Original Paper

Prediction of Reactivation After Antivascular Endothelial Growth Factor Monotherapy for Retinopathy of Prematurity: Multimodal Machine Learning Model Study

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

[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 Pediatric Ophthalmology, Guangzhou Children's Hospital and Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China

[4]Department of Ophthalmology, Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Xiaohe Lu, MD, PhD
Department of Ophthalmology
Zhujiang Hospital
Southern Medical University
No 253 Gongyedadao Middle Road
Guangzhou, 510260
China
Phone: 86 15002000613
Email: luxh63@163.com

Abstract

Background: Retinopathy of prematurity (ROP) is the leading preventable cause of childhood blindness. A timely intravitreal injection of antivascular endothelial growth factor (anti-VEGF) is required to prevent retinal detachment with consequent vision impairment and loss. However, anti-VEGF has been reported to be associated with ROP reactivation. Therefore, an accurate prediction of reactivation after treatment is urgently needed.

Objective: To develop and validate prediction models for reactivation after anti-VEGF intravitreal injection in infants with ROP using multimodal machine learning algorithms.

Methods: Infants with ROP undergoing anti-VEGF treatment were recruited from 3 hospitals, and conventional machine learning, deep learning, and fusion models were constructed. The areas under the curve (AUCs), accuracy, sensitivity, and specificity were used to show the performances of the prediction models.

Results: A total of 239 cases with anti-VEGF treatment were recruited, including 90 (37.66%) with reactivation and 149 (62.34%) nonreactivation cases. The AUCs for the conventional machine learning model were 0.806 and 0.805 in the internal validation and test groups, respectively. The average AUC, sensitivity, and specificity in the test for the deep learning model were 0.787, 0.800, and 0.570, respectively. The specificity, AUC, and sensitivity for the fusion model were 0.686, 0.822, and 0.800 in a test, separately.

Conclusions: We constructed 3 prediction models for ROP reactivation. The fusion model achieved the best performance. Using this prediction model, we could optimize strategies for treating ROP in infants and develop better screening plans after treatment.

(J Med Internet Res 2025;27:e60367) doi: 10.2196/60367

KEYWORDS

retinopathy of prematurity; reactivation; prediction; machine learning; deep learning; anti-VEGF

XSL•FO
RenderX

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 [1-3]. More than 20,000 infants experience blindness, and an estimated 12,300 infants have different levels of visual impairment due to ROP [4]. A dysregulation of vascular endothelial growth factor (VEGF) has been proven to play an important role in the development of ROP [5,6]. Laser photocoagulation and anti-VEGF agents are the mainstay treatments for ROP. The clinical application of intravitreal injection (IVI) with 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 [7-10]. 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 between 6.8% and 64% after treatment [8,9,11-13].

Early detection and timely treatment upon reactivation is critical for infants with ROP 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 [12,14]. Neonatal factors include low gestational age, low birth weight, early postmenstrual age (PMA) at initial treatment, and low Apgar scores [12]. Maternal factors include multiple births [12]. Other risk factors related to neonatal interventions include oxygen requirement before or after treatment and a longer duration of hospitalization [9]. 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 [15-17]. Artificial intelligence has been mainly applied to ROP diagnosis and prediction based on imaging [17-19]. To our knowledge, studies on ROP reactivation after treatment are very limited, and there is no successful prediction model for clinical application. Machine learning is a subset of artificial intelligence and includes conventional machine learning and deep learning [20-22]. In this 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

Ethical Considerations

This study was approved by the study hospitals' institutional ethics committees (2022-KY-143) 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 deidentified before analysis.

Study Population

We collected data retrospectively on infants who received anti-VEGF monotherapy for ROP requiring treatment between April 2016 and November 2022 at Hospital 1 (Zhujiang Hospital, Southern Medical University), Hospital 2 (The Third Affiliated Hospital, Guangzhou Medical University), and Hospital 3 (Guangzhou Children's Hospital and Guangzhou Women and Children's Medical Center), all in Guangzhou, China. Infants with incomplete data or any other ocular diseases besides ROP were excluded. Additionally, 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 the RetCam III digital fundus camera (Natus). The diagnosis of ROP and the indication for its treatment were based on the International Classification of ROP Revisited and the Early Treatment for ROP study, respectively. Treatment-requiring ROP included threshold disease, stage 4 or 5 ROP, and type 1 or aggressive ROP. Ocular examinations were conducted before and on days 1, 7, 14, and 28 after anti-VEGF therapy, and either biweekly or monthly depending on the ocular findings and systemic conditions. A 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 disease, vascular dilation, tortuosity, or new/recurrent neovascularization that required further treatment [13].

IVI of Anti-VEGF

All parents or guardians of the infants were fully informed of the efficacy and possible complications prior to IVI of conbercept, and they 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 aesthesia. Topical tobramycin dexamethasone was administered for 3 days after the injection. All operations were done by a trained pediatric ophthalmologist (author SF).

Clinical Risk Factors

Based on previous studies and clinical experience, the potential clinical risk factors of ROP reactivation extracted from electronic medical records included 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 and embryo transfer. Neonatal factors included gestational age, birth weight, 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

XSL•FO
RenderX

hyperbilirubinemia. Ocular conditions included zone I ROP, preretinal hemorrhage, and aggressive ROP. The hemoglobin concentration before treatment was included among the laboratory factors. Finally, neonatal interventions included mechanical ventilation and oxygen therapy (before treatment) [9,11,12].

In the data processing stage, any missing data were handled according to the type of variable. For continuous data, we imputed missing values by calculating the mean within each category based on the label (0 or 1). For discrete data, we used the mode within each category to fill in missing values. This approach ensured that the values imputed were representative of their respective categories, thus minimizing bias in the analysis.

Image Collection and Case Labeling

All retinal photographs were captured using the commercial RetCam camera. Retinal photographs of poor photographic quality were excluded by 2 experienced ophthalmologists (Authors SF and RW). Since the prediction for ROP reactivation would be performed independently on each infant rather than each retinal photograph, retinal photographs of both eyes from the same infant were labeled as a single case. All cases were labeled independently by 2 clinical ophthalmologists (SF and RW). Based on the ocular findings after anti-VEGF therapy, each case was annotated reactivation or nonreactivation. 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, indicating good agreement between the 2 ophthalmologists in labeling. Moreover, the labels were further confirmed by a senior retinal specialist (author XL) to generate a final annotation. These annotations were used as ground-truth labels in the development and validation of the prediction models for ROP reactivation.

Development and Validation of the 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 for 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; Python Software Foundation).

During the data preprocessing stage, we employed 2 filling strategies. For discrete data, we used mean filling followed by rounding; for continuous data, we used mean filling. Subsequently, the continuous data underwent standardization, transforming the data into a standard normal distribution with a mean of 0 (SD 1), ensuring uniform scales across different features. This helped mitigate potential model biases arising from scale differences, enhancing overall model robustness and performance.

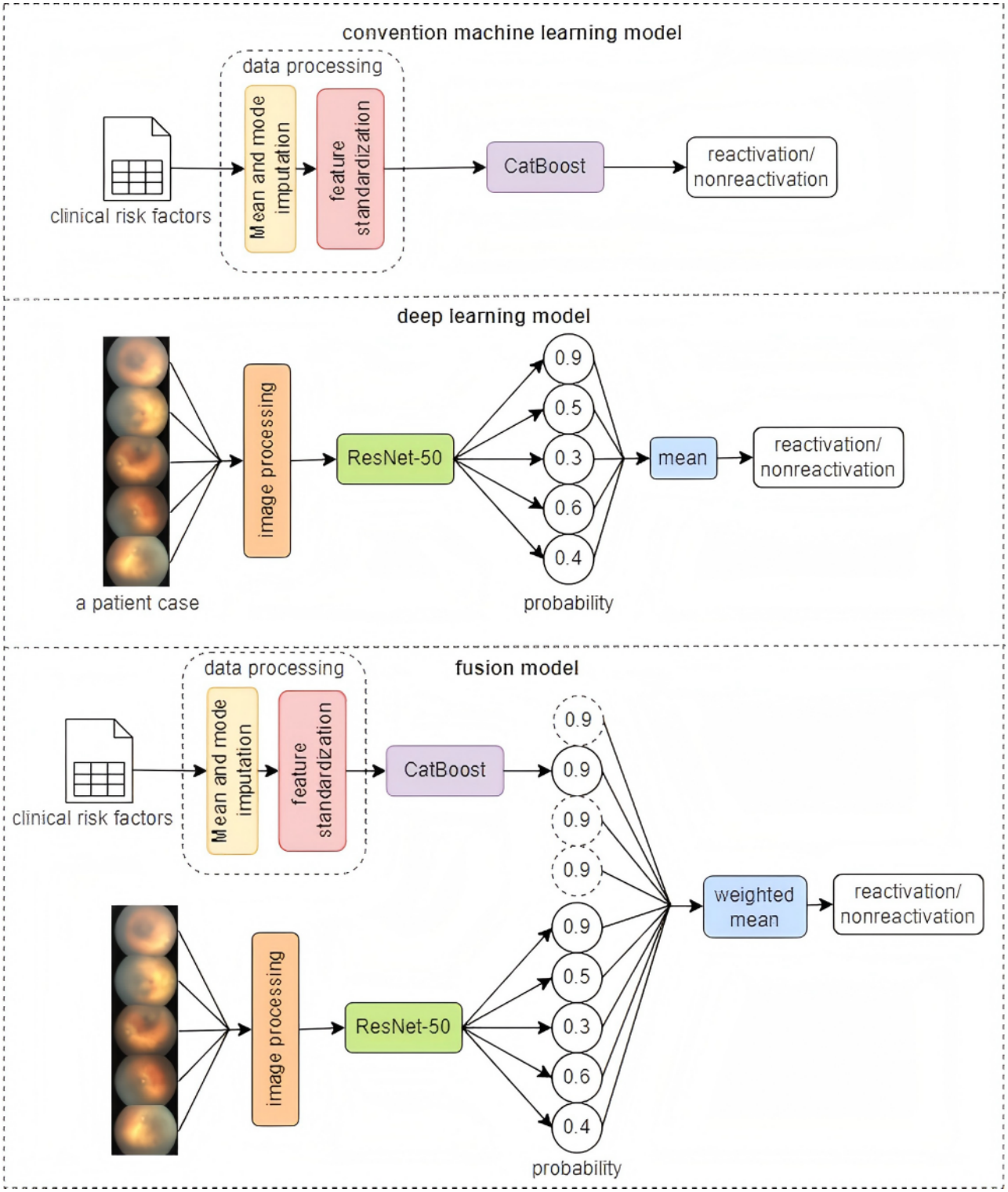
In this study, we employed a comprehensive approach to assess the feature importance in predictive models, utilizing 5 different algorithms: random forest (RF), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), and logistic regression (LR). Table S1 in Multimedia Appendix 1 contains the parameter details. The method involves ranking the feature importance or weights for each algorithm, whereas for LR, feature importance was determined by the absolute values of the weights. Subsequently, these rankings were visualized using scatter plots, with each feature represented along the x-axis and its importance rank along the y-axis. This comprehensive visualization offered a comparative analysis of feature contributions across different algorithms, providing insights into the robustness and consistency of the feature importance.

We conducted an in-depth investigation into the impact of varying the number of selected features (N) on the predictive model performance, employing 5 distinct algorithms. We systematically varied the number of selected features, choosing from the set (5, 10, 15, 20, 25, and 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 the 5 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 data set. 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 10 iterations of training for each of the 5 distinct predictive models and computed their average AUC values on the corresponding validation sets, along with their respective SDs.

XSL•FO
RenderX

Figure 1. Illustration of the proposed algorithmic pipeline. CatBoost: Categorical Boosting.



Deep Learning Model

Figure 1 presents an illustration of the deep learning prediction model. Retinal photographs captured before anti-VEGF therapy of infants with ROP were obtained to develop a prediction model using deep learning algorithms. ResNet-50, a deep convolutional neural network classifier pretrained on the ImageNet database, was fine-tuned with our data set to predict ROP reactivation. The process was implemented with Python (version 3.9) and PyTorch (version 1.11.0).

We designed a series of image preprocessing methods beginning with feature enhancement and noise suppression using median filtering. This technique was chosen for its effectiveness in removing isolated noise pixels while maintaining the clarity of important features such as blood vessel edges. Subsequently, the images were converted to grayscale for processing to reduce computational complexity.

For the training set, we employed a series of preprocessing steps. These included resizing the images to 224 × 224 pixels, introducing RandAugment for random data augmentation, and

XSL•FO
RenderX

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 224 × 224 pixels and transformed them into tensors. This ensured uniformity in the model's input during the validation process. This diversified preprocessing approach helped the model adapt to the distinct characteristics of each data set, ultimately improving its generalization performance.

We employed a pretrained 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 learned to adjust its parameters by minimizing the cross-entropy loss, enabling it to more accurately predict the samples' categories. 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 the model parameters and improve the overall model generalization.

We constructed a data set in which each case consisted of multiple medical images and randomly selected 5 images per patient for the experiment. 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 the ROP reactivation prediction, we applied the Gradient-Weighted Class Activation Mapping (Grad-CAM) technique to generate heat maps of key regions. This technique highlights areas crucial to ROP reactivation prediction by analyzing feature weights in specific network layers, such as layer 4 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 5 conventional machine learning prediction models, the model based on the CatBoost algorithm exhibited the best predictive performance. We weighed and averaged the probability outputs of the optimal conventional machine learning model and the deep learning model to obtain the final prediction result. 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 images, can capture complex patterns that traditional models find challenging.

To fully leverage the strengths of both models, we treated the 5 probability scores generated by the deep learning model for each image as equal in weight and assigned the output of the risk factor model a weight equivalent to that of 4 images, thereby producing a more accurate and robust prediction result. We obtained the deep learning probability scores for each sample through independent training on the training set, 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.

Statistical Analysis

Statistical analyses were performed using R software (version 3.0.2; R Core Team). Categorical variables were expressed as numbers and percentages and analyzed using chi-square tests. If any cell number was less than 5, the Fisher exact test was applied. Continuous variables conforming to normal distribution were expressed as mean and SD and compared using the independent 2-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 using Python (version 3.7.0). $P<.05$ was considered statistically significant.

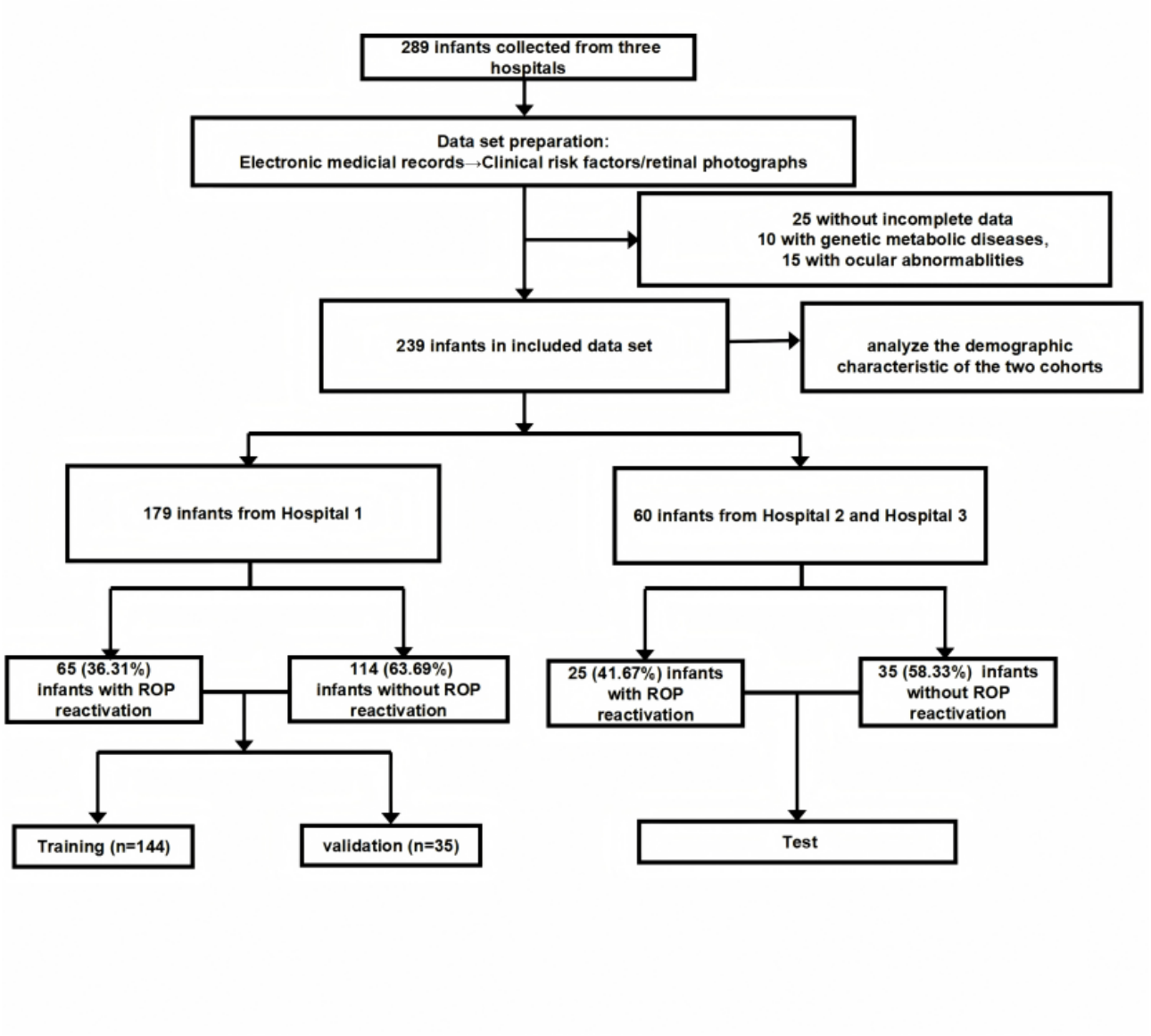
Results

Demographics of the Study Groups

The study group finally included 239 infants. The infant recruitment flow is presented in Figure 2. The detailed demographic characteristics of the enrolled infants are summarized in Table 1. Among the 179 included infants of Hospital 1, the mean gestational age was 28.87 (SD 2.46) weeks, and the mean birth weight was 1.18 (SD 0.35) kg. Among the 60 infants enrolled from Hospital 2 and Hospital 3, the mean gestational age was 28.30 (SD 1.90) weeks, and the mean birth weight was 1.09 (SD 0.28) kg. There was no significant difference between the demographic characteristics of the 2 cohorts (all $P>.05$).

XSL•FO
RenderX

Figure 2. The flow chart of patient cohorts. ROP: retinopathy of prematurity.



XSL•FO
RenderX

Table 1. Demographic characteristics of all the enrolled infants.

Variables	All infants (N=239)	Training (n=179)	Test (n=60)	*P* value
Maternal factors				
Maternal age (years), mean (SD)	30.72 (5.32)	30.65 (5.08)	30.95 (6.03)	.73
In vitro fertilization and embryo transfer, n (%)	41 (17.15)	31 (17.32)	10 (16.67)	.91
Gestational hypertension, n (%)	30 (12.55)	24(13.41)	6 (10)	.49
Gestational diabetes mellitus, n (%)	34 (14.23)	25 (13.97)	9 (15)	.84
Cesarean delivery, n (%)	105 (43.93)	79 (44.13)	26 (43.34)	.91
Premature rupture of membranes, n (%)	37 (15.48)	26 (14.53)	11 (17.74)	.48
Neonatal factors				
Gestational age (weeks), mean (SD)	28.66 (2.38)	28.87 (2.46)	28.30 (1.90)	.07
Birth weight (kg), mean (SD)	1.15 (0.34)	1.18 (0.35	1.09 (0.28)	.05
Multiple gestations, n (%)	58 (24.27)	46 (25.70)	12 (20)	.37
Small for gestational age, n (%)	39 (16.32)	30 (16.76)	9 (15)	.75
Intrauterine distress, n (%)	22 (9.20)	18 (10.06)	4 (6.67)	.43
Postmenstrual age at initial ROP[a] treatment (weeks), mean (SD)	36.31 (2.29)	36.29 (2.31)	36.38 (2.25)	.78
Male sex, n (%)	131 (54.81)	94 (52.51)	37 (61.67)	.22
Apgar scores at 1 minute, mean (SD)	7.35 (2.05)	7.38 (2.07)	7.25 (2.01)	.67
Apgar scores at 5 minutes, mean (SD)	8.58 (1.39)	8.62 (1.43)	8.58 (1.29)	.98
Asphyxia, n (%)	98 (41)	72 (40.2)	26 (43.33)	.67
Respiratory distress syndrome, n (%)	179 (75.24)	130 (72.63)	49 (81.67)	.16
Bronchopulmonary dysplasia, n (%)	146 (61.09)	107 (59.78)	39 (65)	.47
Sepsis, n (%)	65 (27.20)	47 (26.26)	18 (30)	.57
Necrotizing enterocolitis, n (%)	37 (15.48)	27 (15.08)	10 (16.67)	.55
Intraventricular hemorrhage, n (%)	107 (44.76)	80 (44.69)	27 (45)	.97
Hypoxic-ischemic encephalopathy, n (%)	48 (20.08)	35 (19.55)	13 (21.67)	.72
Patent ductus arteriosus, n (%)	90 (37.66)	70 (39.11)	20 (33.33)	.42
Atrial septal defect, n (%)	39 (16.32)	27 (15.08)	12 (20)	.37
Patent foramen ovale, n (%)	183 (76.57)	139 (77.65)	44 (73.33)	.49
Hyperbilirubinemia, n (%)	148 (61.92)	110 (61.45)	38 (63.33)	.80
Pneumonia, n (%)	140 (58.58)	103 (57.54)	37 (61.67)	.58
Hemoglobin concentration (g/L), mean (SD)	125.04 (26.67)	124.85 (25.78)	125.60 (29.38)	.85
Mechanical ventilation, n (%)	214 (89.54)	160 (89.39)	54 (90)	.89
Oxygen therapy (before treatment), n (%)	234 (97.91)	175 (97.77)	58 (96.67)	.64
Ocular conditions				
Zone 1 ROP, n (%)	65 (27.20)	44 (24.58)	21 (35)	.12
Aggressive ROP, n (%)	40 (16.74)	31 (17.32)	9 (15)	.68
Retinal hemorrhage, n (%)	71(29.71)	49(27.37)	22(36.67)	.17

[a]ROP: retinopathy of prematurity.

XSL•FO
RenderX

Predictive Performance of the Conventional Machine Learning Models

Using all clinical risk factors, the ROC curves and AUC values of the models with different algorithms are compared in Figure S1 in Multimedia Appendix 1. CatBoost had the best performance, with a mean AUC of 0.812 (SD 0.012), as shown in Table 2. Through comprehensive visualization offered by a comparative analysis of feature contributions across different algorithms, the top 20 features were ranked. The top 5 predictors were gestational age, birth weight, PMA at the initial IVI treatment, pneumonia, and hemoglobin concentration (Figure S2 in Multimedia Appendix 1).

Table 2. Predictive performances of the 5 models.

Algorithm	AUC[a], mean (SD)
AdaBoost	0.808 (0.008)
Random Forest	0.803 (0.020)
XGBoost[b]	0.780 (0.005)
CatBoost[c]	0.812 (0.012)
Logistic regression	0.720 (0.005)

[a]AUC: area under the curve.

[b]XGBoost: Extreme Gradient Boosting.

[c]CatBoost: Categorical Boosting.

To evaluate the relationship between the number of variables and predictive performance, the top 5, 10, 15, 20, 25, and 30 variables were introduced to 5 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 (Figure S3 in Multimedia Appendix 1). Finally, the CatBoost prediction model with the top 20 variables had the best performance. The AUCs, sensitivities, and specificities of the CatBoost prediction model were 0.806, 0.800, and 0.750 for the internal validation, respectively, and 0.805, 0.800, and 0.657 for the test, respectively (Table 3 and Figures 3 and 4). The confusion matrix is presented in Figure S4 in Multimedia Appendix 1.

Table 3. Performance comparison of the different models.

Model	ACC[a]	SEN[b]	SPE[c]	AUC[d]
CMLM[e]_ val[f]	0.771	0.800	0.750	0.806
CMLM_ test	0.716	0.800	0.657	0.805
DLM[g]_ val	0.771	0.733	0.800	0.767
DLM_ test	0.667	0.800	0.570	0.787
FM[h]_ val	0.742	0.800	0.700	0.823
FM_ test	0.733	0.800	0.686	0.822

[a]ACC: accuracy.

[b]SEN: sensitivity.

[c]SPE: specificity.

[d]AUC: area under the curve.

[e]CMLM: conventional machine learning model.

[f]Val: internal validation.

[g]DLM: deep learning model.

[h]FM: fusion model.

XSL•FO
RenderX

Figure 3. Receiver operating characteristic (ROC) curves of 3 models in the internal validation cohort. CMLM: conventional machine learning model; DLM: deep learning model; FM: fusion model.

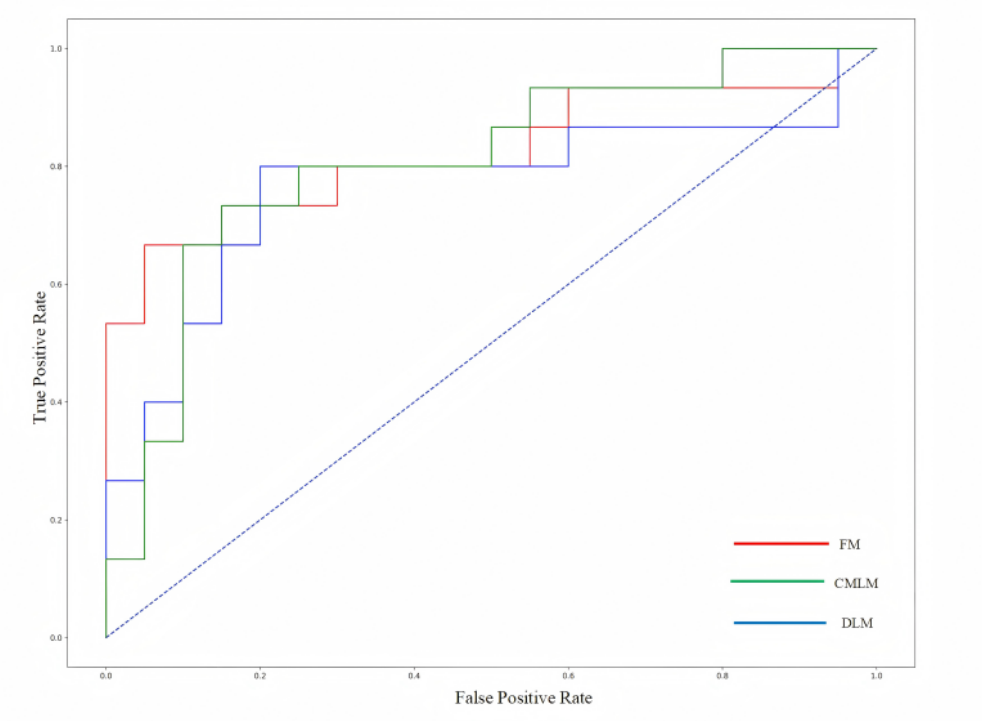
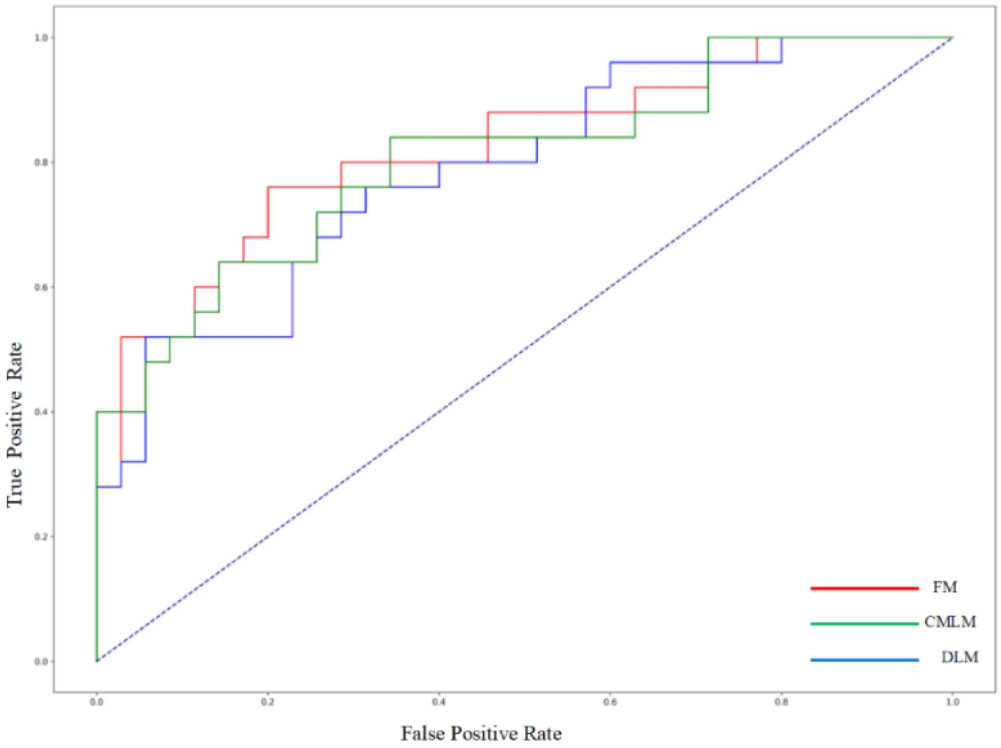


Figure 4. Receiver operating characteristic (ROC) curves of 3 models in the test cohort. CMLM: conventional machine learning model; DLM: deep learning model; FM: fusion model.



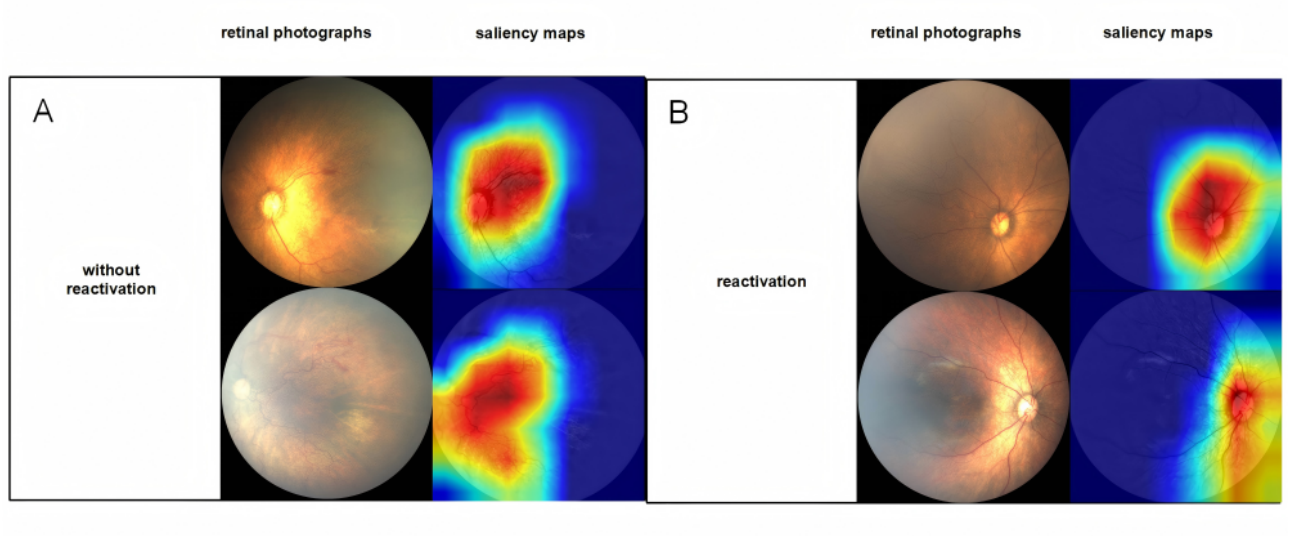
Predictive Performance of the Deep Learning Model

The predictive performances of the deep learning model based on retinal photographs captured before anti-VEGF therapy in internal validation and test are summarized in Table 3 and Figures 3 and 4. The AUCs, sensitivities, and specificities of the deep learning prediction model were 0.767, 0.733, and 0.800 for the internal validation, respectively, and 0.787, 0.800, and 0.570 for the test, respectively. To improve the interpretability of the model, we used Grad-CAM to visualize the key regions in retinal photographs highly associated with 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.

XSL•FO
RenderX

The confusion matrix is presented in Figure S5 in Multimedia Appendix 1.

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 had the best predictive performance among the 5 conventional machine learning models, it was deployed to be combined with the deep learning model. The AUCs of the fusion prediction model were 0.823 for the internal validation and 0.822 for the test. The confusion matrix is presented in Figure S6 in Multimedia Appendix 1.

Discussion

Principal Findings

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 pretreatment 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 different 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 nonperfusion have been recommended for infants receiving anti-VEGF treatment who are at high risk of reactivation [23,24]. 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 recommended to miss even one case of ROP reactivation due to its devastating visual consequences, we do not recommend discontinuing regular follow-up examinations after treating infants at low risk of ROP reactivation.

We employed 5 different algorithms to select the machine learning model with the best predictive performance. Most of them achieved high performance, among which the CatBoost model exhibited optimal performance in the internal validation and test. CatBoost is a modification of the gradient-boosted decision tree algorithm, which is compatible with categorical features [25]. In previous studies, the CatBoost algorithm has shown satisfactory performance in predicting malaria, Parkinson disease, perioperative major adverse cardiovascular events, and in-hospital mortality [26-29]. 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 a problem-solving approach to realize the locally optimal choice at each stage. PMA at the initial IVI treatment was significantly different between the non-ROP reactivation and ROP reactivation groups in this study (non-ROP reactivation: mean 36.93, SD 2.10 weeks versus ROP reactivation: mean 35.15, SD 2.21 weeks; $P<.05$). Similarly, Lyu et al [9] found that early PMA at the initial IVI treatment was a significant independent risk factor for ROP reactivation [12]. Infants with earlier PMA at initial treatment may experience worse systemic conditions and severe ROP that necessitate timely treatment. The presence of pneumonia implies more severe systemic hypoxia that can exacerbate ocular hypoxia and increase the risk of neovascularization [12]. 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 [18,19,30-32]. We previously described the use of a deep learning system to predict ROP and its severity before 45 weeks PMA [19]. 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 has a prognostic significance in the early course of ROP [19]. The ROP vascular severity score derived from a deep learning classifier based on retinal photographs captured within 4 weeks before and after treatment and at the time of treatment could consistently reflect disease progression and posttreatment

XSL•FO
RenderX

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 initial treatment may predict ROP reactivation [10]. 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 [10]. In this 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 ROP reactivation modeling. With the development of artificial intelligence, more studies are combining image features with clinical information. Coyner et al [33] 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 hematoxylin–eosin-stained images and clinical information accurately assessed the risk of recurrence [34]. Eilts et al [10] reported that retinal photographs before and after treatment were predictive for ROP reactivation after ranibizumab injection. Their study evaluated the potential role of an artificial intelligence–derived screening method in predicting ROP reactivation after anti-VEGF treatment. However, their study had some limitations, the main one being the small sample size, as only 19 infants were included. Moreover, only posterior retinal photographs captured from RetCam were analyzed, potentially missing the value from peripheral retinal photographs. On the other hand, in this study, we collected more cases and captured retinal photographs from different angles.

Study Limitations

Our study had several limitations. First, the sample size was relatively small. Machine learning models perform better when trained using large-scale data sets. In addition, the generalizability of the prediction models needs to be validated in prospective multiple-center data sets. Second, this study was conducted in infants receiving IVI of conbercept, whereas reactivation rates may vary with different anti-VEGF agents and laser photocoagulation. Infants treated with other anti-VEGF agents and laser photocoagulation as the initial treatment for ROP should be included in future studies. Third, due to this study's retrospective design, we could not obtain other potential predictive factors, such as posttreatment risk factors. Fourth, the retrospective design may limit the amount and quality of the collected data. A limitation of this model architecture is that it does not have an end-to-end design, which can lead to error accumulation. The errors in the output of each model may be amplified during the weighted averaging, affecting the final prediction results. One potential improvement could be to optimize the model through end-to-end joint training, using an ensemble learning framework or neural network to automatically adjust the weights and reduce error accumulation.

Conclusion

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

Acknowledgments

This study was supported by the Guangzhou Municipal Science and Technology Project (202002020046), the President Funding of Zhujiang Hospital, Southern Medical University (yzjj2023ms19), and the Foshan Self-funded Science and Technology Plan Project (2220001005134) in 2022. 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.

Data Availability

The data sets are not publicly available because they are part of an ongoing study. However, the data are available from the corresponding author on reasonable request.

Authors' Contributions

SF, XL, and NC contributed to the study concept and design. RW, YB, QL, JW, and SW were responsible for data acquisition, analysis, and interpretation. The statistical analysis was carried out by YX, RW, YZ, and PH. Administrative, technical, and material support was provided by XL and SF. RW and YZ drafted the manuscript. SF carried out a critical revision of the manuscript for important intellectual content. XL supervised the study. All authors approved the final manuscript. SF, XL, and NC contributed equally.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1

Supplementary.
[DOCX File , 506 KB-Multimedia Appendix 1]

References

1. Hartnett ME. Retinopathy of prematurity: evolving treatment with anti-vascular endothelial growth factor. Am J Ophthalmol. Oct 2020;218:208-213. [FREE Full text] [doi: 10.1016/j.ajo.2020.05.025] [Medline: 32450064]
2. Krohne TU, Camp AT, Pfeil JM, Müller A, Stahl A, Lagrèze WA, et al. Retinopathy of prematurity in Germany over 13 years: incidences, treatment preferences and effects of national guideline changes. Arch Dis Child Fetal Neonatal Ed. Dec 20, 2024;110(1):37-42. [doi: 10.1136/archdischild-2024-327133] [Medline: 38816192]
3. Bhatnagar A, Skrehot HC, Bhatt A, Herce H, Weng CY. Epidemiology of retinopathy of prematurity in the US from 2003 to 2019. JAMA Ophthalmol. May 01, 2023;141(5):479-485. [FREE Full text] [doi: 10.1001/jamaophthalmol.2023.0809] [Medline: 37052930]
4. Blencowe H, Lawn JE, Vazquez T, Fielder A, Gilbert C. Preterm-associated visual impairment and estimates of retinopathy of prematurity at regional and global levels for 2010. Pediatr Res. Dec 2013;74 Suppl 1(Suppl 1):35-49. [FREE Full text] [doi: 10.1038/pr.2013.205] [Medline: 24366462]
5. Wu Z, Zhao J, Lam W, Yang M, Chen L, Huang X, et al. Comparison of clinical outcomes of conbercept versus ranibizumab treatment for retinopathy of prematurity: a multicentral prospective randomised controlled trial. Br J Ophthalmol. Jul 2022;106(7):975-979. [FREE Full text] [doi: 10.1136/bjophthalmol-2020-318026] [Medline: 33637618]
6. Strawbridge J, Cheng JY, Gundlach BS, Gillespie T, Karmouta R, Khitri M, et al. Short-term reactivation of retinopathy of prematurity after primary ranibizumab treatment. Retina. Nov 01, 2024;44(11):1945-1951. [doi: 10.1097/IAE.0000000000004206] [Medline: 39024625]
7. Jin E, Yin H, Li X, Zhao M. Short-term outcomes after intravitreal injections of conbercept versus ranibizumab for the treatment of retinopathy of prematurity. Retina. Aug 2018;38(8):1595-1604. [doi: 10.1097/IAE.0000000000001763] [Medline: 28699927]
8. Bai Y, Nie H, Wei S, Lu X, Ke X, Ouyang X, et al. Efficacy of intravitreal conbercept injection in the treatment of retinopathy of prematurity. Br J Ophthalmol. Apr 2019;103(4):494-498. [doi: 10.1136/bjophthalmol-2017-311662] [Medline: 30030391]
9. Lyu J, Zhang Q, Chen C, Xu Y, Ji X, Li J, et al. Recurrence of retinopathy of prematurity after intravitreal ranibizumab monotherapy: timing and risk factors. Invest Ophthalmol Vis Sci. Mar 01, 2017;58(3):1719-1725. [doi: 10.1167/iovs.16-20680] [Medline: 28324112]
10. Eilts SK, Pfeil JM, Poschkamp B, Krohne TU, Eter N, Barth T, et al. Assessment of retinopathy of prematurity regression and reactivation using an artificial intelligence-based vascular severity score. JAMA Netw Open. Jan 03, 2023;6(1):e2251512. [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.51512] [Medline: 36656578]
11. Cheng Y, Zhu X, Linghu D, Liang J. Comparison of the effectiveness of conbercept and ranibizumab treatment for retinopathy of prematurity. Acta Ophthalmol. Dec 2020;98(8):e1004-e1008. [FREE Full text] [doi: 10.1111/aos.14460] [Medline: 32385940]
12. Ling K-P, Liao PJ, Wang NK, Chao A-N, Chen K-J, Chen T-L, et al. Rates and risk factors for recurrence of Retinopathy of Prematurity after laser or intravitreal anti-vascular endothelial growth factor monotherapy. Retina. Sep 2020;40(9):1793-1803. [doi: 10.1097/IAE.0000000000002663] [Medline: 31800460]
13. Chiang MF, Quinn GE, Fielder AR, Ostmo SR, Paul Chan RV, Berrocal A, et al. International Classification of Retinopathy of Prematurity, Third Edition. Ophthalmology. Oct 2021;128(10):e51-e68. [FREE Full text] [doi: 10.1016/j.ophtha.2021.05.031] [Medline: 34247850]
14. Iwahashi C, Utamura S, Kuniyoshi K, Sugioka K, Konishi Y, Wada N, et al. Factors associated with reactivation after intravitreal bevacizumab or ranibizumab therapy in infants with Retinopathy of Prematurity. Retina. Nov 01, 2021;41(11):2261-2268. [doi: 10.1097/IAE.0000000000003196] [Medline: 33958533]
15. Decuyper M, Maebe J, Van Holen R, Vandenberghe S. Artificial intelligence with deep learning in nuclear medicine and radiology. EJNMMI Phys. Dec 11, 2021;8(1):81. [FREE Full text] [doi: 10.1186/s40658-021-00426-y] [Medline: 34897550]
16. Wang X, Wang R, Yang S, Zhang J, Wang M, Zhong D, et al. Combining radiology and pathology for automatic glioma classification. Front Bioeng Biotechnol. 2022;10:841958. [FREE Full text] [doi: 10.3389/fbioe.2022.841958] [Medline: 35387307]
17. Du-Harpur X, Watt F, Luscombe N, Lynch M. What is AI? Applications of artificial intelligence to dermatology. Br J Dermatol. Sep 2020;183(3):423-430. [FREE Full text] [doi: 10.1111/bjd.18880] [Medline: 31960407]
18. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol. Jul 01, 2018;136(7):803-810. [FREE Full text] [doi: 10.1001/jamaophthalmol.2018.1934] [Medline: 29801159]
19. Wu Q, Hu Y, Mo Z, Wu R, Zhang X, Yang Y, et al. Development and validation of a deep learning model to predict the occurrence and severity of retinopathy of prematurity. JAMA Netw Open. Jun 01, 2022;5(6):e2217447. [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.17447] [Medline: 35708686]

XSL•FO
RenderX

20. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. May 28, 2015;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
21. Ham Y, Kim J, Luo J. Deep learning for multi-year ENSO forecasts. Nature. Sep 2019;573(7775):568-572. [doi: 10.1038/s41586-019-1559-7] [Medline: 31534218]
22. Li JO, Liu H, Ting DS, Jeon S, Chan RP, Kim JE, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. Prog Retin Eye Res. May 2021;82:100900. [FREE Full text] [doi: 10.1016/j.preteyeres.2020.100900] [Medline: 32898686]
23. Moshfeghi DM. Systemic solutions in retinopathy of prematurity. Am J Ophthalmol. Sep 2018;193:xiv-xviii. [doi: 10.1016/j.ajo.2018.05.013] [Medline: 29792838]
24. Garcia Gonzalez JM, Snyder L, Blair M, Rohr A, Shapiro M, Greenwald M. Prophylactic peripheral laser and fluorescein angiography after bevacizumab for Retinopathy of Prematurity. Retina. Apr 2018;38(4):764-772. [doi: 10.1097/IAE.0000000000001581] [Medline: 28267112]
25. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data. 2020;7(1):94. [FREE Full text] [doi: 10.1186/s40537-020-00369-8] [Medline: 33169094]
26. Lee YW, Choi JW, Shin E. Machine learning model for predicting malaria using clinical information. Comput Biol Med. Mar 2021;129:104151. [doi: 10.1016/j.compbiomed.2020.104151] [Medline: 33290932]
27. Yang B, Wang X, Mo J, Li Z, Gao D, Bai Y, et al. The amplitude of low-frequency fluctuation predicts levodopa treatment response in patients with Parkinson's disease. Parkinsonism Relat Disord. Nov 2021;92:26-32. [doi: 10.1016/j.parkreldis.2021.10.003] [Medline: 34666272]
28. Li L, Zhang Z, Xiong Y, Hu Z, Liu S, Tu B, et al. Prediction of hospital mortality in mechanically ventilated patients with congestive heart failure using machine learning approaches. Int J Cardiol. Jul 01, 2022;358:59-64. [doi: 10.1016/j.ijcard.2022.04.063] [Medline: 35483478]
29. Shen L, Jin Y, Pan A, Wang K, Ye R, Lin Y, et al. Machine learning-based predictive models for perioperative major adverse cardiovascular events in patients with stable coronary artery disease undergoing noncardiac surgery. Comput Methods Programs Biomed. Mar 2025;260:108561. [doi: 10.1016/j.cmpb.2024.108561] [Medline: 39708562]
30. Tan Z, Simkin S, Lai C, Dai S. Deep learning algorithm for automated diagnosis of retinopathy of prematurity plus disease. Transl Vis Sci Technol. Nov 2019;8(6):23. [FREE Full text] [doi: 10.1167/tvst.8.6.23] [Medline: 31819832]
31. Wang J, Ji J, Zhang M, Lin J, Zhang G, Gong W, et al. Automated explainable multidimensional deep learning platform of retinal images for retinopathy of prematurity screening. JAMA Netw Open. May 03, 2021;4(5):e218758. [doi: 10.1001/jamanetworkopen.2021.8758] [Medline: 33950206]
32. Liu Y, Xie H, Zhao X, Tang J, Yu Z, Wu Z, et al. Automated detection of nine infantile fundus diseases and conditions in retinal images using a deep learning system. EPMA J. Mar 2024;15(1):39-51. [doi: 10.1007/s13167-024-00350-y] [Medline: 38463622]
33. Coyner AS, Chen JS, Singh P, Schelonka RL, Jordan BK, McEvoy CT, et al. Single-examination risk prediction of severe retinopathy of prematurity. Pediatrics. Dec 01, 2021;148(6). [FREE Full text] [doi: 10.1542/peds.2021-051772] [Medline: 34814160]
34. Yang J, Ju J, Guo L, Ji B, Shi S, Yang Z, et al. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. Comput Struct Biotechnol J. 2022;20:333-342. [FREE Full text] [doi: 10.1016/j.csbj.2021.12.028] [Medline: 35035786]

Abbreviations

AdaBoost: Adaptive Boosting
AUC: area under the curve
CatBoost: Categorical Boosting
EMA: Exponential Moving Average
Grad-CAM: Gradient-Weighted Class Activation Mapping
IVI: intravitreal injection
LR: logistic regression
PMA: postmenstrual age
RF: random forest
ROC: receiver operating characteristic
ROP: retinopathy of prematurity
VEGF: vascular endothelial growth factor
XGBoost: Extreme Gradient Boosting

XSL•FO
RenderX

Edited by A Coristine; submitted 11.05.24; peer-reviewed by A Jafarizadeh, G Lim; comments to author 14.09.24; revised version received 20.01.25; accepted 31.01.25; published 23.04.25

Please cite as:
Wu R, Zhang Y, Huang P, Xie Y, Wang J, Wang S, Lin Q, Bai Y, Feng S, Cai N, Lu X
Prediction of Reactivation After Antivascular Endothelial Growth Factor Monotherapy for Retinopathy of Prematurity: Multimodal Machine Learning Model Study
J Med Internet Res 2025;27:e60367
URL: https://www.jmir.org/2025/1/e60367
doi: 10.2196/60367
PMID:

©Rong Wu, Yu Zhang, Peijie Huang, Yiying Xie, Jianxun Wang, Shuangyong Wang, Qiuxia Lin, Yichen Bai, Songfu Feng, Nian Cai, Xiaohe Lu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX