

# Early Diagnosis and Quantitative Analysis of Stages in Retinopathy of Prematurity Based on Deep Convolutional Neural Networks

Peng Li<sup>1,2</sup> and Jia Liu<sup>3</sup>

<sup>1</sup> School of Electronic and Information Engineering, Tongji University, Shanghai, China

<sup>2</sup> Department of Electronic and Information Engineering, Tongji Zhejiang College, Jiaying, China

<sup>3</sup> Optometry Center, Jiaying Maternity and Child Health Care Hospital, Jiaying, China

**Correspondence:** Peng Li, 71-4, Green Brook Rose Garden, Nanhu District, Jiaying, Zhejiang Province 314005, China.  
e-mail: [airrob@163.com](mailto:airrob@163.com)

**Received:** November 6, 2021

**Accepted:** April 22, 2022

**Published:** May 17, 2022

**Keywords:** deep convolutional neural networks; retinopathy of prematurity; quantitative analysis; early detection of disease stages

**Citation:** Li P, Liu J. Early diagnosis and quantitative analysis of stages in retinopathy of prematurity based on deep convolutional neural networks. *Transl Vis Sci Technol.* 2022;11(5):17, <https://doi.org/10.1167/tvst.11.5.17>

**Purpose:** Retinopathy of prematurity (ROP) is a leading cause of childhood blindness. An accurate and timely diagnosis of the early stages of ROP allows ophthalmologists to recommend appropriate treatment while blindness is still preventable. The purpose of this study was to develop an automatic deep convolutional neural network-based system that provided a diagnosis of stage I to III ROP with feature parameters.

**Methods:** We developed three data sets containing 18,827 retinal images of preterm infants. These retinal images were obtained from the ophthalmology department of Jiaying Maternal and Child Health Hospital in China. After segmenting images, we calculated the region of interest (ROI). We trained our system based on segmented ROI images from the training data set, tested the performance of the classifier on the test data set, and evaluated the widths of the demarcation lines or ridges extracted by the system, as well as the ratios of vascular proliferation within the ROI on a comparison data set.

**Results:** The trained network achieved a sensitivity of 90.21% with 97.67% specificity for the diagnosis of stage I ROP, 92.75% sensitivity with 98.74% specificity for stage II ROP, and 91.84% sensitivity with 99.29% sensitivity for stage III ROP. When the system diagnosed normal images, the sensitivity and specificity reached 95.93% and 96.41%, respectively. The widths (in pixels) of the demarcation lines or ridges for normal, stage I, stage II, and stage III were  $15.22 \pm 1.06$ ,  $26.35 \pm 1.36$ , and  $30.75 \pm 1.55$ . The ratios of the vascular proliferation within the ROI were  $1.40 \pm 0.29$ ,  $1.54 \pm 0.26$ , and  $1.81 \pm 0.33$ . All parameters were statistically different among the groups. When physicians integrated quantitative parameters of the extracted features with their clinic diagnosis, the  $\kappa$  score was significantly improved.

**Conclusions:** Our system achieved a high accuracy of diagnosis for stage I to III ROP. It used the quantitative analysis of the extracted features to assist physicians in providing classification decisions.

**Translational Relevance:** The high performance of the system suggests potential applications in ancillary diagnosis of the early stages of ROP.

## Introduction

Retinopathy of prematurity (ROP) is a vasoproliferative disorder that occurs in premature infants with lighter weights and shorter gestation periods.<sup>1</sup> This disease is a leading cause of childhood blindness. As the survival rate of preterm infants is increas-

ing, the number of children with ROP is also increasing.<sup>2</sup> In the 1980s, the International Classification of Retinopathy of Prematurity was developed,<sup>3</sup> which was revised in 2005.<sup>4</sup> In 2021, the third edition was published.<sup>5</sup> According to the guide, the diagnosis of ROP involves three dimensions: stages I to V, zones I to III, and the presence of pre-plus disease or plus disease.

It is important to diagnose ROP accurately and timely, by clinical fundus examination or by reading retinal images. However, since classification guidelines provide only qualitative indications, this leads to a diagnostic result that depends mainly on the ophthalmologists' subjective decisions.<sup>6</sup> In addition, diagnostic differences also exist when different experts use different hardware in different regions. All of these lead to inconsistent diagnostic results for ROP.

To address this problem, many experts have developed semiautomated quantitative analysis tools to diagnose ROP more objectively. Their results include ROP Tool,<sup>7</sup> principal spanning forests algorithms,<sup>8</sup> computer-aided retinal image analysis,<sup>9</sup> and so on. However, these methods were not completely automatic, and humans needed to determine features and cut points. In general, the output did not correlate well enough with clinical diagnoses to be widely used.<sup>10</sup>

Deep convolutional neural networks (DCNNs)<sup>11</sup> have shown great advantages in many medical image applications.<sup>12–15</sup> DCNNs provide a fully automated, end-to-end solution and do not need manual input, which is a huge advantage.

Plus disease, which has been studied by many experts, is an important feature in determining the need for treatment for ROP. In 2016, Worrall et al.<sup>16</sup> began to apply DCNNs in the diagnosis of Plus disease for premature infants. Brown et al.<sup>17</sup> studied a completely automatic system, which was able to classify retinal images as normal, pre-plus disease, and plus disease with great accuracy.

In another direction, the diagnosis of early stage ROP has also been researched. This is not only because the diagnosis of stages relies mainly on subjective interpretations,<sup>6</sup> but the diagnoses between stages I and III are also crucial,<sup>4</sup> which allow doctors to recommend the appropriate treatment while blindness is still preventable. In contrast, patients with stages IV to V have already had irreversible damage to the retinas.

In 2018, Hu et al.<sup>18</sup> applied DCNNs in the diagnosis of stage I to III ROP. Mulay et al.<sup>19</sup> and Ding et al.<sup>20</sup> diagnosed stages of ROP using segmented images based on DCNNs. This is not only because stages I to III and normal retinas are more subtly classified by the existence, size, and shape of the demarcation line or ridge as well as vascular proliferation, but also these features are well fitted to be obtained by segmenting using DCNNs.

DCNNs have been found to have improved performance in medical image fields.<sup>21</sup> However, they also have limitations in that the features on which DCNNs rely are not transparent or explainable.<sup>22</sup> The limitation presents great challenges for the adoption of DCNNs because medical accountability is important and may

lead to serious legal consequences. An ideal system should be able to provide not only objective results but also the reasons behind them. Many experts tried to have more explainable DCNNs by combining DCNNs with traditional feature extractions. Similar work had been done by Mao et al.<sup>23</sup> and Yildiz et al.<sup>24</sup> in the diagnosis of plus disease.

In this study, we developed an automated DCNN-based system. Using segmented images, we trained a classifier to categorize images into four categories: normal, stage I ROP, stage II ROP, and stage III ROP. By evaluating the feature parameters extracted by our system, we showed significant differences among different categories. In addition, we showed the role of these parameters in improving the consistency of the manual diagnosis. To the best of our knowledge, this was the first attempt to quantitatively analyze the segmented features for diagnosis of early stage ROP.

## Methods

The study was approved by the Ethics Review Committee of Jiaying Maternal and Child Health Hospital, China, and followed the principles of the Declaration of Helsinki.

## Data

All images of premature infants were collected from January 2018 to December 2020 at the Ophthalmology Department of Jiaying Maternal and Child Health Hospital by the Retcam3 (Natsu Medical, Inc., Pleasanton, CA, USA) imaging system. These images were collected from five standard fields of view (posterior, nasal, temporal, superior, and inferior) and were 1600 × 1200 in size (in pixels).

After abandoning low-quality images, we selected 18,827 retinal images collected from preterm infants with a gestational age of fewer than 37 weeks and a weight less than 2000 g. We invited several experts (more than three) to remove low-quality images by consensus selections according to the following criteria:

1. Less than 25% of the peripheral area of the retina is unobservable due to artifacts, including the presence of foreign bodies, out-of-focus imaging, blurring, and extreme light conditions.<sup>25</sup>
2. Insufficient focus of the images with blood vessels is the reference.

We constructed three data sets: a training data set to train DCNNs, a test data set to test the

**Table 1.** Characteristics of the Three Data Sets

Characteristic	Training Data Set			Test Data Set			Comparison Data Set		
	Patients	Eyes	Images	Patients	Eyes	Images	Patients	Eyes	Images
Gestation (weeks)	31.31 ± 5.42			30.43 ± 5.80			30.82 ± 5.61		
Birthweights (g)	1542.01 ± 435.02			1442.03 ± 517.03			1512 ± 467.00		
Normal	1785	2203	11,544	447	552	2893	—		
Stage I	231	286	1499	58	72	378	—		
Stage II	157	194	1017	40	50	262	—		
Stage III	87	108	566	22	28	147	—		
Total	2260	2791	14,626	567	702	3680	73	97	521

performance of the network, and a comparison data set to compare DCNN predictions with manual diagnosis. We assigned a reference diagnostic criterion (stage I ROP, stage II ROP, stage III ROP, or normal) to each image in training and test data sets. The reference diagnostic criteria were determined by the consensus diagnosis provided by three ROP experts and compared to the previous clinical diagnosis.

Table 1 describes the characteristics of the three data sets, with 14,626, 3680, and 521 retinal images originating from 2260, 567, and 73 different preterm infants, respectively. Multiple images of different standard view fields were acquired for each eye, which led to a significant increase in the number of images.

To train the vessel segmentation network, we selected 1825 (204 infants) retinal images from the training data set and trained and tested the network in a ratio of 1464 (162 infants) to 361 (42 infants). An ophthalmologist annotated the retinal vessels using the brush annotation tool in dedicated annotation software. To train the segmentation network for the demarcation line or ridge, we selected 2738 (306 infants) retinal images from the training data set and trained and tested the network in a ratio of 2196 (243 infants) to 542 (63 infants). An ophthalmologist, using dedicated standard software, drew a boundary polygon around the demarcation line or ridge and labeled the polygon regions.

The test data set consisted of 3680 images (567 infants), of which 2893 were normal (447 infants), 378 were stage I ROP (58 infants), 262 were stage II ROP (40 infants), and 147 were stage III ROP (22 infants). The timing of retinal screening, the time interval of follow-up, and the time points for treatment of threshold ROP and prethreshold ROP were performed in strict accordance with the guidelines. Children were treated as soon as they developed the threshold ROP or prethreshold ROP. Therefore, there were no cases of stage IV and V ROP in this study.

All images in the comparison data set were collected from 73 infants. We extracted feature parameters to analyze the differences among the groups and the correlation between the features and stages. We invited two ophthalmologists with different experience to perform image diagnosis separately, to study the significance of the system based on quantitative analysis in assisting the clinic diagnosis of stages of ROP.

## Network Architectures

The structure of our system is given in Figure 1. Image inputs were segmented by two deep learning networks (Fig. 1A). Then, we calculated the ROI (Fig. 1B) and extracted the feature parameters (Fig. 1C). We trained the classifier (Fig. 1D) with the ROI segmented images to diagnose normal and stage I to III ROP.

## Segmentation

As shown in Figure 1, at the beginning of our system, the input retinal images were segmented by two deep learning networks. One segmented the retinal blood vessels, and the other segmented the demarcation lines or ridge.

We used two open-source Retina U-Nets<sup>26</sup> as the segmentation networks. During the training, we modified the number of convolution kernels between different layers, applied the stochastic gradient optimizer, and set the momentum to 0.9, the learning rate to 0.001, the batch size to 32, and the gradient clipping to 5.0. The Retina U-Nets were initialized from the Gaussian distribution  $N(0, 0.01)$ . Throughout the entire process of segmentation, the size (in pixels) was always maintained at the original size (1600 × 1200). We used whole images instead of slicing or cropping them.

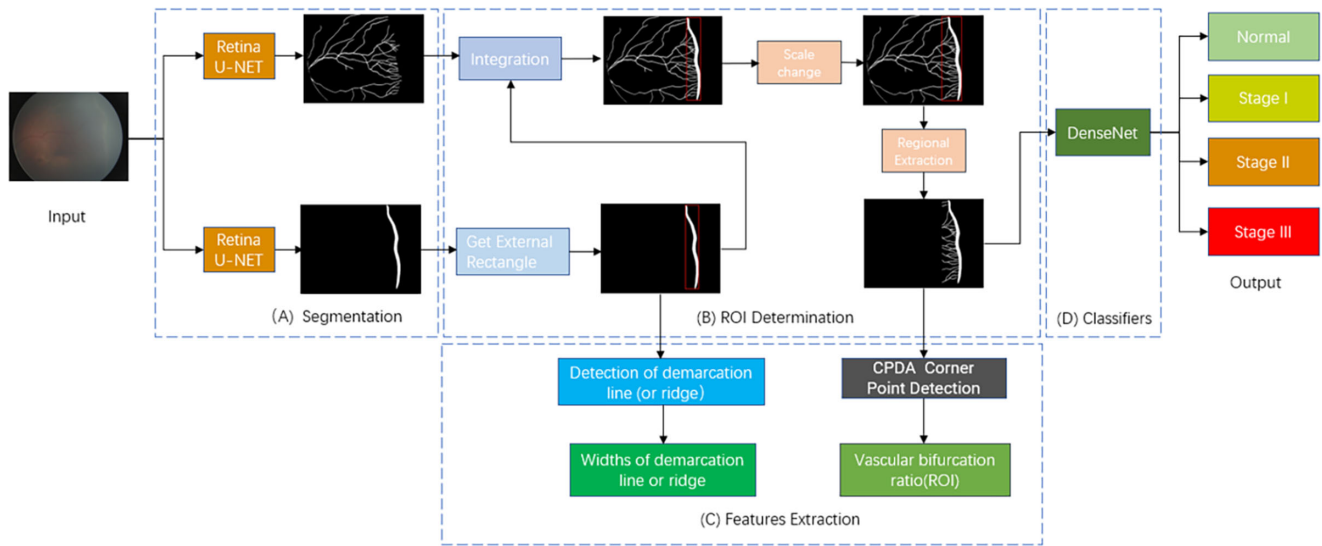


Figure 1. Structure of the proposed system.

### ROI Determination

After binarizing the segmented images using the Otsu algorithm,<sup>27</sup> we used the Canny edge detection algorithm<sup>28</sup> to detect the contours of the demarcation lines or ridge. We also used the Sklansky algorithm<sup>29</sup> to find the convex packets, followed by the rotating caliper algorithm<sup>30</sup> to obtain the minimum outer rectangles. Finally, we used nonmaximum suppression to remove the redundant rectangles.

We used the `cv2.logic_or` function from OpenCV4.5.2 (Open-Source Computer Vision Library) to integrate the segmented vessels with the segmented demarcation lines or ridge. Since they were all 1600 × 1200 in size, we did not need to resize them. We took the lengths and 1.5 times the widths of the minimum external rectangles as parameters to draw a new rectangle. To study the vascular proliferation near the demarcation lines or ridge, we offset the rectangle toward the vascular direction. Finally, we obtained this rectangle region (ROI).

### Feature Extraction

We performed a Euclidean distance transformation on the binarized segmented images. By calculating the distances to the nearest contour pixels, we obtained the widths of the segmented demarcation lines or ridges. We also used the Zhan-Suen fast parallel refinement algorithm<sup>31</sup> to extract centerlines from the segmented vessels. After corner point detection using the Chord-to-Point Distance Accumulation (CPDA) detection method,<sup>32</sup> we obtained all candidate

nodes. We calculated the branching ratios based on the bifurcation points distinguished from candidate points using adaptive rectangular windows. We analyzed only the bifurcation points within the ROI, and the parameters obtained were all averaged.

### Classifier

We used the Dense Net<sup>33</sup> to classify the images into four categories: normal, stage I ROP, stage II ROP, and stage III ROP. Dense Net is an excellent classification network that has several compelling advantages: alleviating the gradient disappearance problem, enhancing feature propagation, encouraging feature reuse, and greatly reducing the number of parameters. We set the output of the final layer to 4 and the batch size to 20. We also used migration learning from the ImageNet<sup>34</sup> data set to initialize the weights of the model. By flipping them horizontally, vertically, and rotating them at six different angles, data augmentation was applied.

### Statistical Analysis

We used the area under the curve (AUC) score under the receiver operating characteristic curve to measure the performance of the classifier during training. To avoid overfitting and underfitting, we divided the training data set into five parts, randomly selected four parts for training, and used the remaining for testing. The cross-validations were repeated five times (fivefold cross-validation) to obtain AUC scores, and 95% confidence intervals were calculated using the formula of Hanley and McNeil.<sup>35</sup> We used the Scikit-Learn



library tools (French Institute for Research in Computer Science and Automation, Rocquencourt, France) to calculate the AUC scores.

On the basis of the AUC scores, we selected the best configuration and conducted performance tests. To measure the performance of the classifier, we calculated the sensitivity and specificity of the results on the test data set.

The guidelines<sup>4</sup> tell us that stage I to III ROP and normal retinas are more subtly classified by the existence, size, and shape of the demarcation line or ridge as well as vascular proliferation. Therefore, our system evaluated the widths of the demarcation line or ridge and the ratios of vessel proliferation within the ROI.

A one-way analysis of variance (ANOVA) was conducted on the extracted feature parameters from images of the comparison data set to obtain the differences between groups of stages I to III. For ANOVA, we performed a  $\chi^2$  test and used S-N-K and Duncan's assumption of equal variance.

Due to the temporal sequential nature of the stages, we conducted ordered logistic regression on the extracted feature parameters. We performed a parallel line test to make sure that the autoregressive coefficients of the independent variables were always constant. We also set 95% confidence intervals, as well as the maximum stepwise quadratic score to 5 and the maximum number of iterations to 100. Finally, we chose the feature parameters of stage III as a reference.

We also invited two ophthalmologists with different experience to perform manual diagnoses on the

comparison data set. We used the results of DCNN predictions and manual diagnosis to calculate  $\kappa$  values to investigate the role of our system in improving the consistency of clinical diagnostic results.

## Experiments

All networks were implemented in Tensorflow1.10 (NVIDIA, Santa Clara, CA, USA) and evaluated on a computer with an NVIDIA GeForce TITAN XP GPU. All statistical analyses were performed using the statistical software SPSS Statistics 26.0 (IBM, Armonk, NY, USA).

## Results

### Automated Segmentation of Blood Vessels and Demarcation Lines or Ridges

The segmentation results of blood vessels and demarcation lines or ridges for four categories are shown in Figure 2. For blood vessels, the sensitivity was 0.82 and specificity was 0.98. To our knowledge, there has been no study on the quantitative analysis of stage I to III ROP using DCNNs. However, previous studies using DCNN methods for segmentation of retinal vessels obtained a sensitivity and specificity of 0.7 to 0.9 and 0.8 to 0.9, respectively.<sup>36</sup> Mao et al.<sup>23</sup> used modified U-Net to segment vessels in retinal images of preterm infants and obtained a sensitivity of 0.72 and a specificity of 0.99. In this work, the sensi-

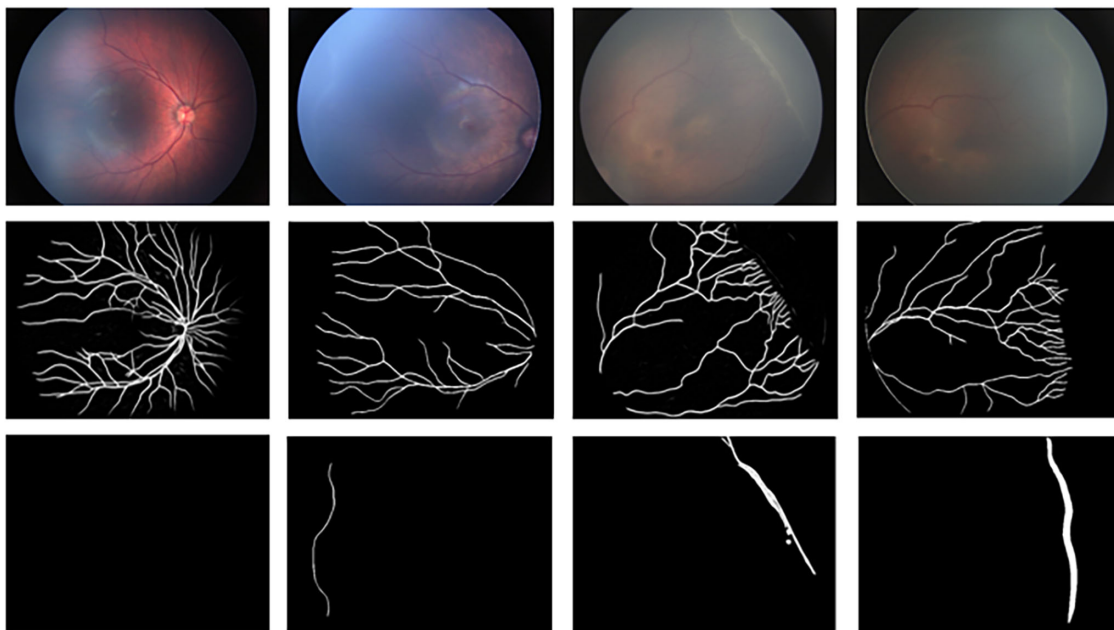


Figure 2. Original images and segmented images.

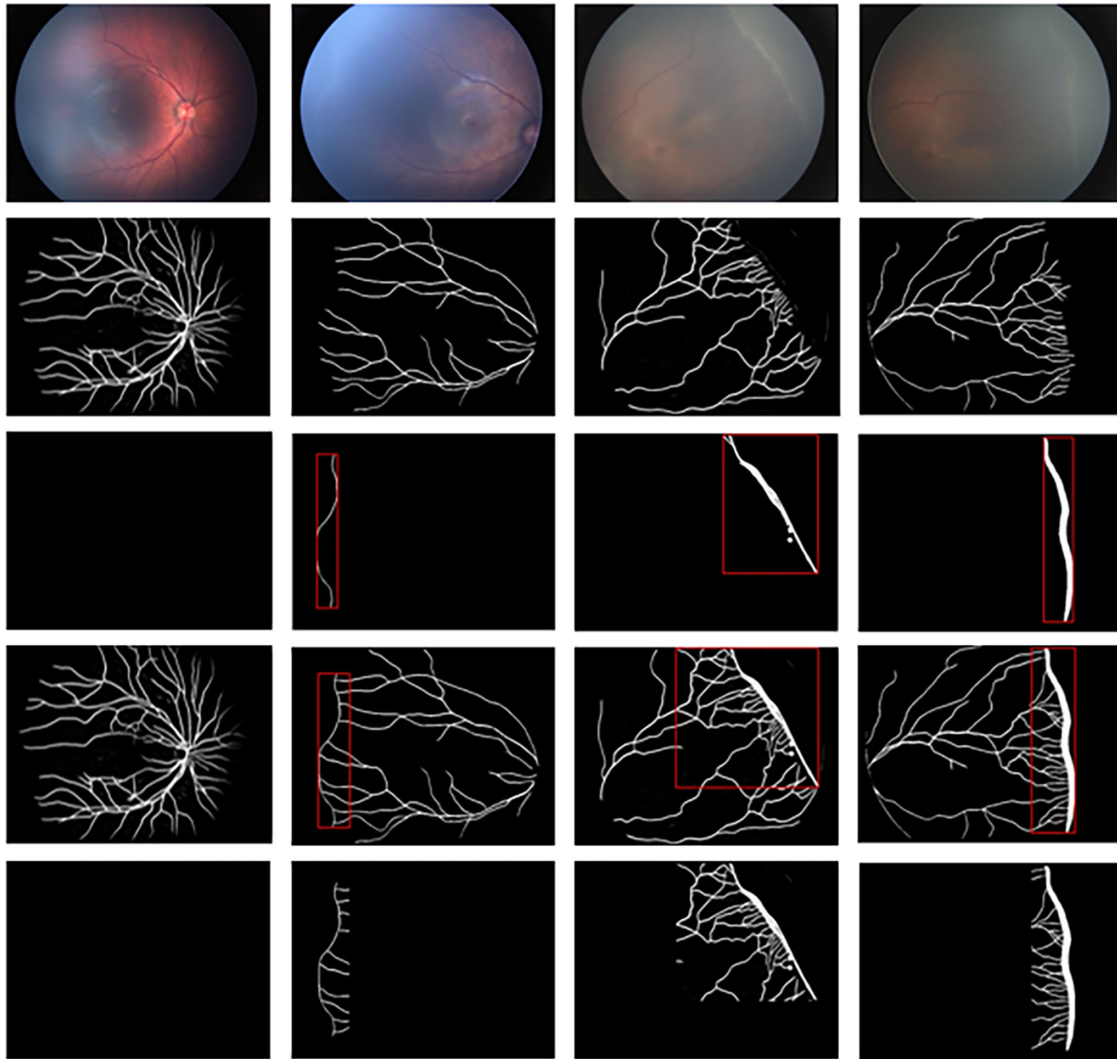


Figure 3. Images in the processing of ROI determination.

tivity of segmentation of the demarcation line or ridge reached 0.93 and the specificity was 0.99.

Figure 2 shows the original images and their segmented images. The images in the first row are the input retinal images, the middle row shows the segmented vessels, and the segmented demarcation lines or ridges are in the last row. Starting from the left column, the images are sorted by different stages (normal, stage I, stage II, and stage III). No demarcation lines or ridges were detected in normal images, so the segmented image in the first column was completely black (background).

### Determination of ROI

As shown in Figure 3, original retinal images are in the first row, segmented images of demarcation lines or ridges with minimum external rectangles (red

rectangles) are in the third row, and the images in the fourth row are the integrated segmented images with red rectangles (ROI). The images in the last row are black (background) except the ROI, and these images are the inputs to the classifier. We also extracted feature parameters from the segmented images to perform statistical analysis, and the relevant data are in the section on statistics regarding extracted features.

### Training and Performance of the Classifier

We applied a convolutional neural network classifier (Dense Net) to diagnose stage I to III ROP, and the training results of the classifier on the training data set are shown in Figure 4. Our classifier reached an average AUC score of 0.9612 for normal (Fig. 4A), 0.9252 for stage I ROP (Fig. 4B), 0.9868 for stage II ROP (Fig. 4C), and 0.9875 for stage III ROP (Fig. 4D).

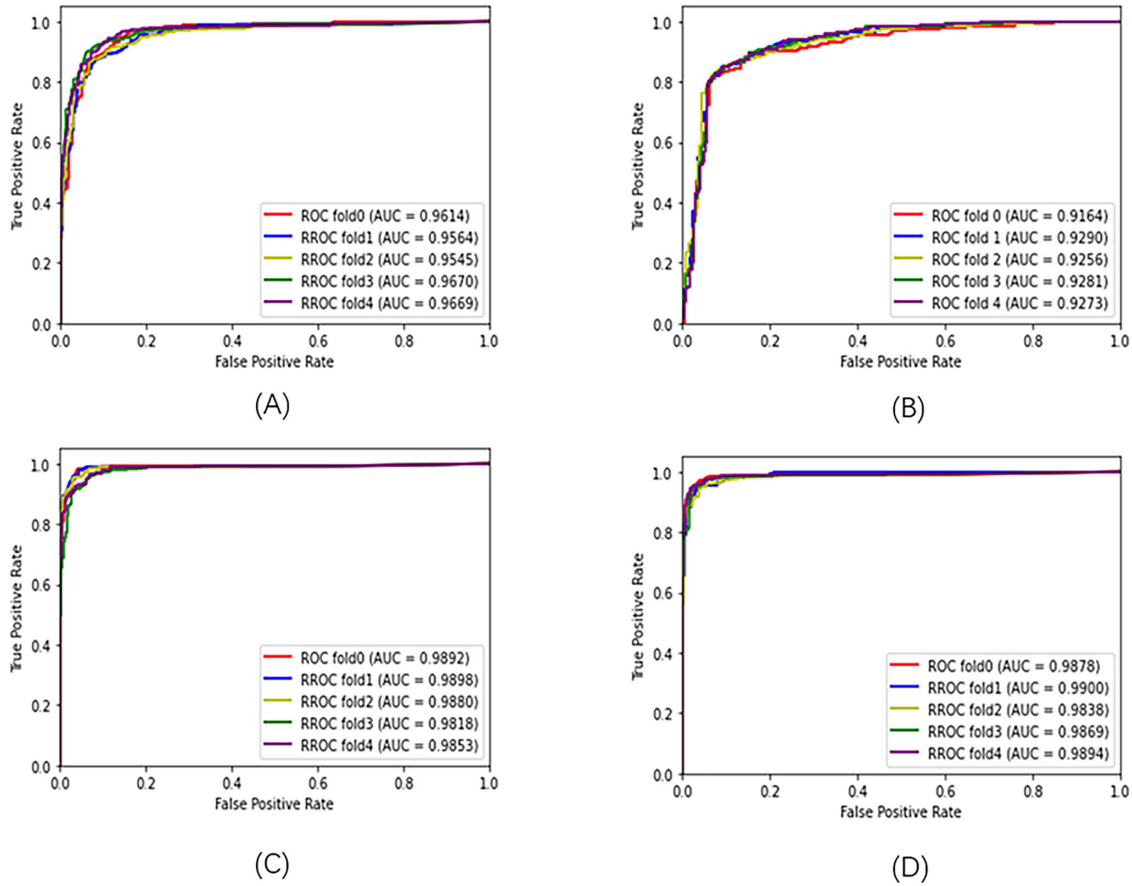


Figure 4. Fivefold cross-validations of the classifier on the training data set.

Table 2. DCNN Predictions and Reference Criteria on the Test Data Set

DCNN Prediction	Reference Criteria				Total
	Normal	Stage I	Stage II	Stage III	
Normal	2789	21	6	5	2821
Stage I	64	341	9	4	418
Stage II	23	12	243	3	281
Stage III	17	4	4	135	160
Total	2893	378	262	147	3680

Based on the results of fivefold cross-validations, we selected the best-performing configuration (blue curve) with an average AUC score of 0.9663 for all different stages.

We tested the performance of the classifier on the test data set, with the results shown in Table 2. The network correctly diagnosed 3508 of 3680 images (95.33%), as shown in Table 2, which achieved a sensitivity of 90.21% and a specificity of 97.67% for the diagnosis of stage I ROP, a sensitivity of 92.75% and a specificity of 98.74% for stage II ROP, and a sensitiv-

ity of 91.84% and a specificity of 99.29% for stage III ROP. When the system diagnosed normal images, the sensitivity and specificity reached 95.93% and 96.41%, respectively.

### Statistics Regarding Extracted Features

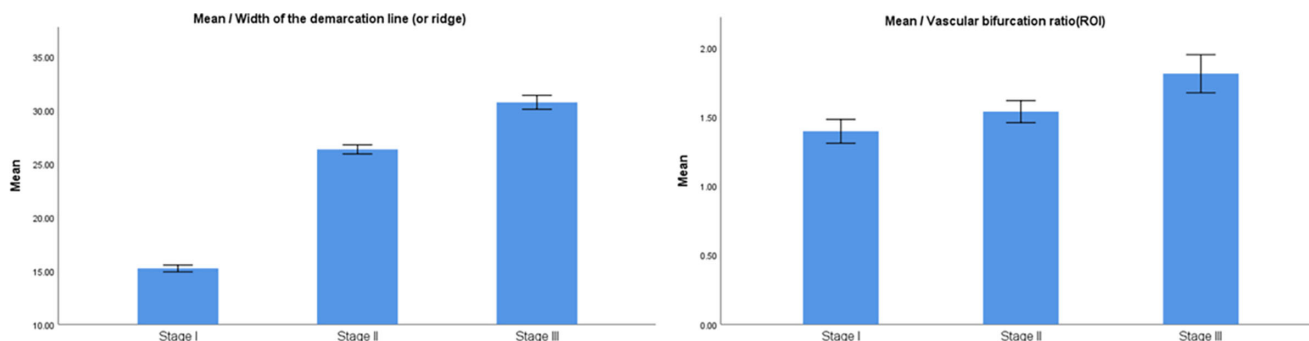
On the basis of the segmented retinal images of the comparison data set, we evaluated the widths of the demarcation line or ridge and the ratios of vessel proliferation within the ROI. We performed a one-way ANOVA on feature data, and the results are shown in Table 3.

The results of the one-way ANOVA test are shown in Table 3 widths (in pixels) of the demarcation line or ridge for stage I, stage II, and stage III ROP were  $15.22 \pm 1.06$ ,  $26.35 \pm 1.36$ , and  $30.75 \pm 1.55$ , respectively, while ratios of the vascular proliferation within the ROI were  $1.40 \pm 0.29$ ,  $1.54 \pm 0.26$ , and  $1.81 \pm 0.33$ , respectively. All values were statistically different among the groups ( $P < 0.001$ , 95% confidence interval). Note that inputs of segmented normal images

**Table 3.** One-Way ANOVA for Feature Parameters

Features	States	<i>n</i>	Mean ± SD	<i>F</i>	Significance
Width of the demarcation line (or ridge)	I	45	15.2180 ± 1.0605 <sup>a</sup>	1390.927	0.000
	II	42	26.3475 ± 1.3552 <sup>b</sup>		
	III	24	30.7455 ± 1.5462 <sup>c</sup>		
Vascular bifurcation ratio (ROI)	I	45	1.3963 ± 0.2879 <sup>a</sup>	16.703	0.000
	II	42	1.5385 ± 0.2566 <sup>b</sup>		
	III	24	1.8134 ± 0.3267 <sup>c</sup>		

Different superscripted letters indicate significant difference between groups.



**Figure 5.** Quantification parameters of extracted features.

**Table 4.** Ordered Logistic Regression for Feature Parameters

Characteristic	Wald	df	<i>P</i>	OR	OR (95% CI)	
					Lower	Upper
Stage 1	20.891	1	0.000	5.371E+24	32.525	81.361
Stage 2	23.972	1	0.000	2.097E+32	44.631	104.215
Widths	23.615	1	0.000	10.892	1.425	3.351
Bifurcation ratio	4.995	1	0.025	45.015	0.468	7.146

CI, confidence interval; OR, odds ratio.

were all black (background) and therefore not counted in Table 3.

The mean parameters of the segmented features are shown in Figure 5. All parameters increased significantly from stages I to III and reached the highest values in stage III.

We also performed ordered logistic regression on feature parameters. The parallel test met the requirements ( $P > 0.05$ ), the model fit  $\chi^2$  value was 429.112 ( $P < 0.001$ ), and the Cox–Snell value was 0.882, which indicated that the regression model explained up to 88.20% of the parameters.

We used the feature parameters of stage III ROP as a reference to dividing the model into two binary logistic models. The results are in Table 4. We found that the classification of stage I to III ROP was related to the widths of the demarcation line or crest and

the ratios of vessel proliferation within the ROI ( $P < 0.001$ , 95% CI confidence interval), and images with larger widths were 10.89 times more likely to be classified as stage II or III. Meanwhile, images with larger vascular bifurcation ratios were 45.02 times more likely to be classified as stage II or III.

We invited two ophthalmologists, one with over 10 years of experience and the other with only 3 years of standardized training, to diagnose retinal images of the comparison data set.

In Table 5, we calculated the  $\kappa$  score for DCNN predictions and the diagnosis from an ophthalmologist with 10 years of experience ( $\kappa = 0.9425$ ), which was close to perfect agreement. We also calculated scores for an ophthalmologist with only 3 years of training experience. When diagnosed with original retinal images (Table 6), the score was 0.8385, and



**Table 5.** DCNN Predictions and Manual Diagnosis (with 10 Years of Experience)

DCNN Predictions	Manual Diagnosis				Total
	Normal	Stage I	Stage II	Stage III	
Normal	406	3	1	0	410
Stage I	2	41	2	0	45
Stage II	0	1	40	1	42
Stage III	0	0	1	23	24
Total	408	45	44	24	521

**Table 6.** DCNN Predictions and Manual Diagnosis (with 3 Years of Training)

DCNN Predictions	Manual Diagnosis				Total
	Normal	Stage I	Stage II	Stage III	
Normal	401	8	1	0	410
Stage I	5	38	2	0	45
Stage II	1	2	34	5	42
Stage III	0	1	6	17	24
Total	407	49	43	22	521

**Table 7.** DCNN Predictions and Manual Diagnosis (images with quantitative parameters)

DCNN Predictions	Manual Diagnosis (with 3 Years of Training)				Total
	Normal	Stage I	Stage II	Stage III	
Normal	405	5	0	0	410
Stage I	3	<b>41</b>	1	0	45
Stage II	0	1	<b>39</b>	2	42
Stage III	0	0	2	<b>22</b>	24
Total	408	47	42	24	521

Note: Bold values were significant growth values.

when diagnosed with images with feature parameters, the score was 0.9268 (Table 7). We found that the ophthalmologists could use the quantitative segmentation features as the basis for their clinic diagnostic decisions, combining manual diagnosis with quantitative parameters to improve the consistency of diagnostic results for early stages of ROP.

## Discussion

In this study, we developed an automatic diagnostic system based on DCNNs. We trained the system using segmented images within the ROI, which could provide a diagnosis of stage I to III ROP with extracted

parameters. We also performed a quantitative analysis of these parameters.

Unlike the Mask R-CNN architecture used by Ding et al.,<sup>20</sup> we used two Retina U-Nets and a Dense Net. Retina U-Net combined the Retina Net,<sup>37</sup> a one-stage detector, and the structure of U-Net,<sup>38</sup> which could preserve the location information in images well. We calculated the ROI to extract the features, and by this method, we completed the data compression. Throughout the whole process, the size of images was always kept to the original, instead of resizing images to  $299 \times 299$  (in pixels) and training the system after randomly slicing images, as Ding et al.<sup>20</sup> did. A study done by Kim et al.<sup>39</sup> showed that retinal appearance assessment based on the whole image provided a more accurate and reliable DCNN classification compared to quadrant-based assessment.

With the Dense Net as a classifier, we achieved an overall accuracy of 97.98% for all four categories of the test data set, with a  $\kappa$  score of 0.9425. In a similar work, Ding et al.<sup>20</sup> obtained an overall accuracy of 67%. Of course, we cannot simply compare the performance metrics because of the different data sets.

More important, we not only provided an automatic classifier based on DCNN but also performed a quantitative analysis of the extracted feature parameters. The results of the statistical analysis of the parameters of the widths of the demarcation lines or ridges and the ratios of the vascular proliferation within the ROI showed that all quantitative parameters increased significantly among different groups. This may help to enhance the explainable DCNN predictions.

The results of the ordered logistic regression for these parameters showed that ratios of the vascular bifurcation within the ROI had a greater odds ratio value (45.015 vs. 10.892), which suggests that ratios played a greater role in the diagnosis of stages II and III. Second, the quantitative parameters of the ratios in Figure 5 showed a smaller difference between stages I and II, suggesting that the system relied more on the quantitative width parameters when diagnosing stage I, which explained why we obtained only 90.21% sensitivity in stage I. Finally, the logistic model explained only 88.2% of all parameters, which indicates that DCNNs learned more features than the two extracted ones. In the future, the dissection and visualization of features learned by DCNNs will be very interesting.

We also performed comparative tests on the comparison data set. Physicians integrated quantitative parameters of the extracted features with their clinic diagnosis, and the  $\kappa$  score was improved from 0.8385 to 0.9268. This suggested that our study may help alleviate the current situation of the insufficient number of hospital ophthalmologists.<sup>40</sup>

## Limitations

The input images were a potential limitation for our system. We used only retinal images of sufficient quality, which were acquired at Jiaying Maternal and Child Health Hospital using Retcam3 from preterm infants with lighter weights and shorter gestation periods. The Retcam3 imaging system is expensive, so many hospitals use other alternative devices such as PanoCam (Visunex Medical Systems, Suzhou, Jiangsu, China), and the different devices introduce differences in images. Second, during the screening of newborns, the ROP has been detected in many heavier and full-term infants. Therefore, it may not be sufficient to extract features from images of preterm infants alone. Finally, the sufficient quality images selected only may not represent reality. We may lose features in the suboptimal images.

## Future Work

In future studies, we will try to collaborate with different regions and hospitals to obtain more retinal images of preterm infants using different devices. We will expand data sets using newborn screening images to extract features to diagnose ROP, rather than just images from preterm infants. We will also investigate more stringent, quantitative image screening criteria and develop preprocessing modules to provide image screening automatically and objectively, and it would also be interesting to dissect and visualize the features learned by the DCNNs in future studies.

## Conclusion

The system we studied was capable of providing an accurate diagnosis of stage I to III ROP. Ophthalmologists can integrate DCNN decisions with the quantitative analysis of features to support them in making the best judgments.

## Acknowledgments

The authors thank Jia Liu at Jiaying Maternal and Child Health Hospital, who provided the retinal images of the preterm infants used in this study. We also thank doctors in ophthalmology who provided much help, including manual diagnosis.

Supported by the Jiaying Science and Technology Project, “Exploration of Eye Disease Screening Model for Infants and Children Aged 0–3 Years and Its Appli-

cation in Primary Eye Care Work” (2019AD32156), and the Zhejiang Province Medical and Health System Science and Technology Project, “Exploration of Eye Disease Screening Model for Infants and Children Aged 0–3 Years and Its Application in Primary Eye Care Work” (2020KY965).

Disclosure: **P. Li**, None; **J. Liu**, None

## References

1. Flynn JT, Bancalari E, Bachynski BN, et al. Retinopathy of prematurity: diagnosis, severity, and natural history. *Ophthalmology*. 1987;94(6):620–629.
2. Gilbert C, Field A, Gordillo L, et al. Characteristics of infants with severe retinopathy of prematurity in countries with low, moderate, and high levels of development: implications for screening programs. *Pediatrics*. 2005;115(5):E518–E525.
3. International Committee for the Classification of Retinopathy of Prematurity. An international classification of retinopathy of prematurity. *Arch Ophthalmol*. 1984;102:1130–1134.
4. International Committee for the Classification of Retinopathy of Prematurity. The international classification of retinopathy of prematurity revisited. *Arch Ophthalmol*. 2005;123(7):991–999.
5. Chiang MF, Quinn GE, Fielder AR, et al. International Classification of Retinopathy of Prematurity, Third Edition. *Ophthalmology*. 2021;128(10):e51–e68.
6. Campbell JP, Aater-Cansizoglu E, Bolon-Canedo V, et al. Expert diagnosis of plus disease in retinopathy of prematurity from computer-based image analysis. *JAMA Ophthalmol*. 2016;134(6):651–657.
7. Wallace DK, Zhao Z, Freedman SF. A pilot study using “ROPtool” to quantify plus disease in retinopathy of prematurity. *J AAPOS*. 2007;11(4):381–387.
8. Bas E, Ater-Cansizoglu E, Erdogmus D, Kalpathy-Cramer J. Retinal vasculature segmentation using principal spanning forests. *9th IEEE International Symposium on Biomedical Imaging (ISBI) - From Nano to Macro, Barcelona, Spain*. IEEE; 2012:1792–1795.
9. Wittenbery LA, Jonsson NJ, Chan RVP, Chiang MF. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus*. 2011;49(1):11–19.
10. Wilson CM, Wong KR, Ng J, Cocker KD, ELLs AL, Fielder AR. Digital image analysis in

- retinopathy of prematurity: a comparison of vessel selection methods. *J AAPOS*. 2012;16(3):223–228.
11. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436.
  12. Ertosun MG, Rubin DL. Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. *AMIA Annu Symp Proc*. 2015;2015:1899.
  13. Sim Y, Chung MJ, Kotter E, et al. Deep convolutional neural network-based software improves radiologist detection of malignant lung nodules on chest radiographs. *Radiology*. 2020;294(1):199–209.
  14. Fujisawa Y, Otomo Y, Ogata Y, et al. Deep learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumor diagnosis. *Br J Dermatol*. 2019;180(2):373–381.
  15. Xiao H, MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44(4):1408–1419.
  16. Worrall DE, Wilson CM, Brostow GJ. Automated retinopathy of prematurity case detection with convolutional neural networks. *2nd International Workshop on Deep Learning in Medical Image Analysis (DLMIA)*. Athens, Greece: Springer, Cham; 2016:68–76.
  17. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136(7):803–810.
  18. Hu J, Chen Y, Zhong J, Ju R, Yi Z. Automated analysis for retinopathy of prematurity by deep neural networks. *IEEE Trans Med Imaging*. 2018;38(1):269–279.
  19. Mulay S, Ram K, Sivaprakasam M, Vinekar A. Early detection of retinopathy of prematurity stage using deep learning approach. *Medical Imaging 2019: Computer-Aided Diagnosis*. San Diego, CA: SPIE; 2019;10950:758–764.
  20. Ding A, Chen Q, Cao Y, Liu BY, et al. Retinopathy of prematurity stage diagnosis using object segmentation and convolutional neural networks. *International Joint Conference on Neural Networks (IJCNN)*. Glasgow, UK: IEEE; 2020:1–6.
  21. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
  22. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res*. 2019;72:100759.
  23. Mao J, Luo Y, Liu L, et al. Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks. *Acta Ophthalmol*. 2020;98(3):E339–E345.
  24. Yildiz VM, Tian P, Yildiz I, et al. Plus disease in retinopathy of prematurity: convolutional neural network performance using a combined neural network and feature extraction approach. *Transl Vis Sci Technol*. 2020;9(2):10–10.
  25. Yuen V, Ran A, Shi J, et al. Deep-learning-based pre-diagnosis assessment module for retinal photographs: a multicenter study. *Transl Vis Sci Technol*. 2021;10(11):16.
  26. Jaeger PF, Kohl SAA, Bickelhaupt S, et al. Retina U-Net: embarrassingly simple exploitation of segmentation supervision for medical object detection. *PMLR*. 2020:171–183.
  27. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybernet*. 2007;9(1):62–66.
  28. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Machine Intell*. 1986;PAMI-8(6):679–698.
  29. Sklansky J, Nahin PJ. A parallel mechanism for describing silhouettes. *IEEE Transactions on Computers*. 1972;100(11):1233–1239.
  30. Robert JM, Toussaint G. Computational geometry and facility location. *Proc. International Conference on Operations Research and Management Science*. Athens, Greece: Wiley-Blackwell; 1990.
  31. Zhang TY, Suen CY. A fast parallel algorithm for thinning digital patterns. *Comm ACM*. 1984;27(3):236–239.
  32. Awrangjeb M, Lu G. Robust image corner detection based on the chord-to-point distance accumulation technique. *IEEE Trans Multimedia*. 2008;10(6):1059–1072.
  33. Huang G, Liu Z, Laurens VDM, Weinberger KQ. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE; 2017:4700–4708.
  34. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115(3):211–252.
  35. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29.
  36. Wang S, Yin Y, Cao G, Wei BZ, Zheng YJ, Yang GP. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing*. 2015;149:708–717.
  37. Lin TY, Goyal P, Girshick R, He KM, Dollar P. Focal loss for dense object detection.

*Proceedings of the IEEE International Conference on Computer Vision*. 2017:2980–2988.

38. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Med Image Comput Comput Assist Intervent*. 2015;9351:234–241.
39. Kim SJ, Campbell JP, Kalpathy-Cramer J, et al. Accuracy and reliability of eye-based vs quadrant-based diagnosis of plus disease in retinopathy of prematurity. *JAMA Ophthalmol*. 2018;136(6):648–655.
40. Gilbert C. Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control. *Early Hum Dev*. 2008;84(2):77–82.