Research Article

An Artificial-Intelligence- and Telemedicine-Based Screening Tool to Identify Glaucoma Suspects from Color Fundus Imaging

Alauddin Bhuiyan ⓘ,[1,2] Arun Govindaiah,[1] and R. Theodore Smith[2]

[1]*iHealthscreen Inc., New York, NY, USA*
[2]*New York Eye and Ear Infirmary, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

Correspondence should be addressed to Alauddin Bhuiyan; bhuiyan@ihealthscreen.org

Background and Objective. Glaucomatous vision loss may be preceded by an enlargement of the cup-to-disc ratio (CDR). We propose to develop and validate an artificial-intelligence-based CDR grading system that may aid in effective glaucoma-suspect screening. *Design, Setting, and Participants*. 1546 disc-centered fundus images were selected, including all 457 images from the Retinal Image Database for Optic Nerve Evaluation dataset, and images were randomly selected from the Age-Related Eye Disease Study and Singapore Malay Eye Study to develop the system. First, a proprietary semiautomated software was used by an expert grader to quantify vertical CDR. Then, using CDR below 0.5 (nonsuspect) and CDR above 0.5 (glaucoma suspect), deep-learning architectures were used to train and test a binary classifier system. *Measurements*. The binary classifier, with glaucoma suspect as positive, is measured using sensitivity, specificity, accuracy, and AUC. *Results*. The system achieved an accuracy of 89.67% (sensitivity, 83.33%; specificity, 93.89%; and AUC, 0.93). For external validation, the Retinal Fundus Image Database for Glaucoma Analysis dataset, which has 638 gradable quality images, was used. Here, the model achieved an accuracy of 83.54% (sensitivity, 80.11%; specificity, 84.96%; and AUC, 0.85). *Conclusions*. Having demonstrated an accurate and fully automated glaucoma-suspect screening system that can be deployed on telemedicine platforms, we plan prospective trials to determine the feasibility of the system in primary-care settings.

1. Introduction

Glaucoma is a group of diseases that damage the eye's optic nerve and result in vision loss and blindness [1]. Glaucoma, with age-related macular degeneration (AMD) and diabetic retinopathy (DR), is one of the three leading causes of blindness in developed countries and is now the second leading cause of blindness globally, after cataracts [2, 3].

Glaucoma is characterized by loss of retinal ganglion cells (RGCs), which results in visual field impairment and structural changes to the retinal nerve fiber layer (RNFL) and optic disc [4]. Glaucoma has few early symptoms; over 3 million Americans have glaucoma, and the number is over 76 million worldwide, with projections showing 111 million by 2040 [5]. About half of those affected do not know it [6]. Most of the time, when detected, it is already late, i.e., with irreversible visual field loss. Therefore, it is essential to identify individuals at the early stages of this disease for treatment. The social and economic costs of vision loss from glaucoma are also extremely high. Early detection of these conditions halts a downward spiral in overall health: depression, loss of independence, need for nursing home care, falls, fractures, and death. These adverse outcomes are also highly costly. The total economic burden in the USA, direct and indirect, of vision loss and blindness from all causes, is now $145 billion, expected to triple by 2050 in real Dollars, with increased longevity generally [7].

The relationship between estimated RGC counts and CDR suggests that assessment of change in CDR is a sensitive method for the evaluation of progressive neural losses in glaucoma; specifically, the retinal cup-disc ratio (CDR) is highly correlated with glaucoma [8–13]. Although several

techniques [8, 9, 12–14] have been proposed to measure the cup-disc ratio, they have not been extensively validated for screening, and the current research is mainly focused on overall glaucoma detection. For example, Saxena et al. [15] proposed a method for glaucoma detection using the standard ORIGA dataset in 2020, whose AUC is 0.82. However, the American Academy of Ophthalmology outlined a set of tests to determine glaucoma, which are eye pressure, eye drainage angle, optic nerve damage, peripheral side vision (visual field test), computerized imaging of optic nerve, and thickness of the cornea [16]. Thus, using color fundus imaging alone is not a standard protocol for glaucoma detection or diagnosis. The CDR can be an effective tool to identify the glaucoma suspect, and our focus is mainly to identify the glaucoma-suspect individuals (as a screening process from primary-care settings), who can be further tested to determine glaucoma and its progression. Thus, starting with cup-to-disc ratio may also be more suited for clinical applicability because of its inherent explicability. At present, although there are techniques to detect CDRs, a full-fledged low-cost automated system that passively detects glaucoma from patients' yearly visits in a primary-care setting is still not widely available.

We have considered the vertical cup-disc ratio and the threshold value to determine the normal and abnormal, based on the following studies. A larger or abnormal CDR is mentioned in [10] and categorized as CDR > 0.5. The same research scheme also suggested that small changes in CDR may be associated with significant losses of RGCs, especially in eyes with large CDRs. Enlarged CDR is one indicator of the risk of glaucoma [11]. Most individuals fall near the average vertical CDR of 0.4, and 2.5% of the population have a cup/disc ratio of over 0.7 [17]. Studies showed that, for the normal (nonglaucoma) population, the horizontal C/D ratio is usually larger than the vertical C/D ratio, but the vertical ratio increases faster in the early and intermediate stages of glaucoma [18]. Also, studies have documented that the normal C/D ratios range from less than 0.3 (66 percent of normal individuals) to greater than 0.5 (only 6 percent of normal individuals). Therefore, we considered CDR 0.5 and above as glaucoma suspect [19, 20] and began exploring deep-learning methods for detecting larger CDRs [21].

We note that we have utilized the quantified vertical CDR when other research schemes used the qualitative assessment (e.g., small, medium, and large). We developed and validated our vertical CDR quantification software to perform this quantified grading. The software demonstrates high repeatability and reliability, which we have also provided in the paper. We developed and validated our AI-based glaucoma-suspect screening results based on the quantified vertical CDR. This should provide higher accuracy and confidence than selective judgment.

(i) This paper describes a method for glaucoma-suspect screening that utilizes a cloud-based system and incorporates telemedicine facilities. Thus, the screening will be available in remote clinics and primary-care settings.

(ii) This paper describes results on a novel automated method that addresses the early screening of glaucoma suspects, which is a major public health concern.

(iii) Therefore, an accurate and efficient screening in remote primary-care settings can provide a mass screening of the population currently dropping from yearly visits to the ophthalmologist.

The rest of the paper describes the development and validation of this glaucoma-suspect screening tool.

2. Materials and Methods

The global strategy of the study is organized as follows:

(2.1) Data sources, describing the various datasets

(2.2) Ground truth, describing manual grading

(2.3) Preprocessing, describing data curation and data processing before training

(2.4) Architecture, describing the technical details of the training and validation

2.1. Data Sources. Fundus images from three sources were used to conduct training experiments and a fourth for external validation. A total of 1546 color fundus images that included the disc were selected randomly from the Age-Related Eye Disease Study (AREDS) [22] and Singapore Malay Eye Study (SiMES) study [23], and all the images from the Retinal Image Database for Optic Nerve Evaluation (RIM-ONE) dataset [24], an ophthalmic reference image database specifically designed for glaucoma analysis, were selected. For external validation, we used the Online Retinal fundus Image database for Glaucoma Analysis (ORIGA) [25]. Although these retinal images had already been graded for glaucoma, we performed our gradings for consistency (Section 2.2).

Briefly, the AREDS is a 13-year study of age-related eye diseases. The participants were of the ages 55 to 80 when they were enrolled. 30-degree fundus photographs were graded as glaucoma present or absent by the AREDS ophthalmic grading center. We used fundus images from those cases as well as from the normal control population for this experiment.

SiMES-1 was a cross-sectional, population-based epidemiological study of eye diseases. It was performed on 3,280 randomly selected Malay adults living in the southwestern part of Singapore. All study participants underwent various questionnaires and detailed eye examinations. We have taken those images for which information about the presence or absence of glaucoma was present.

RIM-ONE is an ophthalmic reference image database specifically designed for glaucoma diagnosis, not only for medical education purposes but also as an evaluation tool for designers of segmentation algorithms. RIM-ONE is available as a free download as part of a research collaboration between three Spanish hospitals: Hospital Universitario de

Canarias, Hospital Clínico San Carlos, and Hospital Universitario Miguel Servet.

The ORIGA-light dataset is an ophthalmic reference image database specifically designed for glaucoma analysis. ORIGA-light serves as a benchmarking resource for researchers to evaluate image processing algorithms that detect and analyze various image signs highly related to glaucoma diagnosis. To facilitate this, the authors of ORIGA used their in-house grading tools to grade several glaucoma-related signs. The publicly available dataset that we used has 680 graded images, out of which 460 are healthy, and the rest are graded as glaucoma, taken from adults aged between 40 and 80 years. Each image is segmented and annotated by trained professionals from the Singapore Eye Research Institute.

The AREDS dataset may be obtained with request on their website "dbgap.ncbi.nlm.nih.gov." All the other datasets are available upon request from the authors of the corresponding datasets.

2.2. Ground Truth.

As noted, we did not use prior annotations of the presence of glaucoma but instead graded each image manually for vertical and horizontal CDR. A proprietary software called "CDR annotator" [26] was used for the purpose. Figure 1 shows the interface for marking the region of the cup and the disc, from which the vertical and horizontal CDRs are automatically generated. Before this, regions of interest (optic disc) were identified and cropped from the fundus image automatically using custom deep-learning methods (unpublished, Figure 2).

Two computer engineers at iHealthScreen Inc. were trained by expert ophthalmologists to grade CDRs in each image. Whenever there was disagreement in grading, the two graders adjudicated and produced uniform grading (CDRs) as ground truth for the images. Before adjudication, 250 images were randomly chosen to evaluate intergrader correlation and 200 for intragrader correlation. The intergrader and intragrader Pearson correlations between their CDR ratio annotations, 0.832 and 0.841, respectively [27], showed good consistency. The vertical CDRs were used to categorize the images into two classes: class 1 (not glaucoma suspect): vertical CDR ≤ 0.5 and class 2 (glaucoma suspect): vertical CDR > 0.5. After the categorization, the final dataset used for training is as shown in Table 1. Similarly, after quality control (removing 42 ungradable images), 638 images were selected out of a total of 680 images in the ORIGA-light dataset and processed and graded. CDRs of 452 images in the ORIGA dataset were graded by the experts to be less than or equal to 0.5, and those of 186 images were graded to be above 0.5.

2.3. Preprocessing.

Two types of input images were used simultaneously: the original RGB and a transformed, or preprocessed, RGB image. The transformed RGB image is a color space-averaged image [28]. We subtracted the local means with a kernel size 8 and Gaussian blurring. Such a preprocessing technique of local color averages (LCAs) is effective when dealing with images from various sources
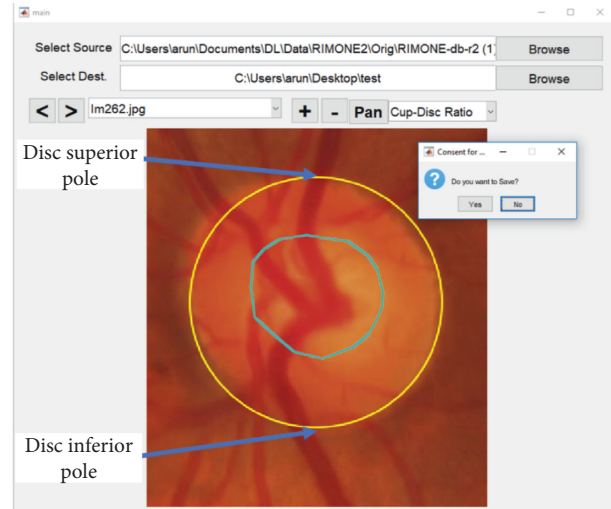

Figure 1: The interface of "CDR annotator," the proprietary cup/disc ratio grading software. The optic cup is circled in blue.

taken under different conditions. Figures 2(c) and 2(d) show an example of such a preprocessing technique.

The images from three datasets were combined to form a unified dataset. The test set was then taken from these images randomly. These cropped images from the original RGB image set and the LCA set form the final input for this experiment.

2.4. The Architecture and the Telemedicine Platform.

The architecture we propose (shown in Figure 3) consists of an ensemble of five neural networks and a classical tree learning algorithm. The overall system is a binary classifier that classifies the images into one of the two categories (CDR ≤ 0.5 and CDR > 0.5).

To build an image classification model robust in terms of image and dataset variations that is capable of learning features on such a wide scale in terms of size and location, the image preprocessing techniques and neural network selections were made carefully [29]. Multiple different neural networks, when used, are hypothesized to learn features from an image differently. Combining the results from different models to produce a final output is a general practice to obtain a better performance than each of the constituent network architectures [30]. To increase the robustness, different input sizes for the networks were chosen. Also, two types of images are fed into the models. One type is regular RGB images, and the other is preprocessed LCA images.

Deep-learning architectures were trained and validated to produce completely automatic outputs. All graded images (1546) were grouped into two categories: CDR ≤ 0.5 (1057 images) and CDR above 0.5 (489 images) in the two-class model.

The initial test models (from which the final models are chosen) were built to evaluate feasibility. The network architectures used are Inception-Resnet-V2 [31], NasNet [32], Xception [33], and Inception [34]. Deep-learning
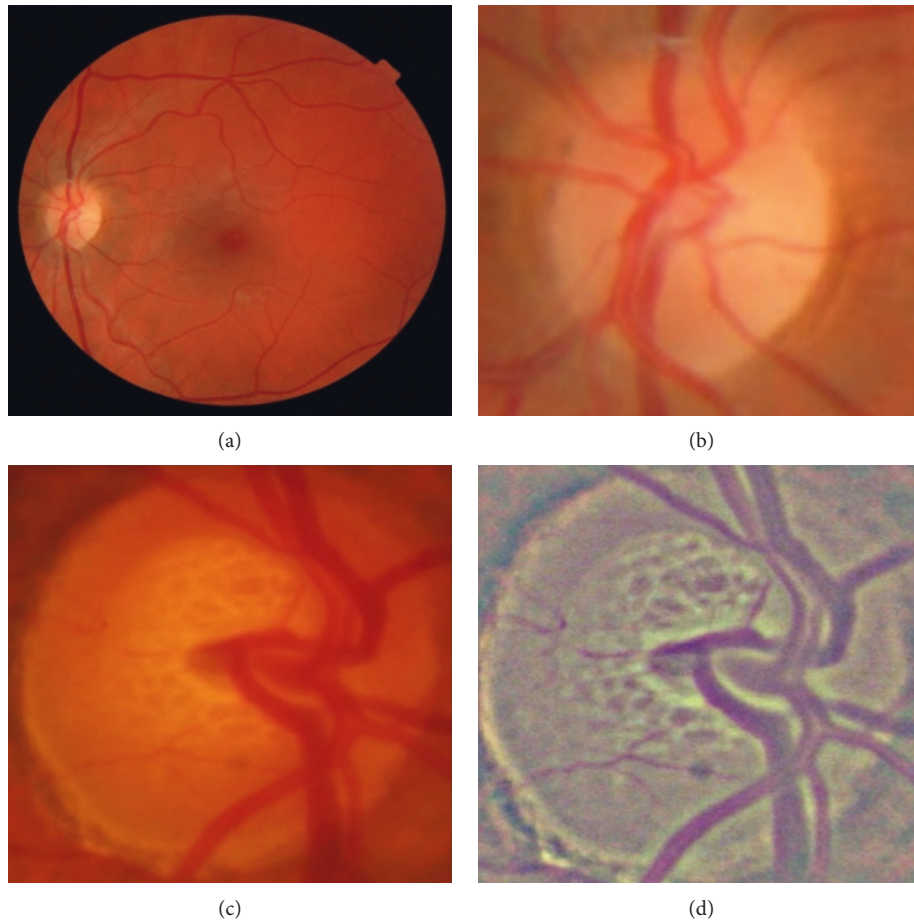
FIGURE 2: (a) A 30-degree fundus image and (b) the automatically cropped optic disc region. (c) Another cropped example that is preprocessed to obtain the (d) final image.

TABLE 1: Number of images taken from AREDS, SIMES, and RIM-ONE1. The table shows the various groups to which the images belong based on their graded cup/disc ratios.

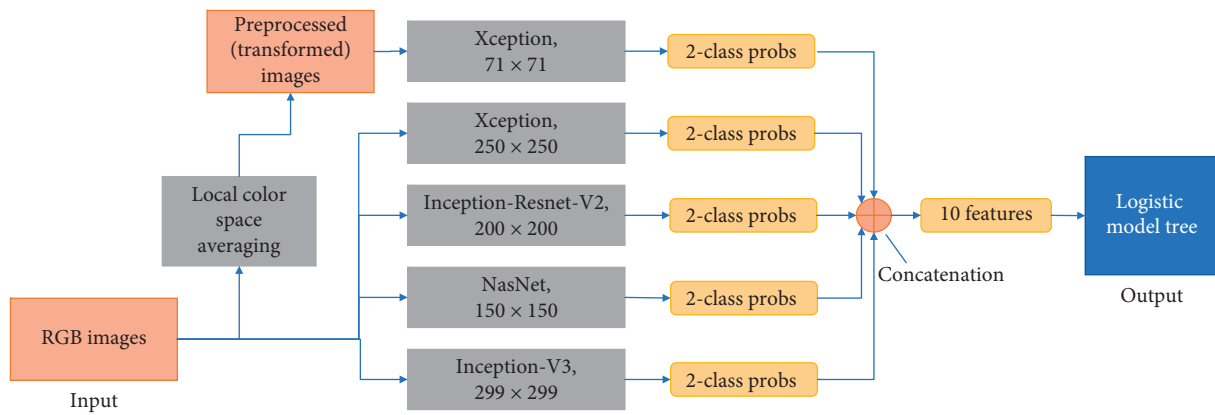| Category (manually graded for vertical cup/disc ratio)	AREDS	SIMES	RIM-ONE1	Total
≤0.5	462	253	342	1057
Above 0.5	131	243	115	489

FIGURE 3: Overall schematic representation of the model building. The two types of input images (RGB and transformed) fed into five neural networks are shown. Each network accepts a rescaled image (shown below the name of the architecture) and produces two probabilities. The ten resulting probabilities are then concatenated to form a feature array of size 10 and fed as input to a logistic model tree which acts as the final classifier in this system.

architectures such as alexnet [35] and VGG networks [36] initially focused on stacking layers deeper and deeper, hoping to get better performance. Inception architecture changed this approach by going "wider" instead of focusing ongoing just "deeper." Every version of Inception optimized speed and accuracy, and the version used in our experiment, Inception-V3, uses an RMSProp optimizer [37], factorized 7×7 convolutions, and added batch normalization [38] in the auxiliary classifiers.

The original Inception networks without depth-wise separable convolutions were modified to include the residual connections, called the Inception-ResNet networks. Inception modules allowed for adding more layers, and residual connections made the network converge faster, which is the basis for Inception-Resnet-V2. Xception is a novel deep convolutional neural network architecture inspired by Inception, where Inception modules have been replaced with depth-wise separable convolutions. This architecture slightly outperforms Inception-V3 on the ImageNet dataset. NASNet learns the model architectures directly on the dataset of interest by searching the best convolutional layer (or "cell") on a small dataset and then applying it to a larger dataset by stacking together more copies of this cell [39].

The various input sizes used range from 71×71 to 399×399. The best models, thus, obtained will be ensembled to form a final architecture. Through experimentation, we developed an ensemble of five networks in the final architecture for the glaucoma screening system. The description of the five models is given below.

(1) Xception- input size: 71×71; input image type: local color averaged (transformed)

(2) Xception- input size: 250×250; input image type: RGB

(3) Inception-Resnet-V2- input size: 200×200; input image type: RGB

(4) NasNet- input size: 150×150; input image type: RGB

(5) Inception-V3- input size: 299×299; input image type: RGB

The full framework for building the model is shown in Figure 3. The networks are trained for 500 epochs. We trained the networks with a batch size of 20 images. This is a high number considering the limitations of GPU memory. The low resolution of cropped images helped achieve a bigger batch size. The Adam [40] optimizer was used with a learning rate of 0.0004. To save time, an early stopping mechanism halted training if there was no improvement for 20 consecutive epochs. Every epoch is monitored for loss, and this value is used for early stopping. The loss function used in this system was categorical cross entropy. This quantity was used to determine the hyperparameters of the networks. SoftMax activation is used as the last layer in each of the architectures. All the networks were trained on NVIDIA Titan V GPU for two weeks with an average time of 20 to 30 minutes per epoch.

Each model gives a probability array of size 2 for 2 classes. The five arrays from five models are concatenated to form a feature array of size 10 that is then used to build a Logistic Model Tree [41] model for final output. Sensitivity, specificity, accuracy, and Cohen's kappa were calculated to evaluate the models.

2.5. The Role of the AI Platform. The CDR carries three advantages: first, it is a single variable that is known to be strongly correlated with the disease, in particular with losses of RGCs as noted and, as such, is inherently explicable and acceptable to the eye community; second, measurement of CDR can be accomplished from a single retinal color photograph obtained by an automated, nonmydriatic camera in a primary-care office and forwarded on a telemedicine platform for expert interpretation with semiautomated methods [26]; and third, that expert interpretation, which is still time consuming and expensive for humans, can be replaced by AI for efficiently and accurately evaluating the images as we propose to demonstrate herein. In fact, we have already introduced such a HIPAA-compliant telemedicine platform, iPredict, with the requisite capabilities of AI solutions and report generation.

A telemedicine platform has been introduced that enables the cloud-based processing of the AI solutions, and report generation can extensively simplify the process of evaluating the images on a mobile/tablet or a low-performance computer, a requirement for the successful glaucoma-suspect screening at primary-care settings. We aim to address this with our HIPAA-compliant telemedicine platform iPredict.

In the future, we propose to use the Software Tool 'iPredict-glaucoma' at the iPredict platform (https://ipredict.health/). An online version of the Glaucoma-Suspect Screening system is available at https://www.ihealthscreen.org/ipredict-glaucoma/(the username and password can be obtained for research purposes through writing to the corresponding author). The AI-based telemedicine platform iPredict developed by iHealthscreen Inc. integrates the server-side programs (the image analysis and deep-learning modules for screening systems) and local remote computer/mobile devices (for collecting and uploading patient data and images). The images are first checked for gradability automatically by an artificial-intelligence-based system developed in house from 3000 fundus images manually graded for gradability, and the system achieved over 99% accuracy. The server analyzes the images, and a report will be sent to the remote clinic with an individual's screening results and further recommendations.

identification of the glaucoma suspects, is proposed to help prevent this sight-threatening disease.

3. Results

The two-class glaucoma model (CDR $\leq$ 0.5 and above 0.5) achieved an accuracy of 89.67% (95% CI −85.65% to 92.87%) with a sensitivity of 83.33% (95% CI −75.44% to 89.51%) and a specificity of 93.89% (95% CI - 89.33% to 96.91%) along with Cohen's kappa of 0.78 (95% CI −0.71 to 0.85) when above 0.5 cases are considered as positive. The AUC for the same data was 0.93 (0.89 to 0.96), as shown in Figure 4. The complete results of the system are detailed in Table 2.

On the external validation dataset, the two-class model achieved a sensitivity of 80.11% (73.64% to 85.59%) and specificity of 84.96% (81.32% to 88.12%) with Cohen's kappa of 0.62 (0.57 to 0.67) on the ORIGA dataset. The AUC for the same data was 0.85 (0.81 to 0.90), as shown in Figure 4. The complete results on the external validation dataset can be found in Table 2.

The cloud-based and HIPAA-compliant telemedicine platform 'iPredict' (https://ipredict.health/) has been validated for image and data transfer accuracy. We have transferred and analyzed nearly 850 images for AMD screening and DR screening from 4 primary-care clinics in Queens and Manhattan, New York, USA. The initial results from utilizing our platform are reported in [42, 43]. We found a 100% correlation between the results obtained from directly evaluated images and the images transferred and processed by iPredict. We have also tested 100 images for vertical CDR computation and received the same accuracy.

4. Discussion

In this study, we have demonstrated an accurate and fully automated deep-learning screening system for glaucoma suspects through retinal photography that may be effective for the identification of glaucoma suspects in primary-care settings.

Glaucoma is a prevalent, blinding disease worldwide with few symptoms until irreversible later stages and is undiagnosed at rates approaching 50% even in developed countries [6]. Hence, the pressing public health calls for effective community screening. The need is even greater in communities of color, with an overall ratio of 8 : 1 for nonwhite to white primary glaucoma blindness, due, at least in part, to receiving medical care later in the disease than whites [44], Compounding the problem, there is also a dramatically earlier age of onset in this group. In an Afro-Caribbean population, glaucoma-suspect status was high across all age groups, with significant prevalence even in populations less than 40 years of age [45].

We have shown on several large datasets that the cup/disc ratio (CDR) can be measured automatically from retinal photography with sufficient accuracy to discriminate suspects from nonsuspects and, thus, potentially facilitate referral of suspects to an ophthalmologist for specialized care. Thus, a future, achievable goal is an AI *telemedicine* platform in which our current methodology will be deployed in

primary-care settings through remote image capture. A prospective trial will be needed to determine the feasibility of the system in clinical settings, with inexpensive, automated nonmydriatic retinal cameras and a telemedicine platform for image transfer to the deep-learning screening system. Such systems have been tested clinically with proven accuracy for screening DR in comparison to expert graders [46]. It is, thus, reasonable to expect that similar success may be achieved with glaucoma.

A National Eye Institute study showed that 90% of glaucoma subjects can be prevented from progression to severe glaucoma through timely identification and intervention [6]. However, nearly sixty percent of Americans with diabetes skip annual sight-saving exams recommended by their Primary-Care Physicians (PCPs). Given such poor compliance by diabetics, who are informed about the risks to their vision, it is likely that compliance with eye exams is even worse in the general population [47]. Therefore, our focus is to identify the suspects in the primary-care settings, not only to get them needed care but also to eliminate large numbers of unnecessary specialist visits for glaucoma screening.

The medical imaging and diagnostics field has been revolutionized by advances in deep learning in recent years. Extensive research interest is being shown in using artificial intelligence for solving medical problems [48]. Ting et al. [49] detailed the potential applications of AI in ophthalmology. Gulshan et al. [50], in their seminal paper, showed the application of AI in diabetic retinopathy from fundus images using deep learning. Recently, we have published two groundbreaking works on late AMD prediction [51] and diabetes screening in primary-care settings [46]. There is also considerable research in other medical areas such as multiple sclerosis [52], neurodegeneration [53], and age-related macular degeneration [54–57]. Several AI techniques [8, 9, 12–14] have been proposed to measure the cup-disc ratio, but they have not been validated for screening glaucoma suspects.

We note that the current research using color fundus imaging is mainly focused on the detection of glaucoma, which we believe is not an appropriate option for clinical settings if we follow the glaucoma detection or diagnosis protocol (https://www.aao.org/eye-health/diseases/glaucoma-diagnosis). Thus, here, we aim to clear the differentiation of the term glaucoma and glaucoma suspect. The glaucoma detection is a diagnosis of glaucoma that requires the structural and functional abnormalities from a glaucoma suspect, which is implied by its name, a category of markers with an increased likelihood of the disease.

Li et al. [58] and Ting et al. [59] trained computer algorithms to detect the glaucoma-like disc, defined as a vertical CDR of 0.7 and 0.8, respectively. In general, an eye with a vertical CDR above 0.5 is considered a glaucoma suspect [60]. In this paper, we introduced an automated cup-disc measurement tool that can determine if the vertical cup-disc ratio is above or below 0.5, in conjunction with a deep-machine-learning-based tool. We have published this approach, an ensemble of deep-learning architectures with a logistic tree at the end, for effective use in AMD screening,

Cup-to-disc model
Receiver operating characteristic

— Two-class model on test data
AUC = 0.93 (0.89 to 0.96)

(a)

Cup-to-disc model
Receiver operating characteristic

— Two-class model on ORIGA-light
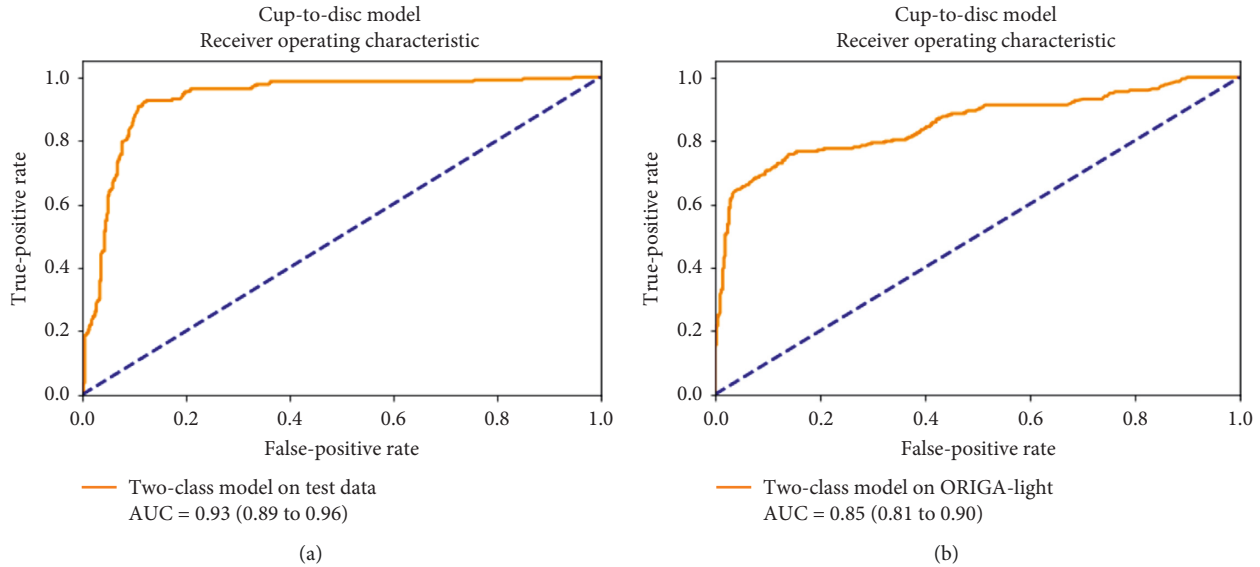AUC = 0.85 (0.81 to 0.90)

(b)

FIGURE 4: Area under ROC of the test data results (a) and the results for the ORIGA-light dataset (b) for the two-class cup-to-disc ratio model.

TABLE 2: Accuracy, sensitivity, specificity, Cohen's kappa, and AUC for the system with 95% confidence intervals on the test data and ORIGA-light (validation data), with the ratio over 0.5 considered as a positive case for the model.

Measure	Test data (with 95% CI)	ORIGA-light (with 95% CI)
Accuracy	89.67% (85.65% to 92.87%)	80.11% (73.64% to 85.59%)
Sensitivity	83.33% (75.44% to 89.51%)	84.96% (81.32% to 88.12%)
Specificity	93.89% (89.33% to 96.91%)	83.54% (80.43% to 86.34%)
Cohen's kappa	0.78 (0.71 to 0.85)	0.62 (0.57 to 0.67)
Area under the curve	0.93 (0.89 to 0.96)	0.85 (0.81 to 0.90)

but it is novel in glaucoma screening. With a telemedicine platform, it could screen glaucoma suspects on a large scale in primary-care settings, conferring substantial public health benefits of reduced vision loss from glaucoma and reduced health-care costs.

5. Limitations and Future Work

In our testing, the sensitivities are somewhat lower than the specificities, with, therefore, a somewhat greater risk of missing true cases. In our future work in this project, we aim to understand and anticipate the doctors' requirements in this new method of screening and tune the system such that the false positives and false negatives are in an acceptable ratio. Small discs with "pathologic cups" are hard to detect. Geometrically, a small disc with a CDR of 0.7 has much less healthy neural rim tissue than a normal-sized disc with the same CDR. Therefore, CDR asymmetry would be a reasonable addition to the screening program. In general, a provisional diagnosis of a glaucoma suspect is generally given with CDR asymmetry ($>/ = 0.20$) [61]. This criterion could be implemented in the next version of the present DL architecture that is already tuned to CDR measures. Asymmetrical cupping of the optic disc was found in 5.6% of normal individuals, in 30% of a group with ocular hypertension without field defect, and in 36% of those with

established chronic open-angle glaucoma and field loss [62]. We note that our algorithm classifies an optic disc based on vertical CDR based on a single retinal image, and asymmetrical cupping may show up as a different reading in the two eyes of the same image, in turn, helping doctors with an additional biomarker for glaucoma. Also, we did not have any data on the refractive errors, high myopia, etc. for the study subjects which could be sources of errors in interpreting disc photos and are more relevant than just assessment of the vertical CDR. We aim to address them in the future study.

6. Strengths

The model was built on several large datasets, with external validation on another. The output is an ideal binary target for glaucoma suspects with a single highly correlated and easily measured/interpreted variable. We have had success with a novel hybrid AI approach to screening for AMD and DR from fundus photos that performs, at least, as good as other techniques in the literature. So, we chose this route again for screening glaucoma suspects: an ensemble of DL techniques is first trained on the image inputs to produce sets of probabilities (one set for each DL technique) for classifying the image into a disease state; these sets of probabilities are then inputted to an independently trained

logistic model tree which acts as the final classifier in this system.

A glaucoma diagnosis requires multiple structural and functional criteria that are not available or appropriate in the primary-care setting. To our knowledge, our system is the only one proposed that is a full-fledged, passive screening system for an adequate screening of glaucoma suspects with a single disease marker that could be easily obtained in the primary-care setting on a telemedicine platform without expensive, specialized equipment or services.

Future work can carry these methods into the primary-care setting to perform annual screening for this silent blinding disease. We propose to address this urgent public health need with future prospective trials of our system for low-cost, rapid glaucoma screening. These trials will be modeled on our current ongoing NIH-funded 3-year trial (SBIR Phase IIb R44EY031202, A Bhuiyan PI, "https://projectreporter.nih.gov/project_info_details.cfm?aid=10010769") for detection and prediction of AMD in primary-care settings with our published DL algorithms [51]. A complete HIPAA-compliant functional AI-based telemedicine platform for real-time diagnosis is already in place, which integrates the server-side screening programs (image analysis and deep-learning modules) and local remote devices (for collecting patient data and images).

7. Conclusions

We have developed an effective deep-learning/logistic model tree hybrid screening tool for the identification of glaucoma suspects by vertical CDR from nonmydriatic retinal photographs.

Building on this tool, a full AI telemedicine platform is envisioned in a future state where our current AI methodology will be deployed in primary-care settings. Fully automated cameras will capture images for transfer through the cloud to the server side for immediate results and further patient referral if needed, with significant public health benefit for early detection and prevention of this sight-threatening disease.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

A. B. and A. G. are paid employees of iHealthscreen Inc. R. T. S. declares no conflicts of interest.

Authors' Contributions

A. B. and A. G. conceptualized the study, formulated the methodology, and performed validation; A. G. conducted formal analysis and investigation; A. B. and R. T. S. curated data; A. G., R. T. S., and A. B. reviewed and edited the manuscript; R. T. S. supervised the work; and A. B. carried out project administration and acquired funding. All authors have read and agreed to the published version of the manuscript.

References

[1] Bright-Focus-Foundation, "Glaucoma fact and figures," 2020, http://www.brightfocus.org/glaucoma/article/glaucoma-facts-figures.

[2] A. Foster, C. Gilbert, and G. Johnson, "Changing patterns in global blindness: 1988–2008," *Community Eye Health*, vol. 21, no. 67, p. 37, 2008.

[3] S. Resnikoff, D Pascolini, D Etya'ale et al., "Global data on visual impairment in the year 2002," *Bulletin of the World Health Organization*, vol. 82, pp. 844–851, 2004.

[4] R. N. Weinreb and P. T. Khaw, "Primary open-angle glaucoma," *The Lancet*, vol. 363, no. 9422, pp. 1711–1720, 2004.

[5] N. E. Institute, "Glaucoma data and statistics," 2020, https://www.nei.nih.gov/learn-about-eye-health/resources-for-health-educators/eye-health-data-and-statistics/glaucoma-data-and-statistics.

[6] D. Gorman, "Glaucoma: preventing the visual damage," 2020, https://www.clinicaladvisor.com/home/features/glaucoma-preventing-the-visual-damage/.

[7] J. S. Wittenborn and D. B. Rein, *Cost of Vision Problems: the Economic Burden of Vision Loss and Eye Disorders in the United States*, NORC at the University of Chicago, Chicago, IL, USA, 2013.

[8] B. Al-Bander, W. Al-Nuaimy, M. A. Al-Taee, and Y. Zheng, "Automated glaucoma diagnosis using deep learning approach," in *Proceedings of the 2017 14th International Multi-Conference on Systems, Signals & Devices (SSD)*, pp. 207–210, Marrakech, Morocco, March 2017.

[9] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 715–718, Milano, Italy, August 2015.

[10] A. J. Tatham, R. N. Weinreb, L. M. Zangwill, J. M. Liebmann, C. A. Girkin, and F. A. Medeiros, "The relationship between cup-to-disc ratio and estimated number of retinal ganglion cells," *Investigative Opthalmology & Visual Science*, vol. 54, no. 5, pp. 3205–3214, 2013.

[11] M. F. Armaly, "Cup/disc ratio in early open-angle glaucoma," *Documenta Ophthalmologica*, vol. 26, no. 1, pp. 526–533, 1969.

[12] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "Convolutional neural network transfer for automated glaucoma identification," in *Proceedings of the 12th International Symposium on Medical Information Processing and Analysis*, Tandil, Argentina, December 2016.

[13] S. Nawaldgi, "Review of automated glaucoma detection techniques," in *Proceedings of the 2016 International Conference on Wireless Communications*, pp. 1435–1438, Fez, Morocco, October 2016.

[14] R. Hemelings, B. Elen, J. Barbosa-Breda et al., "Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer

learning," *Acta Ophthalmologica*, vol. 98, no. 1, pp. e94–e100, 2020.

[15] A. Saxena, A. Vyas, L. Parashar, and U. Singh, "A glaucoma detection using convolutional neural network," in *Proceedings of the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 815–820, Coimbatore, India, April 2020.

[16] K. Boyd, "Glaucoma diagnosis," *American Academy of Ophthalmology*, 2020, https://www.aao.org/eye-health/diseases/glaucoma-diagnosis.

[17] "Glaucoma: What every patient should know." Johns Hopkins Medicine., 2020)., https://www.hopkinsmedicine.org/wilmer/services/glaucoma/book/ch06s03.html.

[18] J. Tsai and M. Forbes, *Medical Management of Glaucoma*, I. West, Ed., Professional Communications, New York, NY, USA, 2nd edition, 2004.

[19] American Academy of Ophthalmology, *Glaucoma Basic Science and Clinical Science Course, 2003-2004*, American Academy of Ophthalmology, San Francisco, CA, USA, 2003.

[20] J. Tsai, "How to evaluate the suspicious optic disc," *Review of Ophthalmology*, vol. 12, no. 6, p. 40, 2005.

[21] A. Bhuiyan, A. Govindaiah, and T. Smith, "An artificial intelligence based screening tool to identify glaucoma suspects from color fundus imaging," *Investigative Ophthalmology & Visual Science*, vol. 61, no. 9, p. PB009, 2020.

[22] Age-Related-Eye-Disease-Study-Research-Group, "The Age-Related Eye Disease Study (AREDS): design implications. AREDS report no. 1." *Control Clinical Trials*, vol. 20, no. 6, pp. 573–600, 1999.

[23] A. W. P. Foong, S.-M. Saw, J.-L. Loo et al., "Rationale and methodology for a population-based study of eye diseases in Malay people: the Singapore Malay eye study (SiMES)," *Ophthalmic Epidemiology*, vol. 14, no. 1, pp. 25–35, 2007.

[24] F. Fumero, S. Alayón, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: an open retinal image database for optic nerve evaluation," in *Proceedings of the 2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 1–6, Bristol, UK, June 2011.

[25] Z. Zhang, "Origa-light: an online retinal fundus image database for glaucoma analysis and research," in *Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 3065–3068, Québec, Canada, September 2010.

[26] iHealthScreen, "Cup-to-Disc ratio (CDR) annotator," iHealthScreen Inc., Richmond Hill, NY, USA, 2020.

[27] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*, vol. 2, pp. 1–4, Springer, Berlin, Germany, 2009.

[28] M. Ebner, "Color constancy based on local space average color," *Machine Vision and Applications*, vol. 20, no. 5, pp. 283–301, 2009.

[29] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: overview, challenges and the future," *Lecture Notes in Computational Vision and Biomechanics*, pp. 323–350, 2018, https://arxiv.org/abs/1704.0682546.

[30] O. Sagi and L. Rokach, "Ensemble learning: a survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018.

[31] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," pp. 1–12, 2016, https://arxiv.org/abs/1602.07261.

[32] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8697–8710, Salt Lake City, UT, USA, June 2018.

[33] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, Honolulu, HI, USA, June 2017.

[34] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, Las Vegas, NV, USA, June 2016.

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *The Proceedings of ICLR*, vol. 1409, pp. 1–14, 2015.

[37] G. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," vol. 14, no. 8, 2012.

[38] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, https://arxiv.org/abs/1502.03167.

[39] A. Nabavizadeh, T. Payen, A. C. Iuga et al., "Noninvasive Young's modulus visualization of fibrosis progression and delineation of pancreatic ductal adenocarcinoma (PDAC) tumors using Harmonic Motion Elastography (HME) in vivo," *Theranostics*, vol. 10, no. 10, pp. 4614–4626, 2020.

[40] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980.

[41] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Machine Learning*, vol. 59, no. 1-2, pp. 161–205, 2005.

[42] S. Alauddin, "A prospective trial of an artificial intelligence based telemedicine platform to stratify severity of age-related macular degeneration (AMD)," *Investigative Ophthalmology & Visual Science*, vol. 61, no. 7, p. 1843, 2020.

[43] A. Bhuiyan, A. Govindaiah, M. Hossain, and R. T. Smith, "Development and validation of an automated AI-based diabetic retinopathy screening tool for primary care setting," *Investigative Ophthalmology & Visual Science*, vol. 61, no. 7, p. 3821, 2020.

[44] R. Hiller and H. A. Kahn, "Blindness from glaucoma," *American Journal of Ophthalmology*, vol. 80, no. 1, pp. 62–69, 1975.

[45] C. L. Bokman, L. R. Pasquale, R. K. Parrish, and R. K. Lee, "Glaucoma screening in the Haitian afro-Caribbean population of South Florida," *PLoS One*, vol. 9, no. 12, Article ID e115942, 2014.

[46] A. Bhuiyan, "Development and validation of an automated diabetic retinopathy screening tool for primary care setting," *Diabetes Care*, vol. 43, no. 10, pp. e147–e148, 2020.

[47] "Sixty percent of Americans with diabetes skip annual sight-saving exams." American Academy of Ophthalmology. (2020), https://www.aao.org/newsroom/news-releases/detail/sixty-percent-americans-with-diabetes-skip-exams.

[48] A. Becker, "Artificial intelligence in medicine: what is it doing for us today?" *Health Policy and Technology*, vol. 8, no. 2, pp. 198–205, 2019.

[49] D. S. W. Ting, L. R. Pasquale, L. Peng et al., "Artificial intelligence and deep learning in ophthalmology," *British Journal of Ophthalmology*, vol. 103, no. 2, pp. 167–175, 2019.

[50] V. Gulshan, L. Peng, M. Coram et al., "Development and validation of a deep learning algorithm for detection of

diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[51] A. Bhuiyan, T. Y. Wong, D. S. W. Ting, A. Govindaiah, E. H. Souied, and R. T. Smith, "Artificial intelligence to stratify severity of age-related macular degeneration (AMD) and predict risk of progression to late AMD," *Translational Vision Science & Technology*, vol. 9, no. 2, p. 25, 2020.

[52] G. Casalino, G. Castellano, A. Consiglio, M. Liguori, N. Nuzziello, and D. Primiceri, "A predictive model for MicroRNA expressions in pediatric multiple sclerosis detection, modeling decisions for artificial intelligence," in *Proceedings of the International Conference on Modeling Decisions for Artificial Intelligence*, pp. 177–188, Milan, Italy, September 2019.

[53] E. Lella and G. Vessio, "Ensembling complex network 'perspectives' for mild cognitive impairment detection with artificial neural networks," *Pattern Recognition Letters*, vol. 136, 2020.

[54] A. Govindaiah, T. Smith, and A. Bhuiyan, "A new and improved method for automated screening of age-related macular degeneration using ensemble deep neural networks," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 702–705, Honolulu, Hawaii, July 2018.

[55] A. Govindaiah, A. Hussain, R. Smith, and A. Bhuiyan, "Deep convolutional neural network-based screening and assessment of age-related macular degeneration from fundus images," *IEEE International Symposium on Biomedical Imaging*, vol. 10, pp. 1525–1528, 2017.

[56] P. M. Burlina, N. Joshi, K. D. Pacheco, D. E. Freund, J. Kong, and N. M. Bressler, "Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration," *JAMA Ophthalmology*, vol. 136, no. 12, pp. 1359–1366, 2018.

[57] F. Grassmann, J. Mengelkamp, C. Brandl et al., "A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography," *Ophthalmology*, vol. 125, no. 9, pp. 1410–1420, 2018.

[58] Z. Li, Y. He, S. Keel, W. Meng, R. T. Chang, and M. He, "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," *Ophthalmology*, vol. 125, no. 8, pp. 1199–1206, 2018.

[59] D. S. W. Ting, C. Y.-L. Cheung, G. Lim et al., "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *Jama*, vol. 318, no. 22, pp. 2211–2223, 2017.

[60] The Meaning of Cupping, "MedRounds," 2020, http://www.medrounds.org/glaucoma-guide/2006/02/section-1-b-meaning-of-cupping.html.

[61] L. Vijaya, R. George, M. Baskaran et al., "Prevalence of primary open-angle glaucoma in an urban south Indian population and comparison with a rural population," *Ophthalmology*, vol. 115, no. 4, pp. 648–654, 2008.

[62] R. S. Fishman, "Optic disc asymmetry," *Archives of Ophthalmology*, vol. 84, no. 5, pp. 590–594, 1970.