Automated Detection of Glaucoma From Topographic Features of the Optic Nerve Head in Color Fundus Photographs

Lipi Chakrabarty, MD, Gopal Datt Joshi, PhD,† Arunava Chakravarty, MD,† Ganesh V. Raman, MD,* S.R. Krishnadas, MD,* and Jayanthi Sivaswamy, PhD†*

Objective: To describe and evaluate the performance of an automated CAD system for detection of glaucoma from color fundus photographs.

Design and Setting: Color fundus photographs of 2252 eyes from 1126 subjects were collected from 2 centers: Aravind Eye Hospital, Madurai and Coimbatore, India. The images of 1926 eyes (963 subjects) were used to train an automated image analysis-based system, which was developed to provide a decision on a given fundus image. A total of 163 subjects were clinically examined by 2 ophthalmologists independently and their diagnostic decisions were recorded. The consensus decision was defined to be the clinical reference (gold standard). Fundus images of eyes with disagreement in diagnosis were excluded from the study. The fundus images of the remaining 314 eyes (157 subjects) were presented to 4 graders and their diagnostic decisions on the same were collected. The performance of the system was evaluated on the 314 images, using the reference standard. The sensitivity and specificity of the system and 4 independent graders were determined against the clinical reference standard.

Results: The system achieved an area under receiver operating characteristic curve of 0.792 with a sensitivity of 0.716 and specificity of 0.717 at a selected threshold for the detection of glaucoma. The agreement with the clinical reference standard as determined by Cohen κ is 0.45 for the proposed system. This is comparable to that of the image-based decisions of 4 ophthalmologists.

Conclusions and Relevance: An automated system was presented for glaucoma detection from color fundus photographs. The overall evaluation results indicated that the presented system was comparable in performance to glaucoma classification by a manual grader solely based on fundus image examination.

Key Words: automated detection, glaucoma, computer-aided diagnosis, fundus photography

(*J Glaucoma* 2015;00:000–000)

Glaucoma, a degenerative optic neuropathy, is the second leading cause of blindness in the world accounting for 12.3% of the total blindness worldwide[1] and projected to affect 79.86 million people by 2020.[2] It is a chronic disease, asymptomatic in its early stages, resulting in a gradual, progressive, and irreversible degeneration of optic nerve, finally leading to blindness. Undiagnosed glaucoma could contribute to a large number of cases of preventable blindness. Known risk factors associated with glaucoma can provide guidelines for targeting at-risk groups for early detection and treatment of glaucoma. This methodology is currently the most effective way for preventing blindness and low vision. However, studies in India and other countries have shown that 50% to 90% of glaucoma cases remain undetected in the population.[3–8] This is due to the lack of a simple, precise, and cost-effective screening tool for glaucoma. Various studies have shown that intraocular pressure (IOP) and visual field test are found to be neither specific nor sensitive enough to be effective for screening.[9,10] Advanced diagnostic imaging such as optical coherence tomography and Heidelberg retinal tomography are not cost-effective for use in detection and hence restricted to clinical settings. Assessment of the optic nerve head (ONH) can be performed by a trained professional. However, these assessments are not well suited for mass screening due to the severe shortage of trained professionals, particularly in countries like India.

In the last decade, digital nonmydriatic color fundus photography has emerged as a suitable imaging modality for diabetic retinopathy screening in the community as fundus imaging is widely available.[11,12] A recent study[13,14] benchmarks ONH examination from monoscopic optic disc photographs among eye-care professionals and gives insights into the decision-making processes that clinicians adopt in determining whether an ONH appearance is glaucoma to us or not. Here, 9 conventional topographic features such as disc features, cup features, peripapillary atrophy (PPA), retinal nerve fiber layer (RNFL) defect, disc hemorrhage, and the vertical cup to disc ratio (CDR) were analyzed based on the grading outcomes by 197 ophthalmic clinicians on 42 monoscopic optic disc photographs of healthy and glaucomatous eyes. The findings of the study underscore the interobserver variability in assessment and provide recommendations to improve ONH examination teaching and medical education modules.

In this paper, we describe an objective decision-making system for the detection of glaucoma, on the basis of the assessment of established topographic features from optic disc fundus photographs using computer-based analytical methodologies.

PRIOR STUDIES

Existing automatic assessment methods focus on disc features[15,16] such as CDR to quantify the glaucomatous disc changes from color fundus images. Statistical features

Received for publication December 24, 2014; accepted October 9, 2015.
From the *Aravind Eye Care System, Madurai, Tamil Nadu; and †International Institute of Information Technology, Hyderabad, India.
Supported by the Department of Science and Technology, Government of India, under Grants SR/S3/EECE/0024/2009 and DST/INT/NL/Biomed/P(3)/2011(G).
Disclosure: The authors declare no conflict of interest.
Reprints: Jayanthi Sivaswamy, PhD, International Institute of Information Technology Hyderabad, Gachibowli, Hyderabad, Andhra Pradesh 500032, India (e-mail: jsivaswamy@iiit.ac.in).

DOI: 10.1097/IJG.0000000000000354

from analysis of fundus images have also been used to classify the corresponding eyes as normal or glaucomatous.[17–20] A recent comprehensive approach[21] combines information from multiple sources such as a patient's personal data, fundus image, and genome information. In this study, we explicitly analyze clinically known topographic features, which include disc and cup features, PPA, and RNFL defect, to arrive at a decision on the presence of glaucoma in a patient's fundus image. The analysis of hemorrhage was excluded because its presence is found to be rare in ONH assessment. The use of personal and genome information used in Liu et al[21] is deliberately excluded because of nonavailability of such information in community screening settings. The new computer-based analytical approach and design of the study distinguishes this study from prior work.

MATERIALS AND METHODS

The presented machine-learning–based automatic glaucoma system requires a set of labeled fundus photograph examples consisting of both normal and glaucomatous eye categories to learn about each individual category to perform classification. Therefore, 2 independent sets of image data, referred to as training data set and clinical reference standard, were collected to train the automated system and evaluate the performance of the system, respectively. A cross-training methodology was implemented during data collection, wherein the training and the reference standard data set were collected from 2 different sites. A common data collection protocol was followed for both data sets as described in the next section. In the subsequent subsections, the detailed description is provided on the reference standard data set, manual grading on reference standard data set, and training data set for the presented system.

Data Collection Protocol

The clinical reference standard and training data sets were collected from subjects who visited Aravind Eye Hospital, Coimbatore, India (Site-A) and Aravind Eye Hospital, Madurai, India (Site-B), respectively, between September 2013 and March 2014. After obtaining informed consent, vision, IOP readings, and family history of glaucoma of each subject were recorded. The pupil was dilated pharmacologically with Tropicamide 0.8% and 5% phenylephrine eye drop or Tropicamide 1% eye drop. Fundus images of the dilated eyes were taken by a trained technician in a dark room using Topcon TRC50EX fundus camera and Zeiss Visucam NM/FA fundus camera at site-A and site-B, respectively.

The ONH region of each eye was captured with a 30-degree field of view with an image resolution of 1900×1600 pixels at site-A and 2588×1958 pixels at site-B. A 30-degree field-of-view fundus image provided sufficient coverage of the ONH with a good portion of the inferior and superior peripheral RNFL. Two fundus images per eye were taken in a sequential order to aid automatic ONH segmentation.[22] All cases were scrutinized to identify adult subjects with clear media, after dilatation. Adult subjects with clear media were included, whereas those with ONH anomalies, major optic nerve tilt, or advanced hypertensive or diabetic retinopathy were excluded. Poor quality of fundus images of the ONH in at least 1 eye also served as an exclusion criterion.

The collection and analysis of image data were approved by the Institutional Research Board of Aravind Eye Care System, which governs the constituent site-A and site-B. The entire study adhered to the tenets set forth in the Declaration of Helsinki.

The Clinical Reference Standard Data Set

Each eye of a subject was classified as normal, glaucoma suspect, or confirmed glaucomatous on the basis of the consensus of masked clinical examination outcomes of 2 independent glaucoma specialists. Here, the glaucoma suspect category is defined for an eye having the following features: IOP consistently >21 mm Hg, appearance of the optic disc or RNFL suggestive of likelihood of glaucomatous damage, diffuse or focal narrowing or sloping of the neuroretinal rim, diffuse or focal abnormalities of the RNFL, disc hemorrhages, asymmetric neural rim within the optic discs or between fellow eyes and visual field abnormalities. This 3-level classification is the clinically followed practice in general for patient diagnosis at site-A and site-B. The examination was based on the subjects' history, age, family history, and clinical examination. The clinical test consisted of slit-lamp examination, IOP measurement with Goldman applanation tonometer, gonioscopy, and visual field examination by HFA 24-2 or 10-2 strategy. Ancillary investigation with optical coherence tomography or Heidelberg retinal tomography was ordered when needed. A total of 170 subjects were enrolled for this study. Seven subjects were excluded due to poor image quality and 6 subjects were excluded due to disagreement in a consensus clinical outcome. The final reference data set was made up of fundus images of a total of 314 eyes (157 subjects), with 145 normal, 64 glaucoma suspect, and 105 confirmed glaucomatous eyes.

This reference data set is used to: (i) evaluate the performance of the presented system and (ii) do a comparative analysis with the manual grading (of images) results for the same data set by 4 glaucoma specialists.

Manual Grading on Reference Standard Data Set

The objective of collecting results of manual grading on reference data set is to assess the performance of the automated system against the results of manual grading. The manual grading was entirely based on fundus photograph examination as other clinical details were not provided similar to information available to the automated system. The manual grading results were collected from 4 glaucoma specialists. Two of the 4 specialists were the ones who provided the reference standard opinion. To avoid any bias the manual classification decision was obtained from these 2 specialists after a gap of more than a month of clinical examination of the subjects. Glaucomatous optic neuropathy was defined using evidence of any of the following indicators: excavation, neuro-retinal rim thinning or notching, RNFL defect, PPA, or an asymmetry in the CDR between the left and right eyes. Each eye was classified into 3 different categories: normal, glaucoma suspect, and confirmed glaucomatous aligned with the reference standard data set categories.

Automated Glaucoma Classification System

The development of the presented system primarily involves the extraction of visual indicators from the fundus image, which are associated with glaucomatous damage in an eye. The presence and the distribution of these extracted visual indicators are quantified represented by a feature vector. The feature vector for an image typically encodes the information

to differentiate between different stages of glaucomatous damage in an eye. A set of images of eyes from different glaucomatous categories along with respective image feature vector are used to build a classification model to differentiate among these categories. The resultant model is then used to classify each image from the reference standard data set into 1 of 2 glaucoma categories. The training data set used to build classification model and a complete system description is described in the subsequent subsections.

Training Data Set

An exclusive set of data consisting of 1926 eyes (963 subjects) were collected at site-B using the aforementioned data collection protocol. The clinical examination-based identification of the category for each eye (as used for the reference data set) was not practically possible. Hence, a different scheme was used as described below.

A customized software tool was designed to visualize the fundus photograph of an eye. Each fundus image present in the training data set was presented to 5 glaucoma specialists independently and their opinion was recorded. Their majority consensus opinion was used to assign a label to each eye. It should be noted that the labels now restricted to 2 categories: normal and glaucomatous. The reason for restricting the opinion to 2 categories is as follows: Our system is only capable, by design, to classify an image as belonging to either normal or glaucomatous category. As the labeled set is used to train our automated system in classifying fundus images, the collected opinions also had to be restricted to 2 categories.

System Description

The presented system uses a set of image visual indicators extracted using different image analysis methods published earlier: (a) the segmentation of ONH,[22,23] (b) The detection of PPA and RNFL defect,[24] and (3) low-level structural clustering.[25] The extracted visual indicators are aimed at capturing the intra and peripheral ONH changes associated with glaucomatous damage in the form of a feature vector. The list below summarizes the attributes of the proposed feature vector:

- CDR: It captures the cup enlargement in the inferior and superior directions. Its value ranges from 0 to 1.
- Cup to disc area ratio: It captures the cup enlargement in all directions. Its value ranges from 0 to 1
- PPA: Probability of PPA presence in both inferior and superior directions.
- RNFL defect: Probability of the RNFL defect presence in both the inferior and superior directions.
- Optic disc changes: Histogram-based vector distributions of low-level structures computed within the optic disc, cup, and rim regions.
- Cup symmetry: Difference of color distributions (histograms) of the cup regions above and below the horizontal axis, which bisects the disc. This difference vector captures the symmetry profile of the cup in the inferior and superior directions, which might change in the local cup notching scenario.

To perform glaucoma classification in a previously unseen fundus image, the following steps are performed:

(1) The amount of cupping is estimated by CDR and cup to disc area ratio after segmentation of the disc and cup regions.
(2) The probabilities of the presence of both the peripheral glaucoma indicators PPA and RNFL are estimated in the superior and inferior directions.
(3) The distribution of low-level structures (within the disc, cup, and rim regions) is computed from the result of structural clustering within the ONH region.
(4) Aforementioned features are concatenated to get a final feature set for classification.
(5) A statistical classifier is trained using image features extracted from labeled examples of the training set.
(6) Image features are computed for each fundus image in the reference standard data set. The classification result

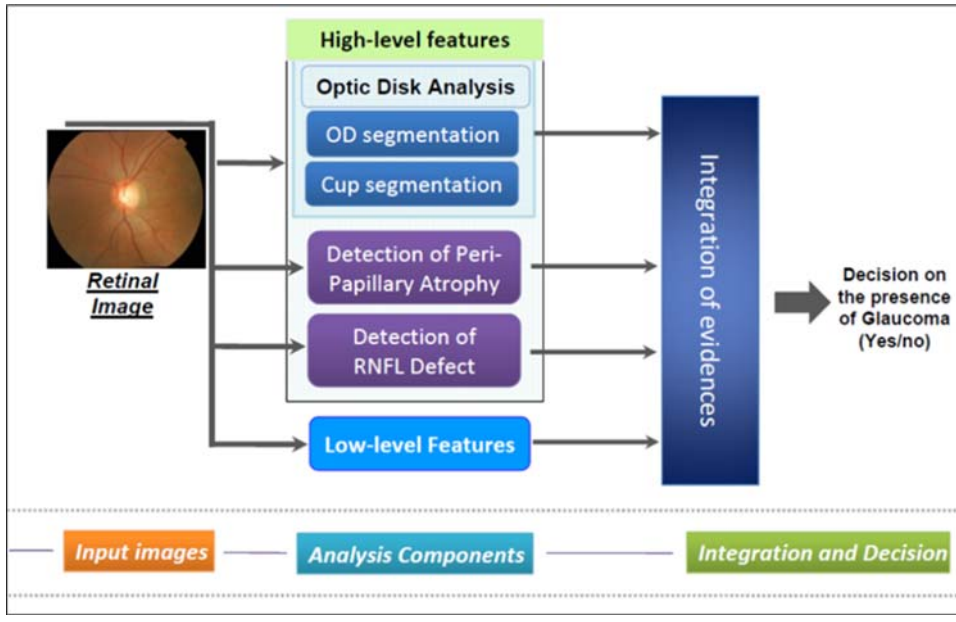

FIGURE 1. An automated system for detection of glaucoma from fundus image. Figure 1 can be viewed in color online at www.glaucomajournal.com

is stored on each eye image of the reference standard data set for later evaluation.

Figure 1 illustrates these steps. For the classification steps (5 and 6), a cascade-forward neural network, which is similar to feed-forward neural networks, except that it has weighted connectivity from the input and every previous layer to the following layers, was used to perform the final classification. A 5-fold cross-validation method on the training set was used to determine the optimal numbers of hidden layers and nodes in each layer, for a given set of input features. The reported results were obtained for 2 hidden layers with 2 and 5 nodes, respectively. The hyperbolic tangent sigmoid was used as a transfer function at each node.

Statistical Evaluation

The presented system is capable of classifying a fundus image as belonging to the normal or glaucomatous category. However, a 3-level opinion (normal, glaucoma suspect, and confirmed glaucomatous) is available for the eye images in the reference standard data set. To evaluate the system performance, the eyes in the last 2 categories (glaucoma suspect and confirmed glaucomatous) from the reference data set were merged to form the "glaucoma" category. This step transforms the 3-level reference data set to a 2-level data set having normal and glaucomatous categories. The merging procedure reflects a situation in a screening setup where suspect glaucomatous cases along with confirm glaucomatous cases shall be referred to a hospital for detailed examination and monitoring. After merging, the reference data set of 314 eyes comprises 145 normal and 169 glaucomatous eyes. The same merging step is applied on the manual grading results to transform 3-level decision on each eye to 2-level decision on each eye in the reference data set. In the presented evaluation, the manual grading results and automated system results obtained on the reference data set are independently evaluated and presented.

Receiver operating characteristic (ROC) curve is used to evaluate system performance on the reference standard data set. ROC curve depicts system sensitivity against (1-specificity) values obtained by varying the threshold applied on the system output, which is on a continuous scale ranging from 0 to 1. The area under the curve (AUC) is a measure of accuracy of the classifier with a value of 1 indicating perfect classification, whereas a value of 0.5 indicates chance and corresponds to a 45-degree diagonal line in the ROC curve. The Cohen κ measure along with 90% confidence interval is used to report the degree of agreement of the system with the reference standard. A κ value of 1 indicates perfect agreement, whereas a value of 0 indicates agreement equivalent to chance.

RESULTS

Table 1 summarizes the obtained results: sensitivity, specificity, Cohen classification agreement κ with the reference standard and hit-miss rates in terms of true positive/negative and false positive/negative. The detection performance of the system depends on the placement of the threshold along the range of output for the classifier. For a fair comparison with manual grading, we report the sensitivity and specificity of the automated system at the average sensitivity value obtained for the 4 specialists who performed manual grading.

Specialist-D has the best grading performance in terms of agreement with the reference data set with the highest κ of 0.59 at a sensitivity of 0.769 and specificity of 0.821. The system has equally good detection sensitivity (0.716) but a lower specificity (0.717). The κ (0.43) value is lowest due to the low specificity of the system.

Figure 2 shows the ROC curve obtained for the system. The performance of each specialist is shown as a single sensitivity/specificity point. The ROC curve can be seen to be below the specialists' data points. In general, a value of 1 for the AUC indicates perfect classification and 0.5 signifies classification that is equivalent to a random assignment. The AUC obtained for the automated system is 0.792.

Next, we present the performance of manual grading on eyes that were misclassified by the system. Both false-positive and false-negative results were considered as misclassification. The system was set to operate with a sensitivity of 0.716 (the average value across specialists) and specificity of 0.717. The bar graphs in Figures 3 and 4 show the decisions of the specialists on the false-positive and false-negative (for the system) cases, respectively. To evaluate the classification performance at a subject level, a set of criteria as described in Table 2 was formulated. A subject-level assessment should ideally require a subject with even 1 glaucomatous eye to be referred to experts for clinical examination. However, as we wish to assess the performance of the system/manual grading, we designed a more stringent set of rules that penalizes the system/specialist if the correct eye of a subject is not identified as glaucomatous. Thus, if according to the system/manual grading, 1 eye of a subject is identified as glaucomatous and

TABLE 1. Performance on Reference Data Set With 314 Eyes

	Manual Grading				
	Specialist-A	**Specialist-B**	**Specialist-C**	**Specialist-D**	**Automated System**
True positive	117	127	118	130	121
True negative	123	117	124	119	104
False positive	22	28	21	26	41
False negative	52	42	51	39	48
Sensitivity	0.692	0.751	0.698	0.769	0.716
Specificity	0.848	0.807	0.855	0.821	0.717
κ	0.53	0.55	0.55	0.59	0.43
95% CI	0.62-0.44	0.64-0.46	0.64-0.45	0.68-0.50	0.53-0.33

Suspect cases are combined with the confirmed cases, sensitivity, specificity, and κ of the automated system and 4 glaucoma specialists against the reference standard.

CI indicates confidence interval.

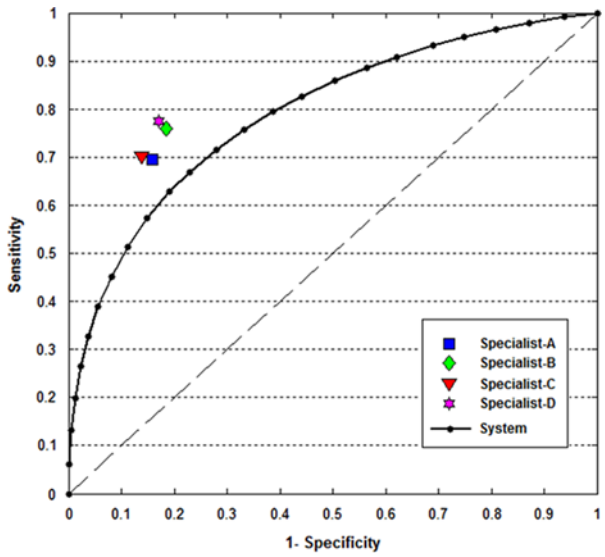

FIGURE 2. Receiver operating characteristic curve for the automated glaucoma classification system on the reference data set. Manual grading performances of 4 specialists are shown as single sensitivity/specificity points with different colors and shapes. Figures 2 can be viewed in color online at www.glaucomajournal.com

there is no match with the reference standard at the eye level, it is deemed to be a false negative by our rules (6 and 7 in Table 2). These rules were used to determine the hit-miss rate for both manual and automated classification. Table 3 presents the performances obtained for both the system and 4 specialists.

DISCUSSION

The automated system presented here analyses fundus images and extracts topographic features of the optic disc

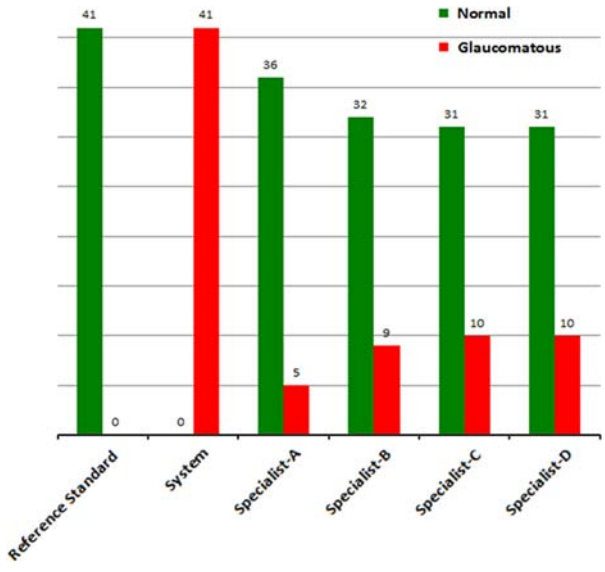

FIGURE 3. Manual grading on systems' false-positive cases. Figures 3 can be viewed in color online at www.glaucomajournal.com

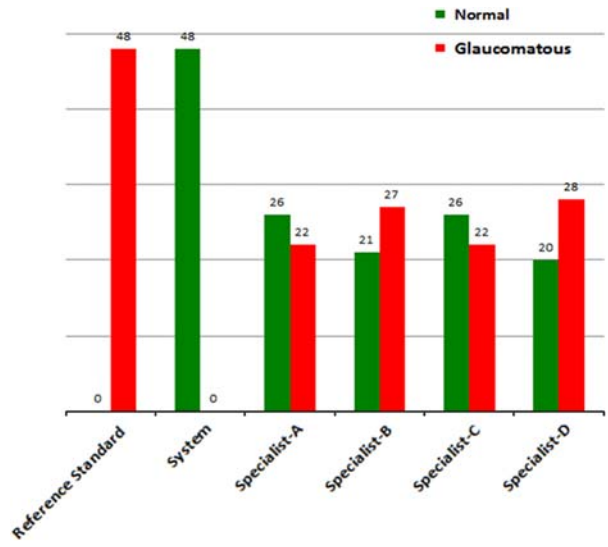

FIGURE 4. Manual grading on systems' false-negative cases. Figures 4 can be viewed in color online at www.glaucomajournal.com

such as the RNFL thinning, neural rim width, and PPA in addition to the CDR to classify a given image as normal or glaucomatous. Prior studies on automated classification of glaucoma have relied on estimation of the CDR from monocular fundus images as the predominant feature for classification. The CDR is a topographical feature and hence ideally requires stereoscopic imaging for accurate estimation. Manual grading of glaucoma, based only on CDR, has been shown to be less reliable when using monocular as compared with stereoscopic images.[26] Exclusive reliance on CDR for automatic classification from monocular fundus images is likely to overestimate the presence of glaucoma with consequent decrease in specificity and enhanced, unnecessary referrals. In contrast, the proposed system, utilizes all known visual, topographic features from monocular fundus images typically used in manual assessment of fundus images. This is a major strength of this study and distinguishes this study from prior studies.

The results of evaluation of the system and manual grading performance in Table 1 indicate that the systems performance is comparable to manual grading in terms of sensitivity but with a lower κ against the reference data set. Specialists C and D were the ones who clinically examined the subjects and provided the diagnosis for deriving the reference data set. Yet, for the image-based manual grading, their average sensitivity/specificity (0.73/0.84) was almost the same as that of specialists A and B (0.72/0.85) who did not examine the subjects' eyes. This suggests that the optic disc assessment based on fundus photographs, by specialists, is handicapped in the absence of ancillary information like visual examination of the eye, IOP, visual field criteria, etc. This handicap is less evident in assessment of eyes with normal optic discs (Fig. 3) as specialists are largely in agreement with Reference Dataset, with at least 3/4 correct decisions. The handicap is more evident in persons with glaucoma (Fig. 4) as at least 40% of true glaucoma were misclassified by specialists as normal. The higher classification error is partly due to the fact that the glaucoma class includes fundus images of eyes that were

TABLE 2. Hit-miss Rate Criteria for Subject-level Performance Evaluation

	Eye-level Reference Standard			Manual Grading/Automated Classification		
Rule No.	Left Eye	Right Eye	Diagnosis Glaucoma (G)/Normal (N)	Left Eye	Right Eye	Category
1	1	0	G	0	1	TP
2	1	0	G	1	0	TP
3	1	1	G	1	1	TP
4	1	1	G	1	0	TP
5	1	1	G	0	1	TP
6	0	1	G	1	0	FN
7	1	0	G	0	1	FN
8	0	0	N	1	1	FP
9	0	0	N	0	1	FP
10	0	0	N	1	0	FP
11	0	0	N	0	0	TN

Values 1 and 0 indicate glaucomatous and normal eye, respectively.
FN indicates false negative; FP, false positive; TN, true negative; TP, true positive.

deemed to be clinically suspect. Thus, such eyes could belong to either normal or glaucoma category when a 2-level decision is required as in manual grading. Interobserver variability in diagnosis of glaucoma has been reported to exist whether decisions are based on clinical examination[26] or fundus images.[14] Our study suggests that the interobserver variance in diagnosis based on manual grading (which is based only on images) is higher for glaucomatous eyes as compared with normal eyes (Figs. 3, 4). Further, these results appear to indicate that the absence of ancillary clinical diagnostic information posed a greater difficulty for specialists in classifying abnormal ONHs as glaucomatous.

At the eye level (Table 1), the average sensitivity/specificity of 4 specialists (based on manual grading) is 0.72/0.83, whereas for the system it is 0.72/0.72. Thus, the system matches the performance of specialists but with a lower specificity. At the subject level (Table 3), the same figures are 0.77/0.77 for specialists and 0.83/0.62 for the system. Thus, the sensitivity of glaucoma detection remains more or less stable for specialists both at eye and subject levels, whereas it is not for the system. The lower specificity of the system, both at eye and subject levels, implies that a higher number of unnecessary referrals would result from automated detection as compared with a manual grading.

Reports on automated systems for glaucoma assessment are very limited. The system presented by Liu et al,[21] which uses multiple sources (image, personal, genetic, etc.) of information, is reported to have a sensitivity/specificity of 0.720/0.662 and an AUC of 0.722. Our system has a sensitivity/specificity of 0.716/0.717 with an AUC of 0.792. However, the 2 studies differ in terms of sample size and populations (Singapore-Malay vs. Indian). Despite the encouraging performance achieved by the presented system, several issues remain and can be improved in the continuation of this study. The system has been tested on a small-sized data set. Therefore, performance in terms of sensitivity, specificity, and κ on a larger prospective data set may not be comparable to these results. The reported performance for manual grading might improve with stereo fundus-based assessment. But this was not possible in this study due to the unavailability of stereo fundus photography. Inclusion of a larger set of fundus images, representing a wider range of disc morphology, is necessary to further improve the performance of the system. The automated system was trained on nearly 2000 images and hence the image labeling was taken to be the consensus among the decisions of multiple experts who did a manual grading of the fundus images. Alternately, the system training can be done with a large data set and the label on each image can be defined on the basis of clinical examination of an eye.

There are some limitations of this study, which need to be addressed in future evaluation studies. The imaging of an undilated eye is the most preferred methodology for screening at a larger scale. However, the findings of this study do not give any insights on the performance of the presented system on undilated eye images. Further, only sufficiently good-quality fundus images are considered in this study with an objective to solely assess the capabilities of the automated glaucoma system. However, image quality might vary in the

TABLE 3. Subject-level Performance on 157 Cases

	Manual Grading				
	Specialist-A	Specialist-B	Specialist-C	Specialist-D	Automated System
True positive	70	78	75	81	80
True negative	47	42	45	47	36
False positive	12	16	13	12	22
False negative	28	21	24	17	19
Sensitivity	0.714	0.788	0.758	0.827	0.808
Specificity	0.797	0.724	0.776	0.797	0.621

TABLE 4. Comparative Evaluation of the Performance of System and Manual Grading on Reference Standard Data Set With Suspect Cases Removed (250 eyes)

	Manual Grading				
	Specialist-A	**Specialist-B**	**Specialist-C**	**Specialist-D**	**Automated System**
True positive	87	96	95	94	87
True negative	123	117	124	119	107
False positive	22	28	21	26	38
False negative	18	9	10	11	18
Sensitivity	0.83	0.91	0.90	0.90	0.83
Specificity	0.84	0.85	0.88	0.85	0.78
κ	0.67	0.70	0.75	0.70	0.55
95% CI	0.58-0.77	0.62-0.79	0.67-0.83	0.61-0.79	0.45-0.66

CI indicates confidence interval.

real population and these findings might not generalize to target population. An automatic image quality assessment or image enhancement module can be planned in the future studies to incorporate images with large quality variations.

In this study, the suspect category was merged with the confirmed glaucomatous category for the evaluation of the proposed system with the assumption that those cases also need a clinical examination and monitoring by a glaucoma expert, to avoid the risk of developing glaucoma in the future. Table 4 presents the system performance on the reference standard after excluding suspect cases. It can be seen that the performance of the system, as well as manual grading, improves significantly. However, we feel that identification of subjects with confirmed glaucomatous or at the risk of developing glaucoma (suspect glaucoma) needs to be addressed in the same manner. This could be a preventive strategy to minimize the risk of missing any subject at risk for glaucoma, during screening. However, this study does not provide any insights to decide on the inclusion and exclusion of suspect glaucomatous cases.

The acceptable sensitivity and specificity of the system, given a low prevalence of glaucoma, cannot be recommended on the basis of the findings of this study. A large-scale evaluation on the system on target population is required with the consideration of other factors such as dilation, inclusion/exclusion of subjects, identification of target population, etc. Furthermore, this study was performed on a particular population in South India and cannot be generalized for populations across different ethnicity/countries.

In summary, we presented an automated system for glaucoma detection, which is comparable in performance to classification based on color fundus images by trained ophthalmologists. Although this indicates the potential for the system to be used in glaucoma screening, for effective population-based screening, however, the system needs to be validated in a larger sample of fundus images. These images need to be with varied optic disc appearances and morphology obtained by nonmydriatic cameras to be representative of the screening populations in future studies.

REFERENCES

1. Resniko S, Pascolini D, Etya'ale D, et al. Global data on visual impairment in the year 2002. *Bull World Health Organ*. 2004;82:844–851.
2. Quigley H, Broman A. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol*. 2006;90: 262–267.
3. Ramakrishnan R, Nirmalan P, Krishnadas S, et al. Glaucoma in a rural population of southern India: the Aravind comprehensive eye survey. *Ophthalmology*. 2003;110: 1484–1490.
4. Vijaya L, George R, Paul P, et al. Prevalence of open-angle glaucoma in a rural south Indian population. *Invest Ophthalmol Vis Sci*. 2005;46:4461–4467, 285.
5. Varma R, Ying-Lai M, Francis B, et al. Prevalence of open-angle glaucoma and ocular hypertension in Latinos: the Los Angeles Latino Eye study. *Ophthalmology*. 2004;111: 1439–1448.
6. de Voogd S, Ikram M, Wolfs R, et al. Incidence of open-angle glaucoma in a general elderly population: the Rotterdam study. *Ophthalmology*. 2005;112:1487–1493.
7. Tielsch J, Sommer A, Katz J, et al. Racial variations in the prevalence of primary open-angle glaucoma—the baltimore eye survey. *JAMA*. 1991;266:369–374.
8. American Academy of Ophthalmology Preferred Practice pattern Committee. Preferred Practice Guidelines. Comprehensive Adult Medical Eye Evaluation. San Francisco, CA: AAO; 2010. Available at: www.aao.org/ppp.
9. Caprioli J. Clinical evaluation of the optic nerve in glaucoma. *Trans Am Ophthalmol Soc*. 1994;92:589–641.
10. Gordon M, Torri V, Miglior S, et al. Validated prediction model for the development of primary open-angle glaucoma in individuals with ocular hypertension. *Ophthalmology*. 2007;114:10–19.
11. Philip S, Fleming A, Goatman K, et al. The efficacy of automated disease/no disease grading for diabetic retinopathy in a systematic screening programme. *Br J Ophthalmol*. 2007; 91:1512–1517.
12. Quellec G, Lamard M, Cazuguel G, et al. Automated assessment of diabetic retinopathy severity using content-based image retrieval in multimodal fundus photographs. *Invest Ophthalmol Vis Sci*. 2011;52:8342–8348.
13. Kong Y, Coote M, O'Neill E, et al. Glaucomatous optic neuropathy evaluation project: a standardized internet system for assessing skills in optic disc examination. *Clin Exp Ophthalmol*. 2011;39:308–317.
14. ONeill E, Gurria L, Pandav S, et al. Glaucomatous optic neuropathy evaluation project factors associated with underestimation of glaucoma likelihood. *JAMA Ophthalmol*. 2014; 132:560–566.
15. Abramoff M, Alward W, Greenlee E, et al. Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. *Invest Ophthalmol Vis Sci*. 2007;48:1665–1673.
16. Corona E, Mitra S, Wilson M, et al. Digital stereo image analyzer for generating automated 3-d measures of optic disc deformation in glaucoma. *IEEE Trans Med Imaging*. 2002; 21:1244–1253.
17. Bock R, Meier J, Michelson G, et al. Classifying glaucoma with image-based features from fundus photographs. *Proc DAGM*. 2007;LNCS 4673:355–364.

18. Bock R, Meier J, Nyl L, et al. Glaucoma risk index: automated glaucoma detection from color fundus images. *Med Image Anal*. 2010;14:471–481.
19. Meier J, Bock R, Michelson G, et al. Effects of preprocessing eye fundus images on appearance based glaucoma classification. *Proc Conf Comp Anal Images Patterns*. 2007;165–172.
20. Yu J, Abidi S, Artes P, et al. Automated optic nerve analysis for diagnostic support in glaucoma. *Proc IEEE Symp Comp Based Med Syst*. 2005;19:97–102.
21. Liu J, Zhang Z, Wong D, et al. Automatic glaucoma diagnosis through medical imaging informatics. *J Am Med Inform Assoc*. 2013;20:1021–1027.
22. Joshi G, Sivaswamy J, Krishnadas S. Depth discontinuity-based cup segmentation from multi-view colour retinal images. *IEEE Trans Biomed Eng*. 2012;59:1523–1531.
23. Joshi G, Sivaswamy J, Krishnadas S. Optic disk and cup segmentation from monocular colour retinal images for glaucoma assessment. *IEEE Trans Med Imaging*. 2011;30: 1192–1205.
24. Joshi G, Prashanth R, Sivaswamy J, et al. Detection of peri-papillary atrophy and Retinal nerve fibre layer defect from retinal images. *Proc. of International conference on Image Analysis and Recognition - Special Session on Retinal Image Analysis*. 2012;400–407.
25. Niemeijer M, Abramoff M, van Ginneken B. Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. *Med Image Anal*. 2006;10:888–898.
26. Varma R, Steinmann W, Scott I. Expert agreement in evaluating the optic disc for glaucoma. *Ophthalmology*. 1992; 99:215–221.