International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No. 1 (Aug-24)

http://dx.doi.org/10.12785/ijcds/160180

Retinopathy of Prematurity Disease Diagnosis using Deep Learning

Elizabeth Ndunge Mutua[1], Bernard Shibwabo Kasamani[2] and Christoph Reich[3]

[1,2]*School of Computing and Engineering Sciences,Strathmore University, Nairobi, Kenya*
[3]*Institute for Data Science, Cloud Computing and IT Security,Furtwangen University, Furtwangen, Germany*

Received 28 Jan. 2024, Revised 26 Apr. 2024, Accepted 22 May 2024, Published 18 Aug. 2024

Abstract: Retinopathy of Prematurity (ROP) is a disease affecting infants born preterm, at birth their retina is not well developed and in most times after birth the veins of the retina do not develop to full term. Sometimes these veins stop growing and then suddenly start growing to the wrong directions and this abnormally causes retina traction, causing blindness. Each country has its own screening guidelines for the disease diagnosis. The disease can be categorized as severe or mild and has five stages. Stage one and two is not severe and can develop and heal unnoticed. Stage three should be diagnosed because it is reversable through treatment but when the disease progresses to Stage four, retina traction occurs causing blindness at stage five. The emergent of digital imaging support has resulted to having hospitals capturing retina images to determine the presence or absence of severe ROP. These images can be used to determine the presence of retinal detachment or lack of growth of the veins. ROP disease diagnosis is expensive with few eye specialists available in hospitals and the process of capturing retina images by non-eye specialists and transmitting them to specialists for diagnosis pauses many issues. Different cameras produce images of different contrast, image transmission may cause quality reduction depending on the channel of transmission. These challenges call for the development of systems to support both image quality assessment and assistive disease diagnosis. This paper proposes a Deep learning model to assist ophthalmologists to determine the presence or absence of the disease as well as diagnosing the disease at stage three. A customized ResNet-50 model was first applied to preprocess the images and separate images of quality from non-quality images. Ninety-one (91) images were obtained from Kaggle database and 11,100 images from HVDROPDB database which were used for model training, testing and validation at a ratio of 0.80 training, 0.20 testing, 0.20 validation. After preprocessing, desired features were extracted and fed into a Deep Neural Network to quickly classify them as either having the disease or not. Those with ROP were further divided into two sub-classes: ROP Stage two or three. The model was able to achieve an accuracy of 92.8%, sensitivity of 94.9%, and precision of 97.3%.

Keywords: Retina image analysis, Retinopathy Classification, Eye Disease Diagnosis, Image Quality, Deep Learning

1. INTRODUCTION

Retinopathy of Prematurity screening in most countries is done for babies of gestational age 28 weeks and or 1500 grams [1]. At gestational age of 16 weeks, retina vessels begin to form and by week 39 they are fully developed [2]. Preterm babies have their retina vessels underdeveloped and, in most cases, they do not grow to full term [3]. As shown in Figure 1, The disease has five stages, stage one and two can develop and heal without any medical treatment but at stage three which is a reversable stage, diagnosis and treatment is required [3]. Stage four and five are severe stages, at stage four retina traction occurs and the baby goes blind at stage five [4], [5]. Stage three of the disease has a unique feature of the development of a well visible ridgeline which we used as a key feature for diagnosis [6], [7]. When the disease is at stage one, a thin white line begins to form because of the stopping of growth of the vessels [4], [5], [6], [7]. This line grows wider and the color changes to pink at stage two [7]. For stage three, the demarcation line is very wide and upon screening it can easily be noticed. In stage four, retinal detachment occurs with the baby going blind at stage five with a symptom of a white tinny spot in the eye [7].

A. *Retinopathy of prematurity in Retina Zones*

The retina has three zones as shown in Figure 2, with zone I being the innermost zone which is circular in shape and is surrounded by inner zone II. Zone II is divided into two, inner zone II and outer zone II. Inner zone II is closer to zone I and outer zone II is closer to zone III. Zone III is the outer zone [2]. Retinopathy of Prematurity disease in zone I is more aggressive and could rapidly progress from mild stages to severe stages [3].

B. *Pre and Plus ROP Diseases*

Pre-Plus ROP is ROP that is mild and not severe and is characterized by abnormal growth of blood veins with

E-mail address: elizabeth.mutua@strathmore.edu,bshibwabo@strathmore.edu,rch@hs-furtwangen.de

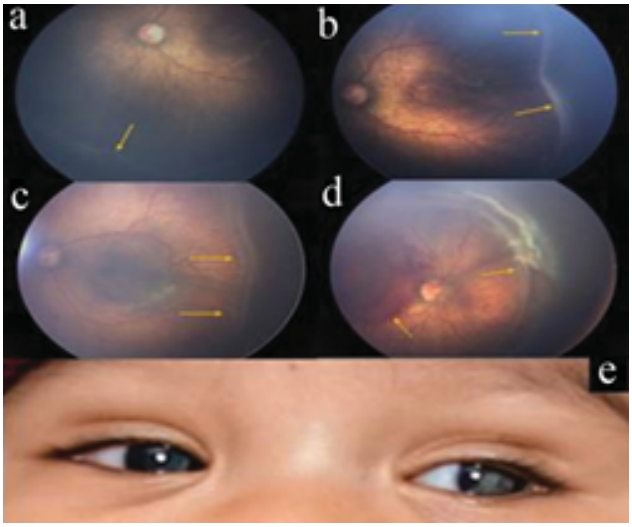

Figure 1. ROP stage Features, stage one (image a), stage two (image b), stage three (image c), stage four (image d), stage five (image e)

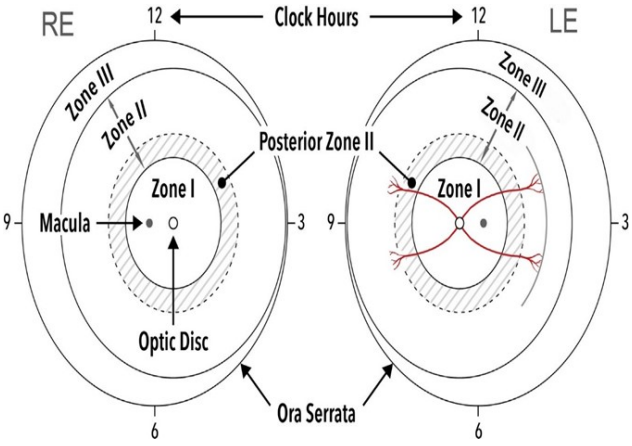

Figure 2. Retina zones [5]

no presence of Plus ROP [8]. Plus ROP is a severe stage of ROP which most times is found in Zone I and can be diagnosed when the pupil is dilated [6].

C. Aggressive ROP (A-ROP)

Retinopathy of Prematurity disease is aggressive if upon occurrence it has the potential to progress rapidly to stage V. This type is found in zone II and the effects of this type of ROP are severe for regions where disease screening is not common [6]. Most times it is diagnosed while at stage V.

D. Challenges with ROP disease diagnosis

To diagnose ROP, an ophthalmologist conducts an eye examination of the baby and where possible captures images for assessment. The procedure for ROP disease diagnosis which requires an eye specialist pauses challenges because of the increasing burden of lack of enough eye specialists [4]. Hospitals without eye specialists use clinicians to capture the images and have them transmitted via emails to ophthalmologists to assist in the diagnosis [5]. Images captured using different devices have different resolutions [9]. Image transfer can also cause distortion and results interpretation becomes a challenge. For these reasons, the use of intelligent systems to support disease diagnosis is important. Stage one and two ROP is able to develop and clear without diagnosis, but stage 3 should be diagnosed and if not, once the disease progresses to stage four, the retina detaches and at stage five the baby goes blind. ROP stage three is treatable and hence our motivation of developing a Deep Learning model for diagnosing this stage of the disease. This work first customizes the ResNet-50 model to first separate images of quality from non-quality ones. Images classified as of quality are used to train a Deep Neural Network for ROP stage II and stage III diagnosis. This paper is divided into five sections: Section one provides background information of ROP disease. Section two presents a review of Literature on Deep Learning applications for ROP disease diagnosis. Section three shows the methodology and the model, section four discusses the model results and section five makes conclusions and recommendations for future work.

2. Literature Review

This section presents a review of recent developments and the available architectures for ROP disease diagnosis.

A. Data Preparation

Eighteen studies were first obtained after initial database screening, seven of which used data from China [10], [11], [12], [13], [14], [15], [16], the study [17] got data from North America, [18] from America and Mexico, [19] from America and Nepal and the study [20] had data obtained from a hospital in New Zealand. The eight studies had their data collected between 2011 and 2020. The studies [14], [15], [16], [17], [20]. Sorted their data into two data sets: quality and non-quality. Six studies [13], [14], [15], [18], [20], [21] focused on detecting the severity of the disease, utilizing datasets larger than 5358 images. Ten studies did data augmentation to increase on the datasets for both training and testing [11], [12], [14], [15], [18], [19], [20], [22], [23], [24]. There was a similarity between the following eleven studies where all babies screened were of a gestational age above 30 weeks and or birth weight above 1500 grams [11], [12], [13], [15], [16], [19], [20], [25], [26], [27], [28]. The eleven studies also used databases of a size whose range was between 2668 and 52,249 images. Twelve studies [10], [11], [12], [16], [19], [22], [23], [24], [25], [26], [27], [28] developed models to detect the presence or absence of ROP disease.

B. Architectures design

Eleven studies [11], [12], [13], [15], [16], [19], [20], [25], [26], [27], [28] developed CNN, U-Net, VGG16, VGG19, ResNet and ImageNet. Nine studies [10], [12], [14], [15], [16], [22], [23], [24], [25] had datasets for training, testing and validation with images more than 10,000.

All studies co-opted one to five image graders as experts to sort quality images to be used for model development. Twelve studies [10], [17], [18], [19], [20], [22], [23], [24], [25], [26], [27], [28] did a comparative analysis of the performance of their models with similar existing architectures. Five studies [10], [11], [12], [16], [19] detected the presence or absence of ROP disease within an average of 0.984 Area Under the ROC Curve (AUC). Six studies [12], [13], [14], [15], [17], [20] detected the Plus disease providing an average of 91.13% sensitivity and 95.92%, specificity. Three studies [10], [11], [16] had an average sensitivity of 95.72% and average specificity of 98.15%. One study [16] worked closely with an ophthalmologist to prepare a report of their model results with his results for ROP disease diagnosis achieving 94.1% sensitivity and 99.3% specificity. Two studies [13], [17] detected the presence of ROP at stage one and two with an average sensitivity of 96.2% and 95.7% specificity.

Sankari et al. [29] developed a hybrid model to segment retina vessels using fundus images. The model was trained with 3200 images and tested with 800 images which had been collected from three hospitals in India. The design of the hybrid model constituted a modification of multiResUNet model for image segmentation and a gray-level cooccurrence matrix was used for feature extraction. Extracted features were used to train the model achieving an accuracy of 94.5%, sensitivity of 94%, and specificity of 93%. Our work advances this development by developing customizing ResNet50 model to first perform image quality assessment and using quality images for training and testing the model. We were also able to detect the presence of ROP as well as ROP stage II and III. Wagner et al. [30] developed a code free Deep Learning model to detect the presence of Pre or Plus ROP disease. Their work used 1370 images obtained at Homerton hospital in the UK. Even though this model was able to detect the presence of Pre and Plus ROP disease, code free Deep Learning models have major challenges of not able to adapt well with the approach of transfer learning. This means that their adaptability to be re-configured is complex. These models work best where the developer has many images and available computing resources. Vijay et al. [31] developed a Deep Convolutional Neural network model for detecting the location of the optical disk as well as detecting ROP at zone I. II and III. The work used 1117 images where the model first sought to identify the location of optical disk then image segmentation and using the segmented images for ROP zone identification. Training this model took 10 hours and the model was able to achieve an accuracy of 98.94% for detecting optical disk and 96.69% for vessels segmentation.

Li and Liu [32] developed a Deep Convolutional neural Network using 18,827 images captured using retcam camera at maternaland hospital in China. The aim of their work was to classify images into four classes: Normal, ROP stage I, stage II, stage III. The design of the architecture had a U-Net module for identifying a Region of Interest (ROI) as the segmented features, and a DCNN for classifying the images into four classes. Their model achieved sensitivity and specificity of 90.21%, 97.67% respectively for ROP stage I, sensitivity and specificity of 92.75%, 98.74% for ROP stage II, and sensitivity and specificity of 91.84%, 99.29% respectively for ROP stage III. Their work recommended the use of images captured from different regions and hospitals to increase diversity of the images used in training the model. Our work advances this work by using images captured from different hospitals and countries which is publicly available in Kaggle database. Our model was also able to achieve an accuracy of 92.8%, sensitivity of 94.9%, and precision of 97.3% for ROP stage III disease diagnosis. Bai et al. [33] developed a Deep Learning algorithm to detect the presence of Pre and Plus ROP disease. The model was trained using 8052 images attaining sensitivity and specificity of 43% and 96% respectively.

Sharafi et al. [34] developed an algorithm for the detection of ROP Plus disease using two features: tortuosity and vessel dilation. The work utilized 76 images, 37 with no ROP Plus disease and 39 with the Plus ROP disease. Non quality images were manually sorted by a human expert. This study was limited because of using few images hence the increase in the chances of model overfitting. Salih et al. [35] did a study for a comparative performance of three architectures: VGG-19, ResNet-50 and EfficientNetB5 to detect the presence of ROP in retina zones. The three models were trained and tested using 1365 images obtained from Amal clinic in Iraq and EfficientNetB5 model had a higher accuracy of 87.27%. Coyner et al. [36] developed a logistic regression model for detecting whether an infant with ROP would require treatment and the data size set was of 3760 images collected from India, Nepal and Mongolia. Rao et al. [37] developed a Deep Learning algorithm for ROP detection using 37,477 images captured using different devices in India and for a period of eleven years. 25,982 images were used to train the model, 4006 for validation and 7,489 for testing. A binary classification was used for ROP detection and customized EfficientNetB0 model for ROP stages I-III detection. The model achieved a sensitivity of 91.46% and a specificity of 91.22% in detecting the presence of ROP.

Struyven et al. [38] developed two models: DenseNet model and code free Machine learning model AutoML model. The two models were trained using 6620 images and tested using 200 images with a ratio of 100 images without ROP disease, 49 images with Pre-plus disease and 51 images with Plus disease. The DenseNet model had a sensitivity and specificity of 39.7%,97.3% respectively while the code free model had a sensitivity and specificity of 60.3%,98.9% respectively. Salih et al. [39] did a performance comparative evaluation of ten models in detecting ROP stages. The models included: DenseNet121, DenseNet169, ResNet50, ResNet101, ResNet152, AlexNet169, VGG16, Inception-

v3, SqueezNet1-0 and SqueezNet1. All models were trained and tested with the same number of images of 3720 in number. Inception-v3 model had the highest accuracy of 99.5% followed by DenseNet169 with an accuracy of 77.14%. Attallah [40] developed an application for ROP diagnosis using Gabor wavelet images. Features were extracted for ROP disease detection achieving an AUC of 0.98. Salih et al. [41] did a performance evaluation of four models: DenseNet161, ResNet50, EfficientNetB5 and a fusion classification algorithm using 2776 images. The fusion classification algorithm had the highest accuracy of 90.2%, followed by EfficientNetB5 (81.29%,), ResNet50 (80.57%) and DenseNet161 (69.78%).

3. Methodology

ResNet-50 architecture [42] is a revised advanced version of the ResNet-34 model which utilizes a bottleneck structure while building the blocks. The blocks are built to allow 1*1 convolutions as the bottleneck to reduce model parameters and the multiplication of the matrix. The design of the architecture enhances the speed of the model training providing a three-layered stack. The architecture consists of a 7*7 Kernel convolution which are two sized strides, one maximum pooling layer with a two sized stride, one average pooling layer with one fully connected layer utilizing SoftMax activation function. The implementation follows a function, Output=F(X)+X where x represents an input from a residual block which is an output from the previous layer and F(X) is a function of the convolutional Neural Network Model built in blocks. The model architecture design steps shown in Figure 3 is designed to receive images as inputs, which are preprocessed by the ResNet-50 pipeline. After preprocessing, desired features are extracted and fed into a Deep Neural Network to quickly sort them as with having the disease or not. Those with ROP are further run under a classification module for ROP stage III disease detection.

A. Data Preparation

There exist only two publicly accessible databases with ROP images: Kaggle database [43] and HVDROPDB database [44]. As shown in Figure 4, a total of ninety-one (91) images were obtained from the Kaggle database [43]. Thirty-nine (39) images had been labeled as not having ROP disease. Nineteen (19) images had been labelled ROP stage one, Twenty-two (22) ROP stage two, eleven (11) images ROP stage three. The database does not contain images of ROP stage four and five. These images were all collected by different eye specialists globally using fundus camera and stored in Kaggle which is an online open-source database for scientific and research use. HVDROPDB database [44] contains images captured using two different cameras: Retcam and Neo. Images collected using Retcam camera were 10,000 in number while using Neo camera were 1,100. We combined them into one database forming a total number of 11,100 images and were able to extract images of quality amounting to 670 images with ROP, out of which 50 had ROP stage 3 and 450 noROP. Data was augmented to increase the volume of the images resulting to a total of 2,300 images with ROP, 1,900 images without ROP, 422 images with ROP stage two and 470 images with ROP stage three as shown in table 1. Open CV library was used to preprocess images through reducing their size to 224*224. To increase the volume of the images data was augmented and all images were rotated setting the range of 2, width of 0.05 and height of 0.05, enabling the Zoom range to be [0.85,1.15]. Selection of the images for training, testing and validation was done with a ratio of 0.80 training, 0.10 testing and 0.10 validation.

B. Algorithm

The below algorithm was used to guide the design steps of the steps of developing the model which helped to observe the expected outputs.

Input Image from database
%Convert image to grayscale
image=rgb2gray(image); print(image)
Print grayscale converted image output
% Apply CLAHE to normalize image contrast
print (CLAHE _Output)
% Extract features
print (Vessel); header ('Vessel resulting output')
%Determine ROP disease
Classify images to have ROP or NO ROP
For ROP case Then
Classify images either ROP stage 2 or 3
END

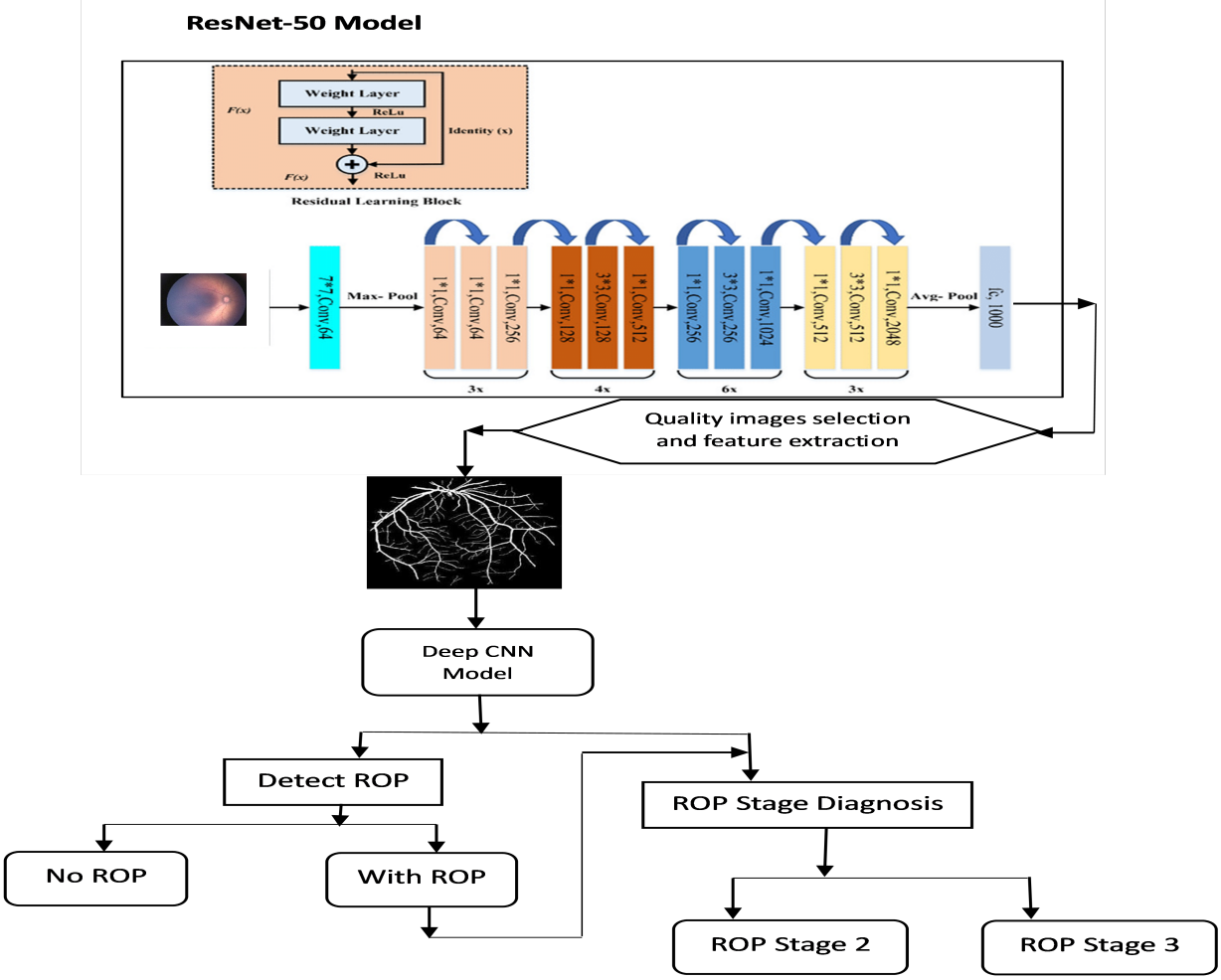

Figure 3. Model Architecture Design Steps

TABLE I. Dataset, Original images (Oimages),Original images with ROP (WithROP), Augmented with ROP (AwithROP), Images without ROP (NoROP), Augmented images without ROP (ANoROP), ROP stage 2 images (Stage2), Augmented ROP stage 2 images (Astage2), ROP stage 3 images (Stage), Augmented ROP stage 3 images (Astage3)

Database	Oimages	Qimages	WithROP	AwithROP	NoROP	ANoROP	Stage2	Astage2	Stage3	Astage3
Kaggle [43]	91	91	52	600	39	400	22	102	11	70
HVDROPDB [44]	11,100	890	670	1,700	450	1,500	170	320	50	400
Totals	11,191	981	722	2,300	489	1,900	192	422	61	470

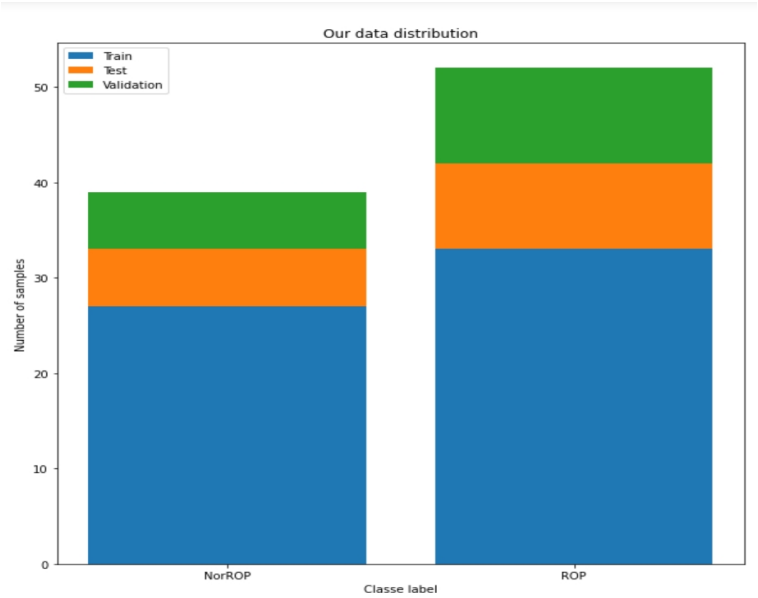

Figure 4. Data class output

4. Results

This section presents the output of the model together with a comparison of the model results with other recent developments.

A. Data Visualization and preprocessing

Before augmentation, data was split the into a ratio of 0.80 training, 0.20 testing and 0.20 validation and produced outputs to view the images as shown by Figure 7. We build a list of indexes for each data set which produced an output of the total number of images used per data set with an output summary graph. Images had been captured from both eyes, left, and right resulting to different structures of the images, hence flipping the images helped to achieve a uniform structure. The design structure of the Deep Convolutional Neural Network consisted of five layers for feature extraction and pretraining, one pooling layer and three fully connected layers with model epochs terminating at 40. Adam optimizer was used to adjust the learning rate of all parameters reducing the loss function. At epoch 32 and 40, we did observe that there were no significant changes producing a stable result of 92.3% accuracy on testing results. The model loss also decreased with the increase of each epoch as shown in Figure 5. An additional function was developed to predict the labels or class group to which each image belonged to. As shown by Figure 6, a confusion matrix was created to show test labels against predicted test labels.

B. Deep Neural Network Model Training

Before training the model, images were first preprocessed using a customized ResNet-50 model to ensure that images whose features were not visible were discarded. As shown by Figure 8, image grayscale conversion was done to reduce any chance of classification errors as a result of color and Contrast Limited AHE (CLAHE) applied to reduce image contrast as shown in Figure 9. Image vascular structure was extracted and used as the model training features. The distinguishing feature of ROP stage II from Stage III is that the ridge line for ROP stage III is longer than that of Stage II as shown in Figure 10. As guided by equation 1-3, we obtained accuracy, sensitivity, and specificity of the VGG19 and the S-Net models and compared their results with our model. We also did a comparison of the results of our model with that of an ophthalmologist as guided by the Kappa method [45]. The approach helped to allow comparison of the model's results with the rating provided by the ophthalmologist. Using equation 1-3, TP (True Positive) is used to denote that ROP disease is present and the model results gives a positive output to confirm the presence of the disease, FN (False Negative) used to denote that ROP disease is present and the model results provides a negative result as disease absent, FP (False Positive) standing for disease absent and the model results showing ROP disease present, TN (True negative) to denote ROP disease absent and the model results showing absence of the disease.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

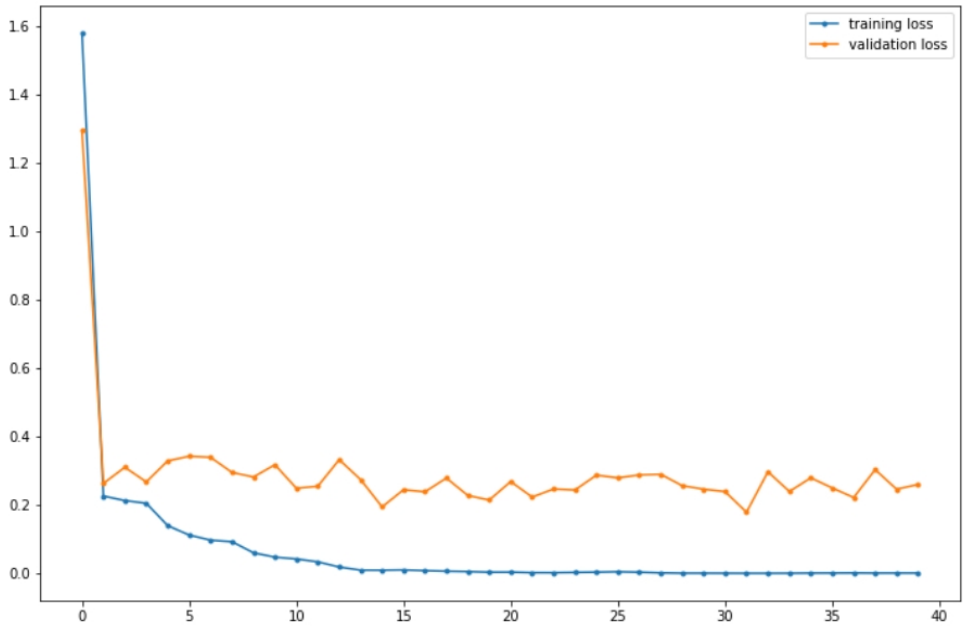

Figure 5. Model Training and Validation Loss

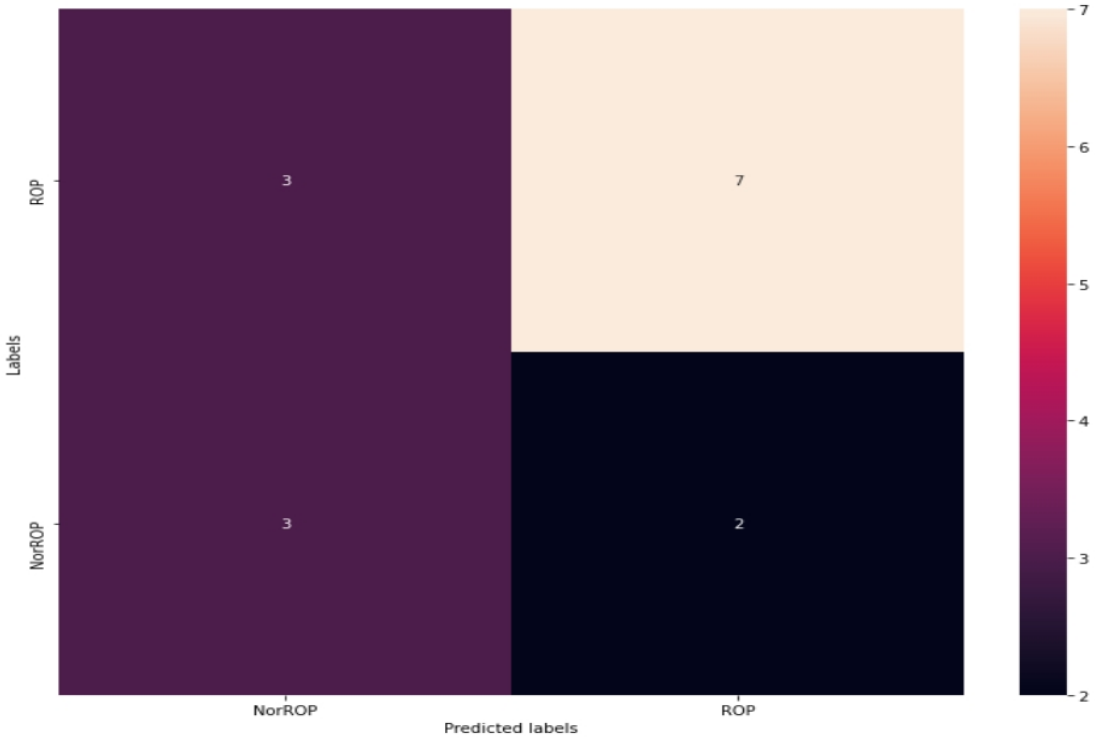

Figure 6. Prediction labels output

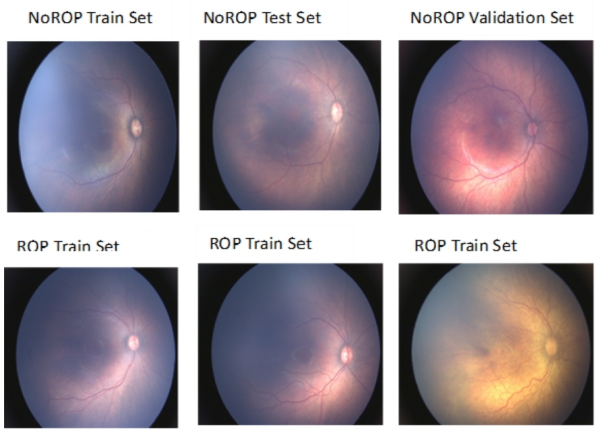

Figure 7. Data Visualization output

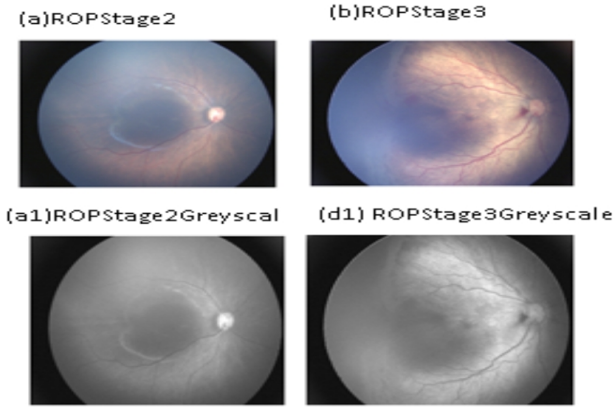

Figure 8. Image colour Normalization

C. Model Design Platforms

Some images classified as not of quality were processed using MATLAB R2023B and were incorporated into the quality database. The computer used for running the deployed model operated under Windows 10 OS with a RAM of 1 Terabyte. With the sponsorship of Google PhD Research Africa Funding, we were able to afford the GeForce RTX 2080 NVIDIA GPU which provided a powerful graphics for better image view. The GPU provides 11GB memory with a 352-bit bus to boost the memory.

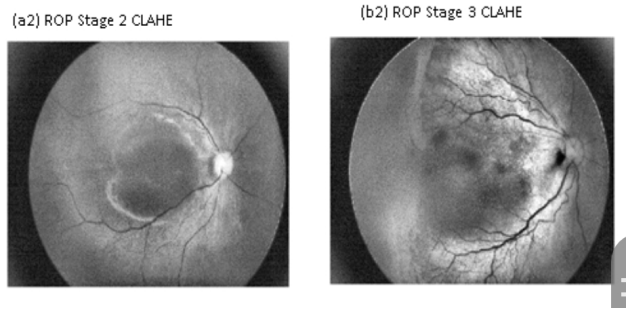

Figure 9. CLAHE output

D. Experimental Setup

The majority of the existing Deep Learning architectures have been developed to detect either the presence or absence of ROP disease or the ROP disease severity with no model for ROP stage III disease diagnosis. We configured the design of two models VGG-19 [46] and S-Net [47] ensuring that the same number of images are used for training, testing and validation. These two models had been built to detect presence or absence of ROP disease where we changed the design to have presence of ROP was named ROP stage II and absence ROP stage III. The design architecture for the two models VGG19 and S-Net had four layers [46], [48] with the VGG19 having the first two layers of size 200 with a dropout rate of 49.5%, the third layer was lesser in size of 68 and with a dropout rate of 49.7%, the fourth layer was applied to produce disease detection. The S-Net model design was different with all layers of different sizes, having the first layer of size 510 with a dropout rate of 49.6%, second layer of size 202 and a dropout rate of 48.9%, the third layer of size 100 with a dropout rate of 50% and the last layer for disease detection giving no drop out.

As shown in Table I, a confusion matrix was created with 0 representing ROP stage two images and 1 representing ROP stage three images. The models were tested using the same image data sizes of one hundred and eighty-four (184). VGG19 model classified one hundred and seventy-five (175) images to their correct class of ROP stage two (TP) and giving nine images (9) as wrong classifications for ROP stage two (TN). The model also correctly classified one hundred and seventy-eight (178) images containing ROP stage three (TP) and six (6) images wrongly classified (TN) for ROP stage three. S-Net model on the other hand classified one hundred and sixty-eight (168) images to their correct class of ROP stage two (TP) and giving sixteen images (16) as wrong classifications for ROP stage two (TN). The model also correctly classified one hundred and seventy (170) images containing ROP stage three (TP) and fourteen (14) images wrongly classified (TN) for ROP stage three. As shown in Table II, VGG19 model had the highest accuracy of ROP disease classification, and the results were compared with the results of our model. Five-fold cross validation was performed, and the results of our model was at an accuracy of 95.8, 96.9% sensitivity and 98.3% precision outperforming that of VGG19 model which had an accuracy of 92.8%, 94.9% sensitivity and 97.3% precision as shown in Table III. As shown in Table IV, we did a comparative analysis of the performance of our model with the results of similar recent developments even though most models were not developed to diagnose ROP stage three but for the presence or absence of the disease which was one of our objectives.

TABLE II. VGG19 and S-NET models confusion matrix

Model		0	1
VGG19	0	175	9
	1	178	6
S-NET	0	168	16
	1	170	14

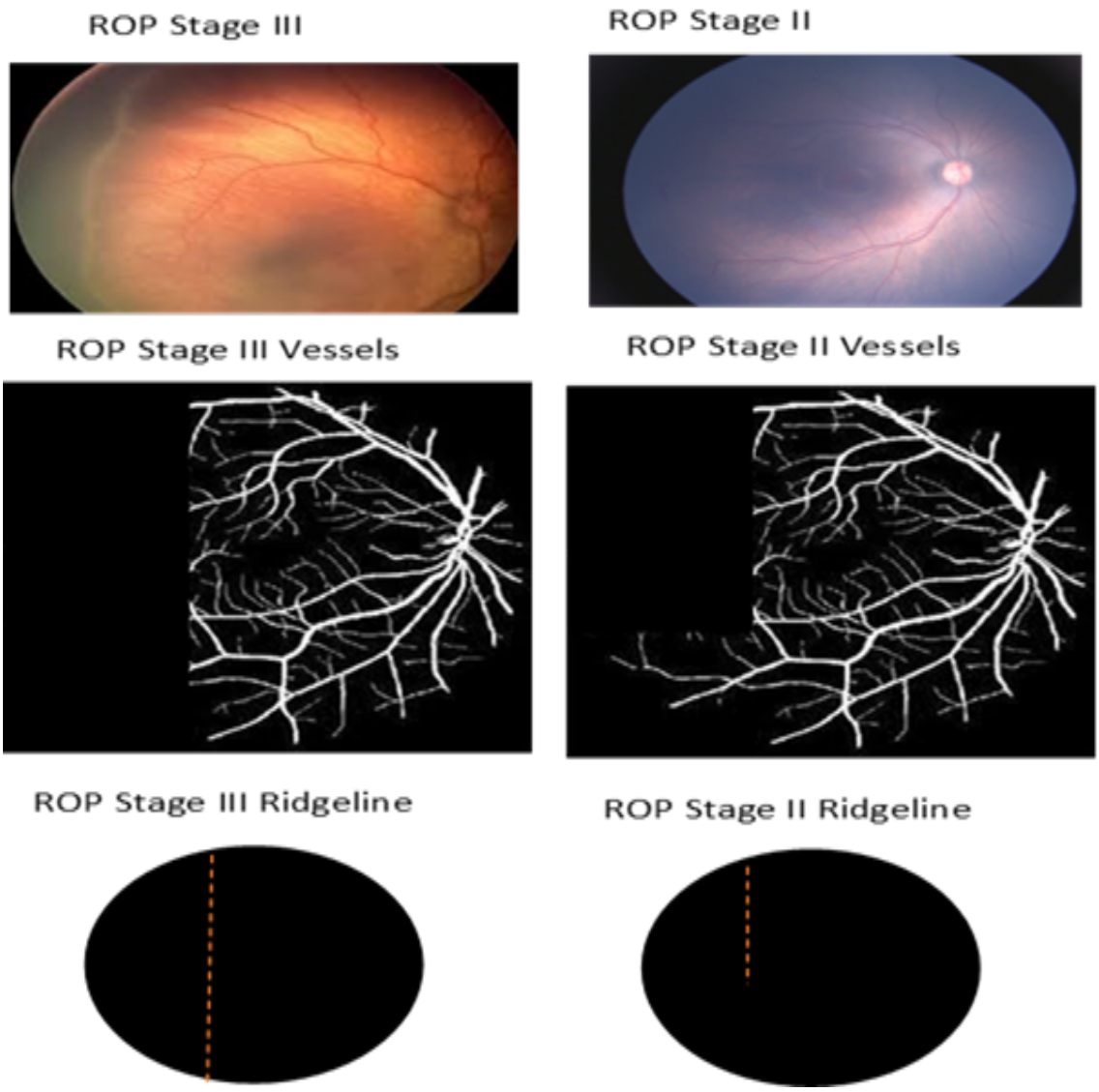

Figure 10. ROP stage III Ridgeline extraction

TABLE III. S-NET AND VGG19 RESULTS

Model	Accuracy	Specificity	Sensitivity	Precision	AUC
S-NET [47]	86.21%	89.50%	95.63%	93.41%	0.90%
VGG19 [46]	91.21%	97.52%	97.72%	90.84%	0.97%

TABLE IV. 5-FOLD CROSS VALIDATION OUTPUT

Fold Number	Accuracy	Sensitivity	Specificity	Precision	AUC
Fold 1	89.30	92.41	93.23	96.72	0.95
Fold 2	89.61	93.20	94.51	96.91	0.96
Fold 3	90.82	93.91	94.92	96.91	0.96
Fold 4	91.73	94.42	95.21	97.12	0.97
Fold 5	92.82	94.90	95.20	97.31	0.98

TABLE V. RECENT DEVELOPMENTS RESULTS

Citation	Publication Year	Classification	No. of Images	Accuracy	Sensitivity	Specificity
Haung et al., [12]	2021	ROP Stage 1and II	11372	99.92%	96.12%	95.90%
Wang et al., [10]	2018	Mild or Severe ROP	20795	None	96.63%	99.32%
Hu et al., [11]	2019	Mild or Severe ROP	3017	97.21%	None	None
Brown et al., [17]	2018	Plus or Pre Plus ROP	5511	None	88.42%	92.32%
Sankari et al., [29]	2022	Presence or absence or ROP	400	92.87%%	none	none
Wagner et al., [30]	2023	Plus or Pre Plus ROP	1370	None	None	None
Vijay et al., [31]	2023	ROP in Zones	1117	98.94%	None	None
Li and Liu., [32]	2022	ROP stage I-III	18827	None	91.84%	99.29%
Bai et al., [33]	2023	Plus or Pre Plus ROP	8052	97.3%	96.60%	98%
Sharafi et al., [34]	2024	Plus or Pre Plus ROP	Not specified	86%	None	None
Salih et al., [35]	2023	ROP in Zones	1365	88.82%	None	None
Coyner et al., [36]	2022	Mild or Severe ROP	16344	80%	None	None
Rao et al., [37]	2023	Presence or absence of ROP	37477	None%	91.46%	91.22%
Struyven et al., [38]	2022	Plus or Pre Plus ROP	6620	None	39.7%,60.3%	97.3%,98.9%
Salih et al., [39]	2023	ROP stages	3720	99.5%	None	None
Attallah et al., [40]	2023	Presence or absence of ROP	Not specified	None	None	None
Salih et al., [41]	2023	Plus or Pre Plus ROP	2776	90.2%	None	None

5. Discussion

From this study it is observable that studies on ROP disease can be grouped into five classes: Detection of ROP stages [12], [32], [39], Detection of mild or severe ROP symptoms [10], [11], [36], detection of the Pre or Plus ROP disease [17], [30], [33], [34], [38], [41], Presence or absence of ROP [29], [37], [40] and ROP in zones [31], [35]. Studies investigating ROP at stage three and presented data, model design and results of the model accuracy was one [32]. This study was limited because data was collected from one hospital and concluded the need to expand their work by using images collected from diverse countries which we have achieved. In our study we had a module for image quality assessment of the images before use for the model training. Twelve studies [11], [12], [13], [15], [16], [19], [20], [25], [26], [27], [28], [39] used human graders to sort images of quality from non-quality and this can result to human errors. Two studies [29], [31] did image segmentation to obtain features of interest for model development which we also did and do agree that this step works to maximize on the model accuracy and minimize errors.

Two studies [35], [39] did a comprehensive work of conducting a performance comparative study of many architectures such as the DenseNet121, DenseNet169, ResNet50, ResNet101, ResNet152, AlexNet169, VGG16, Inception-v3, SqueezNet1-0 and SqueezNet1. However, their work did not present details of model's customization, design and challenges. Their focus was more aimed to determine which model outperforms the other without analysing model design limitations which could be a contribution to the results. To enhance features visibility and reduce image contrast, our work applied CLAHE which made the vessels more visible for extraction. Colour was also normalised by converting all images to grayscale. Our novel contribution and advancement to the development of Deep learning models for ROP stage II and III diagnosis, was the ability of our model to distinguish in length the ridgeline for ROP stage II which was short than that of ROP stage III as shown in figure 10. The develop of our model was done a time when a new ROP database (HVDROPDB) had been published last year 2022 and this helped us obtain 11, 191 images which made the model training and testing images adequate. Data from the two databases used to build our model are publicly accessible and this is an advantage as compared to most of all other studies which are privately owned.

6. Study Limitation

The design of our model architecture allows a sequential flow which begins by the application of the ResNet50 model to separate images of quality from non-quality ones, features are then extracted and fed into the Deep Convolutional Neural Network for ROP stage II or III diagnosis. This design sequence could be a limitation to ophthalmologists who would only one the model to separate images of quality from non-quality without doing the disease diagnosis. Our future work will now focus one developing an algorithm for retina quality assessment without ROP stage diagnosis. Data used to train, test and validate the model was obtained from two public databases: Kaggle [43] and HVDROPDB [44]. The images were well labelled by ophthalmologists with information about the device used to capture the images, image size available. This was a big step which made the development of the model easy however we seek to obtain data captured with different devices such as mobile devices and use it for testing our model as we continue to observe the results.

7. Conclusion

Retinopathy of Prematurity (ROP) disease affects babies born preterm and characterized with low birthweights. ROP disease is preventable when diagnosed before progressing from stage three to four. The disease which currently is termed as third pandemic has a high contribution to preventable blindness among children. High income countries have national screening guidelines and screening is mandatory to all babies born preterm which reduces the chances of blindness. The disease remains a burden for low-income countries with only two countries in Africa: South Africa and Kenya being the only two countries with national screening guidelines. The screening criteria for countries without screening guidelines is determined through a discussion and agreement of ophthalmologists available for the disease screening. The increase in the cost burden for the disease diagnosis with only few ophthalmologists available for the disease screening creates the need for technological experts to develop Artificial Intelligence systems for assistive diagnosis. Deep Learning applications have recently been developed to accurately provide a classification of many disease stages or detection. These systems results accuracy relies mostly on the quantity of data used for training paying attention on the model testing and evaluation by an expert. There exist three key challenges associated with Deep Learning models development which this work managed to overcome.

A. Data quality

Data from the two databases had labelling on the camera used and we do note that different cameras produced images of different contrast affecting clarity and resolution. Transfer of images in some cases was causing distortion and we managed to overcome this challenge by building a module for image quality assessment then using quality images to train the model.

B. Data Imbalance

The dataset with ROP disease and the one without ROP were not equal in numbers creating imbalance. Imbalanced data increases the chances of misclassification. To avoid these challenges, we were able to augment the data to increase on volume of the images and obtained the Area Under Curve (AUC) to show the model loss results as shown in Figure 8.

C. Model development

Developing a Deep Learning architecture from scratch is a tedious task and we do appreciate the developers of the existing architectures which can be customized and applied to solve many problems. In this work, we were able to customize ResNet-50 architecture for detecting the presence of ROP disease and a Deep Neural Network model for ROP stage three disease diagnosis. Transfer learning was applied to quickly adapt the training of the customized architectures to use the new available datasets successfully. This work developed a Deep Learning model to first detect the presence or absence of ROP disease then diagnose ROP stage three which is a critical stage. The ResNet-50 architecture is customized to preprocess the images for ROP disease detection. Features are extracted where the ridgeline for ROP stage three is done for model training. Extracted features are used to train a Deep Neural Network (DNN) to class ROP stage three. 91 images were obtained from Kaggle database, and 50 images were obtained from a recently published ROP database the HVDROPDB. The data sets were combined and augmented to achieve adequate datasets for training, testing and validation. The performance of our model was compared with the VGG19 and S-Net architectures where our model results had the highest accuracy of 92.8%, 94.9% sensitivity and 97.3% precision. To support future developments, we do recommend that there is need for more population-based research to be done on ROP and allow the data to be public which will support innovative developments for the disease diagnosis.

8. Availability of data and Materials

Data used in this work is publicly available on Kaggle database and the HVDROPDB database. No license is required for use or during publication of articles.

9. Funding

This work received funding from Google PhD research in the year 2023.

10. Ethical approval

This study was reviewed and approved by Strathmore University Institutional Scientific and Ethical Review Committee (SU-ISERC), Certificate number (SU-ISERC1534/22) and approved by the Kenya National Commission for Science, Technology and Innovation (NACOSTI), license number (NACOSTI/P/23/23702).

11. Consent to participate

This work did not involve direct participation of human subjects or animals. Data and human images used in this study had been collected by ophthalmologists, properly labelled and stored in databases for research purposes and use.

12. Competing interest

All Authors declare no Competing Interests

References

[1] K. Sabri, A. L. Ells, E. Y. Lee, S. Dutta, and A. Vinekar, "Retinopathy of prematurity: A global perspective and recent developments," 2022.

[2] A. Thanos, Y. Yonekawa, B. Todorich, and D. M. Moshfeghi, ""screening and treatments using telemedicine in retinopathy of prematurity," vol. 8, p. 147–512, 2016.

[3] A. S. H. Tsai, L. Acaba-Berrocal, and M. Sobhy, ""current management of retinopathy of prematurity," " *Curr Treat Options Peds*, vol. 8, 2022.

[4] T. Chan-Ling, G. A. Gole, G. E. Quinn, S. J. Adamson, B. A. Darlow, and "Pathophysiology, *screening*. and treatment of ROP: a multi-disciplinary perspective, " Prog Retin Eye Res, 62:77–119, 2017.

[5] M. D. Abràmoff, M. K. Garvin, and M. Sonka, ""retinal imaging and image analysis," 2010.

[6] A. S. Tsai, H. D. Chou, X. C. Ling, T. Al-Khaled, N. Valikodath, E. Cole, V. L. Yap, M. F. Chiang, R. Chan, and W. C. Wu, ""assessment and management of retinopathy of prematurity in the era of anti-vascular endothelial growth factor (VEGF) Progress in retinal and eye research," 2021.

[7] M. Yanoff and J. S. Duker, *Ophthalmology*. Elsevier Health Sciences, 2022.

[8] O. Onyango, S. Sitati, L. Amolo, F. Murila, S. Wariua, G. Nyamu, M. Lango, and A. Patel, "Retinopathy of prematurity in Kenya: prevalence and risk factors in a hospital with advanced neonatal care," *The Pan African medical journal*, vol. 29, no. 152., 2018.

[9] M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, K. H. Parker, and N. W. Witt, ""evaluation of a portable retinal imaging device: towards a comparative quantitative analysis for morphological measurements of retinal blood vessels," p. 6, 2023.

[10] J. Wang, R. Ju, Y. Chen, L. Zhang, J. Hu, and Y. Wu, ""automated retinopathy of prematurity screening using deep neural networks," pp. 1–8, 2018.

[11] J. Hu, Y. Chen, J. Zhong, R. Ju, and Z. Yi, ""automated analysis for retinopathy of prematurity by deep neural networks," pp. 269–279, 2019.

[12] Y. Huang, H. Basanta, E. Kang, K. Chen, Y. Hwang, and C. Lai, "Automated detection of early-stage ROP using a deep convolutional neural network," pp. 1099–1103, 2021.

[13] J. Mao, Y. Luo, L. Liu, J. Lao, Y. Shao, and M. Zhang, ""automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks," " *Ophthalmol, Vol*, 2020.

[14] Y. Tong, W. Lu, Q. Deng, C. Chen, and Y. Shen, ""automated identification of retinopathy of prematurity by image-based deep learning," pp. 7–40, 2020.

[15] J. Wang, M. Zhang, J. Lin, G. Zhang, and W. Gong, ""automated explainable multidimensional deep learning platform of retinal images for retinopathy of prematurity screening," 2021.

[16] Y. Zhang, L. Wang, Z. Wu, J. Zeng, Y. Chen, and R. Tian, ""development of an automated screening system for retinopathy

of prematurity using a deep neural network for wide-angle retinal images," pp. 32–41, 2019.

[17] J. Brown, A. Campbell, K. Beers, S. Chang, and Ostmo, ""automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks," *" JAMA Ophthalmol, Vol 36*, vol. 03-10, 2018.

[18] M. Yildiz, P. Tian, I. Yildiz, M. Brown, and J. Kalpathy, ""plus disease in retinopathy of prematurity: Convolutional neural network performance using a combined neural network and feature extraction approach," *" Transl Vis Sci Technol*, vol. 9, p. 10, 2020.

[19] J. Chen, A. Coyner, S. Ostmo, K. Sonmez, S. Bajimaya, and E. Pradhan, ""deep learning for the diagnosis of stage in retinopathy of prematurity: Accuracy and generalizability across populations and cameras," pp. 27–35, 2021.

[20] Z. Tan, S. Simkin, C. Lai, and S. Dai, ""deep learning algorithm for automated diagnosis of retinopathy of prematurity plus disease," *" Transl Vis Sci Technol, Vol 8*, vol. 23, 2019.

[21] S. Ramachandran, P. Niyas, A. Vinekar, and R. John, ""a deep learning framework for the detection of Plus disease in retinal fundus images of preterm infants," pp. 62–75, 2021.

[22] J. Peter, F. Michael, S. J. Chiang, and M. D. Chen, ""artificial intelligence for retinopathy of prematurity," 2022.

[23] A. Bai, C. Carty, and S. Dai, ""performance of deep-learning artificial intelligence algorithms in detecting retinopathy of prematurity: A systematic review," pp. 296–307, 2022.

[24] H. Adam, T. Peng, E. Deniz, K. C. Jayashree, O. Susan, J. Karyn, C. Paul, F. Michael, and C. Peter, *"Improved Training Efficiency for Retinopathy of Prematurity Deep Learning Models Using Comparison versus Class Labels.* " America academy of ophthalmology, 2022.

[25] S. A. Coyner, S. J. Chen, K. Chang, P. Singh, S. Ostmo, R. V. Chan, M. F. Chiang, C. J. Kalpathy, and J. Campbell, ""synthetic medical images for robust," *Privacy-Preserving Training of AI: Application to Retinopathy of Prematurity Diagnosis, " Mendeley Data*, vol. 10., 2022.

[26] Q. Wu, Y. Hu, and Z. Mo, *"Development and Validation of a Deep Learning Model to Predict the Occurrence and Severity of Retinopathy of Prematurity.* " JAMA Netw Open, 2022.

[27] J. Well, J. Lie, M. Tong, G. Zonayman, H. J. C.Y.Uzhong, and Z. Peiquan, ""automatic retinopathy of prematurity staging with a semi supervised deep learning method based on dual cosistency regularization," *" Research square, Vol 18*, vol. 10, 2022.

[28] V. M. Raja, U. Snekhalatha, C. Ashok, and B. Prabhu, ""automated diagnosis of Retinopathy of prematurity from retinal images of preterm infants using hybrid deep learning techniques," *" Biomedical Signal Processing and Control, Vol 85*, vol. 1746-8094, 2023.

[29] V. M. R. Sankari, U. Umapathy, S. Alasmari, and S. M. Aslam, "Automated detection of retinopathy of prematurity using quantum machine learning and deep learning techniques." IEEE Access, 2023, vol. 11, pp. 94 306–94 321.

[30] S. K. Wagner, B. Liefers, M. Radia, G. Zhang, R. Struyven, L. Faes, J. Than, S. Balal, C. Hennings, C. Kilduff, P. Pooprasert, S. Glinton, M. Arunakirinathan, P. Giannakis, I. Z. Braimah, I. S. H. Ahmed, M. Al-Feky, H. Khalid, D. Ferraz, J. Vieira, and K. Balaskas, "Development and international validation of custom-engineered and code-free deep-learning models for detection of plus disease in retinopathy of prematurity: a retrospective study." *The Lancet. Digital health*, vol. 5, no. 6, (2023).

[31] K. Vijay, P. Het, P. Kolin, and A. Shorya, "Acm trans. comput. healthcare," *Deep Learning-assisted Retinopathy of Prematurity (ROP) Screening*, vol. 4, p. 3, 2023.

[32] P. Li and J. Liu, ""early diagnosis and quantitative analysis of stages in retinopathy of prematurity based on deep convolutional neural networks," 2022.

[33] A. Bai, S. Dai, J. Hung, A. Kirpalani, H. Russell, J. Elder, S. Shah, C. Carty, and Z. Tan, ""multicenter validation of deep learning algorithm ROP," *AI for the Automated Diagnosis of Plus Disease in ROP*, vol. 12, p. 13, 2023.

[34] S. M. Sharafi, N. Ebrahimiadib, and R. Roohipourmoallai, *"Automated diagnosis of plus disease in retinopathy of prematurity using quantification of vessels characteristics.* " Sci Rep, Vol 14, 2024.

[35] N. Salih, M. Ksantini, and N. Hussein, *"Prediction of ROP Zones Using Deep Learning Algorithms and Voting Classifier Technique.* " Int J Comput Intell Syst, Vol 16, 2023.

[36] A. S. Coyner, P. K., Shah *et al.*, "External validation of a retinopathy of prematurity screening model using artificial intelligence in 3 low- and middle-income populations," 2022.

[37] D. P. Rao, F. M. Savoy, J. Z. Tan, B. P. Fung, C. M. Bopitiya, A. Sivaraman, and A. Vinekar, "Development and validation of an artificial intelligence based screening tool for detection of retinopathy of prematurity in a South Indian population," *Frontiers in Pediatrics*, vol. 11, 2023.

[38] R. R. R. Struyven, S. Wagner, B. Liefers, G. Zhang, S. Glinton, M. Kelly, M. Radia *et al.*, "Deep learning-based detection of Plus disease of Retinopathy of Prematurity," *Invest. Ophthalmol. Vis. Sci., vol. 63, no*, vol. 63, no. 7, pp. 212–F0059, Jun. 2022.

[39] N. Salih, M. Ksantini, N. Hussein, D. B. Halima, A. A. Razzaq, S. A., and Mahmood, "Detection of retinopathy of prematurity stages utilizing deep neural networks," *in Proceedings of Seventh International Congress on Information and Communication Technology, Springer Nature*, vol. 10., p. 699–706, 2023.

[40] O. Attallah, "Gabrop: Gabor Wavelets-Based CAD for Retinopathy of Prematurity Diagnosis via Convolutional Neural Networks," *Diagnostics*, vol. 13, 2023.

[41] N. Salih, M. Ksantini, N. Hussein, and D. B. Halima, "Deep learning models and fusion classification technique for accurate diagnosis of retinopathy of prematurity in preterm newborn," *Baghdad Sci. J. doi:*, vol. 10., 2023.

[42] K. He, X. Zhang, S. Ren, and J. Sun, *"Deep Residual Learning for Image Recognition.* " IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 770-778, 2024.

[43] J. S. Hosia, P. S. Angelica, Z. Annisa, N. E. Arundhati, and E. Latifa, *"Detecting Retinopathy of Prematurity Disease in Latifa.* " Baby Based on Fundus Image Data with CNN Model using VGG19's Architecture, ", 2022.

[44] R. Agrawal, S. Kulkarni, R. Walambe, and K. Kotecha, ""assistive

framework for automatic detection of all the zones in retinopathy of prematurity using deep learning," *" J Digit Imaging, Vol 34*, vol. 4, 2021.

[45] L. Ma, H. Zhang, and X. Xie, *"Reverberation Level Recognition by Formants Based on 10-Fold Cross Validation of GMM Digital TV and Wireless Multimedia Communication.* " IFTC 2017 Communications in Computer and Information Science, 2017.

[46] J. Choi, T. Yoo, J. Seo, U. M. Kwak, and T. Rim, ""multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database," *" Vol 2*, vol. 11, 2017.

[47] G. Huang, Z. Liu, D. M. Van, K. Q. Weinberger, and C. C. Networks, "IEEE conference on computer vision and pattern recognition (CVPR)," 2017.

[48] A. Tanvir, F. A. Muhammed, A. Fayadh, A. Samet, and A. Aydin, "Diagnostic biomedical signal and image processing applications with deep learning methods," 2023.

Elizabeth Ndunge Mutua is a Doctoral Candidate at Strathmore University's School of Computing and Engineering Sciences. Her research interest include Machine Perception, Deep Learning for assistive diseases diagnosis and the development of AI Applications.

Dr. Bernard Shibwabo Kasamani is a Senior Lecturer at Strathmore University's School of Computing and Engineering Sciences. He also serves as the Director of Graduate Studies and is an active researcher in the areas of Machine Learning, Embedded Systems, Full Stack Applications, Dataspaces and Data Engineering.

Prof. Christoph Reich is a Professor at Furtwangen university, Institute for Data Science, Cloud Computing and IT Security, Areas of specialization, Machine Learning, Data science and computer security.