Supplemental Table 1: Performance of Machine Learning Models for 1-Step Strategy on both the Validation and Independent Test Sets after Feature Selection					
Model (in ascending order of performance on recall)	**Accuracy on the Validation Set**	**Accuracy on the Test Set**	**Precision**	**Recall / Sensitivity**	**AUC**
LightGBM	94.1%	91.4%	0.55	0.67	0.61
Bagging Classifier	95.7%	94.7%	0.74	0.72	0.73
Random Forest	96.1%	94.1%	0.69	0.72	0.71
XGBoost	94.7%	93.1%	0.64	0.72	0.67
Decision Tree	94.5%	94.4%	0.70	0.75	0.72
SVM (RBF kernel)	85.1%	80.5%	0.31	0.78	0.44
Logistic Regression	82.2%	79.6%	0.30	0.81	0.44
Linear SVC	81.8%	78.4%	0.29	0.82	0.43

Abbreviations: AUC, area under the receiver operating characteristics curve; SVC, support vector classifier; SVM, support vector machine

Supplemental Table 2: Descriptive comparison of selected non-image ROP prediction and screening approaches

Model/study	Setting and cohort	Prediction time	Inputs	Target outcome	Validation design	Reported metrics	Relevance to the present study
WINROP[1-5]	Sweden; single-center development cohort, N=353	Weekly from birth; alarm before PMA 32 weeks	GA and serial postnatal weight gain	Treatment-requiring or sight-threatening ROP	Retrospective development; later prospective validations in multiple countries	Development sensitivity 100%; later sensitivity 85%-100% and specificity 25%-89%	Early growth-based alarm; performance varied across settings and oxygen practices
G-ROP[6]	United States/Canada; 41 hospitals, prospective validation N=3,981	Birth through postnatal days 10-39	GA, birthweight, weight gain in three age windows, hydrocephalus	Type 1 ROP	Prospective multicenter validation in a new at-risk cohort	Sensitivity 100% (219/219); 35.6% fewer infants would undergo examinations	Validated screening criteria rather than a directly comparable probabilistic ML model
DIGIROP-Screen v1.0[7,8]	Swedish registry development N=6,991; three international validation cohorts N=1,241	Birth and postnatal weeks 6-14	Birth characteristics plus age and status at first diagnosed ROP	ROP treatment	Registry development plus external validation; later Greek validation N=640	External cohorts: 100% sensitivity except one infant with severe malformations; Greek birth model sensitivity 81.4%, AUC 0.82	Uses serial ROP examination data; Greek results show the need for local validation and updating
DIGIROP 2.0[9,10]	Swedish registry N=11,139; contemporary Swedish external cohort N=1,530	Prescreen and serial screen models	Birth/ROP data plus parenteral nutrition duration >=14 days	ROP treatment	Internal/external validation; later national external validation with model updating	2023 models: sensitivity 100%; specificity 46.6% prescreen and 76.9% screen. 2026 prescreen AUC 0.89; sensitivity 100%	Broadens eligibility and adds a treatment-exposure variable; still population-specific
NLP-ROPCare[11]	China; single hospital, retrospective cohort N=3,922	At admission	Unstructured free-form admission notes; pretrained language models	ROP occurrence and severity	Retrospective internal evaluation; external validation not yet reported	Occurrence: AUC 0.90, F1 89.35%; severity: AUC 0.91, F1 78.44%	Uses unstructured text rather than the structured two-step inputs used here
Takeda et al.[12]	Japan; four neonatal units in one region, N=215	At first ROP screening	Thirty-five non-imaging clinical variables	Any ROP during hospitalization	Repeated 10-fold cross-validation; no independent external cohort	Random Forest: AUC 0.93, sensitivity 95.7%, specificity 66.0%; Naive Bayes: AUC 0.94, sensitivity 94.6%, specificity 73.6%	Non-imaging clinical model, but with a smaller cohort and a different endpoint
Present study	India; 25 newborn care units. STEP 1 N=5,926; STEP 2 N=1,361; prospective cohort N=515 from the same population	STEP 1 during NICU care; STEP 2 after structured ophthalmic assessment	STEP 1: demographic, perinatal, laboratory, and NICU variables. STEP 2: adds structured ophthalmic findings	Any ROP (STEP 1); treatment-requiring ROP among assessed infants (STEP 2)	Retrospective training/testing followed by prospective evaluation in the same population	AUC 0.86 (STEP 1) and 0.89 (STEP 2); sequential prospective identification 84.9% (437/515)	Workflow-aligned structured-data approach in an Indian public-health network; not an image-pixel model

Note: Reported metrics are reproduced from the cited studies and are not directly comparable because populations, eligibility criteria, prediction times, inputs, outcomes, thresholds, and validation designs differ.
Abbreviations: AUC, area under the receiver operating characteristic curve; F1, F1 score; GA, gestational age; ML, machine learning; NICU, neonatal intensive care unit; PMA, postmenstrual age; ROP, retinopathy of prematurity.

REFERENCES

1. Hellström, A. et al. Early weight gain predicts retinopathy in preterm infants: new, simple, efficient approach to screening. *Pediatrics* **123**, e638–e645 (2009).
2. Zepeda-Romero, L. C. et al. Prediction of retinopathy of prematurity using the screening algorithm WINROP in a Mexican population of preterm infants. *Arch. Ophthalmol*. **130**, 720–723 (2012).
3. Sun, H. et al. The use of the WINROP screening algorithm for the prediction of retinopathy of prematurity in a Chinese population. *Neonatology* **104**, 127–132 (2013).
4. Piyasena, C. et al. Prediction of severe retinopathy of prematurity using the WINROP algorithm in a birth cohort in south east *Scotland. Arch. Dis. Child Fetal Neonatal Ed*. **99**, F29–F33 (2014).
5. Ko, C. H. et al. Using WINROP as an adjuvant screening tool for retinopathy of prematurity in southern Taiwan. *Am. J. Perinatol*. **30**, 149–154 (2015).
6. Binenbaum, G. et al. Postnatal Growth and Retinopathy of Prematurity (G-ROP) Study Group. Validation of the postnatal growth and retinopathy of prematurity screening criteria. *JAMA Ophthalmol*. **138**, 31–37 (2020).
7. Pivodic, A. et al. Development and validation of a new clinical decision support tool to optimize screening for retinopathy of prematurity. *Br. J. Ophthalmol*. **106**, 1573–1580 (2022).
8. Moutzouri, S. et al. Predicting the risk of treatment-requiring retinopathy of prematurity in preterm infants in Greece: external validation of DIGIROP prognostic models. *Acta Ophthalmol*. **103**, 272–280 (2025).
9. Pivodic, A. et al. Prognostic value of parenteral nutrition duration on risk of retinopathy of prematurity: development and validation of the revised DIGIROP clinical decision support tool. *JAMA Ophthalmol*. **141**, 716–724 (2023).
10. Pivodic, A. et al. External validation and updating of DIGIROP prediction models including parenteral nutrition for treatment-requiring retinopathy of prematurity in a Swedish national cohort. *BMJ Open Ophthalmol*. **11**, e002727 (2026).
11. Zhang, Y. et al. NLP-ROPCare: predicting retinopathy of prematurity with admission notes using natural language processing. *BMJ Open Ophthalmol*. **11**, e002385 (2026).
12. Takeda, Y. et al. Prediction models for retinopathy of prematurity using nonimaging machine learning approaches: a regional multicenter study. *Ophthalmol. Sci*. **5**, 100715 (2025).