RimNet: A Deep Neural Network Pipeline for Automated Identification of the Optic Disc Rim

Haroon Adam Rasheed, B.S. 1, Tyler Davis, M.S. 2, Esteban Morales5, MS, Zhe Fei, M.S., Ph.D.3,4, Lourdes Grassi, M.D.5, Agustina De Gainza, M.D.5, Kouros Nouri-Mahdavi, M.D., M.S.5, Joseph Caprioli, M.D.5

Key Words: Glaucoma, DDLS, segmentation, mRDR, neural network

Abbreviations/Acronyms: MAE – Median Average Error, mRDR – minimum Rim-To-Disc Ratio, CDR - Cup-to-disc-ratio, ISNT - Inferior>superior>nasal>temporal, IoU – Intersection over Union, RimIoU – Rim Intersection over Union, CupIoU – IoU of the optic cup, DiscIoU – IoU of the optic disc, CupDice – Dice score of the optic cup, DiscDice – Dice score of the optic disc, DDLS – disc damage likelihood scale, ARW – Absent Rim Width, RDAR — Rim To Disc Area Ratio

Financial Support: Unrestricted grant from Research to Prevent Blindness, The National Institutes of Health Grant R01-EY029792 (KNM), Payden Glaucoma Fund, Simms/Mann Family Foundation and an unrestricted research grant from Heidelberg Engineering (KNM)

The sponsor or funding organization had no role in the design or conduct of this research.

Meeting Presentation: Presented as a poster at The Association for Research in Vision and Ophthalmology Annual Meeting, 2022

Conflict of Interest: No conflicting relationship exists for any author.

Running Head: RimNet: Automated Rim Identification with mRDR and ARW calculation

Address for Reprints: 100 Stein Plaza 2-118 Los Angeles CA 90095

1. University of California Los Angeles David Geffen School of Medicine, 2. Computer Science, University of California Los Angeles, Los Angeles, CA, United States, 3. University of California Los Angeles Jonathan and Karin Fielding School of Public Health, Los Angeles, CA, United States, 4. Biostatistics, University of California Los Angeles, Los Angeles, CA, United States, 5. Glaucoma, Jules Stein Eye Institute, Los Angeles, CA, United States
Abstract

Purpose
Accurate neural rim measurement based on optic disc imaging is important to glaucoma severity grading and often performed by trained glaucoma specialists. We aim to improve upon existing automated tools by building a fully automated system (RimNet) for direct rim identification in glaucomatous eyes and measurement of the minimum rim-to-disc ratio (mRDR) in intact rims, the angle of absent rim width (ARW) in incomplete rims, and the rim-to-disc-area ratio (RDAR) with the goal of optic disc damage grading.

Design
Retrospective cross-sectional study.

Participants
1208 optic disc photographs with evidence of glaucomatous optic nerve damage from 1021 eyes of 903 patients with any form of primary glaucoma were included. Mean age was 63.7 (±14.9) years. The average mean deviation of visual fields is -8.03 (±8.59).

Methods
The images were required to be of adequate quality, have signs of glaucomatous damage, and be free of significant concurrent pathology as determined independently by glaucoma specialists. Rim and optic cup masks for each image were manually delineated by glaucoma specialists. The database was randomly split into 80/10/10 for training, validation, and testing, respectively. RimNet consists of a deep learning rim and cup segmentation model, a computer vision mRDR measurement tool for intact rims, and an ARW measurement tool for incomplete rims. mRDR is calculated at the thinnest rim section while ARW is calculated in regions of total rim loss. The rim-to-disc area ratio (RDAR) was also calculated. Evaluation on the Drishti-GS dataset provided external validation (Sivaswamy 2015).

Main Outcome Measures
Median Absolute Error (MAE) between glaucoma specialists and RimNet for mRDR and ARW.

Results
On the test set, RimNet achieved a mRDR MAE of 0.03 (0.05), ARW MAE of 31 (89) degrees, and an RDAR MAE of 0.09 (0.10). On the Drishti-GS dataset, an mRDR MAE of 0.03 (0.04) and an mRDAR MAE of 0.09 (0.10) was observed.

Conclusions
RimNet demonstrated acceptably accurate rim segmentation and mRDR and ARW measurements. The fully automated algorithm presented here would be a valuable component in an automated mRDR-based glaucoma grading system. Further improvements could be made by improving identification and segmentation performance on incomplete rims and expanding the number and variety of glaucomatous training images.
Introduction

Glaucoma is the leading cause of irreversible blindness and the second leading cause of blindness worldwide. Roughly half of glaucoma cases are undiagnosed according to population-based studies. Early treatment preserves patient quality of life and reduces disease burden. Therefore, identification of early glaucoma is key to preventative care.

Glaucoma diagnosis and grading are performed, in part, by evaluation of the optic nerve head’s neuroretinal rim of the optic disc. Metrics often include cup-to-disc ratio (CDR), rim-to-disc ratio (mRDR), and the ISNT rule, which compares the regional width of the neuroretinal rim. Recent studies have shown the advantages of mRDR compared to ISNT and CDR for glaucoma classification accuracy.

The mRDR cannot adequately account for the degree of damage in optic discs with localized rim loss where the neuroretinal rim is noncontinuous or ‘incomplete’. A solution can be found in the Disc Damage Likelihood Scale (DDLS) proposed by Spaeth et al. DDLS accounts for incomplete rims by measuring the angle for which a rim is absent. This is called the absent rim width (ARW). Additionally, the scale accounts for disc size which affects the significance of the mRDR or ARW. It is commonly accepted and has been incorporated into eye health guidelines for optometrists and ophthalmologists. DDLS is limited as a diagnostic tool by the need for expert time to accurately grade images. Automated high-efficacy DDLS grading could offer a powerful screening method.

In recent years, a confluence of several factors has led to efforts in automated glaucoma diagnosis and grading. First, studies have shown that automated algorithms can offer more consistent and reliable grading than human graders. Second, there has been a rapid advancement in image segmentation, image processing, and deep learning neural networks. In other fields, several neural networks outperformed human graders in image classification tasks. This could allow for unprecedented accuracy in optic rim segmentation and glaucoma grading. Finally, the optic disc exhibits characteristic alterations in glaucomatous patients, a prime candidate for automated segmentation and analysis. Together, these factors make automated glaucoma diagnosis and grading a possibility.

While DDLS also requires disc size analysis, automated rim segmentation with mRDR calculation for intact neuroretinal rims and ARW calculation for incomplete neuroretinal rims offers a step towards creating an efficacious, high-throughput diagnostic system for glaucomatous disc damage. Such a segmentation algorithm would need to be broadly applicable. Additionally, it would require an expansive learning capacity that could be applied to a variety of fundus images taken with different imaging modalities and with concurrent pathologies and normal variations. Convolutional neural networks offer such an approach.

The goal of this paper is to present a novel convolutional neural network algorithm for neuroretinal rim segmentation, automated mRDR calculation for intact rims, and ARW calculation for incomplete neuroretinal rims. This neural network algorithm offers an important step towards building an automated DDLS screening tool.

Methods

The study adhered to the tenets of the Declaration of Helsinki, was approved by UCLA’s Human Research Protection Program, and conformed to the Health Insurance Portability and Accountability Act (HIPAA) policies.
Optic disc photographs were taken from the UCLA Stein Eye Glaucoma database. The images were of varied magnifications and taken from slides and three different digital fundus cameras. All cameras were visible light cameras. No infra-red, laser scanning, red-free, autofluorescence, or hand-held smartphone-based cameras were used. Slide films were scanned and digitized at a third-party location.

The enrolled images met the following inclusion and exclusion criteria as deemed by two board-certified glaucoma specialists. Inclusion criteria include: (i) evidence of glaucomatous damage in the posterior pole; (ii) images had to be in focus, with discernible posterior pole and vasculature details. Exclusion criteria were concurrent non-glaucoma disease including optic neuritis, optic disc neovascularization, and vitreous hemorrhage that would impair visualization of the posterior pole. Globally, it was ensured that the full spectrum of glaucomatous damage, from early-stage intact neuroretinal rims to late-stage incomplete rims, were included while abiding by the inclusion and exclusion criteria. Figure 4 shows the mRDR distributions of our train, validate, and test set. The neuroretinal rim and optic cup were then manually segmented by one of three glaucoma specialists with a smart tablet and the image editing program GIMP. These masks were used as ground truth. The diagnostic categories for patients are shown in Table 1.

RimNet Model and Hyperparameter Architecture

A deep learning model for rim segmentation was developed as the centerpiece of the RimNet pipeline. The model was developed with Python 3.9.7. Libraries used include TensorFlow 2.6.0, Segmentations Models 1.0.1, Keras Tuner 1.04, OpenCV Python 4.5.3, Numpy 1.19.5, Scipy 1.7.1, and Scikit-learn 0.24.2.

Optimizing the deep learning model requires a careful choice of model architecture and hyperparameters. The choice of hyperparameters can greatly influence the prediction speed, processing requirements, and accuracy of a neural network model. These hyperparameters include the decoder, learning rate, optimizer, and loss function as shown in Table 3. The optimal combination of these parameters is task-dependent. Whereas trial and error has been used in the past, newer architecture search techniques allow for the rapid evaluation of combinations of hyperparameters with the goal of optimizing a selected metric.

To narrow the search space, an encoder of InceptionV3 was chosen based on literature review and computational efficiency. InceptionV3 was first published in 2015, outperforming popular encoders at the time with a fraction of the computation costs. It has previously been used for medical segmentation. Our workstation uses NVIDIA 2080 RTX Ti graphics cards. Therefore, with limited computational efficiency, the selection of InceptionV3 was appropriate.

Transfer learning with ImageNet weights was used to initialize InceptionV3. No transfer learning was done for the decoder. Augmentations were used including a 20-degree rotation, a 10% vertical shift, a 10% horizontal shift, a horizontal flip, a vertical flip, up to a 30% random crop, a brightness change by +/-50 units, and a contrast limited adaptive histogram equalization (CLAHE) filter. Image down sampling was completed via a nearest neighbor algorithm. Color information was encoded using RGB channels with 8 bits per channel. The encoder and decoder were coupled using the Segmentation Models 1.0.1 library. The total number of trainable parameters was 29,896,979. No dropout layers were manually added.
Finally, a random search was performed using the Keras Tuner library\textsuperscript{16}. The search parameters included the decoder, loss function, learning rate, and the optimizer. The rim Intersection over Union (IoU) was used as the segmentation metric. The full search space is documented in Table 3.

End-to-End mRDR Calculation Procedure

mRDR, ARW, and rim-to-disc-area (RDAR) measurements are the final output of RimNet, which can be accomplished by accurate rim segmentation followed by image analysis. These two steps, along with preprocessing, led to the final framework for RimNet as shown in Figure 1.

The optic disc photographs were first resized to 224x224 with nearest neighbor interpolation in order to meet model specifications. A contrast limited adaptive histogram equalization (CLAHE) filter was then applied to highlight distinctive features. The preprocessed image was submitted to the neural network model which generated a segmentation mask of the optic rim and cup. While a segmentation of the optic cup is not directly needed for mRDR or RDAR calculations, it was found that training the model to identify and segment the optic cup improved identification of incomplete rims and ARW calculations. Finally, the rim segmentation mask was resized to the dimensions of the original image to allow for accurate mRDR calculation and submitted to image analysis algorithms.

For mRDR, the algorithm first identified the center of the segmented optic cup using OpenCV. Vectors were created from the center of the cup to the boundary points. Boundary points were found using OpenCV. The number of vectors depended on the number of boundary points detected in the segmented rim. The intersection between the vectors and the segmented rim was taken as the rim width. The shortest rim width was identified and, through boundary point analysis of the rim, the disc diameter was found. Hence, the mRDR was calculated by dividing the rim width by the diameter. The RDAR was calculated by dividing the number of pixels of the segmented rim by the number of pixels in optic disc.

The Absent Rim Width (ARW) was calculated by first applying contour hierarchies to identify shapes within the rim segmentation. We rely on the fact that intact rims will have a ‘second shape’ within the segmentation, the elliptical or circular form of the optic cup. Incomplete rims will not have this second shape. If the rim is classified as broken, 360 radial segments from the center are drawn to the edge of the rim. The radial segments that do not intersect the rim are those within the ‘broken’ segment of the neuroretinal rim. The number of radial segments within the incomplete segment are added to give the ARW, one radial segment for each degree. If there were two breaks in a neuroretinal rims, the angles were added together and reported as one ARW. Examples of this can be found in the neuroretinal rim segmentation shown in Figure 2.

External Validation

The Drishti-GS database is a publicly available dataset of retinal images of glaucomatous eyes with manual cup and disc segmentations\textsuperscript{21,22}. The images were first cropped around the optic disk, as they are available at a field of view of 30 degrees. Then, by subtracting the Drishti-GS cup segmentations from the disc segmentations, rim segmentations were acquired. These were used as ‘ground truth’ for validation testing. The database has been used to compare performance between published optic cup and disc segmentation models through metrics such as IoU and Dice coefficient\textsuperscript{10,23–27}. Few investigators have attempted rim segmentations on the Drishti-GS database\textsuperscript{26}. Therefore, RimNet rim segmentations were used to recreate cup segmentations to allow for comparison with other segmentation models.
intersection over union for cup segmentations (CupIoU) and disc segmentations (DisclIoU) were reported. Additionally, the Dice scores for the cup (CupDice) and disc (DiscDice) were reported.

Evaluation Criteria

The main outcome measures are the median absolute error (MAE) difference between the glaucoma specialists and RimNet for three metrics: mRDR, RDAR, and ARW. A secondary measure is the RimIoU, the IoU of the RimNet rim segmentation compared to that of the glaucoma specialists.

The mRDR, RDAR, and ARW have been explained above. Two measures of segmentation accuracy are also reported: Intersection over Union and Dice scores. The Intersection over Union (IoU), also known as the Jaccard distance, is a measure of segmentation accuracy. It compares the ground truth with the segmentation by reporting the ratio of the intersection area over the union area. The Dice score for cup and disc segmentations are reported for the Drishti-GS dataset to compare segmentation performance. The Dice score compares the ground truth with the segmentation by reporting the ratio of two times the intersection area over the summed area of the ground truth and segmentation.
Results

A database of 1208 optic disc photographs of 1-21 eyes from 903 glaucoma patients were used for training, validation, and testing in an 80/10/10 split. Both scanned slides and original digital images were represented in the dataset. The average (±SD) age of the patients was 63.7 (± 14.9) with a 43:57 male-to-female ratio. Full demographics including gender, age, and race are listed in Table 1. The average (±SD) visual field mean deviation (MD) was –8.03 ± 8.59 dB (range: -31.64, 3.59). Of the 1208 optic disc photographs, 340 had incomplete neuroretinal rims. The diagnoses for the patients are listed in Table 2.

Hyperparameter Architecture

Optimized parameters were found through the random search of 64 model combinations, detailed in Table 318,28–38. The combination of the InceptionV3 backbone and LinkNet architecture proved to be the most accurate18,29. 18LinkNet is a lightweight decoder first published in 201729. Other parameters identified include the loss function of binary cross-entropy, learning rate of 10⁻³, and the Adam optimizer38.

Segmentation Network Results

The code used to train, run, and evaluate RimNet can be found on our public repository at https://github.com/TylerADavis/GlaucomaML. On the test set, a mRDR MAE (IQR) of 0.03 (0.05) was achieved on the intact rims while a ARW MAE (IQR) of 31 (89) degrees was achieved on the incomplete rims. 22 of 34 eyes with incomplete rims were correctly identified as incomplete on segmentation. A RDAR MAE (IQR) of 0.09 (0.10) was achieved on all images. A RimIoU of 0.68 was achieved on intact rims, while a RimIoU of 0.45 was achieved on incomplete rims. The results of RimNet are presented in Table 4. Figure 2 demonstrates examples of RimNet segmentation results. To better examine the accuracy of the mRDR and RDAR calculations, the difference between the estimated values and the ground truths were calculated. Bland-Altman plots comparing the estimated and ground truth mRDR and RDAR are shown in Figure 3.

A comparison of RimNet segmentation on the Drishti-GS dataset to other published works is presented in Table 5. The mRDR MAE (IQR) was 0.03 (0.04) and the RDAR MAE (IQR) was 0.09 (0.10). The IoU of the optic cup (CupIoU) was 0.77 and a IoU of the optic disc (DiscIoU) was 0.91. The Dice score of the cup (CupDice) was 0.86 and Dice score of the optic disc (DiscDice) 0.95 was achieved.

Discussion

These results demonstrate that RimNet is capable of reasonably accurate segmentation and analysis of optic discs with both intact and incomplete rims. Spaeth et al. distinguished different DDLS grades by mRDR steps of 0.15. The MAE of the mRDR is well within this value, showing that RimNet segmentations are clinically relevant. For more advanced glaucoma with DDLS grades of 6 and above, the neuroretinal rim is incomplete and Spaeth et al. uses the ARW to distinguish grades. The five categories are less than 45 degrees, 45 degrees to 90 degrees, 90 degrees to 180 degrees, 180 to 270 degrees, and greater than 270 degrees. The minimum step is 45 degrees; the MAE falls slightly below that category at 31 degrees with 22 of 34 total incomplete rims correctly identified as incomplete. However, the IQR demonstrates a broad range of ARW. The error echoes the difficulties faced by the glaucoma specialists. While segmenting these severely glaucomatous rims to create the ‘ground truth’ masks, glaucoma specialists often differed regarding where rims were interrupted and if rims were incomplete or intact. Though a forced consensus was eventually reached, this demonstrates the difficulty of the task and the variability of this ‘ground truth’. RimNet offers 65% accuracy in identifying incomplete rims and a relatively low ARW MAE. To our knowledge, RimNet is the first to offer such capabilities in published literature.
This work offers three improvements in the current landscape of optic disc segmentation. First, we utilized a dataset of 1208 images with external validation on Drishti-GS. Second, while we have still reported IoU and Dice scores, we have focused on more clinically relevant metrics such as mRDR, RDAR, and ARW. Third, our study is the first to focus on accurate segmentation of incomplete rims. RimNet is a useful step towards completely automating the DDLS algorithm.

Automated segmentation of the optic disc and cup have been previously explored. The original studies were initially based on image processing functions such as thresholding, level set, active contour, clustering, and component extraction with success on local and publicly available datasets. As early as 2001, Chrástek et al. offered an automated method of optic disc segmentation with filtering and edge detection, which achieved a segmentation accuracy of 82%. In 2008, Liu and collaborators used level set and thresholding methods to achieve 97% accuracy when comparing algorithm-determined CDR ratio to clinical CDR ratio on a dataset of 73 images from the Singapore Eye Research Centre. In 2015, Lotankar et al. used active contouring to achieve a 99% pixel-to-pixel accuracy on a private database of 150 images. However, each of these approaches were limited in scope. Level-setting and thresholding would fail with images with decreased or increased intensity caused by pathological findings, which can be commonly seen on optic disc photographs such as peripapillary atrophy. This leads to overestimating or underestimating CDRs. Active contouring may similarly be affected by abnormal pathology or bright artifacts fixating on local maxima or minima within the image. Therefore, though these methods have proven efficacy, they can be improved upon.

An automated grading system for glaucoma diagnosis and progression needs a high efficiency, broadly applicable segmentation algorithm with an expansive learning capacity which could be applied to a variety of fundus images acquired with different imaging modalities with concurrent pathologies and variations. Though further work must be done, deep learning and convolutional neural networks may play an important role in the solution. They have an enormous learning capacity relative to their size. Rapid advances in computational memory and processing speed have made neural networks more accessible for optic disc segmentation. Zilly et al. used ensemble learning to achieve 89% IoU on disc segmentation and 84% IoU on cup segmentation on the Drishti-GS dataset. Sevastopolsky and coworkers furthered this work by using a modified U-Net to achieve a comparable accuracy in less than a tenth of the time. More on segmentation efforts, both image processing functions and neural network attempts, can be found on a review article by Thakur and Juneja.

Several groups have pursued automated mRDR and RDAR calculations. In 2019, Kumar et al. proposed using an imaging processing technique called active discs to segment the optic disc and cup and perform general glaucoma classification (normal, moderate, severe) based on mRDR. Though direct mRDR accuracy was not reported, an mRDR-based approach demonstrated high classification accuracy. In 2020, Martins et al. proposed a smartphone-based glaucoma diagnosis pipeline, which focuses on glaucoma classification and calculates RDAR. However, RDAR results were not directly reported. More recently, Pachade et al. proposed an NENet model consisting of EfficientNetB4 and adversarial learning that achieved an area-under-the-curve (AUC) of 0.901 on RDAR calculation for Drishti-GS.

To the best of our knowledge, RimNet is the first engineering attempt to pursue segmentation and glaucoma grading efforts with incomplete neuroretinal rims. Thus, direct comparison of RimNet to other segmentation models is difficult. However, through the Drishti-GS dataset, an artificially-derived segmentation comparison is possible by recreating cup and disc masks from the RimNet rim segmentations. Table 5 demonstrates that RimNet performed well overall compared to recent segmentation models on the Drishti-GS dataset. While it outperformed several other models in CupDice, DiscIoU, and DiscDice segmentations, it was below average in CupIoU. These results must be understood...
in the context of three factors. First, the Drishti-GS images were available as 30-degree field of views. However, RimNet requires images centered and cropped near the optic disc margin. Therefore, RimNet has a significant information loss compared to other models that use the 30-degree field of view.

Second, RimNet is unique in that it has been trained on both complete and incomplete rims. The models compared to RimNet have been trained only on complete rims. It is reasonable to expect a higher segmentation accuracy in these cases. Finally, the cup and disc segmentations produced by RimNet were artificially derived from the RimNet’s rim segmentation. By not directly predicting on the cup and disc, accuracy was lost. Considering these three factors, RimNet’s performance on Drishti-GS is acceptable.

This is corroborated by the Drishti-GS mRDR MAE of 0.03 (0.04) and RDAR MAE of 0.09 (0.10), both of which are low.

The findings of this study need to be interpreted with the shortcomings in mind. First, the hyperparameter architecture search was limited by the computing and memory limits of our workstation, which uses NVIDIA RTX 2080 Ti graphics cards. We could not include larger models such as ResNet152 into our search due to these memory constraints. Second, the number of ground truth masks and optic disc images, particularly those of more severe glaucoma is limited. Greater numbers of diverse samples would allow RimNet to better learn mRDR and ARW calculations.

RimNet brings glaucomatous detection and DDLS grading a step closer to full automation. Automated grading of disc size is a necessary step to fully autonomous DDLS grading. A future goal would be to not only pursue full automation of DDLS grading, but to test their capabilities as diagnostic tools. One promising avenue for further investigation would be screening with smartphone fundoscopy. The increasing quality of smartphone cameras have made smartphone fundoscopy viable as a screening method. This, combined with automated DDLS grading, could provide a powerful screening tool to revolutionize glaucoma detection.

In conclusion, RimNet provides a method for high efficacy rim segmentation, mRDR, and ARW calculation. It also provides an example of how ophthalmic care be augmented by artificial intelligence. Though more work remains to be done, we believe that detection, diagnosis, and care of glaucoma can integrate with approaches such as these and aid ophthalmologists in decision making to provide higher quality care for a global population of patients.
References


40. Thakur N, Juneja M. Survey on segmentation and classification approaches of optic cup and optic
2022].
of glaucoma. Med Image Anal 2005;9:297–314. Available at:
42. Liu J, Wong DWK, Lim JH, et al. Optic cup and disk extraction from retinal fundus images for
determination of cup-to-disc ratio. 2008 3rd IEEE Conference on Industrial Electronics and Applications,
43. Lotankar M, Noronha K, Koti J. Detection of optic disc and cup from color retinal images for
automated diagnosis of glaucoma. 2015 IEEE UP Section Conference on Electrical Computer and
44. Martins J, Cardoso JS, Soares F. Offline computer-aided diagnosis for Glaucoma detection using
[Accessed June 12, 2022].
Technological Advances and Their Implications for Worldwide Healthcare. Telemedicine Journal and e-
Health 2016;22:198. Available at: /pmc/articles/PMC4790203/ [Accessed March 12, 2022].
47. Nazari Khanamiri H, Nakatsuka A, El-Annan J. Smartphone Fundus Photography. JoVE (Journal of
Visualized Experiments) 2017:e55958. Available at: https://www.jove.com/v/55958/smartphone-
fundus-photography [Accessed March 12, 2022].

Figure legends:

Figure 1: RimNet Pipeline. The raw image first undergoes preprocessing, where CLAHE is applied and the image is resized to model specifications. A mask is generated after applying the segmentation model to the preprocessed image. Image analysis then calculated the RDAR. If the rim is intact, the mRDR is calculated. If the rim is incomplete, ARW is calculated.

Figure 2: Segmentation Results. This figure demonstrates several examples of RimNet segmentation compared to physician segmentation. The left-most column shows the raw image. The middle column overlays the physician segmentation (white) over the raw image. The right-most column overlays the RimNet segmentation (white) over the raw image. In intact rims, green line shows the diameter and the dark blue shows the thinnest rim. In incomplete rims, the dark blue shows the edges of the segmentation.

Figure 3. Bland-Altman plots showing the agreements in mRDR and RDAR between clinician and RimNet in test images. Red dashed lines indicate 95% confidence limits.

Figure 4: Distribution of mRDRs for Train, Validation, and Test Datasets. For each dataset, a frequency histogram is shown above with a box plot corresponding to the dataset below.
Table 1: Demographic Data for Dataset. Lists the gender distribution, age distribution, and racial distribution by camera type.

<table>
<thead>
<tr>
<th>Gender Distribution</th>
<th>Slide Images</th>
<th>Digital Camera 1</th>
<th>Digital Camera 2</th>
<th>Digital Camera 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>407</td>
<td>119</td>
<td>55</td>
<td>12</td>
</tr>
<tr>
<td>M</td>
<td>302</td>
<td>85</td>
<td>44</td>
<td>11</td>
</tr>
</tbody>
</table>

Age Distribution

<table>
<thead>
<tr>
<th></th>
<th>Slide Images</th>
<th>Digital Camera 1</th>
<th>Digital Camera 2</th>
<th>Digital Camera 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>60.72</td>
<td>67.13</td>
<td>72.80</td>
<td>66.92</td>
</tr>
<tr>
<td>SD</td>
<td>13.48</td>
<td>17.43</td>
<td>12.75</td>
<td>17.71</td>
</tr>
<tr>
<td>Median</td>
<td>61.87</td>
<td>71.06</td>
<td>73.91</td>
<td>72.37</td>
</tr>
<tr>
<td>IQR</td>
<td>15.79</td>
<td>16.33</td>
<td>12.86</td>
<td>22.54</td>
</tr>
<tr>
<td>Min</td>
<td>9.36</td>
<td>6.92</td>
<td>16.19</td>
<td>17.48</td>
</tr>
<tr>
<td>Max</td>
<td>90.05</td>
<td>96.10</td>
<td>94.41</td>
<td>86.17</td>
</tr>
</tbody>
</table>

Race Distribution

<table>
<thead>
<tr>
<th></th>
<th>Slide Images</th>
<th>Digital Camera 1</th>
<th>Digital Camera 2</th>
<th>Digital Camera 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian</td>
<td>90</td>
<td>34</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>Black</td>
<td>63</td>
<td>22</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic</td>
<td>66</td>
<td>20</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>White</td>
<td>366</td>
<td>100</td>
<td>45</td>
<td>12</td>
</tr>
<tr>
<td>Other</td>
<td>53</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Unknown</td>
<td>71</td>
<td>22</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 2: Glaucoma Diagnosis for all 1208 patients included in the RimNet dataset.

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Open-Angle Glaucoma</td>
<td>530</td>
</tr>
<tr>
<td>Glaucoma Suspect</td>
<td>403</td>
</tr>
<tr>
<td>Chronic Angle-Closure Glaucoma</td>
<td>71</td>
</tr>
<tr>
<td>Low-Tension Glaucoma</td>
<td>47</td>
</tr>
<tr>
<td>Secondary Open-Angle Glaucoma</td>
<td>35</td>
</tr>
<tr>
<td>Capsular glaucoma with psuedoexfoliation</td>
<td>33</td>
</tr>
<tr>
<td>Anatomical Narrow Angle</td>
<td>27</td>
</tr>
<tr>
<td>Glaucoma secondary to Eye Infection</td>
<td>24</td>
</tr>
<tr>
<td>Pigmentary Glaucoma</td>
<td>15</td>
</tr>
<tr>
<td>Secondary Angle Closure</td>
<td>11</td>
</tr>
<tr>
<td>Congenital glaucoma</td>
<td>7</td>
</tr>
<tr>
<td>Juvenile Glaucoma</td>
<td>3</td>
</tr>
<tr>
<td>Acute angle-closure glaucoma</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 3: Hyperparameter search space for the RimNet included the encoders, decoders, loss functions, learning rates, and optimizer. The optimized metric was intersection over union of the rim.

<table>
<thead>
<tr>
<th>Hyperparameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoders</td>
<td>MobileNetV2, ResNet34, EfficientNetB0, InceptionV3, ResNet101, VGG16, ResNet50</td>
</tr>
<tr>
<td>Decoders</td>
<td>U-Net, FPN, LinkNet, PSPnet</td>
</tr>
<tr>
<td>Loss Function</td>
<td>Binary_Crossentropy, Binary_Focal_Loss</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>$10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam, SGD</td>
</tr>
</tbody>
</table>
Table 4: RimNet Results on internal test set and Drishti-GS dataset. The ARW cannot be calculated on the Drishti-GS dataset because all rims are intact.

<table>
<thead>
<tr>
<th></th>
<th>Internal</th>
<th>Drishti-GS</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRDR</td>
<td>0.04 (±0.03)</td>
<td>0.04(±0.04)</td>
</tr>
<tr>
<td>ARW</td>
<td>48.9 (±35.9)</td>
<td>N/A</td>
</tr>
<tr>
<td>RDAR</td>
<td>0.10 (±0.09)</td>
<td>0.10(±0.08)</td>
</tr>
<tr>
<td>RimIoU</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td>N</td>
<td>120 (87 intact, 33 Incomplete)</td>
<td>101 (101 intact, 0 Incomplete)</td>
</tr>
</tbody>
</table>
Table 5: DRISHTI-GS segmentation performance of RimNet compared to published segmentation models.\textsuperscript{11,17–21}

<table>
<thead>
<tr>
<th>Model</th>
<th>CupIoU</th>
<th>DiscIoU</th>
<th>CupDice</th>
<th>DiscDice</th>
</tr>
</thead>
<tbody>
<tr>
<td>RimNet</td>
<td>0.77</td>
<td>0.91</td>
<td>0.86</td>
<td>0.95</td>
</tr>
<tr>
<td>Zilly et al. (2017)</td>
<td>0.85</td>
<td>-</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>Sevastopolsky (2017)</td>
<td>0.75</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Edupuganti et al. (2018)</td>
<td>0.81</td>
<td>0.69</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Al-Bander and Zheng et al. (2018)</td>
<td>-</td>
<td>-</td>
<td>0.83</td>
<td>0.95</td>
</tr>
<tr>
<td>Joshua et al. (2019)</td>
<td>0.79</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yu et al. (2019)</td>
<td>-</td>
<td>-</td>
<td>0.88</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Precis:

The proposed deep learning pipeline identifies and calculates glaucoma severity metrics of the optic disc rim. It achieves high agreement with clinicians and is the first engineering attempt pursuing glaucoma grading in incomplete neuroretinal rims.
Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: