Detection and Classification of Retinopathy of Prematurity on Multi-source Fundus Images using DenseROPNet

***Abstract*—AI-based disease detection and stage classification is becoming a norm in today's technological world. Unlike the traditional diagnosis, modern diagnostics are showing better accuracy, faster results and accurate predictions using different CNN models. This study focuses on developing and training a CNN model with attention mechanisms which is capable of diagnosing Retinopathy of Prematurity (ROP), a major reason of blindness in infants. Accurate detection of this disease and its stages can make the treatment of it more effective. More importantly, manual detection of early stages of ROP can be quite challenging as the signs are subtle and ROP diagnosis relies on qualitative analysis which may create confusion while taking decisions. Keeping all these in mind, DenseROPNet has been trained and tested on two different fundus image datasets of subjects with two different ethnicities. The proposed model reached the highest accuracy of 99.13% overcoming the accomplishment of most of the existing studies. The motto of this study has been to propose and develop a CNN model which beside performing significantly well, will be explainable. With that in mind, analysis using Explainable AI has also been done to ensure the model's understanding of the dataset is not only accurate but also meaningful.**

***Index Terms*—Retinopathy of Prematurity (ROP), fundus images, retinal images, retinography, Convolutional Neural Network (CNN), Explainable AI (XAI)**

I. Introduction

Retinopathy of Prematurity (ROP) is disease affecting mostly preterm babies with low birth-weights. Being the leading cause of vision loss in case of children worldwide [1], it even caused epidemics in the 20th century [2]. Since it shows no early symptoms [25], it becomes tough to diagnose ROP at an early stage. The stages of ROP depend on the level of the abnormal growth of the blood vessels of the retina. More importantly, detecting the stage of ROP is very difficult as there is no clear demarcation and the level of detachment or disruption of retina is hard to comprehend seeing fundus/retinal images. In the recent years, deep learning models and algorithms have proven their potentiality in early detection of several diseases and classification of stages based on medical images [26]–[28].

A plentiful of studies and research have been conducted to detect and/or classify ROP and its stages using machine learning or deep-learning models. Mustafa Yurdakul et al. [6] proposed ROPGCViT which showed an accuracy of 94.69% accuracy on the dataset by X. Zhao et al. [3] outperforming 50 DL models. Yurdakul et al. managed to cross the accuracy of the models used by X. Zhao et al. on the dataset by X. Zhao et al. Mahboob et al. [7] tested 5 different models and achieved the highest accuracy of around 95.27% with the ConvNeXt-Base model on the dataset by X. Zhao et al. [3] only. N. Mowla et al. [10] developed LightEyeNet model which showed an accuracy of 96.28% on the aforementioned dataset.

A. Krishnan et al. [8] tested machine-learning models and algorithms on the dataset by J. Timkovič et al. [4]. The random forest model showed better performance with an accuracy of 78.62% on the dataset. Kariakin et al. [11] achieved Deep Neural Network (DNN) models where an accuracy of 97% was achieved on the dataset by J. Timkovič et al. [4].

Peng et al. [12] worked on a private dataset. Though Yenice et al. [16] worked on a private dataset, the work was focused on binary classification rather than multi-stage classification. In addition, Tong et al. [13] achieved 90.3% accuracy on a private dataset. Yang et al. [17] similarly worked on a single private dataset. On a different note, Huang et al. [18], Wang et al. [19] and Li et al. [20] focused on early detection and diagnosis of ROP. Huang et al. dealt with healthy cases, Stage 1 and Stage 2 ROP. Wang et al. and Li et al. in their studies worked with Stage 3 ROP additionally. Agrawal et al. [15] used HVDROPDB dataset achieving 90% accuracy with their proposed model.

Very few studies considered to work on more than one different datasets. Chen et al. [9] used two datasets, one dataset was procured using images from the North American institutions while the other was procured using images from the Nepali institutions. They used ResNet-152 architecture for their objective. They found that if training and testing are done on the same population, the sensitivity and specificity remains good compared to the case when testing and training are done on different datasets. So, to get the best result, they trained on both populations and then tested which not only generalized their model across two different populations but also ensured better performance. However, the datasets are private which limits the chances of researchers potentially create a more generalized and robust model.

Finally, a conclusion can be drawn that most of the existing works are trained on one dataset or datasets with same configuration and settings of the camera and the environment or same population. This equips the models with limited generalizability and so these models perform best on a specific type or configuration of datasets or a specific ethnicity.

II. Proposed Model Architectures

With an objective to detect and classify ROP across different populations, DenseROPNet was developed which is a customized and modified of DenseNet121.

DenseROPNet is a judicious combination of Convolutional Network & Channel-wise Attention Network. Convolutional networks are good at extracting the hidden image features taking both spatial and channel wise information in to consideration. But with increasing number of channels per convolutional layer it becomes hard to extract & learn the important features from the overwhelming number of filter channels. To guide the model at a filter channel level we introduce Channel wise Attention through **SE Blocks (Squeeze & Excitation Blocks)**, a proven way of incorporating attention to channel level [14].

Attention can be incorporated in a way of biasing the model towards the more informative components of different blocks. Using Attention in tasks like sequence-to-sequence learning, retaining global context in visual and non-visual tasks are nothing new. But generally using attention mechanism for retaining context is computationally expensive and takes up much time during training. So rather than paying attention on the whole image, we used attention on the channels of the image. Because channels carry vital information about the whole image. Finding the channels with more information is a crucial task. With this goal in mind, we incorporated Squeeze & Excitation Block which assigns weights against every channels of a certain layer and find the more informative channels. Squeeze & Excitation Blocks enhances the learning of convolutional layers explicitly handling the interdependencies of the channels. We implement **SE** Blocks at different convolutional layers. In earlier layers, it excites informative features in a class-agnostic manner, strengthening the shared low-level representations. In later layers, the SE blocks become increasingly specialized, and respond to different inputs in a highly class-specific manner. As a consequence, the benefits of the feature recalibration performed by SE blocks can be accumulated through the network.

The Squeeze & Excitation Operation consists of three main blocks namely Squeeze, Excitation & Scale.

- Squeeze Operation Squeeze block accumulates spatial information through Global Average Pooling.

$$z_c = F_{sq}(\mathbf{x}_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{1}$$

where:

$$\mathbf{z_c} \text{ is squeezed channel c} \tag{2}$$

$$\mathbf{x_c} \text{ is the input channels} \tag{3}$$

- The Excitation Operation learns channel-wise dependencies through MLP.

$$\mathbf{s} = F_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \mathbf{z})) \tag{4}$$

- The Last Operation scales the feature map by learned weights.

$$\tilde{x}_c = F_{scale}(x_c, s_c) = s_c \cdot x_c \tag{5}$$

The Residual Attention Network Block (RANB) does channel wise and spatial attention to keep keep original information while injecting channel emphasis and thus boosts salient features by helping the model to understand where to look and what to look [23]. DenseROPNet also uses Global Max Pooling and Global Average Pooling and finally concatenates to get the overall and strongest evidence of a feature which is advantageous [21], [22]. The model architecture is given in the Fig-1. The SE (Squeeze & Excitation) block and RANB (Residual Attention Network Block) are shown in Fig-2 in detail.

III. Datasets

A. IROPS and RIDRIP datasets

For training, validating and testing our proposed architectures, we used two datasets. The first one is the IROPS fundus image dataset by X. Zhao et al. [3] and the other dataset is the retinal image RIDIRP dataset by J. Timkovič et al. [4].

IROPS dataset contains 1,099 fundus images in both from 483 premature infants.The dataset contained images of five different classes titled Normal, Stage 1 ROP, Stage 2 ROP, Stage 3 ROP and Laser scars. The RIDRIP dataset contains 6,004 retinal images of 188 newborns. Most of the newborns were premature. This dataset contains classes like Normal, Stage 0, Stage 1, Stage 2, Stage 3, Stage 4, Stage 5 and Aggressive Posterior Retinopathy of Prematurity (AP-ROP). Since the objective was to build a model to detect and/or classify the ROP stage in premature babies, these two datasets proved to be the best fit as the first dataset contains images of premature newborns only [3] and the second dataset mostly contains images of premature newborns [4]. These two datasets contain high-quality images with proper classification and relevant details. More importantly, due to numerous anatomical differences between the two different ethnic groups of these two datasets, the datasets have been chosen to increase variation in data.

B. Pre-processing

All the images were converted to 299X299 pixels. After that, all the images were pre-processed using CLAHE. CLAHE which stands for Contrast Limited Adaptive Histogram Equalization can help to adaptively adjust image contrast locally without amplifying or enhancing noise [29]. This helps to locate and observe the abnormal ridges and vessels with more ease which can help in medical image based classification or computer vision related tasks [30]. The effect of CLAHE can be understood from the Fig-3.

C. Merging and Augmentation

Only five different classes were selected to be used for training and testing the model. These are Normal, Stage 0 ROP, Stage 1 ROP, Stage 2 ROP and Stage 3 ROP. The

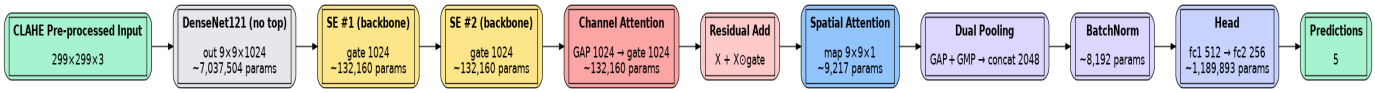
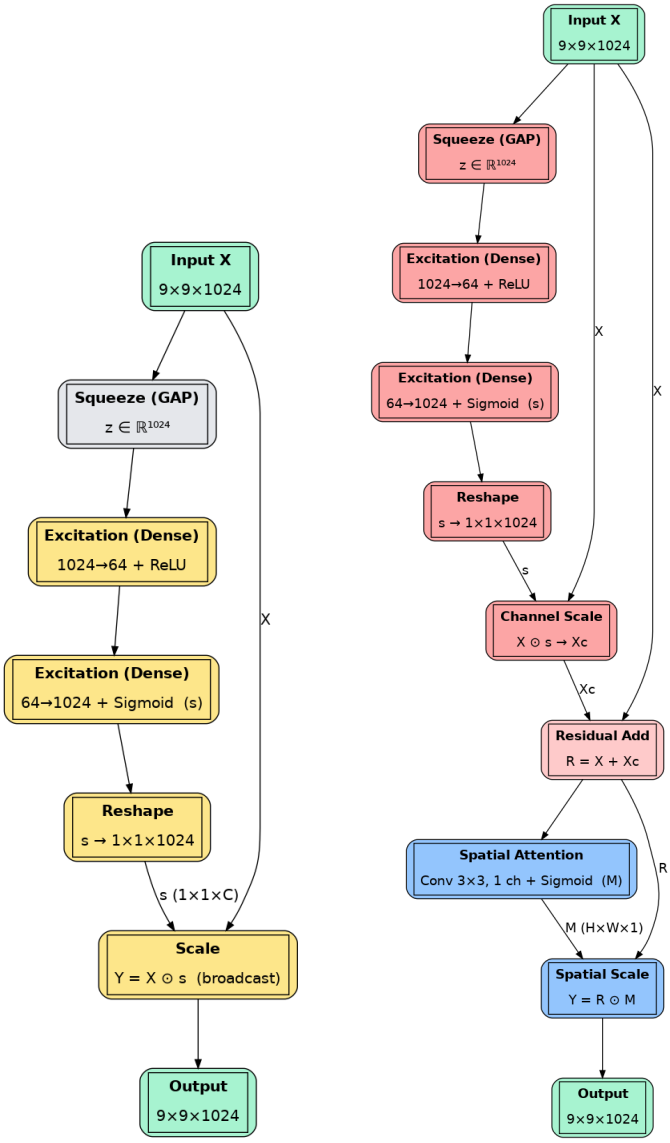


Fig. 1: Model Architecture



(a) SE block (b) RANB

Fig. 2: Attention Blocks

other classes were omitted as the other stages are later and more severe stages of ROP and so there is little benefit in identifying or detecting them using our model as it is too late for recovery in the most of the cases of these stages [5]. The combined dataset had 3216 images under 'Normal' class, 1340 images under 'Stage 0 ROP' class, 251 images under 'Stage 1 ROP' class, 458 images under 'Stage 2 ROP' class and 125 images under 'Stage 3 ROP' class. To counter class-imbalance, data augmentation was done. So, geometric transformations

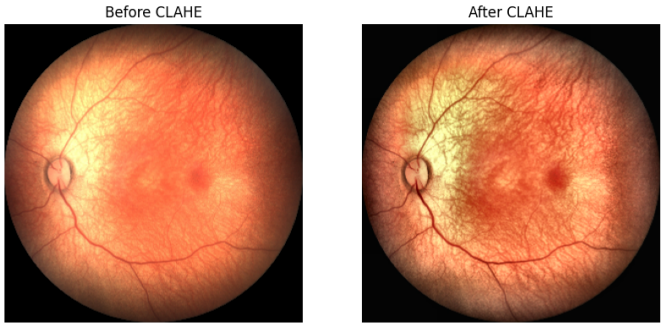


Fig. 3: Effect of CLAHE Pre-processing

and color-space adjustments were done so that each class has 3216 images. After that the dataset were split into training, validation and testing dataset. Training dataset contained 2573 images for each class except 'Stage 2 ROP' which had 2572 images. Validation dataset contained 322 images for each class except 'Normal' and 'Stage 1 ROP' which had 321 images each. Testing dataset had 322 images as well for each class but 'Stage 0 ROP' and 'Stage 3 ROP' had 321 images each. This split was done randomly.

IV. Training & Testing the Proposed Model

A. Training

For training, all the images were normalized to [0,1] with a batch size of 32. Adam Optimizer and Sparse Categorical Cross-Entropy loss function were used. The training metrics can be found in Table-I below.

Model	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
DenseROPNet	0.0716	99.84%	0.1169	98.69%

TABLE I: Training and validation outcomes of DenseROPNet

B. Testing

1) Classification Report: The classification report of DenseROPNet can be seen in Table-II. The average testing loss was found to be 0.0985.

TABLE II: Classification Report of DenseROPNet

Class	Precision	Recall	F1-Score	Accuracy
Normal	0.99	0.99	0.99	99.0%
Stage 0 ROP	0.99	0.99	0.99	99.0%
Stage 1 ROP	0.99	1.00	1.00	100.0%
Stage 2 ROP	0.99	0.98	0.99	98.0%
Stage 3 ROP	0.99	1.00	1.00	100.0%
Average	0.99	0.99	0.99	99.13%

2) Confusion Matrix, ROC Curve & AUC Table: The confusion matrix present in Fig. 4, the ROC curve in Fig. 5 and the table for AUC values found in Table-III suggest that the proposed model can almost perfectly distinguish the different classes.

TABLE III: AUC values across different classes

Class	DenseROPNet
Normal	0.9998
Stage 0 ROP	0.9998
Stage 1 ROP	1.0000
Stage 2 ROP	0.9998
Stage 3 ROP	1.0000
Average	0.9999

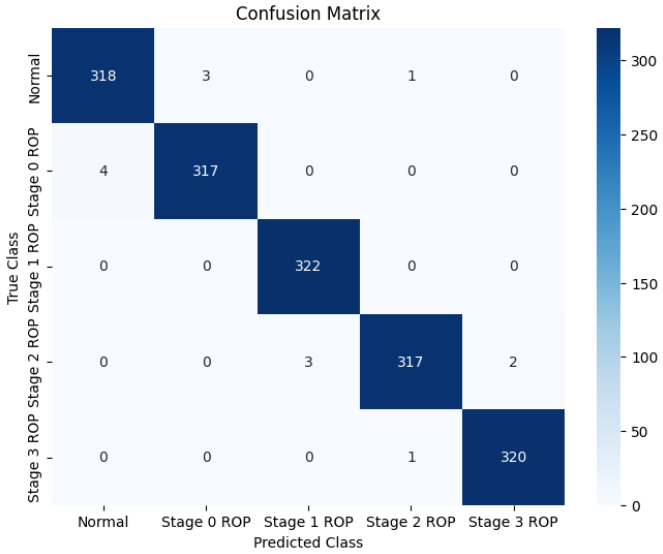


Fig. 4: Confusion Matrix of DenseROPNet

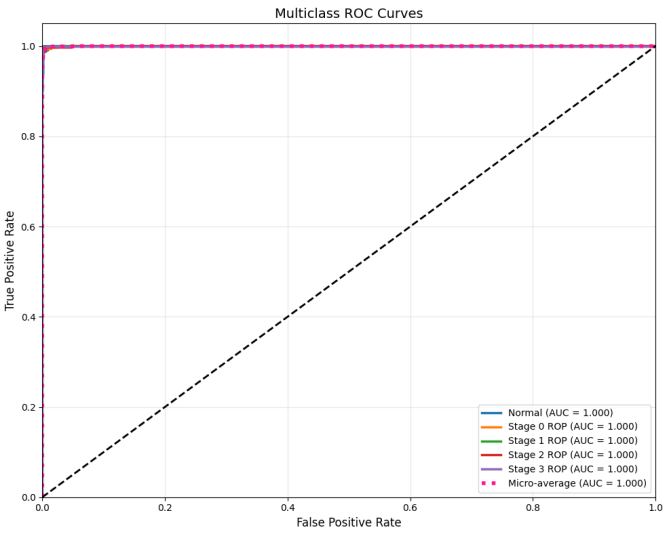


Fig. 5: ROC Curve of DenseROPNet

3) Explainability of the Proposed Model: Explainable AI helps us to understand how deep learning models work. For example, Explainable AI (XAI) helps to understand the Region of Interest (ROI) for image-based classification. That is, it helps to understand which part of the image is focused on more to perform the classification by the deep learning models. Grad-CAM heatmaps helps us understand this perfectly as its potentiality is huge [24]. The redder the zones are, the more important the region is for classification. Fig. 6 shows a sample of Grad-CAM Heatmaps of the proposed model based on testing accuracy. Since the images contain the healthy cases, we see that the redness is not prominent as there is no abnormal vessel. So, it detects the image correctly as that of a healthy subject. For images of healthy eyes, the heatmap shows redness suggesting higher importance or concentration around the macula and optic disc. These are main reference points where healthy nerves and vessels are found [6].

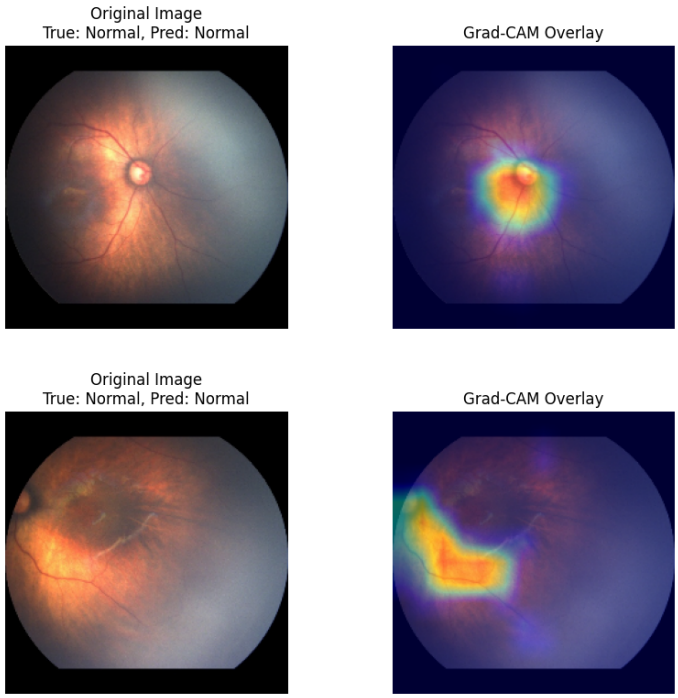


Fig. 6: Grad-CAM Heatmaps

V. Comparison with Existing Works

Most of the existing works were trained on one dataset. Almost all of the works where more than one dataset were used are private and procured using similar configuration of devices and scenarios or have data from same population. In this regard, our study involved using two different public datasets containing fundus images from two different groups having notable anatomical differences between them. In terms of accuracy, the most notable work done on IROPS dataset is the LightEyeNet model by N. Mowla et al. [10] achieved 96.28% accuracy which is trained on apparently only one dataset. More importantly, the model designed by Kariakin et al. [11] managed to gain higher accuracy but it was also trained

on only one dataset, the RIDRIP dataset. That is why, these models lack the generalizability and adaptability. A model is as good as its ability to handle variation, ensure generalization and adapt to changing scenario in data. Table-IV details the comparison with existing works. It is safe to say, the proposed model DenseROPNet performed best when it comes to studies based on dataset by Zhao et al or dataset by Timkovič et al.

TABLE IV: Comparison of Metrics of Different Studies

Model/Author	Dataset Used	Accuracy
ROPGCViT	IROPS Dataset	94.69%
ConvNeXt	IROPS Dataset	95.27%
LightEyeNet	IROPS Dataset	96.28%
Krishnan et al.	RIDIRP Dataset	78.62%
ViT	RIDIRP Dataset	97%
Agrawal et al	HVDROPDB dataset	90%
Tong et al.	private dataset	90.3%
Peng et al.	private dataset	98.27%
Yang et al.	private dataset	69%
Huang et al.	private dataset	92.23%
Wang et al.	private dataset	98.29%
Li et al.	private dataset	97.98%
Chen et al.	two private datasets	∼96.4%
DenseROPNet	*IROPS Dataset & RIDIRP Dataset*	*99.13%*

More importantly, when the LightEyeNet Model [10] which was originally trained on IROPS dataset, when tested on the dataset by RIDIRP dataset, it achieved a test accuracy of 34.14%. Similarly the ViT Model [11] which was originally trained on RIDRIP dataset, when tested on IROPS dataset achieved a test accuracy of 34.78%.

To ensure a fair comparison, the best performing models, LightEyeNet and ViT on IROPS and RIDRIP datasets respectively were trained and tested on the merged dataset following the methodologies of the corresponding authors. The Table-V shows a comparison of the test metrics.

TABLE V: Comparison after Training Best Existing Models on the Merged Dataset

Model	Testing Accuracy	F1-score
LightEyeNet	98.45%	0.98
ViT	96.02%	0.96
DenseROPNet	*99.13%*	*0.99*

VI. Ablation studies

Ablation studies are necessary to justify the modifications done in a pre-exisiting CNN model and the relevant data pre-processing techniques in terms of improving performance metrics. A table containing the ablation studies of the proposed model is present in Table-VI.

TABLE VI: Ablation studies of the model

Case	Testing Accuracy	F1-score
DenseNet121 Base Model	82.59%	0.82
With RANB Blocks	98.51%	0.99
with RANB Blocks and SE blocks	99.13%	0.99

Without the CLAHE pre-processing, DenseROPNet seems to be performing fairly well with 98.82% testing accuracy and 0.99 F1-score. But with CLAHE pre-processing, the performance becomes slightly better where the model achieves an accuracy of 99.13% and F1-score of 0.99.

VII. Conclusion & Future Works

Detection of ROP is important as blindness can reduce an infant's chances of survival [1]. Moreover, early detection of ROP can potentially help the medical professionals track its development and thus reduce the risk of blindness [25]. This study aimed at providing a solution to detect and identify ROP and its stages. Two datasets of two different sources containing subjects of two different ethnicities were used in order to ensure robustness and generalizability of the solution. This is because when training was done on IROPS dataset and tested on RIDRIP dataset, validation accuracy was 95.50% & testing accuracy was 41.78%. Similarly, when the training was done on RIDRIP dataset and tested on IROPS dataset, validation accuracy was 99.75% & testing accuracy was 38.30%.

In order to increase the proposed model's effectiveness, more datasets can be used from variety of sources with more than two different populations. Datasets which contain subjects with different ethnicity and datasets which contain images captured by different fundus cameras can be used to make the solution globally acceptable. Furthermore, steps can be taken to make the model lightweight and thus reduce computational complexity. Another notable approach is to enable federated learning. This way, the model will proceed further towards generalization as many hospitals around the world can use their variety of data across different ethnicities and facial features to fine-tune the model without the breach of data privacy. More importantly, it is a must to develop and use models which are explainable and learn the necessary, useful and general features.

References

[1] Solebo, A. L., Teoh, L., & Rahi, J. (2017). Epidemiology of blindness in children. *Archives of Disease in Childhood*, 102, 853–857. https://doi.org/10.1136/archdischild-2016-310532.

[2] Hong, E. H., Shin, Y. U., & Cho, H. (2022). Retinopathy of prematurity: A review of epidemiology and current treatment strategies. *Clinical and Experimental Pediatrics*, 65(3), 115–126. https://doi.org/10.3345/cep.2021.00773. Epub 2021 Oct 12. PMID: 34645255; PMCID: PMC8898617.

[3] X. Zhao, S. Chen, S. Zhang, *et al.*, "A fundus image dataset for intelligent retinopathy of prematurity system," *Sci Data*, vol. 11, no. 543, 2024, doi: https://doi.org/10.1038/s41597-024-03362-5.

[4] J. Timkovič, J. Nowaková, J. Kubíček, *et al.*, "Retinal Image Dataset of Infants and Retinopathy of Prematurity," *Sci Data*, vol. 11, no. 814, 2024, doi: https://doi.org/10.1038/s41597-024-03409-7.

[5] E. H. Hong, Y. U. Shin, and H. Cho, "Retinopathy of prematurity: A review of epidemiology and current treatment strategies," *Clin. Exp. Pediatr.*, vol. 65, no. 2, pp. 115–126, 2022, doi: https://doi.org/10.3345/cep.2021.00773.

[6] M. Yurdakul, K. Uyar, Ş. Taşdemír and Í. Atabaş, "ROPGCViT: A Novel Explainable Vision Transformer for Retinopathy of Prematurity Diagnosis," *IEEE Access*, vol. 13, pp. 77064-77079, 2025, doi: https://doi.org/10.1109/ACCESS.2025.3564213.

[7] Z. Mahboob, A. O. Ahsan,M. Wilms, "Using an adult retinal image analysis foundation model for retinopathy of prematurity staging: are there benefits?," *Proc. SPIE 13407, Medical Imaging 2023: Image Processing*, 1340703, 2023, doi: https://doi.org/10.1117/12.3047303.

[8] Anantha Krishnan, Md.Salman Sarkar, Laxman Badavath,(2025) "Machine Learning Applications in Retinopathy of Prematurity Diagnosis Using the ROP Retinal Image Dataset," *Journal of Neonatal Surgery*,14(1s), 820-827.

[9] J. S. Chen, A. S. Coyner, S. Ostmo, K. Sonmez, S. Bajimaya, E. Pradhan, N. Valikodath, E. D. Cole, T. Al-Khaled, R. V. P. Chan, P. Singh, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "Deep Learning for the Diagnosis of Stage in Retinopathy of Prematurity: Accuracy and Generalizability across Populations and Cameras," *Ophthalmol. Retina*, vol. 5, no. 10, pp. 1027–1035, Oct. 2021, doi: https://doi.org/10.1016/j.oret.2020.12.013.

[10] N. Mowla, M. N. Mowla, K. Rabie and B. Alsinglawi, "A Lightweight Deep Learning Model for Retinopathy of Prematurity Classification in eHealth Applications," *2025 International Wireless Communications and Mobile Computing (IWCMC)*, Abu Dhabi, United Arab Emirates, 2025, pp. 227-232, doi: https://doi.org/10.1109/IWCMC65282.2025.11059591.

[11] M. Kariakin et al., "Automated Classification of Prematurity Retinopathy Using ViT, ResNet, and EfficientNet," *2025 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, Yekaterinburg, Russian Federation, 2025, pp. 65-68, doi: https://doi.org/10.1109/USBEREIT65494.2025.11054213.

[12] Peng Y, Zhu W, Chen Z, Wang M, Geng L, Yu K, Zhou Y, Wang T, Xiang D, Chen F, Chen X. "Automatic Staging for Retinopathy of Prematurity With Deep Feature Fusion and Ordinal Classification Strategy", *IEEE Trans Med Imaging*, 2021 Jul;40(7):1750-1762. doi: https://doi.org/10.1109/TMI.2021.3065753. Epub 2021 Jun 30. PMID: 33710954.

[13] Tong, Y., Lu, W., Deng, Qq. et al. "Automated identification of retinopathy of prematurity by image-based deep learning", *Eye and Vis* 7, 40 (2020). https://doi.org/10.1186/s40662-020-00206-2.

[14] Hanming Wang, Yunlong Li, Zijun Wu, Huifen Wang, Yuan Zhang, SASE: A Searching Architecture for Squeeze and Excitation Operations, arXiv preprint arXiv:2411.08333, 2024, https://arxiv.org/abs/2411.08333.

[15] R. Agrawal, S. Kulkarni, M. Deshpande, A. Gaikwad, R. Walambe, and K. V. Kotecha, "Classification of Stages 1, 2, 3 and Preplus, Plus disease of ROP using MultiCNN_LSTM classifier," *MethodsX*, vol. 14, p. 103182, Jan. 2025. doi: https://doi.org/10.1016/j.mex.2025.103182. PMID: 39944106; PMCID: PMC11815698.

[16] K. Yenice, E., C. Kara, and Ç. B. Erdaş, "Automated detection of type 1 ROP, type 2 ROP and A-ROP based on deep learning," *Eye*, vol. 38, pp. 2644–2648, 2024. doi: https://doi.org/10.1038/s41433-024-03184-0.

[17] W. Yang, H. Zhou, Y. Zhang, L. Sun, L. Huang, S. Li, X. Luo, Y. Jin, W. Sun, W. Yan, J. Li, J. Deng, Z. Xie, Y. He, and X. Ding, "An interpretable system for screening the severity level of retinopathy in premature infants using deep learning," *Bioengineering*, vol. 11, no. 8, p. 792, 2024. doi: https://doi.org/10.3390/bioengineering11080792.

[18] Y. P. Huang, H. Basanta, E. Y. Kang, K. J. Chen, Y. S. Hwang, C. C. Lai, J. P. Campbell, M. F. Chiang, R. V. P. Chan, S. Kusaka, Y. Fukushima, and W. C. Wu, "Automated detection of early-stage ROP using a deep convolutional neural network," *Br. J. Ophthalmol.*, vol. 105, no. 8, pp. 1099–1103, Aug. 2021. doi: https://doi.org/10.1136/bjophthalmol-2020-316526. Epub 2020 Aug 23. PMID: 32830123; PMCID: PMC7900257.

[19] D. Wang, W. Qiao, W. Guo, and Y. Cai, "Applying novel self-supervised learning for early detection of retinopathy of prematurity," *Electronics Letters*, vol. 60, no. 10, p. e13267, 2024. doi: https://doi.org/10.1049/ell2.13267.

[20] P. Li and J. Liu, "Early diagnosis and quantitative analysis of stages in retinopathy of prematurity based on deep convolutional neural networks," *Transl. Vis. Sci. Technol.*, vol. 11, no. 5, p. 17, May 2022. doi: https://doi.org/10.1167/tvst.11.5.17. PMID: 35579887; PMCID: PMC9123509.

[21] X. Zhang, Q. Nie, Z. Xiao, J. Zhao, X. Wu, P. Guo, R. Li, J. Liu, Y. Wei, and Y. Pan, "Dual-View Pyramid Pooling in Deep Neural Networks for Improved Medical Image Classification and Confidence Calibration," *arXiv*, vol. 2408.02906, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2408.02906.

[22] Z. Li, S. Xuan, X. He, and L. Wang, "Global weighted average pooling network with multilevel feature fusion for weakly supervised brain tumor segmentation," *IET Image Process.*, vol. 16, no. 12, pp. 3300–3313, Dec. 2022. doi: [10.1049/ipr2.12642](https://doi.org/10.1049/ipr2.12642).

[23] W. Wu, S. Liu, Y. Xia, and Y. Zhang, "Dual residual attention network for image denoising", *Pattern Recognition*, vol. 149, p. 110291, 2024.

[24] Suara, S., Jha, A., Sinha, P., and Sekh, A. A. (2024). "Is Grad-CAM Explainable in Medical Images?", *Computer Vision and Image Processing* (pp. 124–135). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-58181-6_11.

[25] National Eye Institute. (n.d.). *Retinopathy of prematurity*. U.S. Department of Health and Human Services. Available: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/retinopathy-prematurity.

[26] Manchikatla, S., Srisevitha, M., Anitha, T., Kumari, J. P., & Uday, K. (2025), "Deep learning for early diagnosis: Enhancing medical imaging for disease detection,"*Cuestiones de Fisioterapia*, 54(3). Available: https://doi.org/10.48047/vs1v9834.

[27] Laçi, H., Sevrani, K., & Iqbal, S., "Deep learning approaches for classification tasks in medical X-ray, MRI, and ultrasound images: A scoping review", *BMC Medical Imaging*, 25(1), 156, 2025. https://doi.org/10.1186/s12880-025-01701-5. PMID: 40335965; PMCID: PMC12057223.

[28] Aftab, J., Khan, M. A., Arshad, S., & et al. (2025). "Artificial intelligence-based classification and prediction of medical imaging using a novel framework of inverted and self-attention deep neural network architecture", *Scientific Reports*, 15, 8724. https://doi.org/10.1038/s41598-025-93718-7.

[29] Stephen M. Pizer, E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B. Zimmerman, Karel Zuiderveld, "Adaptive histogram equalization and its variations", *Computer Vision, Graphics, and Image Processing*, Volume 39, Issue 3, 1987, Pages 355-368, ISSN 0734-189X, https://doi.org/10.1016/S0734-189X(87)80186-X.

[30] Nguyen, T. P. H., Cai, Z., Nguyen, K., Keth, S., Shen, N., & Park, M. (2020). "Pre-processing image using Brightening, CLAHE and RETINEX". *arXiv preprint* arXiv:2003.10822. https://arxiv.org/abs/2003.10822