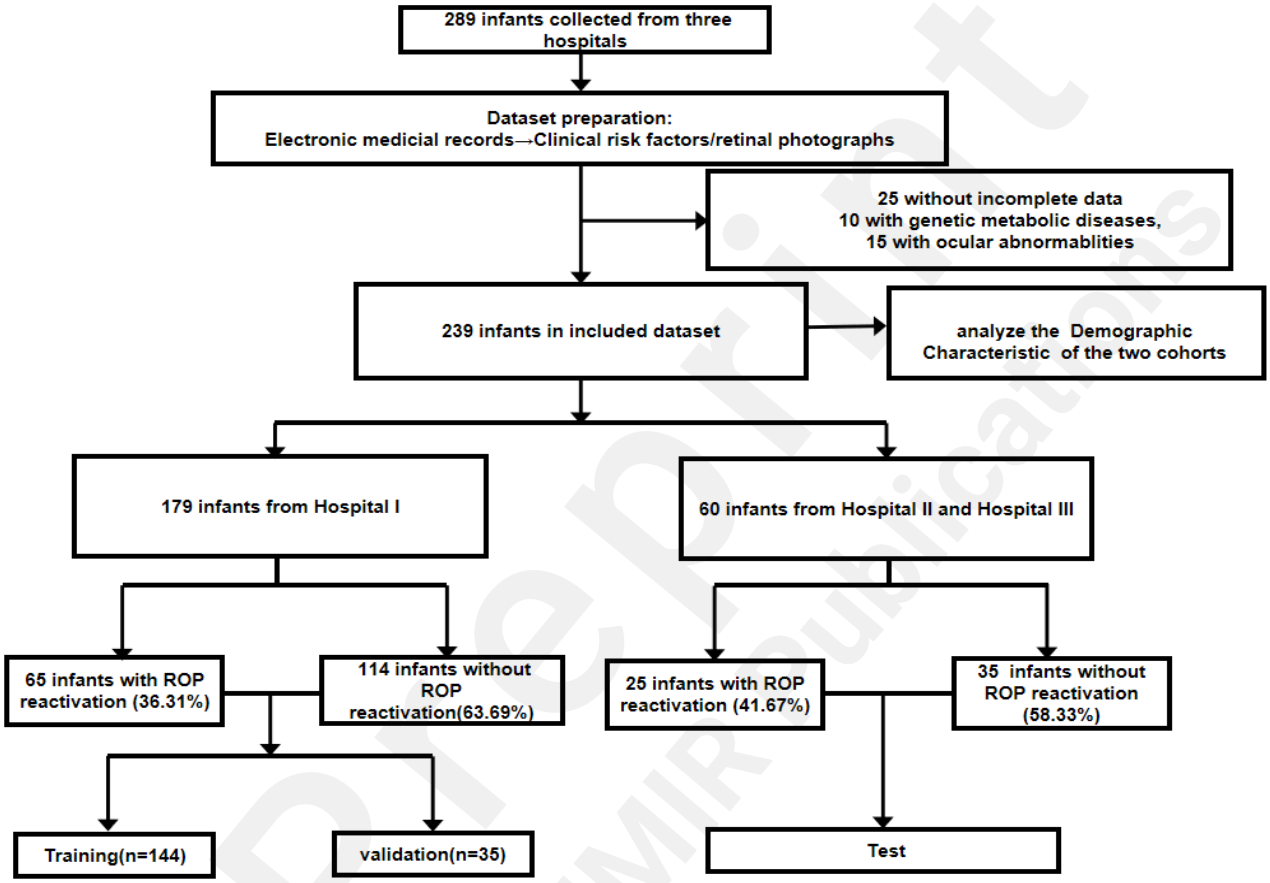


Figure 2. The flow chart of patient cohorts

Table 1. Demographic Characteristic of all infants

Variables	All infants (n=239)	Training (n=179)	Test (n=60)	P
Maternal factors				
Maternal age, mean±SD, years	30.72±5.32	30.65±5.08	30.95±6.03	0.728
In vitro fertilization-embryo transfer, n (%)	41 (17.15)	31 (17.32)	10(16.67)	0.908
Gestational hypertension, n (%)	30 (12.55)	24(13.41)	6(10.00)	0.490

Gestational diabetes mellitus, n (%)	34 (14.23)	25 (13.97)	9(15.00)	0.843
Cesarean delivery, n (%)	105 (43.93)	79 (44.13)	26(43.34)	0.914
Premature rupture of membranes, n (%)	37(15.48)	26(14.53)	11(17.74)	0.480
Neonatal factors				
Gestational age, mean ± SD, weeks	28.66±2.38	28.87±2.46	28.30±1.90	0.068
Birth weight, mean ± SD, kg	1.15±0.34	1.18±0.35	1.09±0.28	0.051
Multiple gestations, n (%)	58(24.27)	46(25.70)	12(20.00)	0.373
Small for gestational age, n (%)	39(16.32)	30(16.76)	9(15.00)	0.750
Intrauterine distress, n (%)	22(9.20)	18(10.06)	4(6.67)	0.432
Postmenstrual age at initial ROP treatment, mean ± SD, weeks	36.31±2.29	36.29±2.31	36.38±2.25	0.780
Male sex, n (%)	131(54.81)	94 (52.51)	37(61.67)	0.218
Apgar scores at 1 minute, mean ± SD	7.35±2.05	7.38± 2.07	7.25±2.01	0.670
Apgar scores at 5 minutes, mean ± SD	8.58±1.39	8.62±1.43	8.58±1.29	0.980
Asphyxia, n (%)	98(41.00)	72(40.22)	26(43.33)	0.672
Respiratory distress syndrome, n (%)	179(75.24)	130(72.63)	49(81.67)	0.162
Bronchopulmonary dysplasia, n (%)	146(61.09)	107(59.78)	39(65.00)	0.473
Sepsis, n (%)	65(27.20)	47 (26.26)	18(30.00)	0.573
Necrotizing enterocolitis, n (%)	37(15.48)	27 (15.08)	10(16.67)	0.548
Intraventricular hemorrhage, n (%)	107(44.76)	80 (44.69)	27(45.00)	0.967
Hypoxic-ischemic encephalopathy, n (%)	48(20.08)	35(19.55)	13(21.67)	0.724
Patent ductus arteriosus, n (%)	90(37.66)	70(39.11)	20(33.33)	0.424
Atrial septal defect, n (%)	39(16.32)	27(15.08)	12(20.00)	0.372
Patent foramen ovale, n (%)	183(76.57)	139(77.65)	44(73.33)	0.494
Hyperbilirubinemia, n (%)	148(61.92)	110(61.45)	38(63.33)	0.795
Pneumonia, n (%)	140(58.58)	103(57.54)	37(61.67)	0.575
Hemoglobin concentration, mean ± SD, g/L	125.04±26.67	124.85±25.78	125.60±29.38	0.850
Mechanical ventilation, n (%)	214(89.54)	160(89.39)	54(90.00)	0.893
Oxygen therapy (before-treatment), n (%)	234(97.91)	175(97.77)	58(96.67)	0.643
Ocular conditions				
Zone 1 ROP, n (%)	65(27.20)	44(24.58)	21(35.00)	0.117
Aggressive ROP, n (%)	40(16.74)	31(17.32)	9(15.00)	0.677
retinal hemorrhage, n (%)	71(29.71)	49(27.37)	22(36.67)	0.173

Predictive Performance of the Conventional Machine Learning Models

Using all clinical risk factors, the ROC curves and AUC values of models with different algorithms are compared in **eFigure 1**, the catboost achieved the best performances (AUC=0.812, SD=0.012, **Table 2**). Through comprehensive visualization offered by a comparative analysis of feature contributions across different algorithms, the top 20 features were ranked, the top five

predictors were gestational age, birth weight, PMA at the initial IVI treatment, Pneumonia and hemoglobin concentration (**eFigure 2**).

In order to evaluate the relationship between the number of the variables and predictive performance, the top 5/10/15/20/25/30 variables were introduced to five algorithms successively. The performance of most models plateaued when 20 variables were introduced. When more than 20 variables were introduced, the performance of Catboost continued to improve slightly with the addition of variables, while the performance of RF and LR began to deteriorate (**e Figure** 3). Finally, the CatBoost prediction model consisted of the top 20 variables achieved the best performances, the AUCs, sensitivities, and specificities of the CatBoost prediction model were, respectively, 0.806, 0.800, and 0.750 for the internal validation; 0.805, 0.800, and 0.657 for the test (**Table 3, Figure 3 and Figure 4**), the confusion matrix is presented in **eFigure4**.

Table 2 predictive performances (AUC) of five models

	AUC	
Algorithm	Mean	SD
AdaBoost	0.808	0.008
Random Forest	0.803	0.020
XGBoost	0.780	0.005
CatBoost	0.812	0.012
Logistic Regression	0.720	0.005

AUC: area under the curve; SD: standard derivation

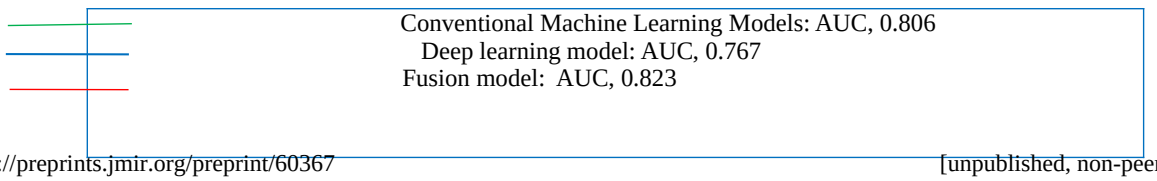
Predictive Performance of the Deep Learning Model

The predictive performance of the deep learning model based on retinal photographs captured before anti-VEGF therapy in internal validation and test are summarized in **Table 3, Figure 3 and Figure 4**. The AUCs, sensitivities, and specificities of the deep learning prediction model were, respectively, 0.767, 0.733 and 0.800 for the internal validation; 0.787, 0.800, and 0.570 for the test. To improve the interpretability of the model, we used Grad-CAM to visualize the key regions in retinal photographs highly associated with the ROP reactivation. Several representative examples of retinal photographs with accompanying saliency maps are shown in **Figure 5**. The optic and the retinal vessels were mainly used to predict reactivation, the confusion matrix is presented in **eFigure 5**.

Table 3 the performances comparison of different models

	ACC	SEN	SPE	AUC
CMLM_ val	0.771	0.800	0.750	0.806
CMLM_ test	0.716	0.800	0.657	0.805
DLM_ val	0.771	0.733	0.800	0.767
DLM_ test	0.667	0.800	0.570	0.787
FM_ val	0.742	0.800	0.700	0.823
FM_ test	0.733	0.800	0.686	0.822

CMLM: Conventional Machine Learning Model; DL: deep learning model; FM: fusion model
Val: internal validation



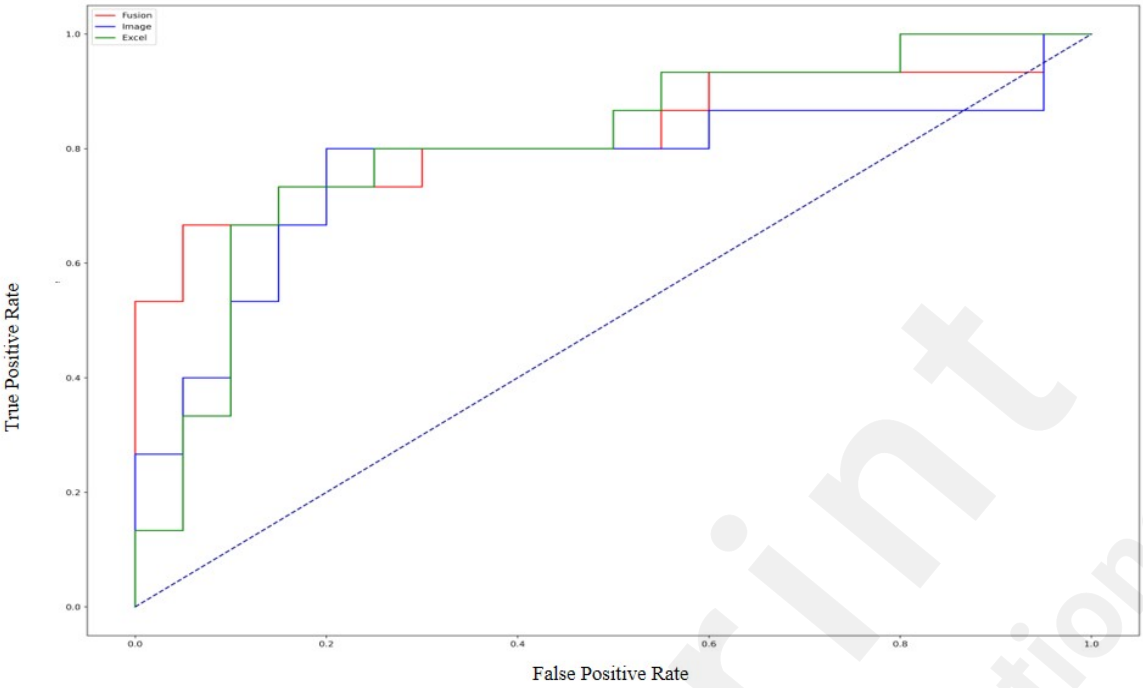
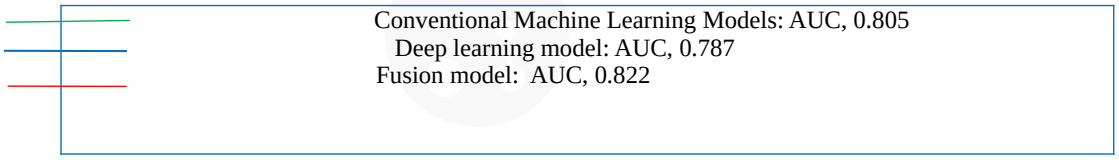


Figure 3. Receiver operating characteristic curves of three models in internal validation cohort



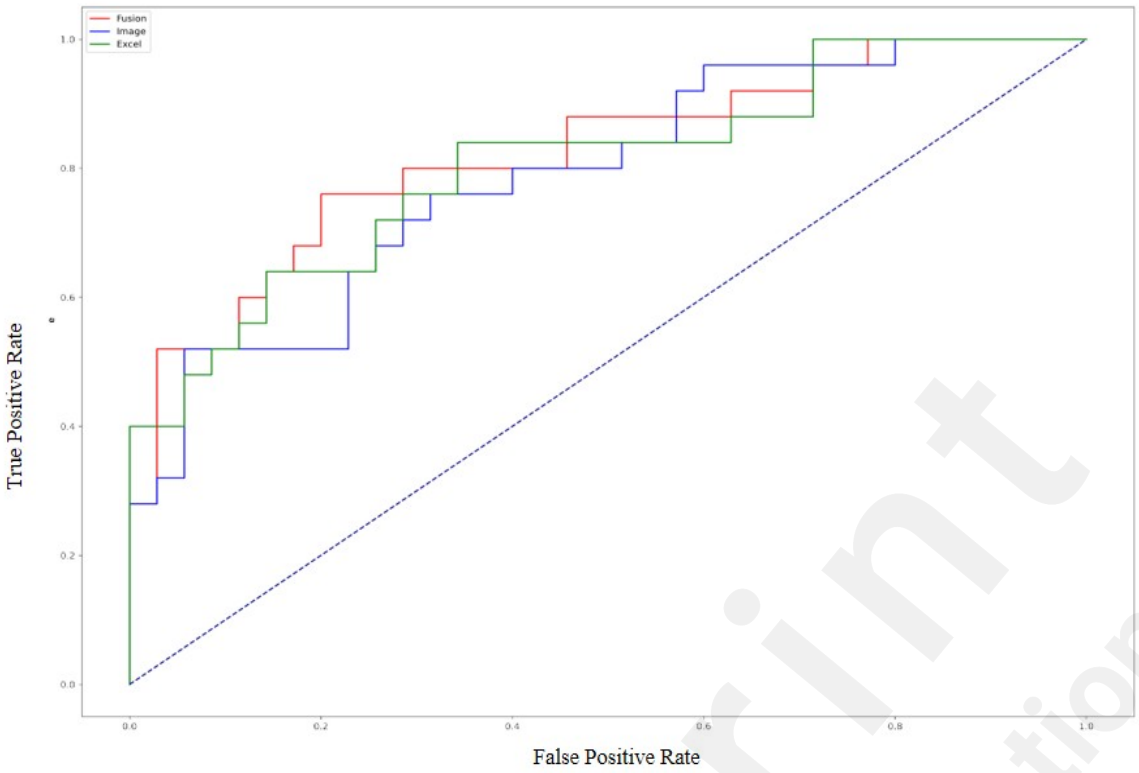


Figure 4. Receiver operating characteristic curves of three models in test cohort

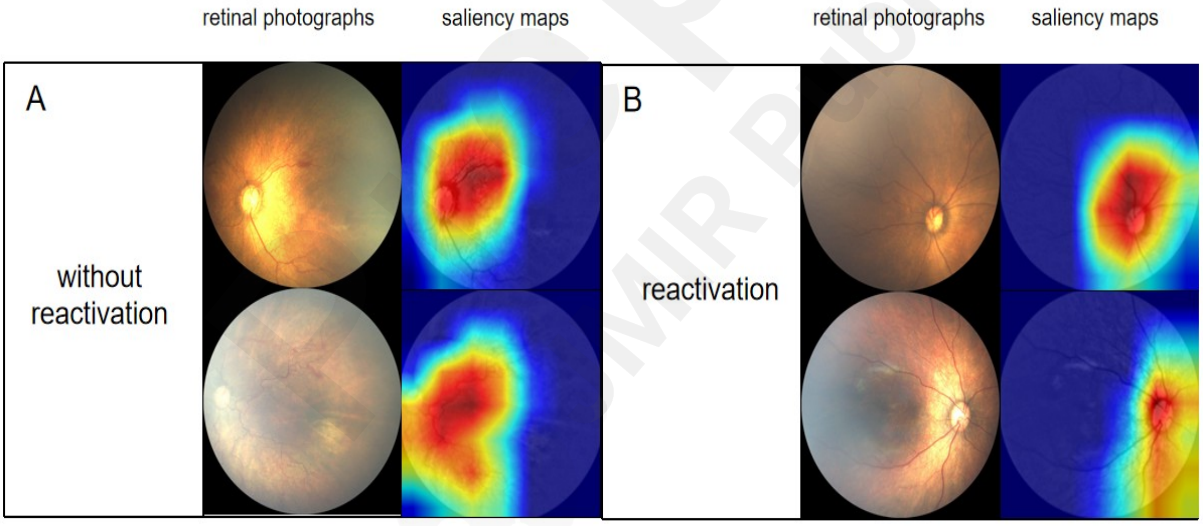


Figure 5 Original Retinal Photographs and Saliency Maps of reactivation and without reactivation Cases from Grad-CAM

Predictive Performance of the Fusion Model

Since the CatBoost model obtained the best predictive performance among the five conventional machine learning models, it was deployed to be combined with the deep learning model. The predictive performance of the fusion model in internal validation and test are summarized in **Table 3, Figure 3 and Figure 4.**. The AUCs of the fusion prediction model were 0.823 for the internal validation; 0.822 for the test, respectively, the confusion matrix is presented in **eFigure 6**.

Discussion

Given the risk of ROP reactivation after anti-VEGF therapy and the crucial need for timely intervention, this study established prediction models for ROP reactivation after anti-VEGF treatment using machine learning algorithms based on pre-treatment data, including clinical risk factors and retinal images. The machine learning prediction models achieved promising performance and were externally validated in a group of infants from other hospitals.

The ROP reactivation prediction models can optimize ophthalmologists' clinical decision-making before anti-VEGF therapy. Fluorescein angiography under anesthesia and, if necessary, laser treatment for residual areas of non-perfusion have been recommended for infants receiving anti-VEGF treatment who are at high risk of reactivation [20, 21]. In addition, more frequent retinal examinations are needed after anti-VEGF treatment. Meanwhile, a small number of infants predicted to be at low risk of ROP reactivation experienced reactivation, and treatment was repeated in this study. Since it is not acceptable to miss even one case of ROP reactivation due to devastating visual consequences, we do not recommend the discontinuation of regular follow-up examinations after treating infants at low risk of ROP reactivation.

We employed five different algorithms to select the machine learning model with the best predictive performance. Most of them achieved high performances, among which the CatBoost model exhibited optimal performance in internal validation and test. CatBoost is a modification of gradient boosted decision tree algorithm, which is compatible with categorical features [22]. In previous studies, the CatBoost algorithm has exhibited satisfactory performances in predicting malaria, Parkinson's disease, and in-hospital mortality [23-25]. In our study, the CatBoost algorithm was combined with the greedy algorithm to identify the most predictive combination of clinical risk factors for ROP reactivation. The greedy algorithm is an algorithmic paradigm that follows the problem-solving method to realize the locally optimal choice at each stage. PMA at the initial IVI treatment was significantly different between NO ROP reactivation and ROP reactivation groups in this study (NO ROP reactivation: 36.93 (2.10) weeks, ROP reactivation: 35.15 (2.21) weeks, $P<0.05$). Similarly, Lyu *et al.* [6] found that early PMA at the initial IVI treatment was a significant independent risk factor for ROP reactivation [9]. Infants with earlier postmenstrual age at initial treatment may suffer from worse systemic conditions and more severe ROP that necessitates timely treatment. The presence of pneumonia implies more severe systemic hypoxia that can exacerbate ocular hypoxia and increase the risk of neovascularization [9]. Longer oxygen treatment for pneumonia may also increase the risk of ROP reactivation after treatment with IVI.

Recently, deep learning technology has been applied to accurately diagnose and predict ROP and its severity based on retinal photographs [15, 16, 26, 27]. We previously described the use of a deep learning system to predict ROP and its severity before 45 weeks postmenstrual age [16]. Retinal status is important for predicting ROP progression and reactivation. Previous studies have demonstrated that early vascular dilation and tortuosity are insufficient to predict ROP. In addition, the extent of temporal retinal blood vessel immaturity at the first screening had prognostic significance in the early course of ROP [16]. The ROP vascular severity score derived from a deep learning classifier based on retinal photographs captured within four weeks before and after treatment and at the time of treatment can consistently reflect disease progression and post-treatment regression of ROP. The ROP vascular severity score at the time of initial treatment and progression rate after initial screening were associated with ROP reactivation, indicating that the features of retinal photographs before the initial treatment may predict ROP reactivation [7]. Although the ROP vascular severity score can help monitor ROP over time, without guidance on optimal cut points, clinicians might not be able to use the vascular severity score for ROP management [7]. In the present study, retinal photographs captured before the initial anti-VEGF therapy were analyzed using deep learning algorithms and showed promising predictive value for ROP reactivation. The optic disc and the areas around the optic disc and retinal vessels may be the potential regions for ROP reactivation,

Furthermore, we integrated the prediction model based on retinal photographs and the optimal model based on clinical risk factors. The fusion model achieved the best performance for predicting

ROP reactivation, suggesting that clinical information and biomarkers from retinal photographs play important roles in the ROP reactivation modeling. With the development of artificial intelligence, more studies combine image features with clinical information. Coyner [28] *et al.* successfully improved the specificity of the ROP prediction model by adding retinal photograph features. The model combining gestational age with vascular severity score was the best-performing model. The breast cancer recurrence model using H&E images and clinical information accurately assessed the risk of recurrence [29].

Our study had several limitations. First, the sample size was relatively small. Machine learning models would be better trained using large-scale datasets. The performance of the prediction models could be improved using larger datasets. In addition, the generalizability of the prediction models needs to be validated in prospective multiple-center datasets. Second, this study was conducted in infants receiving intravitreal injection of conbercept, whereas reactivation rates vary among different anti-VEGF agents and laser photocoagulation. Infants receiving different anti-VEGF agents and laser photocoagulation as the initial treatment for ROP should be included in future studies. Third, due to retrospective design, we could not obtain other potential predictive factors, such as post-treatment risk factors.

Conclusion

In conclusion, we successfully developed and validated machine learning prediction models based on clinical risk factors and pre-treatment retinal photographs for ROP reactivation. The promising results of the prediction models might aid early detection of ROP reactivation and decision-making processes in clinical practice.

Author Contributions:

Concept and design: Songfu Feng, Xiaohe Lu and Nian Cai. Acquisition, analysis, or interpretation of data: Rong Wu, Yichen Bai, Qiuxia Lin, Jianxun Wang and Shuangyong Wang. Statistical analysis: Yiling Xie, Rong Wu, Yu Zhang and Peijie Huang. Administrative, technical, or material support: Xiaohe Lu and Songfu Feng. Drafting of the manuscript: Rong Wu and Yu Zhang. Critical revision of the manuscript for important intellectual content: Songfu Feng. Study supervision: Xiaohe Lu. Final approval of the manuscript: All authors.

Funding

This work was supported by Guangzhou Municipal Science and Technology Project. (202002020046), the President Funding of Zhujiang Hospital, Southern Medical University. (No. yzjj2023ms19) and Foshan Self-funded Science and Technology Plan Project in 2022. (2220001005134).

Acknowledgements

The authors would like to express gratitude to all participants who have agreed to participate in this study and EditSprings for the expert linguistic services provided.

Competing interest Declarations

The authors declare that they have no known competing financial interests or personal relationships

that could have appeared to influence the work reported in this paper.

Availability of data and materials

The datasets are not publicly available due to the reason that the data is also part of an ongoing study but are available from the corresponding author on reasonable request.

Reference

[1] Hartnett M E. Retinopathy of Prematurity: Evolving Treatment With Anti-Vascular Endothelial Growth Factor[J]. Am J Ophthalmol, 2020,218:208-213.

[2] Blencowe H, Lawn J E, Vazquez T, et al. Preterm-associated visual impairment and estimates of retinopathy of prematurity at regional and global levels for 2010[J]. Pediatr Res, 2013,74 Suppl 1(Suppl 1):35-49.

[3] Wu Z, Zhao J, Lam W, et al. Comparison of clinical outcomes of conbercept versus ranibizumab treatment for retinopathy of prematurity: a multicentral prospective randomised controlled trial[J]. Br J Ophthalmol, 2022,106(7):975-979.

[4] Jin E, Yin H, Li X, et al. SHORT-TERM OUTCOMES AFTER INTRAVITREAL INJECTIONS OF CONBERCEPT VERSUS RANIBIZUMAB FOR THE TREATMENT OF RETINOPATHY OF PREMATURITY[J]. Retina, 2018,38(8):1595-1604.

[5] Bai Y, Nie H, Wei S, et al. Efficacy of intravitreal conbercept injection in the treatment of retinopathy of prematurity[J]. Br J Ophthalmol, 2019,103(4):494-498.

[6] Lyu J, Zhang Q, Chen C L, et al. Recurrence of Retinopathy of Prematurity After Intravitreal Ranibizumab Monotherapy: Timing and Risk Factors[J]. Invest Ophthalmol Vis Sci, 2017,58(3):1719-1725.

[7] Eilts S K, Pfeil J M, Poschkamp B, et al. Assessment of Retinopathy of Prematurity Regression and Reactivation Using an Artificial Intelligence-Based Vascular Severity Score[J]. JAMA Netw Open, 2023,6(1): e2251512.

[8] Cheng Y, Zhu X, Linghu D, et al. Comparison of the effectiveness of conbercept and ranibizumab treatment for retinopathy of prematurity[J]. Acta Ophthalmol, 2020,98(8):e1004-e1008.

[9] Ling K P, Liao P J, Wang N K, et al. RATES AND RISK FACTORS FOR RECURRENCE OF RETINOPATHY OF PREMATURITY AFTER LASER OR INTRAVITREAL ANTI-VASCULAR ENDOTHELIAL GROWTH FACTOR MONOTHERAPY[J]. Retina, 2020,40(9):1793-1803.

[10] Chiang M F, Quinn G E, Fielder A R, et al. International Classification of Retinopathy of Prematurity, Third Edition[J]. Ophthalmology, 2021,128(10): e51-e68.

[11] Iwahashi C, Utamura S, Kuniyoshi K, et al. FACTORS ASSOCIATED WITH REACTIVATION AFTER INTRAVITREAL BEVACIZUMAB OR RANIBIZUMAB THERAPY IN INFANTS WITH RETINOPATHY OF PREMATURITY[J]. Retina, 2021,41(11):2261-2268.

[12] Decuyper M, Maebe J, Van Holen R, et al. Artificial intelligence with deep learning in nuclear medicine and radiology[J]. EJNMMI Phys, 2021,8(1):81.

[13] Wang X, Wang R, Yang S, et al. Combining Radiology and Pathology for Automatic Glioma Classification[J]. Front Bioeng Biotechnol, 2022,10:841958.

[14] Du-Harpur X, Watt F M, Luscombe N M, et al. What is AI? Applications of artificial intelligence to dermatology[J]. Br J Dermatol, 2020,183(3):423-430.

[15] Brown J M, Campbell J P, Beers A, et al. Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks[J]. JAMA Ophthalmol, 2018,136(7):803-810.

[16] Wu Q, Hu Y, Mo Z, et al. Development and Validation of a Deep Learning Model to Predict the Occurrence and Severity of Retinopathy of Prematurity[J]. JAMA Netw Open, 2022,5(6):e2217447.

[17] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015,521(7553):436-444.

[18] Ham Y G, Kim J H, Luo J J. Deep learning for multi-year ENSO forecasts[J]. Nature, 2019,573(7775):568-572.

[19] Li J O, Liu H, Ting D, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective[J]. Prog Retin Eye Res, 2021,82:100900.

[20] Moshfeghi D M. Systemic Solutions in Retinopathy of Prematurity[J]. Am J Ophthalmol, 2018,193:xiv-xviii.

[21] Garcia G J, Snyder L, Blair M, et al. PROPHYLACTIC PERIPHERAL LASER AND FLUORESCEIN ANGIOGRAPHY AFTER BEVACIZUMAB FOR RETINOPATHY OF PREMATURITY[J]. Retina, 2018,38(4):764-772.

[22] Hancock J T, Khoshgoftaar T M. CatBoost for big data: an interdisciplinary review[J]. J Big Data, 2020,7(1):94.

[23] Lee Y W, Choi J W, Shin E H. Machine learning model for predicting malaria using clinical information[J]. Comput Biol Med, 2021,129:104151.

[24] Yang B, Wang X, Mo J, et al. The amplitude of low-frequency fluctuation predicts levodopa treatment response in patients with Parkinson's disease[J]. Parkinsonism Relat Disord, 2021,92:26-32.

[25] Li L, Zhang Z, Xiong Y, et al. Prediction of hospital mortality in mechanically ventilated patients with congestive heart failure using machine learning approaches[J]. Int J Cardiol, 2022,358:59-64.
[26] Tan Z, Simkin S, Lai C, et al. Deep Learning Algorithm for Automated Diagnosis of Retinopathy of Prematurity Plus Disease[J]. Transl Vis Sci Technol, 2019,8(6):23.
[27] Wang J, Ji J, Zhang M, et al. Automated Explainable Multidimensional Deep Learning Platform of Retinal Images for Retinopathy of Prematurity Screening[J]. JAMA Netw Open, 2021,4(5):e218758.
[28] Coyner A S, Chen J S, Singh P, et al. Single-Examination Risk Prediction of Severe Retinopathy of Prematurity[J]. Pediatrics, 2021,148(6).
[29] Yang J, Ju J, Guo L, et al. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning[J]. Comput Struct Biotechnol J, 2022,20:333-342.

Supplementary Files

Multimedia Appendixes

Supplemental information (tables and figures).
URL: http://asset.jmir.pub/assets/c3be37bf2392fdbbcc9ee7f2273b08a2.docx