Application of deep learning and image processing analysis of photographs for amblyopia screening

Kaushik Murali, Viswesh Krishna¹, Vrishab Krishna¹, B Kumari

Purpose: Photo screeners and autorefractors have been used to screen children for amblyopia risk factors (ARF) but are limited by cost and efficacy. We looked for a deep learning and image processing analysis-based system to screen for ARF. **Methods:** An android smartphone was used to capture images using a specially coded application that modified the camera setting. An algorithm was developed to process images taken in different light conditions in an automated manner to predict the presence of ARF. Deep learning and image processing models were used to segment images of the face. Light settings and distances were tested to obtain the necessary features. Deep learning was thereafter used to formulate normalized risks using sigmoidal models for each ARF creating a risk dashboard. The model was tested on 54 young adults and results statistically analyzed. **Results:** A combination of low-light and ambient-light images was needed for screening for exclusive ARF. The algorithm had an F-Score of 73.2% with an accuracy of 79.6%, a sensitivity of 88.2%, and a specificity of 75.6% in detecting the ARF. **Conclusion:** Deep-learning and image-processing analysis of photographs acquired from a smartphone are useful in screening for ARF in children and young adults for a referral to doctors for further diagnosis and treatment.

Key words: Amblyopia, deep learning, mobile phone, screening

Amblyopia is a frequently observed visual disorder in children that can lead to permanent visual impairment. Significant refractive errors and strabismus are important amblyogenic risk factors (ARF).[1,2] Amblyopia is present in 1.1–5% of the general population, making it important to screen for amblyogenic factors.[3-5] However, there is limited availability of qualified or trained personnel for primary screening. Hence, photo screening has proven to be an effective method to objectively screen for both refractive errors and amblyogenic factors as defined by 2003 the American Association for Pediatric Ophthalmology and Strabismus (AAPOS) Referral Criteria given below:[6-8]

1. Anisometropia (spherical or cylindrical) >1.5D
2. Significant refractive error
   a. Hyperopia >3.5D in any meridian
   b. Myopia >3.0D in any meridian
   c. Astigmatism >1.5D at 90° or 180°; >1.0D in oblique axis (more than 10° from 90° or 180°).
3. Any manifest strabismus
4. Any media opacity >1 mm in size
5. Ptosis ≤1 mm margin reflex distance.

Non-invasive digital imaging can provide millions of morphological features that can be analyzed in a comprehensive manner using artificial intelligence (AI).[9] Methods based on machine learning (ML) and particularly deep learning (DL) have been used in the screening of various ocular conditions. They are effective to identify, localize, and quantify pathological features in a variety of retinal diseases such as diabetic retinopathy.[10-12]

Combining principles of photo screening and DL, we looked to develop a simple photography-based solution, called Kanna, which would help detect amblyogenicity based on the risk factors as defined by the 2003 AAPOS referral criteria.

A preliminary analysis of images showed that refractive error and media opacities could be studied using a red reflex image which required low lighting conditions and an increased distance between the patient and the image capture device for effective measurements. However, a decrease in light intensity and an increase in the distance caused the generation of a lower resolution image which could not be used for accurate ptosis and strabismus measurements. Therefore, two separate images in ambient and dark surroundings were obtained to focus on the detection of separate ARF.

Methods

The study was conducted at Sankara Eye Hospital and College of Optometry Bangalore. Fifty-four optometry students were recruited for data acquisition. Optometry students were considered instead of children or a general population as the goal of the study was to check if a deep-learning algorithm coupled with an android smartphone was an effective screening modality for ARF. A comprehensive eye exam was performed by an ophthalmologist and, subsequently, images of their face were taken using a smartphone. As part of the eye exam,

Sankara Academy of Vision, Sankara Eye Hospital, ¹National Public School Indiranagar, Bengaluru, Karnataka, India

Correspondence to: Dr. Kaushik Murali, Sankara Eye Hospital, Varthur Road, Kundalahalli Gate, Bengaluru - 560 037, Karnataka, India. E-mail: kaushik@sankaraeye.com; murali.kaushik@gmail.com

refractive error measurement of both eyes, the Hirschberg and cover-uncover tests, the assessment of squint and ptosis evaluation was performed (Margin Reflex Distance 1 and 2). As part of the smartphone imaging procedure, facial images were captured with flash in dark (3–10 lm) and ambient (60–800 lm) light conditions separately at the distances of 0.5 m, 1 m, and 1.5 m each (based on the retinoscopy principle). As both low-light and ambient-light images were to be acquired with flash-enabled, low-light images are to be acquired before the ambient-light images to prevent unwanted pupillary contraction in the red reflex image and decrease the time for pupillary dilation in low-light surroundings.

The light intensity of the ambient room was determined on the basis of the best image quality and resolution as well as the absence of the red reflex. Conversely, the dim-lit room luminosity was predetermined based on the presence of red reflex. An android application was written to modify the smartphone camera flash settings to remove the inbuilt pre-flash that removes red-eye images in low-light conditions.

The Kanna algorithm was developed using DL and image-processing models to conduct measurements and predictions for the presence of ARF in an automated fashion in the following manner.

All images were preprocessed using the Gaussian Blur algorithm and converted to grayscale for the application of DL cascade models.[12] The eye was localized using facial landmarks predicted by DL models [Fig. 1].[13,14] A Convolutional Neural Network (CNN) was trained to detect six iris landmarks along the iris boundary using the UnityEyes dataset.

Parameters like red reflex localization, undilated pupil radius, the hue of the iris region and red reflex, and crescent width were measured using the dark-lit image.[15] These parameters were used to develop the algorithm for the prediction of refractive error as well as media opacities. The eyelid contour, corneal light reflex (CLR), and iris center were determined from the ambient-light image. The angle of squint was calculated mathematically using the CLR position, iris center, and the biometric ratio of the eyeball radius to the iris diameter [described in depth in Appendix 1]. Ptosis (MRD 1 and 2) was measured using the eyelid contour and CLR [Fig. 2].

Based on the 2003 AAPOS referral criteria, we devised a risk prediction system. If any of the five ARF (Anisometropia, Isoametropia, Strabismus, Ptosis, or Media Opacities) exceeded the prescribed thresholds, the image was flagged as ARF positive (predicted with an ARF risk of 1), else it was predicted with ARF risk of 0.

Statistical analysis was done with Microsoft Excel and the NumPy Library in the Python programming language.[16] A confusion matrix was created and phi coefficients were calculated to analyze the correlation between the clinical and Kanna predictions. Based on the degree of freedom (D.F.) and sample size (N = 54) P value at a 0.05 significance threshold was considered.

Results

The 54 optometry students were aged 18–23. The low-light image was able to detect media opacities, isometropia, and anisometropia. External facial characteristics such as ptosis and strabismus could not be determined with a high degree of accuracy from the low-light photography and were studied in the ambient-light photograph. The distribution of patients with various ARF detected clinically and by Kanna algorithm is described in Table 1.

The sharpness and quality of the ambient-light image were best obtained when captured at 0.5 m and that for the low-light image at 1 m.

The confusion matrix of the Kanna algorithm has been illustrated in Table 2. The P value was calculated to be 0.00011 which was less than 0.05 (statistically significant) was calculated using McNemar's Test.
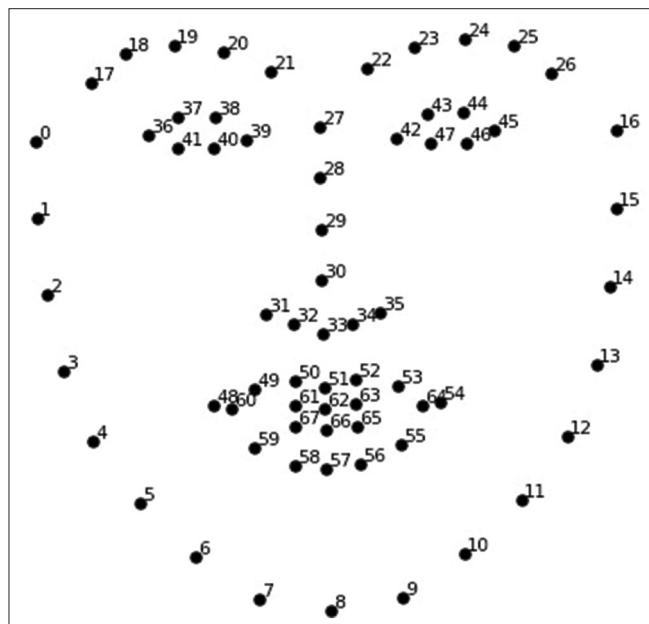

Figure 1: Position of 68 facial landmarks detected (image at bit.ly/2Jgdar0)

Table 1: Dataset and prediction composition

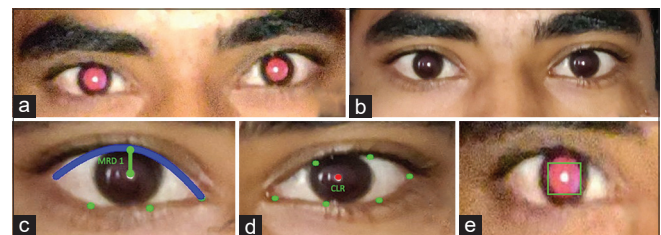|	Clinical	Kanna
Normal	37	30
Anisometropia	7	13
Isoametropia	9	11
Strabismus	1	1
Ptosis	0	0
Media opacities	0	0
Total	54	54

Figure 2: Stages of processing: (a) red reflex image (b) ambient image (c) ptosis measurement (d) strabismus measurement (e) red reflex measurement

The algorithm had an F-Score of 73.2% which should be strongly considered due to the imbalance of the confusion matrix. The sensitivity was 88.2% and the specificity was 75.6% as shown in Table 3. The algorithm detected strabismus (n = 1/1) and refractive errors (n = 14/16) with high accuracy. The correlation coefficient value of strabismus predictions was 0.88, isoametropia was 0.82, and that of anisometropia was 0.79.

Discussion

DL and image processing algorithms effectively detect the described ARF. Ambient-light and low-light photographs screen for exclusive ARF which together enable screening for all ARF described.

Most photo screeners currently available involve customized hardware and software which hinder cost and efficacy for screening.[17] GoCheckKids (Gobiquity Mobile Health, Scottsdale, AZ, USA) enables screening via a preinstalled app on a smartphone but includes a customized hardware component.[18] Our solution involves only a downloadable application for Android mobile phones with no added hardware component.

The advantages of our solutions are:
1. It does not require any specialized hardware/equipment
2. The image can be captured by a layman—does not require any technician as in operating the current auto refractometer
3. Low-cost and widespread availability and utility of smartphones in the current era would help to potentially expand the screening services in every corner of the world
4. Early identification of ARF would help in early referral; thus, helping to reduce the amblyopia incidence
5. In places with a lack of eye care facilities, parents can themselves test their children at home as a preliminary screening for ARF.

Our algorithm also has some limitations:
1. The sample size used for the algorithm was small
2. While pupillary size did not affect the prediction of ptosis and strabismus from the ambient-light image, extreme variation in the pupillary size affected the prediction of

refractive error and media opacities. In cases of extremely small undilated pupil size in a dark-lit room, the red reflex was not visible and the determination of refractive errors and media opacities was not possible. This resulted in two cases being missed by the algorithm
3. Four cases (isoametropia and anisometropia) were falsely predicted as positive due to the blurred margins of the CLR overlapping into the crescent zone causing the algorithm to predict a larger refractive error than the actual. The development of more sophisticated DL models on a larger dataset could potentially resolve this issue
4. Astigmatism affects refractive error measurements due to eccentric photorefraction. The possible differences in axes of cylindrical refractive errors cause varying effects on the red reflex which are difficult to quantify through a single approach (as the camera is held in one particular axis). It is possible that capturing multiple images with a camera placed horizontally and vertically may enable precise astigmatism measurements through the application of convolutional neural networks.

Spherical equivalent values have a stronger correlation than spherical or cylindrical refractive errors as they combine the effect of both spherical and cylindrical refractive errors in the red reflex image. Low positive predictive value (PPV) is within expectations as it is a screening device and was strongly affected by astigmatism. High sensitivity and specificity strongly indicate the potential of this approach for screening amblyopia at a larger scale. Further, it is affordable and scalable as it involves just a mobile phone for screening and uploading photographs for processing with the algorithm.

To the best of our knowledge, this is the first time; a DL model has been developed to identify amblyogenic risk factors.

Conclusion

DL and image processing of facial photographs is capable of screening young children with amblyopia risk factors for a referral to ophthalmologists for further diagnosis and treatment. It is advantageous over using traditional screeners, as it is easily accessible, low-cost, and requires minimal training.

Conflicts of interest
There are no conflicts of interest.

Table 2: Confusion matrix for the classification algorithm

ARF	Kanna positive	Kanna negative	Row total
Clinical positive	15	2	17
Clinical negative	9	28	37
Column total	24	30	54 (N)

ARF: Amblyopia risk factors

Table 3: Accuracy metrics

Accuracy metrics		95% CI
Accuracy	79.6%	66.47-89.37%
Sensitivity	88.2%	63.56-98.54%
Specificity	75.6%	58.80-88.23%
Positive predictive value	62.5%	47.92-75.12%
Negative predictive value	93.3%	78.99-98.12%
F-score	73.2%	54.64-85.25%

(P-value=0.00011 <0.05 significance)

References

1. Rajavi Z, Parsafar H, Ramezani A, Yaseri M, Is non-cycloplegic photorefraction applicable for screening refractive amblyopia risk factors? J Ophthalmic Vis Res 2012;7:3-9.

2. Paff T, Oudesluys-Murphy AM, Wolterbeek R, Swart-van den Berg M, Tijssen E, Schalij-Delfos NE. Screening for refractive errors in children: The PlusoptiX S08 and the Retinomax K-plus2 performed by a lay screener compared to cycloplegic retinoscopy. J AAPOS 2010;14:478-83.

3. Eibschitz-Tsimhoni M, Friedman T, Naor J, Eibschitz N, Friedman Z. Early screening for amblyogenic risk factors lowers the prevalence and severity of amblyopia. J AAPOS 2000;4:194-9.

4. Newman DK, East MM, Prevalence of amblyopia among defaulters of preschool vision screening. Ophthalmol Epidemiol 2006;7:67-71.

5. Karki KJD. Prevalence of amblyopia in ametropias in a clinical set-up. Kathmandu Univ Med J 2006;4:470-3.

6. Donahue SP, Baker JD, Scott WE, Rychwalski P, Neely DE, Tong P, et al. Lions clubs international foundation core. Four photoscreening: Results from 17 programs and 400,000 preschool children. J AAPOS 2006;10:44-8.

7. Salcido AA, Bradley J, Donahue SP. Predictive value of photoscreening and traditional screening of preschool children. J AAPOS 2005;9:114-20.

8. Donahue SP, Arnold RW, Ruben JB, for AAPOS Vision Screening Committee. Preschool vision screening: What should we be detecting and how should we report it? Uniform guidelines for reporting results of preschool vision screening studies. J AAPOS 2003;7:314-6.

9. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. Prog Retin Eye Res 2018;67:1-29.

10. Birtel J, Lindner M, Mishra DK, Müller PL, Hendig D, Herrmann P, et al. Retinal imaging including optical coherence tomography angiography for detecting active choroidal neovascularization in pseudoxanthoma elasticum. Clin Exp Ophthalmol 2019;47:240-9.

11. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 2016;57:5200-6.

12. Zelinsky A. Learning OpenCV---Computer vision with the OpenCV library (Bradski, G.R. et al.; 2008) [On the Shelf]. IEEE Robot Autom Mag 2009;16:100.

13. King DE. Dlib-ml: A machine learning toolkit. J Mach Learn Res 2009;10:1755-8.

14. Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 1867-1874.

15. Bobier WR, Braddick OJ. Eccentric photorefraction. Opt Vis Sci 1985;62:614-20.

16. Oliphant TE. A guide to NumPy. Vol. 1. USA: Trelgol Publishing; 2006.

17. Sanchez I, Ortiz-Toquero S, Martin R, de Juan V. Advantages, limitations, and diagnostic accuracy of photoscreeners in early detection of amblyopia: A review. Clin Ophthalmol 2016;10:1365-73.

18. Peterseim MMW, Rhodes RS, Patel RN, Wilson ME, Edmondson LE, Logan SA, et al. Effectiveness of the GoCheck kids vision screener in detecting amblyopia risk factors. Am J Ophthalmol 2018;187:87-91.

Appendix 1

Manifest strabismus was measured using the ambient-light photograph through the Hirschberg test using the CLR. The deviation of the CLR was quantified by considering its distance from the center of the iris. However, the deviation of the iris results in the apparent center of the iris not lying on the optical center. In order to account for this, we make use of the corneal radius and eyeball radius (calculated by scaling the iris diameter using the ratios of their mean values) to calculate the exact angle. The expression used to calculate the final strabismic deviation is derived from Fig. 3 and is provided below. Pseudo-strabismus was accounted for by considering the difference in the side of the deviation of the CLR with respect to the center of the iris.

$$R \sin \theta = d + c \sin \theta$$
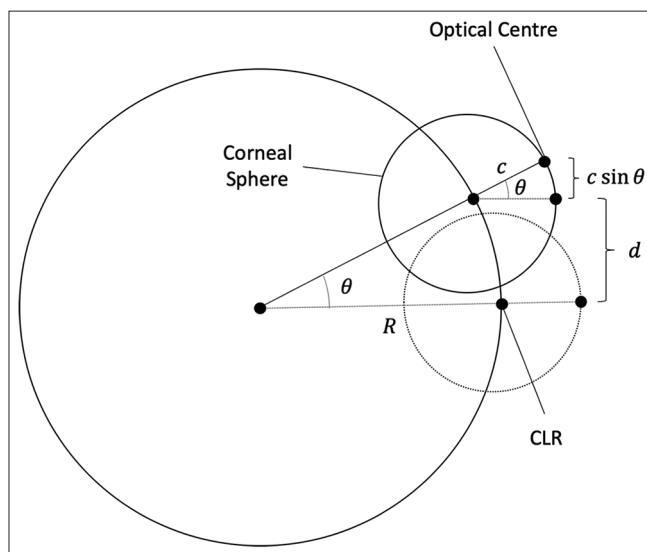
$$\Rightarrow \theta = \sin^{-1} \frac{d}{R-c}$$

Figure 3: Geometric description of the strabismic deviation