



# The Performance of two Well-Known Segmentation Convolutional Neural Networks, Unet and Segnet, for the Segmentation of Blood Vessels, Optic Disc and Demarcation Line in Retinopathy of Prematurity Retcam Images

Carolyn YTW<sup>1,2,3\*</sup>, Timing L<sup>1,2</sup>, Tin LW<sup>3</sup>, Henry HWL<sup>4</sup> and Pearse AK<sup>1,2</sup>

<sup>1</sup>Institute of Ophthalmology, University College London, United Kingdom

<sup>2</sup>Moorfields Eye Hospital NHS Foundation Trust, United Kingdom

<sup>3</sup>Faculty of Medicine, The Chinese University of Hong Kong, China

<sup>4</sup>Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, China

## Research Article

Volume 9 Issue 1

Received Date: May 14, 2024

Published Date: May 28, 2024

DOI: 10.23880/oajo-16000312

\*Corresponding author: Carolyn Yu Tung Wong, Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, New Territories, Hong Kong, Email: carolcarol.carolyn@gmail.com

## Abstract

**Background:** Screening is the predominant strategy for early detection of ROP. However, image analysis is dependent on the experience of ophthalmologists, which introduces subjective variables into the interpretation of data. Computer-aided diagnostics has recently become more widely employed as a second-reader tool, reducing doctors' diagnostic uncertainties. Based on this, the development of such systems is critical, and it may be done by refining its processes, namely, more accurate segmentation and disease detection. Given this circumstance, the current study presents the use of two well-known convolutional neural networks (CNNs) developed for segmentation, U-Net and SegNet, to segment blood vessels (BV), optic discs (OD), and demarcation lines/ridges in ROP fundus images. Defining which architecture is most suited for such a task can assist in reducing the stress on ophthalmologists for ROP screenings and triage.

**Methods:** CNNs are used in segmentation to classify each pixel in an image using self-trained weights. Using three data subsets, each containing 50 RetCam images of ROP and their corresponding masks regarding the BV, OD, or ridge of interest, retrieved from the only publicly available ROP segmentation dataset, i.e. HVDROPDB, we compared the automatic segmentation performed by the U-Net and SegNet using different configurations and in different segmentation tasks related to ROP diagnosis.

**Results:** Among the two proposed architectures, U-Net obtained better overall results in all segmentation tasks, obtaining accuracies of 0.933-0.8751, dice coefficients of 0.49-0.648, and took less training time than the SegNet, which achieved accuracies of 0.933-0.80 and dice scores of 0.40-0.648.

**Conclusion:** The two networks can segment RetCam images of ROP with useful accuracy depending on their configuration, with U-Net being generally faster to train and more accurate.

**Keywords:** Artificial Intelligence; Image Segmentation; Retinopathy of Prematurity; Vasoproliferative

**Abbreviations:** ROP: Retinopathy of Prematurity; CAS: Computer-Aided Systems; ML: Machine Learning; DL: Deep Learning; CNN: Convolution Neural Networks; TP: True Positives; FP: False Positives; TN: True Negative.

## Introduction

Retinopathy of Prematurity (ROP) is a vasoproliferative illness that primarily leads to visual impairment in preterm neonates that weigh less than 1750 g and are born before 34 weeks of gestation [1]. Premature births are susceptible to delays in the normal physiologic process of retinal vascular development [2]. These immature vessels could grow entirely without ROP development or progress to various ROP stages [3].

In ROP, immature vessels are linked to oxygen-induced vascular damage and abnormal neovascularization that develops in the vitreous rather than the retina [2]. ROP in stages 1 (demarcation line), 2 (ridge), or 3 (extraretinal fibrovascular proliferation) might regress spontaneously with subsequent natural vessel growth in certain newborns. When ROP reaches stage 3 with plus disease, it can advance to stages 4 (partial retinal detachment) and 5 (complete retinal detachment), resulting in permanent blindness [4]. Stage 3 ROP with plus disease has to be addressed [2]. Currently, 53,000 children from 15 million preterm births worldwide require ROP treatment [3]. Primary and secondary prevention of ROP are critical in reducing the worldwide burden of ROP and preventing ROP-related early childhood blindness [5]. Primary prevention requires better neonatal care, whereas secondary prevention entails early screening and identification of the disease [5].

Screening at-risk newborns with proper scheduling of examinations and follow-up is crucial to track the progress of the phases and identify infants in need of treatment [6]. Typically, ophthalmologists ought to screen “at risk” babies with a specialized digital wide-field camera, like the RetCam camera [7], or a binocular indirect ophthalmoscope. The advantage of camera-based screening is that it can be done by trained paramedics. Ophthalmologists may evaluate these photographs at base hospitals rather than neonatal care units, where preterm babies are commonly hospitalised. Early detection by ROP specialists is beneficial in terms of disease therapy and management [8].

Unfortunately, there is now a poor doctor-to-patient ratio in ROP assessments. There are very few ROP experts available in diverse locations, including India [9], to cope with the increased occurrence of premature newborns [3]. Even among expert diagnosis, determining ROP can be highly subjective and dependent on experience. In such conditions, it is exceedingly difficult to obtain an accurate and consistent

diagnosis for ROP-affected neonates and those who require ROP therapy. Computer-aided systems (CAS) are necessary to assist medical practitioners with ROP screenings [10]. An automated ROP screening system enables ophthalmologists to save time and concentrate on critical patients that require treatment. An automated method can screen babies efficiently and effectively, ensuring that they receive appropriate, timely treatment and avoid blindness.

With CAS for ROP screening gaining popularity among researchers worldwide, traditional image processing and machine learning (ML) methods began to be used for ROP screening on small image datasets [11,12]. Their shortcoming was that they were time-consuming and not accurate enough for complex problems since they used hand-crafted features [3]. Deep learning (DL) is gaining popularity in medical imaging as the number of images in the dataset grows and GPUs become more widely available. Unlike typical ML methods, DL allows machines to solve complicated problems with greater accuracy, even if the data source is greatly varied and unstructured [3]. A DL network learns high-level features of data sequentially, eliminating the need for human knowledge [3].

To aid the human expert in ROP screening, deep learning algorithms were put forward to identify ROP [13-15]. To help physicians diagnose ROP, the first automated method was built utilising convolution neural networks (CNN) [16]. A unique CNN architecture was presented for detecting ROP and its severity [17]. It is made up of a sub-network for extracting features and a feature aggregation operator for combining these features from diverse input photographs. DeepROP, an automated ROP detection system, was established [18] employing two distinctive DNN models: Id-Net was created to identify ROP. Gr-Net was designed for ROP grading activities. The models were constructed with two independent datasets: identification and grading. The DNN classifier was trained using a transfer learning technique [19]. For classification, the authors employed features that were automatically extracted by pre-trained models, including AlexNet, VGG-16, and GoogLeNet trained on ImageNet. A deep learning framework was developed as well to recognize zone I in RetCam photographs that included both the optic disc and the macula [20].

However, CNNs designed for simple image classification of ROP’s RetCam images might struggle to perform well. This lies within the inherent weaknesses of the RetCam images. The low contrast between blood vessels and retina, variable illumination in the image due to wide-field view, relatively low image resolution, the presence of choroidal blood vessels interfering with the detection of the hallmark ROP vessels in RetCam ROP images, and the inherent presence of noise may impede the accurate localization of ROP pathologies using

simple classification networks [21]. Image segmentation allows for more accurate and easy scrutiny of blood vessel morphology and retinal abnormalities [22]. Segmentation is a vital component in the CAS for ROP screening because separating the lesions from the background allows for more precise ROP differentiating and classifying practices.

Research is currently underway to improve the accuracy of ROP image segmentation using various architectures, functions, and methods. The ultimate goal is to improve the analytical ability and accuracy of CADx systems for ROP diagnosis and classification. The current work takes a first step towards that objective by examining two widely used and credible segmentation CNNs (U-Net and SegNet) that have been tailored to our ROP image dataset. We aim to accurately segregate the BV and optic disc areas in ROP RetCam photos. To determine how CNN performs better on this task, we trained both networks with different activation units and loss functions and compared the results against segmentation labels manually produced by experienced ROP specialists.

## Methods

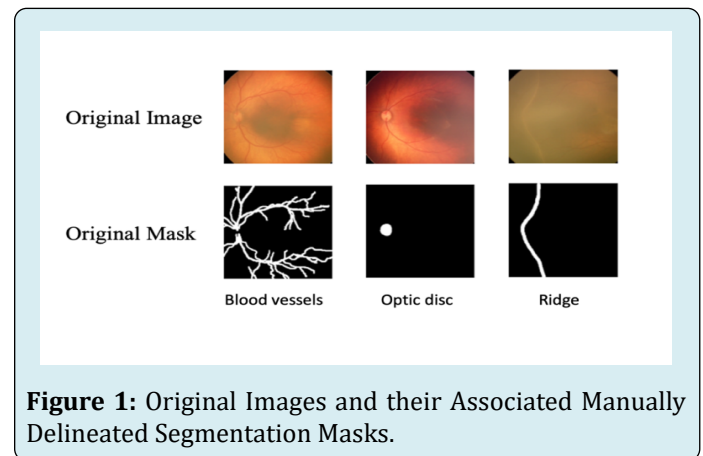
We attempted to conduct the study in accordance with Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis Standards [23]. As of the time of writing, specific reporting guidelines for AI were still being developed, and none were available [24]. To correctly present the key words and results, we referred to the recently released Consolidated Standards of Reporting Trials-AI extension (developed for clinical trial reporting using AI) [25].

## Dataset

We retrieved three sets of 50 images and their corresponding masks from the HVDROPDB database [26], the only publicly available ROP database designed for segmentation, to create data subsets for training three distinct ROP segmentation tasks: BV segmentation, OD segmentation, and ridge/demarcation line segmentation. The HYVDROPDB-BV, HYVDROPDB-OD, and HVDROPDB-RIDGE RetCam image datasets [26] were used to train the segmentation networks separately.

The images are  $256 \times 256$  pixels, captured directly from an 8-bit video signal (256 grayscale levels), and saved in TIFF format. Trained optometrists acquired the photos utilising two Neo or RetCam cameras with a  $120^\circ$  field of view. A team of ROP experts with at least 5 years of experience annotated them, with supervision from a senior ROP expert with 25 years of experience. Prior to annotation, an interobserver variability test was conducted (Kappa

=0.92). However, the possibility of subjective bias cannot be discounted because no external expert was involved in the annotation. To save training time, the images were resized to  $224 \times 224$  pixels during pre-processing. We didn't employ contrast enhancement or speckle reduction during pre-processing since we aimed to measure the networks' ability to learn image features as independently and authentically as possible. We also avoided the practice of re-annotation since this approach is prone to bias in the absence of external annotators and blinding. Figure 1 shows examples of an image-mask pair.



**Figure 1:** Original Images and their Associated Manually Delineated Segmentation Masks.

The data was divided into three sets: training, validation, and test, with 70%, 10%, and 20% allocations, respectively. Data augmentation was not employed to minimise computing costs and avoid introducing more noise to the original set. We were merely interested in assessing the performance of different networks in segmenting BV, OD, and demarcation line or ridge, not to develop a generalizable model for deployment. The three sets were normalised using the training set's mean and standard deviation, as the test set should be reasonably representative of the training set; hence, we must assume that the mean and standard deviation are the same.

## Hardware and Software

Python 3.6.8 was employed as the programming language in this work, coupled with the Tensor Flow and Keras libraries for DL. The testing platform was a LINUX-based PC with an Intel Core i7-7740X CPU @ 4.30GHz, 32 GB of RAM, and a GeForce GTX 1080 Ti graphics card running at a clock speed of 33 MHz. Nvidia's driver version was 390.59. The statistical analysis was carried out using MATLAB R2015a.

## U-Net

U-Net, founded in 2015 at the University of Freiburg's Computer Science Department, is a CNN engineered to

provide swift and precise image segmentation in an array of biomedical applications. It features a contracting path for capturing context and a symmetric expanding path for exact localization. The encoder receives the input image and repeatedly applies two unpadded convolutions, followed by an activation unit and a max pooling operation with stride 2 for downsampling. These operations are repeated four times before passing onto the decoder.

The steps throughout the expansive path are similar. It comprises an upsampling of the feature map, a concatenation with the corresponding feature map of the contracting path, and two convolutions followed by activation. The U-Net's merging with the map from the contraction map sets it apart from other networks. The goal of this expanding path is to provide exact localization in conjunction with contextual information from the contracting path. This contributes to predicting a good segmentation map. The last layer generates an output containing the segmentation result.

## SegNet

SegNet is a CNN with a similar general architecture to U-Net that was created for multi-class pixel-wise segmentation by members of the University of Cambridge's Computer Vision and Robotics Group. The network is based on the layout of the VGG-16 network and has been adapted to pixel-wise segmentation. It is capable of achieving excellent scores while using less memory.

The number of convolutions differs from the U-Net. In the first two steps, there are two convolutions followed by batch normalisation and activation before the pooling process, and three convolutions follow in the next three steps. The decoder portion mirrors the encoder, with the only distinctions being upsampling instead of downsampling and a  $1 \times 1$  last convolution for pixel-wise categorization. SegNet's primary innovation is the reuse of pooling indices. Instead of transferring the full map like U-Net, the decoder upsamples its input feature maps using the encoder's max pooling indices, which are memorised. Reusing memorised indices requires less memory from the system.

## Network Training, Validation, Testing and Results Extraction

During the segmentation model construction process, we trained the U-Net and SegNet networks using three RetCam image data subsets (HYVDROPDB-BV, HYVDROPDB-OD, and HVDROPDB-RIDGE). We made adjustments to the kernel size ( $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$ ) and learning rate (10-3, 10-5, and 10-6), as well as the functions between ReLU and PReLU for activation, and between dice and cross-entropy for loss. In their original architectures, both networks employ ReLU and

cross-entropy; however, subsequent studies recommended the usage of PReLU for boosting accuracy [27] and dice loss as the degree of pixel imbalance grows [28]. We also used 1,000 epochs in U-Net and 2,000 in SegNet because SegNet took much longer to stabilize.

We used 35 photos for training and 5 images for validation, based on our initial dataset split. While the model attempts to fit the training data by minimising the loss function, the validation set refines the parameters. In each training epoch, the model is saved if its validation score is higher than the previous one, and the training data is shuffled to remove any bias in the presentation order. Following that, we assess each segmentation model on a test set of 10 photos.

Each experiment aims to produce a model that performs better on the test set than the previous one. We experimented with different combinations of the batch size, kernel size, dropout rate, and learning rate, until we found the ideal configuration for each network using its original architecture. Then we compared the outcomes of different loss and activation functions.

The segmentation evaluation metrics indicate the correctly segmented image pixels relative to the ground truth. We calculated a few well-known evaluation metrics, including accuracy, precision, recall, and dice score. True positives (TP) are pixels that have been successfully recognised as foreground (white) pixels, but false positives (FP) are the ones detected incorrectly. True negative (TN) are pixels correctly segmented as background (black) pixels, but false negative pixels (FN) are the ones wrongly segmented. Since a class imbalance issue tends to exist for the segmentation BVs, OD and ridge due to the few foreground pixels compared to the background pixels, even 90% accuracy will struggle to accurately label the class. As a result, the dice coefficient (F1-score) is the most often used measure for colour image segmentation, calculating the similarity between ground truth images (manually segmented images) and automatically segmented images. The dice coefficient, also known as the overlap index, is calculated using the equation  $2xTP/(2xTP+FP+FN)$  and runs from 0 to 1, with 0 indicating no overlap and 1 indicating complete overlap between predicted and ground truth.

## Results

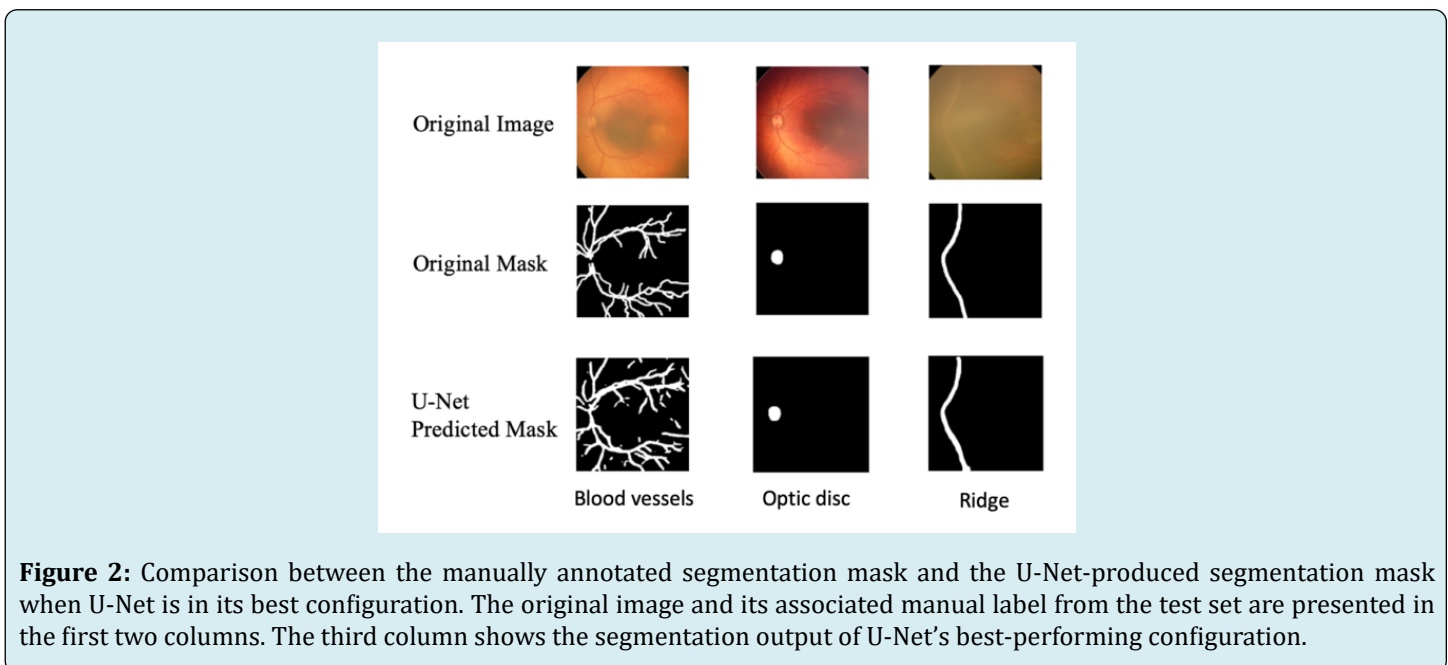
When assessing the performance of the U-Net and SegNet on the HVDROPDB RetCam image subsets, the configurations for the U-Net that generally yielded the best results were found to have a learning rate of 10-5, a  $3 \times 3$  kernel size, a batch size of 8, and a 50% dropout rate. Average dice scores for the testing output images and the training time for each activation-loss duo in the U-Net architecture are presented

in (Table 1 & Figure 2) shows the comparisons between the manually delineated segmentation masks and the U-Net predicted segmentation masks for BV, OD and ridge segmentation using U-Net's best performing configuration. On the other hand, SegNet performed best with the following settings:  $5 \times 5$  kernel size, 10-6 learning rate, 50% dropout

rate, and 32 batch size. (Table 2) shows the average dice score in the training set, as well as training duration in the SegNet architecture. (Figure 3) illustrates the comparisons between the manually delineated segmentation masks and the SegNet predicted segmentation masks for BV, OD and ridge segmentation using SegNet's best performing configuration.

Experiment	Activation Unit	Loss function	Average dice score	Training time (h)
1	PReLU	Cross-entropy	0.741±0.131	6.78
2	PReLU	Dice	0.742±0.133	6.78
3	ReLU	Cross-entropy	0.744±0.137	4.34
4	ReLU	Dice	0.751±0.134	4.34

**Table 1:** U-Net experimental results. Each row considers an experiment using a different pair of activation and loss functions, and the fourth and fifth columns show, respectively, the average dice score, with standard deviation, in the test set and the time to train the model in hours.



**Figure 2:** Comparison between the manually annotated segmentation mask and the U-Net-produced segmentation mask when U-Net is in its best configuration. The original image and its associated manual label from the test set are presented in the first two columns. The third column shows the segmentation output of U-Net's best-performing configuration.

Experiment	Activation Unit	Loss function	Average dice score	Training time (h)
5	PReLU	Cross-entropy	0.681±0.154	10.12
6	PReLU	Dice	0.690±0.140	10.12
7	ReLU	Cross-entropy	0.681±0.150	7.42
8	ReLU	Dice	0.690±0.145	7.42

**Table 2:** Seg-Net experimental results. Each row considers an experiment using a different pair of activation and loss functions, and the fourth and fifth columns show, respectively, the average dice score, with standard deviation, in the test set and the time to train the model in hours.

### Retinal Vessel Segmentation

(Table 3) shows the performance metrics of U-Net and SegNet-based segmentation on the HVDRPDB-BV dataset, using the top-performing setups. U-Net had an accuracy of 0.8751, precision of 0.64, recall of 0.51, and a dice score of 0.49. SegNet's segmentation yielded an accuracy of 0.80, precision of 0.56, recall of 0.45, and dice score of 0.40.

	Accuracy	Precision	Recall	Dice score
U-Net	0.8751	0.64	0.51	0.49
Seg-Net	0.8	0.56	0.45	0.4

**Table 3:** U-Net and Seg-Net performance on blood vessel segmentation in retinopathy of prematurity.

### Optic Disc Segmentation

(Table 4) depicts the comparison of performance parameters of U-Net and SegNet segmentation on the HVDRPDB-OD dataset, best-performing configurations. U-Net achieved an accuracy of 0.933, precision of 0.788, recall of 0.55, and a dice score of 0.648. SegNet's segmentation obtained an accuracy of 0.933, precision of 0.788, recall of 0.55, and dice score of 0.648.

	Accuracy	Precision	Recall	Dice score
U-Net	0.933	0.788	0.55	0.648
Seg-Net	0.933	0.788	0.55	0.648

**Table 4:** U-Net and Seg-Net performance on optic disc segmentation in retinopathy of prematurity.

### Demarcation line/ Ridge Segmentation

Table 5 shows the performance of U-Net and SegNet segmentation on the HVDRPDB-RIDGE dataset, using their top-performing setups. U-Net had an accuracy of 0.9, precision of 0.71, recall of 0.52, and a dice score of 0.5. SegNet's segmentation yielded an accuracy of 0.85, precision of 0.67, recall of 0.47, and dice score of 0.45.

	Accuracy	Precision	Recall	Dice score
U-Net	0.9	0.71	0.52	0.5
Seg-Net	0.85	0.67	0.47	0.45

**Table 5:** U-Net and Seg-Net performance on demarcation line/ ridge segmentation in retinopathy of prematurity.

### Discussion

To automatically detect plus diseases, zones, and stages for ROP diagnosis and classification in AI systems, retinal

structures (BV, OD, and demarcation line/ridge) ought to be segmented. Automated segmentations might also be useful in explaining the ROP diagnosis to ophthalmologists. In our study comparing the performance of a deep U-Net with a SegNet segmentation network, we found that both U-Net and SegNet were capable of accurately segmenting retinal structures of interest for ROP diagnosis and staging, although our deep U-Net performed better overall in ROP segmentation tasks.

It was discovered that utilising PReLU as an activation unit significantly slowed down the training stage for both networks, without providing a meaningful advantage, i.e., enhancing segmentation accuracy. This is especially evident when analysing U-Net performance, which increased training time by about 56.2%. While transferring feature maps from encoder to decoder was intended to slow down U-Net training, slowing the learning rate and utilizing a heavier filter in SegNet models resulted in a greater training time. We highlight that the adoption of various loss functions did not affect training times. When analysing test time, all models

were able to segment the test images in less than a second.

When comparing label accuracy, U-Net outperformed SegNet in nearly all configurations, with the slightest difference being 7.39% in dice scoring. The great majority of the poorest segmentation outcomes, such as the lowest dice coefficient values, occurred in photos with minimal contrast between the lesion and the background, as well as in images that featured various dark areas. SegNet tends to overlook sharp edges, resulting in rounded segmentation maps. Analyzing each model separately while looking at the dice scores in Table 1, we can determine that the largest difference across various U-Net model versions is as low as 0.93%, whereas the difference in SegNet is 1.30%, as illustrated in Table 2.

Our results were compared to other state-of-the-art segmentation models, however, we were unable to acquire the highest scores in BV, OD, and ridge segmentations. According to Table 6, Gojic G, et al. [29] obtained accuracies of 96.6-100% using 1156 fundus images to train a segmentation model to segment ODs in ROP, while Nisha KL,

et al. [30] achieved a 0.723 dice coefficient and an accuracy of 97.5% using 178 fundus images to train a machine learning model to segment BVs. Raja S, et al. [4] used a MultiResUNet on a dataset of 4000 fundus images for BV segmentation, yielding a dice score of 0.824 and an accuracy of 93.4%. The comparatively low segmentation performance of our models might be attributed to input image size limits, with only 50 ROP images available for training in each batch. Future efforts to increase the size of the training ROP image set may improve our models' segmentation ability.

Clinically, our findings could serve as the foundation for future research into the development of AI-aided systems for ROP screening and staging, where the segmentation of BV, optic disc, and demarcation line/ridge is critical to describe ROP characteristics in these systems for diagnosis, severity triaging, monitoring, and individualized management plans and follow-ups [31]. These AI systems could address existing roadblocks in ROP diagnosis and treatment, such as a paucity of ROP experts and subjective variation in the clinical identification of zones, stages, and plus attributes critical to ROP diagnostics [31].

Study	Segmentation Task	Dataset	Architecture	Accuracy	Precision	Recall	Dice Score
Gojic G, et. al. [28]	BV segmentation	STARE, FIRE, DRIVE, HRF, IDRiD and AIIMS datasets (total 1156 fundus images)	YOLO model	OD segmentation: 96.6%-100%	NA	NA	NA
Nisha KL, et. al. [29]	BV segmentation	178 fundus images	Machine learning model	BV segmentation: 97.473%	NA	NA	BV segmentation: 0.72298
Raja S, et. al. [4]	BV segmentation	4000 fundus images	Modified MultiResUnet	BV segmentation: 93.37%	NA	NA	BV segmentation: 0.824

**Table 6:** Performance of other segmentation models designed for ROP segmentation tasks.

## Limitations

However, the input images presented a possible constraint for our system. There were only 50 ROP images per subgroup, which might have limited the segmentation models' capacity to reach their full performance potential. We only employed high-quality retinal images from H.V. Desai Eye Hospital in Pune, India, taken with RetCam on preterm newborns with shorter gestation periods and lower weights. Since the RetCam imaging equipment is costly, many hospitals could opt for less expensive systems such

as PanoCam (Visunex Medical Systems, Suzhou, Jiangsu, China), and image variations may occur as a result of device differences [32]. Second, during newborn screening, ROP was discovered in many heavier and full-term infants [32]. As a result, just extracting features from premature newborn photos may not be adequate. Third, the photographs chosen are of satisfactory quality and may not accurately portray reality. We may lose features in inferior images. Finally, there are concerns about the quality of the prepared ground facts (i.e. segmentation masks created by ROP specialists) because the specifics of the annotators' qualifications and the process

of consensus for inter-observer variability in annotating were not given. Ground truths used might have carried suboptimal quality and questionable reliability.

### Future Work

In future ROP investigations, it is critical to urge for more collaboration across regions and hospitals to gather more retinal images of preterm newborns using various equipment. Instead of merely images from preterm newborns, data sets should be expanded to also include full-term and heavier newborn screening images to extract a boarder spectrum of features for ROP diagnosis. We will also look into more stringent, quantitative image screening criteria and develop preprocessing modules to provide image screening automatically and objectively. It would also be interesting to dissect and visualize the features learned by segmentation networks in future studies.

### Conclusion

We successfully built two convolutional neural networks to segment BV, OD, and ridge in RetCam ROP images, with good accuracies. With our recommended designs, we discovered that the U-Net excels over the SegNet for our ROP RetCam datasets' segmentation. The SegNet takes longer to train and exhibited a lower dice score, with the risk of rounding edges and jeopardising a subsequent diagnosis. Though U-Net is a popular model for medical image segmentation, it has not received the best performance in our BV, OD and ridge segmentation for ROP images. This is mostly owing to the properties of our RetCam image sets. To obtain more accurate segmentation, refinements to the ROP RetCam datasets' size and diversity would be needed.

As stated in the introduction, few studies have been presented on the segmentation of BV, OD and ridge found in ROP RetCam images using U-Net and/or SegNet, and to the best of our knowledge, no comparison of the performance of those networks with different loss functions and activation units has been conducted. Our current work aims to analyse the capabilities of two well-established segmentation networks based on their configurations for the three critical segmentation tasks in ROP diagnostics and staging, with the goal of providing a reference for future researchers interested in developing segmentation-based ROP diagnostic systems. In fact, in the near future, it will be valuable to compare several segmentation approaches and networks, such as GAN and attention-gated U-Net, using the same ROP dataset to determine the best segmentation network for accurate ROP segmentation.

### References

1. Sen P, Jain S, Bhende P (2018) Stage 5 retinopathy of prematurity: An update. *Taiwan J Ophthalmol* 8(4): 205-215.
2. Krista H (2024) Retinopathy of Prematurity - EyeWiki.
3. Agrawal R, Kulkarni S, Walambe R, Kotecha K (2021) Assistive Framework for Automatic Detection of All the Zones in Retinopathy of Prematurity Using Deep Learning. *J Digit Imaging* 34(4): 932-947.
4. Raja SVM, Snehalatha U, Chandrasekaran A, Baskaran P (2023) Automated diagnosis of Retinopathy of prematurity from retinal images of preterm infants using hybrid deep learning techniques. *Biomed Signal Process Control*, pp: 104883.
5. Ju RH, Zhang JQ, Ke XY, Lu XH, Liang LF, et al. (2013) Spontaneous regression of retinopathy of prematurity: incidence and predictive factors. *Int J Ophthalmol* 6(4): 475-480.
6. Charles JB, Ganthier RJr, Appiah AP (1991) Incidence and characteristics of retinopathy of prematurity in a low-income inner-city population. *Ophthalmology* 98(1): 14-17.
7. Singh H, Kaur R, Gangadharan A, Pandey AK, Kumar P, et al. (2020) Neo-Bedside Monitoring Device for Integrated Neonatal Intensive Care Unit (iNICU). *IEEE Access* 3(1): 21-30.
8. Jefferies A (2010) Retinopathy of prematurity: Recommendations for screening. *Paediatr Child Health* 15(10): 667-674.
9. Badarinath D, Chaitra S, Bharill N, Ningappa A (2018) Study of Clinical Staging and Classification of Retinal Images for Retinopathy of Prematurity (ROP) Screening. *International Joint Conferences on Neural Network*, pp: 1-6.
10. Razzak MI, Naz S, Zaib A (2017) Deep Learning for Medical Image Processing: Overview, Challenges and Future.
11. Sivakumar R, Eldho M, Jiji CV, Vinekar A, John R, (2024) Computer aided screening of retinopathy of prematurity-A multiscale Gabor filter approach, pp: 259-264.
12. Rani P, Rajkumar ER (2016) Classification of retinopathy of prematurity using back propagation neural network. *Int J Biomed Eng Technol. Inderscience Publishers* 22(4): 338-348.



13. Greenspan H, Ginneken BV, Summers RM (2016) Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Trans Med Imaging* 35(5): 1153-1159.
14. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42: 60-88.
15. Geethalakshmi K (2018) A survey on deep learning approaches in retinal vessel segmentation for disease identification. *IOSR Journal of Engineering (IOSRJEN)*, pp: 47-52.
16. Worrall DE, Wilson CM, Brostow GJ (2016) Automated Retinopathy of Prematurity Case Detection with Convolutional Neural Networks. Springer International Publishing, pp: 68-76.
17. Hu J, Chen Y, Zhong J, Ju R, Yi Z (2019) Automated Analysis for Retinopathy of Prematurity by Deep Neural Networks. *IEEE Trans Med Imaging* 38(1): 269-279.
18. Wang J, Ju R, Chen Y, Zhang L, Hu J, et al. (2018) Automated retinopathy of prematurity screening using deep neural networks. *EBioMedicine* 35: 361-368.
19. Zhang Y, Wang L, Wu Z, Zeng J, Chen Y, et al. (2019) Development of an Automated Screening System for Retinopathy of Prematurity Using a Deep Neural Network for Wide-Angle Retinal Images. *IEEE Access* 7: 10232-10241.
20. Zhao J, Lei B, Wu Z, Zhang Y, Li Y, et al. (2019) A Deep Learning Framework for Identifying Zone I in RetCam Images. *IEEE Access* 7: 103530-103537.
21. Agrawal R, Agrawal MA, Kulkarni S, Kotecha K, Walambe R (2021) Quantitative analysis of research on Artificial Intelligence in Retinopathy of Prematurity. *Library Philosophy and Practice*, pp: 5342
22. Liu R, Pu W, Nan H, Zou Y (2023) Retina image segmentation using the three-path Unet model. *Sci Rep* 13(1): 22579.
23. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 350: g7594.
24. Collins GS, Moons KGM (2019) Reporting of artificial intelligence prediction models. *The Lancet* 393(10181): 1577-1579.
25. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK (2020) Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 26(9): 1364-1374.
26. Agrawal R, Walambe R, Kotecha K, Gaikwad A, Deshpande CM, et al. (2023) HVDROPDB datasets for research in retinopathy of prematurity. *Data Brief* 52: 109839.
27. He K, Zhang X, Ren S, Sun J (2015) Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification.
28. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M (2017) Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support* 2017: 240-248.
29. Gojic G, Petrovic V, Turovic R, Dragan D, Oros A, et al. (2023) Deep Learning Methods for Retinal Blood Vessel Segmentation: Evaluation on Images with Retinopathy of Prematurity. *ArXiv* pp: 131-136.
30. Nisha KL, Sreelekha G, Sathidevi PS, Mohanachandran P, Vinekar A (2019) A computer-aided diagnosis system for plus disease in retinopathy of prematurity with structure adaptive segmentation and vessel based features. *Comput Med Imaging Graph* 74: 72-94.
31. Agrawal R, Kulkarni S, Walambe R, Deshpande M, Kotecha K (2022) Deep dive in retinal fundus image segmentation using deep learning for retinopathy of prematurity. *Multimed Tools Appl* 81(8): 11441-11460.
32. Li P, Liu J (2022) Early Diagnosis and Quantitative Analysis of Stages in Retinopathy of Prematurity Based on Deep Convolutional Neural Networks. *Transl Vis Sci Technol* 11(5): 17.