



DDLSNet: A Novel Deep Learning-Based System for Grading Fundusoscopic Images for Glaucomatous Damage

Haroon Adam Rasheed, BS,¹ Tyler Davis, MS,² Esteban Morales, MS,⁵ Zhe Fei, MS, PhD,^{3,4} Lourdes Grassi, MD,⁵ Agustina De Gainza, MD,⁵ Kouros Nouri-Mahdavi, MD, MS,⁵ Joseph Caprioli, MD⁵

Purpose: To report an image analysis pipeline, DDLSNet, consisting of a rim segmentation (RimNet) branch and a disc size classification (DiscNet) branch to automate estimation of the disc damage likelihood scale (DDLS).

Design: Retrospective observational.

Participants: RimNet and DiscNet were developed with 1208 and 11 536 optic disc photographs (ODPs), respectively. DDLSNet performance was evaluated on 120 ODPs from the RimNet test set, for which the DDLS scores were graded by clinicians. Reproducibility was evaluated on a group of 781 eyes, each with 2 ODPs taken within 4 years apart.

Methods: Disc damage likelihood scale calculation requires estimation of optic disc size, provided by DiscNet (VGG19 network), and the minimum rim-to-disc ratio (mRDR) or absent rim width (ARW), provided by RimNet (InceptionV3/LinkNet segmentation model). To build RimNet's dataset, glaucoma specialists marked optic disc rim and cup boundaries on ODPs. The "ground truth" mRDR or ARW was calculated. For DiscNet's dataset, corresponding OCT images provided "ground truth" disc size. Optic disc photographs were split into 80/10/10 for training, validation, and testing, respectively, for RimNet and DiscNet. DDLSNet estimation was tested against manual grading of DDLS by clinicians with the average score used as "ground truth." Reproducibility of DDLSNet grading was evaluated by repeating DDLS estimation on a dataset of nonprogressing paired ODPs taken at separate times.

Main Outcome Measures: The main outcome measure was a weighted kappa score between clinicians and the DDLSNet pipeline with agreement defined as ± 1 DDLS score difference.

Results: RimNet achieved an mRDR mean absolute error (MAE) of 0.04 (± 0.03) and an ARW MAE of 48.9 (± 35.9) degrees when compared to clinician segmentations. DiscNet achieved 73% (95% confidence interval [CI]: 70%, 75%) classification accuracy. DDLSNet achieved an average weighted kappa agreement of 0.54 (95% CI: 0.40, 0.68) compared to clinicians. Average interclinician agreement was 0.52 (95% CI: 0.49, 0.56). Reproducibility testing demonstrated that 96% of ODP pairs had a difference of ≤ 1 DDLS score.

Conclusions: DDLSNet achieved moderate agreement with clinicians for DDLS grading. This novel approach illustrates the feasibility of automated ODP grading for assessing glaucoma severity. Further improvements may be achieved by increasing the number of incomplete rims sample size, expanding the hyperparameter search, and increasing the agreement of clinicians grading ODPs. *Ophthalmology Science* 2023;3:100255 © 2022 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Supplemental material available at www.ophtalmologyscience.org.

Glaucoma is the leading cause of irreversible blindness worldwide, with an estimated 80 million people affected in 2020 and a projected rise to 111.8 million people by 2040.^{1,2} Glaucoma is asymptomatic in the early stages; untested individuals often remain undiagnosed until advanced symptoms are present. In developed countries, up to 70% of patients with glaucoma are undiagnosed, a number that rises in areas with less access to screening.³ While patients with mild glaucoma have a quality of life comparable to that of healthy patients, the quality of life drastically decreases with more advanced glaucoma.⁴

Early diagnosis and treatment allow for preservation of patient quality of life and are at the forefront of strategies for reducing disease burden.⁵

Glaucoma diagnostic methods can be grouped into 2 categories: (1) techniques that evaluate structural changes in the eye and (2) techniques that evaluate functional changes in vision. Among those assessing structural changes, OCT and fundus photography are most often used in clinical practice. While OCT has been shown to have a high sensitivity for detection of structural glaucomatous changes, the high cost of the technique often restricts the device to

large eye clinics or centers.^{6,7} This is especially problematic given that developing regions have the highest rates of undiagnosed glaucoma.³ Moreover, the World Glaucoma Association considers the largest barrier to glaucoma screening to be cost.⁸ In contrast to OCT, fundus photography stands as a lower cost option; new advances such as telemedicine screening and smartphone funduscopy have made fundus photography a feasible and financially viable option even in remote locations.^{9,10}

While OCT and fundus photography allow for the structural findings to be captured, a mechanism is needed to classify such changes and correlate them with functional glaucomatous damage. The disc damage likelihood scale (DDLS) is one such approach. The DDLS is a well-established grading scale to correlate glaucomatous damage with progression of fundus photographs.^{3,11,12} The DDLS has been incorporated into the eye health professional guidelines for optometrists and ophthalmologists.^{12,13} The interobserver agreement of DDLS, even among glaucoma specialists, can vary from 85% based on optic disc photographs (ODPs) to 70% based on clinical examination, although intraobserver reliability is high.¹⁴ This is especially troubling as DDLS scores can be used as the basis for referral by a variety of eye health professionals, and improper grading may result in missed opportunities for early intervention.¹²

An ideal screening tool would be high-throughput, accurate, and reliable with high specificity. With the advent of neural network models and an increase in image processing capabilities, high specificity with acceptable sensitivity, together with high throughput, may be achieved with a neural network-based pipeline.¹⁵ In this paper, we present DDLSNet, a neural network pipeline which aims to accurately grade DDLS based on ODPs with a combination of a rim segmentation neural network (RimNet) and a disc size classification network (DiscNet).

Methods

The DDLS grading criteria was created by Spaeth et al.¹¹ Their original grading schema is shown as [Supplemental Figure 1](#). The DDLS score is determined by 2 features of the optic disc: (1) the disc size and (2) the narrowest rim width. Progression of glaucomatous damage is seen as enlargement of the optic disc cup and subsequent thinning of the optic disc rim. This thinning can be measured by the minimum rim-to-disc ratio (mRDR) in intact rims. However, in severe glaucoma, the rim can be completely absent in certain areas. In these cases, the angle for which the rim is completely lost is measured. We call this the "absent rim width" (ARW) and we call these rims "incomplete." These 3 features, mRDR, ARW, and disc size, are the metrics needed to calculate DDLS. The latter is crucial as the significance of mRDR or ARW varies depending on disc size.¹¹ Therefore, the DDLSNet pipeline consists of 2 components: (1) RimNet, which performs rim and cup segmentation and calculates mRDR or ARW, and (2) DiscNet, which classifies the size of the optic disc into small, average, and large. This 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. Informed consent was waived by the UCLA Institutional Review Board.

Database

Our image database was based on a collection of all the ODPs available in the University of California, Los Angeles Stein Eye Glaucoma Division. For the RimNet database, 3 glaucoma specialists manually created a mask of the optic disc rim and optic disc cup for each funduscopy image using the image editing program GIMP. These masks were used as the ground truth. The RimNet dataset had 2 inclusion criteria. The images had to show signs of glaucomatous damage and the images had to be in focus and with discernable posterior pole and vasculature details, both as deemed by 2 board-certified glaucoma specialists. The exclusion criteria were concurrent non-glaucoma disease including optic neuritis, optic disc neovascularization, and vitreous hemorrhage that would impair visualization of the posterior pole. The demographic information for the RimNet dataset is presented in [Table 1](#). [Table 2](#) presents the glaucoma diagnoses for the RimNet dataset. These requirements result in a database that displays the full range of glaucomatous changes to the optic disc rim, ranging from mild optic disc rim narrowing in early-stage glaucoma to absent optic disc rim in severe glaucoma.

The DiscNet database consisted of ODPs with available corresponding Cirrus high-definition OCT Optic Disc Cubes (200 × 200). The size of the Bruch's membrane as measured by Cirrus OCT was used as a proxy for disc area and was used to categorize the disc size into small, average, or large optic discs. The ODPs had to be of good quality—in focus with an unobstructed view of the posterior pole—as determined by a board-certified glaucoma specialist. The OCT images were required to have a good quality (signal strength > 6) and be free of artifacts based on the review of printouts. To examine reliability, a database of nonprogressing glaucomatous eyes was created. Each eye had 2 ODPs available taken < 4 years apart, which were deemed stable as confirmed by a glaucoma specialist. The time restriction was imposed to increase the population included but decrease the chance of glaucoma progression between the 2 photos.

RimNet

RimNet consists of a preprocessing step of contrast enhancement, an optic disc rim and cup segmentation model, and an image analysis step to calculate the mRDR for intact rims and ARW for incomplete rims. This latter case occurs in eyes with severe glaucomatous damage. The model was optimized by submitting it to a hyperparameter search with rim intersection over union (RimIoU) as the metric. The included hyperparameters were the neural network structure, learning rate, loss function, and optimizer.^{16–27} [Table 3](#) lists the hyperparameter search space. Fifty total hyperparameter combinations were trained with the Keras Tuner Library with RimNet proficiency, measured as an IoU. The RimNet model was trained, validated, and tested on a database of images from the University of California, Los Angeles Stein Eye Glaucoma Division with an 80/10/10 split.

DiscNet

DiscNet is a deep neural network developed to assign disc size as small, average, or large as an essential process in DDLS grading. The disc photographs included scanned digitized slides and digital photographs. Disc size information taken from paired OCT data was used as the ground truth. While the original DDLS grading defined small, average, and large discs as diameters of < 1.50 mm, between 1.50 mm and 2.00 mm, and > 2.00 mm respectively,¹¹ we modified the cutoffs slightly to ≤ 1.44 mm, 1.44 mm to 2.28 mm, and ≥ 2.28 so that the 3 disc size categories had more

Table 1. Demographic Data for Dataset Listing the Gender, Age, and Racial Distribution by Camera Type

	Slide Images	Digital Camera 1	Digital Camera 2	Digital Camera 3
Gender distribution				
Female	407	119	55	12
Male	302	85	44	11
Age distribution				
Mean	60.72	67.13	72.80	66.92
SD	13.48	17.43	12.75	17.71
Median	61.87	71.06	73.91	72.37
IQR	15.79	16.33	12.86	22.54
Min	9.36	6.92	16.19	17.48
Max	90.05	96.10	94.41	86.17
Race distribution				
Asian	90	34	24	2
Black	63	22	8	1
Hispanic	66	20	16	6
White	366	100	45	12
Other	53	5	3	0
Unknown	71	22	3	2

IQR = interquartile range; SD = standard deviation

evenly distributed sample sizes. This sorted our available data into a 15/70/15 split for small, average, and large discs.

DiscNet was first built with transfer learning with model weights from ImageNet before submitting it to training.²⁸ In transfer learning, a subsection of the model’s layers are "unlocked" to hone the transferred model performance for a specified task. Often, only the last layer of the model is unlocked but more can be unlocked if needed. The proportion of the model layers allowed to be updated is termed the "tuning fraction" of the model. DiscNet was trained in 2 phases, each with a unique learning rate. The first phase allowed for only the last layer of the model to be trained, while the second phase allowed the weights in a subsection of the model layers, the tuning fraction, to be updated.

A hyperparameter search was completed to select the optimal learning rates in both phases, the tuning fraction, the optimizer, and the network architecture. Table 3 lists the hyperparameter search space, from which each hyperparameter was selected from. Thirty

total hyperparameter combinations were trained with the Keras Tuner library, with classification accuracy as the optimized metric.²⁹

DDLSNet Pipeline

The mRDR and ARW from RimNet and the disc size from DiscNet were used to calculate the DDLS score. A full diagram of our pipeline is shown in Figure 1. DDLSNet was evaluated against a ground truth database of ODPs, which 3 glaucoma specialists had graded with DDLS. The weighted kappa agreement ± 1 DDLS grade between the DDLSNet’s output and the average of the grades of 3 glaucoma specialists was measured. The average of the interobserver agreement for clinicians was also measured.

Evaluating DDLSNet reliability is necessary, as physician intraobserver accuracy for DDLS grading should be matched by our proposed system for it to be clinically useful. A database of pairs of fundus photos taken no more than 4 years apart of 781 nonprogressing glaucomatous eyes was used to test DDLSNet reliability. Each image was graded via DDLSNet, and the difference between the 2 images for each eye was recorded. Glaucoma specialists verified that the eyes were nonprogressing, based on evaluation of the disc photos and the visual fields.

Table 2. RimNet Database Diagnosis Listing the 1208 ODPs Used in RimNet Training, Validation, and Testing in an 80/10/10 Split Respectively

Diagnosis	Count
Primary open-angle glaucoma	530
Glaucoma suspect	403
Chronic angle-closure glaucoma	71
Low-tension glaucoma	47
Secondary open-angle glaucoma	35
Capsular glaucoma with pseudoexfoliation	33
Anatomical narrow angle	27
Glaucoma secondary to eye infection	24
Pigmentary glaucoma	15
Secondary angle closure	11
Congenital glaucoma	7
Juvenile glaucoma	3
Acute angle-closure glaucoma	2

ODPs = optic disc photographs; RimNet = rim segmentation.

Evaluation Criteria

The main evaluation criterion was the weighted kappa agreement between DDLSNet and physicians with the ground truth database. Interobserver and intraobserver agreement was also measured as secondary evaluation criteria.

Results

RimNet was trained, validated, and tested on 1208 ODPs with an 80/10/10 split respectively. The mean age was 63.7 (± 14.9) years with a male:female ratio of 43:57. DiscNet was trained, validated, and tested on a database of 11 536 eyes in an 80/10/10 split. The mean age was 67.6 (± 14.5) and had a male:female ratio of 58:42. DDLSNet was tested on 120 ODPs from the RimNet test set manually graded

Table 3. Hyperparameter Search Space for RimNet: An InceptionV3/LinkNet Architecture, Binary Cross-Entropy Loss Function, Learning Rate of 10^{-3} , and Adam Optimizer Were Selected, RimIoU was the Optimized Metric

Hyperparameters	
Encoders	MobileNetV2, ResNet34, EfficientnetB0, InceptionV3, ResNet101, VGG16, ResNet50
Decoders	U-Net, FPN, LinkNet, PSPnet
Loss function	Binary_Crossentropy, Binary_Focal_Los
Learning rate	10^{-3} , 10^{-4} , 10^{-5} , 10^{-6}
Optimizer	Adam, SGD

RimIoU = rim intersection over union; RimNet = rim segmentation model.

based on DDLS by 3 glaucoma specialists. Reproducibility of DDLSNet was evaluated on 781 eyes, each with 2 ODPs available (mean age = 73.8 (± 11.4) years, male:female ratio = 43:57). The eyes were all classified as non-progressing by a glaucoma specialist based on review of the ODPs. The demographic data for the 4 cohorts are presented in Table 4. The code used to train, run, and evaluate DDLSNet can be found on our public repository at <https://github.com/TylerADavis/GlaucomaML>.

Model Architecture and Hyperparameter Search

After exploring 30 different combinations of hyperparameters through random search, the following hyperparameters were identified as providing the highest classification accuracy for DiscNet: VGG19 architecture,

phase 1 learning rate of 1^{-4} , phase 2 learning rate of 1^{-5} , tuning fraction of 0.5, and Adam optimizer.^{18,22-24,27,30,31} VGG19 is a 19-layer convolutional neural network published in 2015 that has previously been used in medical image analysis.^{22,32,33} For RimNet, 50 different combinations were examined through a random search, which resulted as follows: (1) InceptionV3/LinkNet architecture, (2) binary cross-entropy loss function, (3) learning rate of 10^{-3} , and (4) Adam optimizer.¹⁶⁻²⁷ InceptionV3 was first published in 2015, outperforming popular encoders at the time with a fraction of the computation costs.²⁷ It has been previously used in medical segmentation.^{34,35} LinkNet is a lightweight decoder first published in 2017.²⁶ Given the computational restrictions of our workstation, which uses NVIDIA RTX 2080 Ti graphics cards, these were appropriate choices.

RimNet

The RimNet evaluation criteria were the mean absolute error (MAE) for mRDR for intact rims and the MAE for ARW for incomplete rims between physician grading and RimNet grading, with a secondary evaluation criterion of the RimIoU. The intersection over union (IoU) is a commonly used measure for segmentation accuracy. RimNet achieved an mRDR MAE of 0.04 (± 0.03), an ARW MAE of 48.9 (± 35.9), and a RimIoU of 0.68.

DiscNet

DiscNet raw classification accuracy was found to be 73% (95% confidence interval [CI]: 70, 75) across a test set of

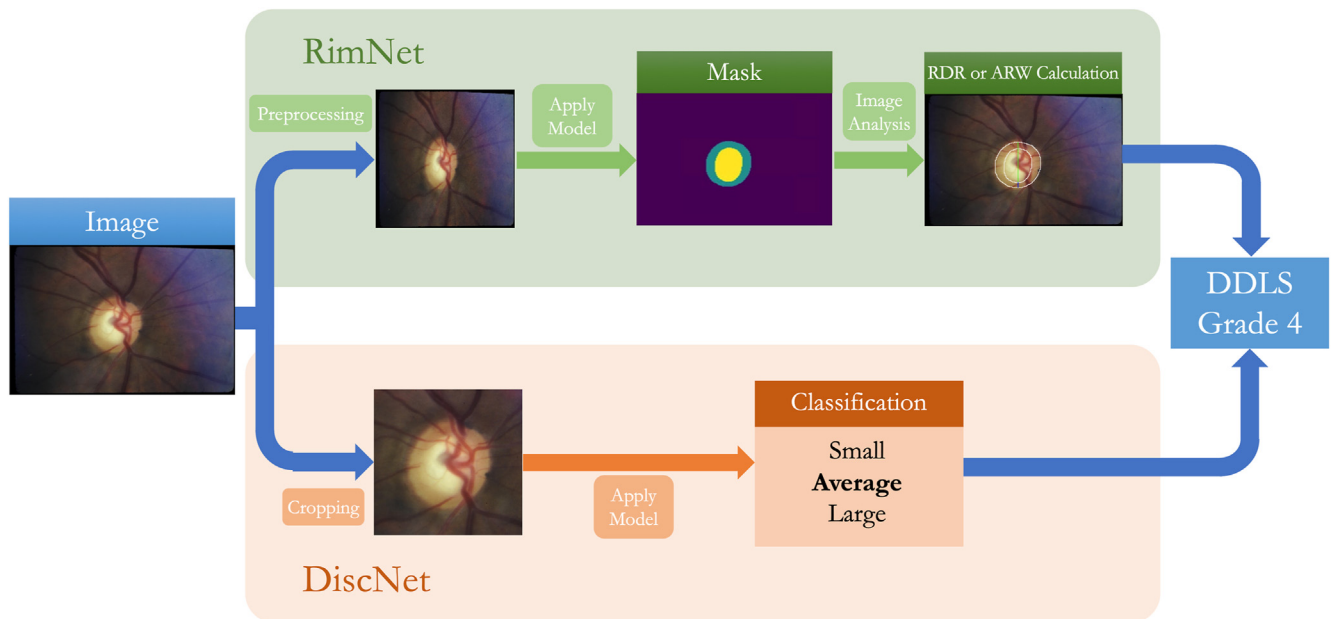


Figure 1. DDLSNet pipeline showcasing both the rim segmentation and disc size classification results. Once optic disc photographs are submitted to the pipeline, rim segmentation (RimNet) calculates the minimum rim-to-disc ratio (mRDR) for intact rims, or absent rim width (ARW) for incomplete rims, while disc size classification (DiscNet) estimates the disc size. The disc size and mRDR ratio or ARW are then used to calculate the disc damage likelihood scale (DDLS) score.

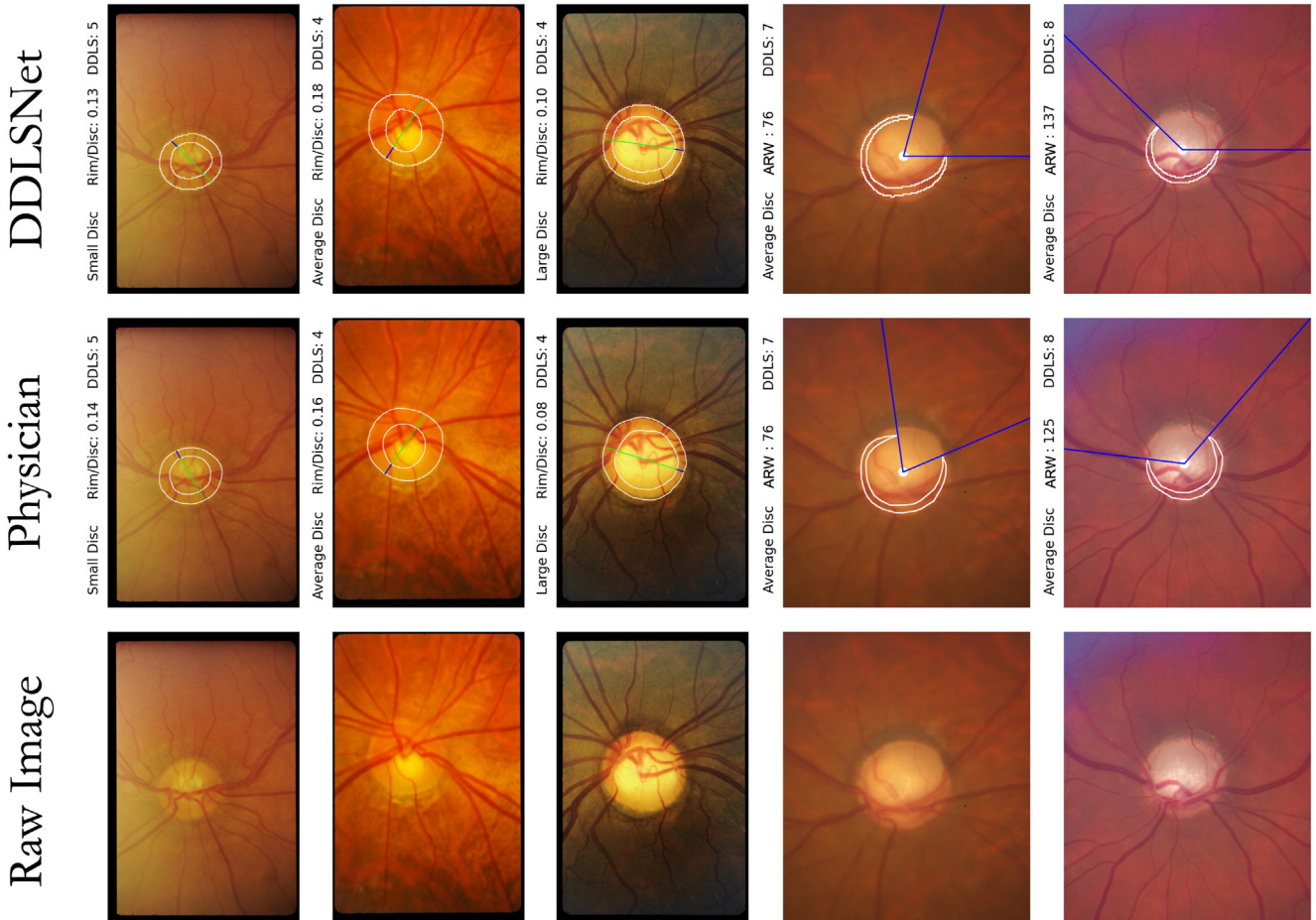


Figure 2. DDLSNet Results on sample optic disc photographs. The white indicates the physician or DDLNet rim segmentation. The blue line on the rim indicates the shortest rim width detected. The green line indicates the disc diameter. Left-most column: raw optic disc photographs. Middle column: delineation of the disc and cup margin by clinicians. Right-most column: DDLSNet grading. The absent rim width is calculated in eyes with complete rim loss in certain regions. The minimum rim-to-disc ratio is calculated in eyes with intact optic disc rim. The annotations above the photographs on the middle and right-most columns represent the disc size, absent rim width or rim-to-disc ratio, and disc damage likelihood scale (DDLS) grade.

1137 images, which included both scanned slides and digitally acquired ODPs. Broken down by category, DiscNet had a classification accuracy of 62% (95% CI: 55, 70) for small discs, 77% (95% CI: 74, 80) for average discs, and 60% (95% CI: 52, 68) for large discs. Notably, only 3 small discs out of 234 (1.2%) were mistakenly classified as large and only 2 large discs out of 146 (1.3%) were mistakenly classified as small.

DDLSNet

DDLSNet was evaluated on a testing database of 120 ODPs. Results for a representative sample of ODPs are displayed in Figure 2. Three glaucoma specialists also graded the same 120 funduscopy images with DDLS. The weighted kappa agreement between the average grading of the 3 glaucoma specialists and DDLSNet was 0.54 (95% CI: 0.4, 0.68). A full breakdown of the results can be found in Table 5. The model matched the kappa scores between physicians, which included 0.49, 0.52, and 0.56, averaged at 0.52. DDLSNet

reproducibility was measured by evaluating pairs of nonprogressing ODPs. Of the 781 pairs of eyes, 485 (62%) had DDLS difference of 0, 267 (34%) had a DDLS difference of 1, 28 (4%) had a DDLS difference of 2, and 1 (0.1%) had a DDLS difference of 3 (Table 6).

Discussion

We present an automated pipeline for estimating the DDLS score with ODPs in patients with suspected or established glaucoma to facilitate detection and monitoring of the disease. The DDLSNet weighted kappa agreement of 0.54 (95% CI 0.40–0.68) demonstrated moderate agreement with clinician grading and matching interclinician agreement. Moreover, the DDLSNet reproducibility was high with 96% of 781 nonprogressing eyes found to have ± 1 DDLS grade difference on stable pairs of ODPs.

Automated glaucoma grading with ODPs has been evolving. Most experimental approaches focus on accurate

Table 4. Demographics Characteristics for the Datasets Used for RimNet, DiscNet, DDLSNet, and the DDLSNet Reliability

	DDLSNet Test Set	RimNet	DiscNet	DDLSNet Reliability
Total no. of images	120	1208	11 536	1562
Total no. of eyes	109	1021	5213	781
Gender: male/female	45:55	43:57	58:42	43:57
Age: mean (SD)	65.9 (± 14.8)	63.7 (± 14.9)	67.6 (± 14.5)	73.8 (± 11.4)
DDLS grading by physician				
1	0			
2	12			
3	29			
4	30			
5	12			
6	19			
7	12			
8	4			
9	2			
10	0			

DDLS = disc damage likelihood scale; DiscNet = disc size classification; RimNet = rim segmentation; SD = standard deviation. The DDLS distribution for our test set of 120 images, graded by glaucoma specialists.

detection of the cup-to-disc ratio with techniques ranging from thresholding to level setting to artificial intelligence models.³⁶ 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 71% with accuracy subjectively defined as "good" or "very good."³⁷ More recently, Kumar and Bindu used U-Net,³⁸ a segmentation neural network architecture, to achieve an IoU of 87.9% in optic disc segmentation.³⁹ Our algorithm for measuring mRDR, RimNet, combines both the image processing techniques used in older segmentation studies and the artificial intelligence of newer studies to achieve a high-efficacy segmentation on a variety of ODPs.

Cup-to-disc ratio has been repeatedly shown to be inferior to DDLS in grading glaucomatous damage.⁴⁰ Several papers addressed detection of the minimum optic disc rim width, an important component of calculating the DDLS score.^{41–43} However, few have used automated DDLS calculation due to the complexity of the challenge. Two studies examined the results of proprietary software built into a 3-dimensional stereographic camera (Kowa Nonmyd WX 3D, Kowa).^{44,45} The camera automatically displays the DDLS grade in its final report. The study by Han et al showed moderate agreement (weighted kappa value, 0.59) with 1 glaucoma specialist.⁴⁴ This study has 2 limitations compared to our study. First, the study only evaluates the

camera against 1 glaucoma specialist rather than the 3 in our study. Second, such camera-specific software does not offer the generalizability of DDLSNet. While functional on certain cameras, such software would not offer the generalizability of DDLSNet. A third study provided clinical validation for RIA-G, an automated cloud-based optic nerve head analysis software that has been reported to be able to grade ODPs based on DDLS.⁴⁶ This study showed a moderate Kappa agreement of 0.62 (0.55, 0.69) between 3 glaucoma specialists and the software. However, the validation set favored photographs of mild glaucoma (average DDLS grade 3, DDLS 1–7 included) and required fundus photographs with a 30-degree field of view.⁴⁶ Our validation set has a wider spectrum of glaucomatous damage (average DDLS grade 4.5, DDLS 2–9) and DDLSNet does not require a 30-degree field of view. Moreover, the RIA-G optic disc cup and disc detection software operates based on contrast detection which would be impaired in photographs with bright artifacts and abnormal pathology.⁴⁶ A fourth study implemented a partial-DDLS grading using active discs, where a circular disc shape was assumed and DDLS grades were grouped into normal, moderate, and severe categories.⁴⁷ The model achieved a category accuracy of 89%.⁴⁷ DDLSNet improves upon this study by directly comparing 10 DDLS grades rather than 3 categories. Additionally, our network

Table 5. Kappa Agreement between DDLSNet and Glaucoma Specialist Grading

Graders	Kappa (95% CI)
Grader 1 versus Grader 2	0.52 (0.32, 0.72)
Grader 1 versus Grader 3	0.56 (0.35, 0.77)
Grader 2 versus Grader 3	0.49 (0.29, 0.7)
Grader average versus DDLSNet	0.54 (0.4, 0.68)

CI = confidence interval; DDLS = disc damage likelihood scale.

Table 6. Difference in DDLSNet Grading between Paired Images of Nonprogressing Optic Disc Photographs.

DDLS Difference	Number of Images
0	481
1	267
2	28
3	1

DDLS = disc damage likelihood scale. All photographs were taken within 4 years of each other (total number of images = 781).

accounts for disc size variations through DiscNet and intact and incomplete rims through RimNet. It is unclear if and to what extent the above studies included optic discs with areas of absent optic disc rim widths, which constitute the most severe DDLS grades.

DDLSNet is the most accurate and generalizable approach developed to date for several reasons. First, it was validated on ODPs with a wide breadth of glaucomatous damage. This included ODPs with areas of absent optic disc rims. Second, it makes no assumptions of the size or shape of the optic disc when grading size. Third, it is built on a neural network model rather than thresholding or contrast-based algorithms, which are limited in learning capacity. Finally, it is not restricted to specific fundus cameras, making it more amenable for use in mobile settings where smartphones or portable fundus cameras can be used for fundus photography.

The shortcomings of our study need to be considered. Expanding the dataset could improve performance of both RimNet and DiscNet. The models will also have to be trained on images with significant concurrent pathologies, such as severe diabetic retinopathy and macular degeneration. The hyperparameter search was limited by the processing power and memory constraints of our NVIDIA RTX 2080 Ti graphics cards, which were used to train the model.

A more extensive hyperparameter search can be done using larger architectures, such as ResNet152, with more powerful computing hardware. Following the hyperparameter search, the selected DiscNet model and RimNet model had the highest accuracy and RimIoU respectively. However, their loss functions had evidence of possible overfitting. This would need to be addressed in a future study. Finally, the number of physicians grading and segmenting fundus images could be increased to allow DDLSNet to learn a wider consensus of gradings.

In conclusion, DDLSNet offers a unique, high-efficacy, high-throughput, reliable DDLS grading system, which is well-suited to perform as a screening, diagnostic, and prognostic tool for identifying and classifying glaucomatous damage and monitoring disease progression. DDLSNet is also well-suited for mobile applications in a variety of settings, including use by individuals without extensive ophthalmic training, such as a neurology resident using a phone camera attachment or optometrists seeking to better evaluate their patients' fundus images. Future study directions include increasing the number of physician graders and examining the implementation in remote areas with limited access. With powerful computing technology, glaucoma screening could be enhanced and widely disseminated, improving clinical outcomes for patients.

Footnotes and Disclosures

Originally received: July 5, 2022.

Final revision: October 3, 2022.

Accepted: November 3, 2022.

Available online: November 12, 2022. Manuscript no. XOPS-D-22-00144

¹ University of California Los Angeles David Geffen School of Medicine, Los Angeles, California.

² Department of Computer Science, University of California Los Angeles, Los Angeles, California.

³ University of California Los Angeles Jonathan and Karin Fielding School of Public Health, Los Angeles, California.

⁴ Department of Biostatistics, University of California Los Angeles, Los Angeles, California.

⁵ Glaucoma Division, Jules Stein Eye Institute, Los Angeles, California.

Disclosure:

All authors have completed and submitted the ICMJE disclosures form.

The author has made the following disclosure:

Financial Support: Unrestricted departmental grant from Research to Prevent Blindness, NIH grant R01-EY029792 (K.N.M.), Payden Glaucoma Fund, Simms/Mann Family Foundation and an unrestricted research grant from Heidelberg Engineering (K.N.M.).

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

HUMAN SUBJECTS: Human Subjects were used in this study. 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. Informed consent was waived by the UCLA Institutional Review Board.

No animal subjects were used in this study.

Meeting Presentation: The Association for Research in Vision and Ophthalmology Annual Meeting, 2022.

Author Contributions:

Conception and design: Rasheed, Davis, Morales, Fei, Grassi, De Gainza, Nouri-Mahdavi, Caprioli

Analysis and interpretation: Rasheed, Davis, Morales, Fei, Grassi, De Gainza, Nouri-Mahdavi, Caprioli

Data collection: Rasheed, Davis, Morales, Fei, Grassi, De Gainza, Nouri-Mahdavi, Caprioli

Obtained funding: N/A

Overall responsibility: Rasheed, Davis, Morales, Fei, Grassi, De Gainza, Nouri-Mahdavi, Caprioli

Abbreviations and Acronyms:

ARW = absent rim width; **CI** = confidence interval; **DDLS** = disc damage likelihood scale; **DiscNet** = disc size classification model; **MAE** = mean average error; **ODP** = optic disc photograph; **mRDR** = minimum rim-to-disc ratio; **RimIoU** = rim intersection over union; **RimNet** = rim segmentation model.

Keywords:

DDLS, Glaucoma, Neural network, Segmentation.

Correspondence:

Joseph Caprioli, MD, Glaucoma Division, Jules Stein Eye Institute, 100 Stein Plaza, Los Angeles, CA 90095. E-mail: caprioli@jsei.ucla.edu.

References

1. Giangiacomo A, Coleman AL. The epidemiology of glaucoma. In: *Glaucoma*. Berlin, Heidelberg: Springer Berlin Heidelberg; 13–21.
2. Tham YC, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:2081–2090.
3. Tan NYQ, Friedman DS, Stalmans I, et al. Glaucoma screening: where are we and where do we need to go? *Curr Opin Ophthalmol*. 2020;31:91–100.
4. Goldberg I, Clement CI, Chiang TH, et al. Assessing quality of life in patients with glaucoma using the glaucoma quality of life-15 (GQL-15) questionnaire. *J Glaucoma*. 2009;18(1):6-12.
5. Leske MC, Heijl A, Hussein M, et al. Factors for glaucoma progression and the effect of treatment: the early manifest glaucoma trial. *Acta Ophthalmol*. 2003;121(1):48-56.
6. Shelton RL, Jung W, Sayegh SI, et al. Optical coherence tomography for advanced screening in the primary care office. *J Biophotonics*. 2014;7:525–533.
7. Kim S, Crose M, Eldridge WJ, et al. Design and implementation of a low-cost, portable OCT system. *Biomedical Optics Express*. 2018;9(3):1232–1243.
8. Weinreb RN, Healey PR, Topouzis F. *Glaucoma screening: the 5th consensus report of the world glaucoma association*. Kugler Publications; 2008. Available at: <https://wga.one/wgal/consensus-5/>. Accessed March 13, 2022.
9. Panwar N, Huang P, Lee J, et al. Fundus photography in the 21st century—a review of recent technological advances and their implications for worldwide healthcare. *Telemed J E Health*. 2016;22:198.
10. Nazari Khanamiri H, Nakatsuka A, El-Annan J. Smartphone fundus photography. *J Vis Exp*. 2017:e55958.
11. Spaeth GL, Henderer J, Liu C, et al. The disc damage likelihood scale: reproducibility of a new method of estimating the amount of optic nerve damage caused by glaucoma. *Trans Am Ophthalmol Soc*. 2002;100:181.
12. Formichella P, Annot R, Zeri F, Tatham AJ. The role of the disc damage likelihood scale in glaucoma detection by community optometrists. *Ophthalmic Physiol Opt*. 2020;40:752–759.
13. Kara-José AC, Melo LAS, Esporcate BLB, et al. The disc damage likelihood scale: diagnostic accuracy and correlations with cup-to-disc ratio, structural tests and standard automated perimetry. *PLoS One*. 2017;12:e0181428.
14. Henderer JD, Liu C, Kesen M, et al. *Reliability of the Disk Damage Likelihood Scale*. *Am Journal Ophthalmol*; 2003;135:44–48.
15. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019;25:954–961.
16. Peleg B, Rubinstein R. *The cross entropy method for classification*. *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*. ICML; 2005:561–568. https://icml.cc/Conferences/2005/proceedings/papers/071_CrossEntropy_MannorEtAl.pdf. Accessed June 10, 2022.
17. Ruder S. An overview of gradient descent optimization algorithms. Available at: <https://arxiv.org/abs/1609.04747v2>; 2016. Accessed June 10, 2022.
18. Kingma DP, Ba JL. Adam: a method for stochastic optimization. 3rd international conference on learning representations, ICLR 2015 - conference track proceedings. Available at: <https://arxiv.org/abs/1412.6980v9>; 2014. Accessed June 10, 2022.
19. Zhao H, Shi J, Qi X, et al. *Pyramid Scene Parsing Network. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2016*. IEEE Computer Science; 2017:6230–6239. Available at: <https://arxiv.org/abs/1612.01105v2>. Accessed June 10, 2022.
20. Lin T-Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. <https://arxiv.org/abs/1612.03144v2>; 2016. Accessed June 10, 2022.
21. Weng W, Zhu X. U-net: convolutional networks for biomedical image segmentation. *IEEE Access*. 2015;9:16591–16603.
22. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd international conference on learning representations, ICLR 2015 - conference track proceedings 2014. Available at: <https://arxiv.org/abs/1409.1556v6>. Accessed June 10, 2022.
23. Tan M, Le Qv. EfficientNet: rethinking model scaling for convolutional neural networks. 36th International Conference on Machine Learning, ICML 2019;2019-June:10691–10700. Available at: <https://arxiv.org/abs/1905.11946v5>. Accessed June 10, 2022.
24. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015;2016-December:770–778. Available at: <https://arxiv.org/abs/1512.03385v1>. Accessed June 10, 2022.
25. Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2018:4510–4520. <https://arxiv.org/abs/1801.04381v4>. Accessed June 10, 2022.
26. Chaurasia A, Culurciello E. LinkNet: exploiting encoder representations for efficient semantic segmentation. 2017 IEEE visual communications and image processing, VCIP 2017 2017;2018-January:1–4. Available at: <http://arxiv.org/abs/1707.03718>. Accessed March 13, 2022.
27. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. Proceedings of the IEEE computer society conference on computer vision and pattern recognition 2015;2016-December:2818–2826. <https://arxiv.org/abs/1512.00567v3>. Accessed March 13, 2022.
28. Deng J, Dong W, Socher R, et al. *ImageNet: A large-scale hierarchical image database*. 2010, 248–255. Available at: https://image-net.org/static_files/papers/imagenet_cvpr09.pdf. Accessed March 13, 2022.
29. O'Malley T, Bursztein E, Long J, et al. *Keras Tuner*. GitHub Repository; 2019. https://keras.io/keras_tuner/.
30. Hinton G, Srivastava N, Swersky K. Neural Networks for Machine Learning Lecture 6a Overview of mini-batch gradient descent. Available at: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf Accessed June 14, 2022.

31. Sutskever I, Martens J, Dahl G, Hinton G. *On the importance of initialization and momentum in deep learning*. PMLR; 2013. <https://proceedings.mlr.press/v28/sutskever13.html>. Accessed June 14, 2022.
32. Ahuja S, Panigrahi BK, Gandhi T. *Transfer Learning Based Brain Tumor Detection and Segmentation using Superpixel Technique*. 2020 International Conference on Contemporary Computing and Applications, IC3A 2020. IEEE; 2020: 244–249. <https://ieeexplore.ieee.org/document/9077067>. Accessed September 18, 2022.
33. Mateen M, Wen J, Nasrullah, et al. Fundus image classification using VGG-19 architecture with PCA and SVD. *Symmetry*. 2019;11:1.
34. Shoaib M, Sayed N. YOLO object detector and inception-V3 convolutional neural network for improved brain tumor segmentation. *Traitement Du Signal*. 2022;39:371–380.
35. Salama WM, Aly MH. Deep learning in mammography images segmentation and classification: automated CNN approach. *Alexandria Eng J*. 2021;60:4701–4709.
36. Thakur N, Juneja M. Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma. *Biomed Signal Process Control*. 2018;42:162–189.
37. Chrástek R, Wolf M, Donath K, et al. Automated segmentation of the optic nerve head for diagnosis of glaucoma. *Med Image Anal*. 2005;9:297–314.
38. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Lecture Notes Computer Sci*. 2015;9351:234–241.
39. Kumar ES, Bindu CS. Two-stage framework for optic disc segmentation and estimation of cup-to-disc ratio using deep learning technique. *J Ambient Intell Humaniz Comput*. 2021. <https://doi.org/10.1007/s12652-021-02977-5>; Accessed March 12, 2022.
40. Cheng KKW, Tatham AJ. Spotlight on the disc-damage likelihood scale (DDLS). *Clin Ophthalmol*. 2021;15: 4059–4071.
41. Haleem MS, Han L, van Hemert J, Li B. Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: a review. *Comput Med Imaging Graph*. 2013;37: 581–596.
42. Choudhary K, Tiwari S. ANN glaucoma detection using cup-to-Disk ratio and neuroretinal rim. *Int J Comput Appl*. 2015;111:975–8887.
43. Issac A, Partha Sarathi M, Dutta MK. An adaptive threshold based image processing technique for improved glaucoma detection and classification. *Comput Methods Programs Biomed*. 2015;122:229–244.
44. Han JW, Cho SY, Kang KD. Correlation between optic nerve parameters obtained using 3D nonmydriatic retinal camera and optical coherence tomography: interobserver agreement on the disc damage likelihood scale. *J Ophthalmol*. 2014;2014: 931738.
45. Gnaneswaran P, Devi S, Balu R, et al. Agreement between clinical versus automated disc damage likelihood scale (DDLS) staging in Asian Indian eyes. *Invest Ophthalmol Vis Sci*. 2013;54:4806.
46. Singh Di, Gunasekaran S, Hada M, Gogia V. Clinical validation of RIA-G, an automated optic nerve head analysis software. *Indian J Ophthalmol*. 2019;67:1089–1094.
47. Kumar JRH, Seelamantula CS, Kamath YS, Jampala R. Rim-to-Disc ratio outperforms cup-to-disc ratio for glaucoma pre-screening. *Sci Rep*. 2019;9:1–9.