IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 4, 2020, accepted September 13, 2020, date of publication September 21, 2020,
date of current version October 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3025346

Unified Diagnosis Framework for Automated Nuclear Cataract Grading Based on Smartphone Slit-Lamp Images

SHENMING HU[1,2], XIAOTING WANG[2,4], HONG WU[3], XINZE LUAN[2], PENG QI[2,6], YI LIN[2], XIANGDONG HE[5], AND WEI HE[1]

[1]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110000, China
[2]He University, Shenyang 110000, China
[3]Shenyang Eyerobo Company Ltd., Shenyang 110000, China
[4]Renmin University of China, Beijing 100000, China
[5]Shenyang Eye Industry Technology Research Institute Company Ltd., Shenyang 110000, China
[6]He Eye Specialist Hospital, Shenyang 110000, China

Corresponding author: Xiaoting Wang (wangxiaoting@huh.edu.cn)

This work was supported in part by the Hand-held Ocular Surface Screening Project under Grant SJ201905, and in part by the Construction of Shenyang Eye Industry Technology Research Institute (Phase II) under Grant 19-301-4-02.

ABSTRACT Cataract constitutes half of the blindness cases worldwide; hence, detecting and treating cataracts in a timely manner are effective strategies for blindness prevention. Recently, methods of detecting cataracts through deep learning are flourishing; however, the task of improving the grading mechanism is still the priority in the research field. This study evaluates the classification capability of the automated nuclear cataract detection algorithm using ocular images captured by smartphone-based slit-lamp. The task of the algorithm is to automatically detect cataract severity in terms of the photometric appearance of the nuclear region of the crystalline lens of the eyes. The nuclear region of the ocular lens was localized by YOLOv3. Subsequently, the combination of a deep learning network, ShuffleNet, and a support vector machine (SVM) classifier was used to grade cataract severity, evaluating the gray conjugate features of the nuclear region. Using the trained algorithm, 819 anterior ocular images captured by smartphone-based slit-lamp were utilized to evaluate the algorithm's performance. The accuracy was 93.5% with Kappa of 95.4% and F1 of 92.3%. The AUC was 0.9198. The proposed validation method could evaluate a cataract severity in 29 ms and the entire classification process in less than 1s. This study can improve the accuracy of the examination, reduce misdiagnosis rate and the difficulty of the doctor's examination. The addition of scoring system can improve the quality of pictures obtained by non-ophthalmologists. The method is especially suitable for cataract screening in the underdeveloped areas or areas which are in shortage of ophthalmic resources. It can also improve the accessibility of ophthalmic medical treatment.

INDEX TERMS Images captured by smartphone-based slit-lamp, automated cataract detection, grade cataract, deep learning.

I. INTRODUCTION

Today, the world has approximately 400 million vision impairment and 40 million blind population [1]. As a country with the largest population, China accounts for a quarter of the world's vision impairment and blindness. In addition, among the blind population, more than half are caused by cataracts [2]. American Academy of Ophthalmology defined cataract as the clouding of the lens [3]. The lens opacity also significantly reduced one's visual acuity and quality of life; hence, the delay in detecting and treating cataracts also cause critical burden to the society [4], [5]. Studies show that the risk factors of cataract include ageing, diabetes, hypertension, smoking, and the exposure to radiation [6].

There is no definite method to prevent cataract [6]. The treatment can be restored by installing artificial lens surgically. The recovery effect is obvious, and it almost achieves the normal visual function. Therefore, early and accurate detection is critical to preventing vision loss.

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

Cataract can be generally divided into congenital cataract and acquired cataract. Acquired cataract is mostly senile cataract, and occurs above the age of 40 years, except cataract caused by external factors such as trauma, poisoning and radiation. The human lens is an oval structure, which is located behind the iris and can be divided into three layers: cortical layer, nucleus and posterior capsule. Clinically, senile cataract is usually divided into cortical cataract, nuclear cataract and posterior cystic cataract according to its stages. Cortical cataract is mainly due to the gray-white turbidity of the lens cortex layer, which can gradually develop into excessive maturity from the initial stage. Nuclear cataracts usually start in the center of the lens, that is, the nuclear layer. It gradually worsens and expands to all sides at the same time. The color changes from light to deep, initially showing light yellow, then gradually becoming dark yellow and dark brown yellow as the disease worsens and the color deepens. People with cataracts are relatively more likely to develop nuclear cataracts, so nuclear cataracts are also most common. Posterior cystic cataract is a lesion in the posterior capsule of the lens. If it is located in the visual axis region, it can affect vision in the early stage. Therefore, how to accurately diagnose and treat cataract in time is important to avoid blindness caused by cataract. Currently, cataract is diagnosed by doctors, observing the state of the lens through a slit lamp or a hand-held slit lamp under a slit lamp microscope. Doctors can diagnose by comparing the examinees' crystal images with standard graded images. These standard graded images include the Lens Opacity Classification System III(LOCS) [7]. The Wisconsin Cataract Classification System [8]. However, such graded assessments are subjective. Automatic grading will improve diagnostic efficiency and avoid the involvement of subjective factors. It can effectively improve the clinical management of cataract disease, and provide theoretical basis for the epidemiology [9]. Especially in the less developed areas with scarce ophthalmologists, the automatic grading system can allow ordinary doctors to screen cataracts and provide doctors with expertise in artificial intelligence-assisted diagnosis treatment.

Recently, multiple computer aided diagnostics in disease detection appear in the field of ophthalmology. But most of the technologies are focused to detect posterior eye diseases [10]. Hence, this study proposes a cataract severity classification algorithm based on deep learning, to provide a basis for automated disease detection of anterior eye diseases.

In this work, our primary contribution is to propose an innovative framework for automated classification using artificial intelligence. This framework integrates three different artificial intelligence networks by means of logical regression, which improves the accuracy of diagnosis and reduces the false positive rate in the process of screening, so that doctors can easily obtain the crystal picture of the patient. This paper also innovatively uses yolov3 for pupil location and cataract identification, instead of simply classifying pictures. Because of the high accuracy of yolov3, it can effectively reduce the difficulty of doctors taking pictures and make it easier for doctors to obtain lens images of subjects, so that through the optimization of the framework, the difficulty of examination of cataract can be reduced.

The article is structured as follows: The second section reviews the relevant automated cataract classification algorithms in the literature. The third section introduces the proposed algorithm. The fourth section evaluates the proposed algorithm using a testing dataset. The final section discusses the results.

II. RELATED WORKS

This chapter summarizes automated cataract detection algorithm that is recently documented in the literature.

The majority of the day's automatic cataract detection algorithms was developed using ocular images captured by slit-lamps as the study datasets. Reference [11], [12] used image rankings on neighbor markers and optimized the learning functions to make cataract severity predictions. Reference [13] designated the color of the crystalline lens as the key feature to provide feedback and support clinicians to make diagnosis. Reference [14] developed an algorithm to classify cataract severity by evaluating visual axis and extracted critical features and their significance from slit-lamp images. With the development of research, the researchers in reference [15]–[22] found that it can effectively extract the features of cataract images, and adequate usage of classification algorithm could accomplish automatic classification. The basic principle is to extract global features or local features of the image to use support vector machine (SVM) or vector regression (SVR) to complete the classification tasks. The algorithms accuracy is up to 90%. Reference [23] used sparse linear regression constraints to perform feature and parameter selection together to complete the grading task. Reference [24] first used the enhanced texture features of extracted cataract images, and the team used statistical data of the enhanced texture features to train linear discriminant analysis (LDA) to detect cataracts. The study tested the algorithm using 4,545 clinical images and obtained an accuracy of 84.8%. Reference [25] extracted features from fundus images based on wavelet transform or discrete cosine transformation and used classification algorithm to complete classification tasks. Although nuclear cataract is the most common type of cataract, multiple studies were developing algorithms in detecting cortical cataracts. Reference [26], [27] used nonlinear least square method and edge detection method to detect cortical opacity and judge the grade of cortical cataract. Other studies distinguished cortical cataract by separating cortical opacity from other opacity types based on different characteristics. The mentioned classification methods utilized traditional methods. Today, with the rapid development of deep learning, researches on cataract classification using deep learning approaches are carried out widely. In reference [28], convolution neural network (CNN) was used to extract the features of pediatric ocular images, and SVM algorithm or softmax classifier are combined to achieve automatic classifications.

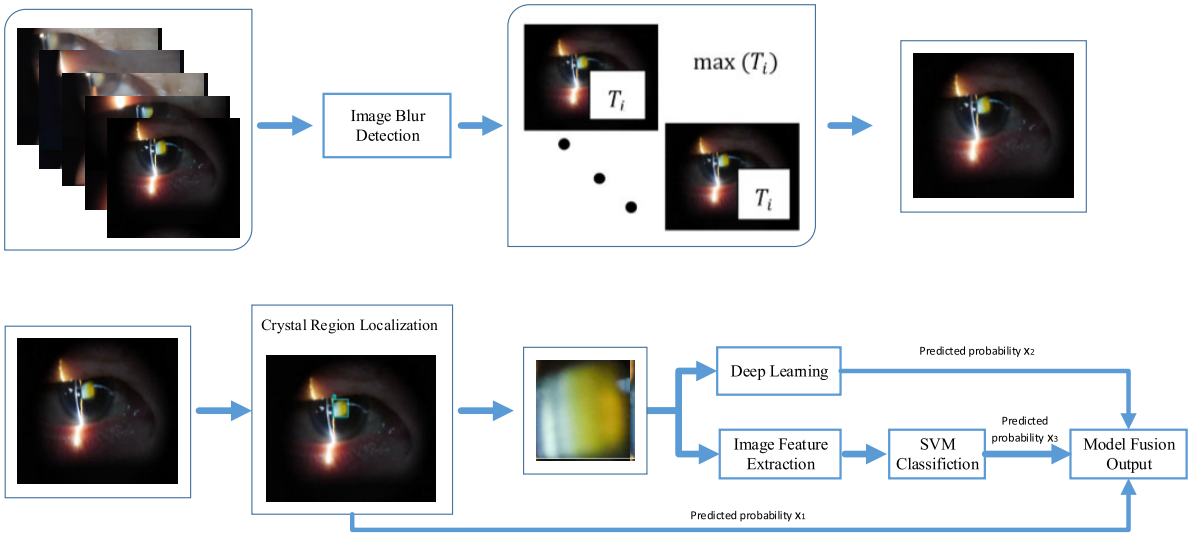


FIGURE 1. An illustration of the proposed UDFA based solution for slit-lamp based nuclear cataract grading.

Compared with conventional methods, the suggested deep learning approach was more effective on automatic classification. Reference [29] also used deep convolution neural network to complete cataract classification. The study trained an algorithm using fundus images and achieved a grading accuracy of 86.7%. Reference [30] published a paper at MICCAI conference to better study the classification of nuclear cataract. By using Faster R-CNN to locate its nuclear region and taking the nuclear region as input, the classification model based on ResNet-101 was trained. Although the method of automated cataract classification was continuously proposed, the approach to improving the accuracy of cataract classification is still a problem that needs to be addressed. This study proposes a combination of three machine learning methods to finally complete cataract classification prediction. The detailed methods are presented in the next section of the paper.

III. PROPOSED METHODOLOGY

Smartphone-based slit-lamp photos from individuals with different cataract severity were used in this study, and the algorithm's main task is to automatically classify the cataract severity based on the photometric appearance of the crystalline lens. Figure 1 represents the complete framework of the proposed algorithm.

A. IMAGE BLUR DETECTION

Multiple images of individuals would be captured while users of the smartphone-based slit-lamp are capturing ocular images for eye screening. It is likely that the images are blurry due to shiver and lack of focus when the images are captured. To resolve this issue, Tenengard [31] algorithm is used in the automated nuclear cataract classification framework to calculate the T-values of multiple images taken by the slit-lamps. Among all, the image with the largest T-value is selected as the output and entered into the next nuclear region detection algorithm. Set the dataset I be a plurality of images taken and Ii is the i-th image among them.

Tenengard [31] function is a gradient-based image definition evaluation function. During the image processing mechanism, an image with sharper edges has larger gradient function values. Tenengrad function uses Sobel operator to extract gradient values in horizontal and vertical directions. The larger the average gray value of the image processed by Sobel operator, the clearer the image. The specific process is shown below:

Set the Sobel convolution kernel be G_x, G_y, and the gradient of the image I_i at the point (x, y):

$$S_i(x, y) = \sqrt{G_x \times I_i(x, y) + G_y \times I_i(x, y)} \tag{1}$$

The Tenengrad value of an image is defined as:

$$Ti = \frac{1}{n} * \sum_x \sum_y s(x, y)^2 \tag{2}$$

In formula (2), n is the total number of pixels of the image. The value T_i of each image in the data set I s calculated. Fig.2 shows the values of image in (a) $\sim$ (d) useing equation (1), and (2) to find the i-th image corresponding to the final return value. max(T_i), $i \in I$ of T value would be used in the next step.

B. NUCLEAR REGION LOCALIZATION BY YOLOv3

Reference [32] proposed YOLO v3. The new feature extraction network used by YOLO v3 integrates the network

Darknet-19 in YOLO v2 and the new popular residual network (residual structure of ResNet). The network uses a large number of 3 $\times$ 3 and 1 $\times$ 1 convolution layers to connect and add shortcut connections between the nodes, so its network structure is more complex with 53 convolution layers. Therefore, YOLO v3 feature extraction network is

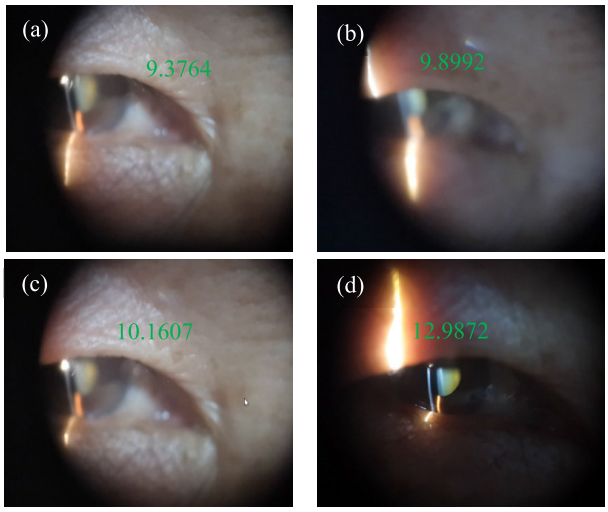


FIGURE 2. Illustrations of anterior ocular image captured by physicians using smartphone-based slit-lamps in practice (images from the same eye). The green values shown on the top left corner of the images (a)-(d) represent the Tenengrad score evaluating the image clarity. The higher the value, the clearer the image.

called Darknet-53. YOLO v3 improves the prediction accuracy on the premise of maintaining the speed advantage, especially strengthening the recognition ability for small objects. We mainly evaluate our models on the Marked Slit Lamp Picture Project (MSLPP) classification dataset. We follow most of the training settings and hyper-parameters used in most papers, and we set the weight decay to 0.0005 and the learning rate is set to 0.001. And the momentum is set to 0.9. The input image size is 416 × 416. It takes 1 day to train a model on 1 GPU, whose batch size is set to 64. Leaky is used to activate the function. We use exactly the same settings for all models to ensure fair comparisons.

When loss converges, the training model is output. During the test, the image of the test set is taken as the input, and the input image size is defined as 416 × 416. According to the training model, the position coordinates of the nuclear region are confirmed, and the required data set is obtained according to the corresponding operations. This algorithm shrinks the scale of the original images. Even though the coordinates are accurate, the pixels of the obtained intercepted image data will be lost. Therefore, after obtaining the coordinates, they are converted to coordinates that can be intercepted on the basis of the original image in proportion k_1, k_2. The specific operations are as follows:

$$k_1 = \frac{image.size\,[0]}{model\_image\_size\,[0]} \tag{3}$$

$$k_2 = \frac{image.size\,[1]}{model\_image\_size\,[1]} \tag{4}$$

$image.size\,[0]$, $image.size\,[1]$ are the length and width of the original image. The values of $model\_image\_size\,[0]$ and $model\_image\_size\,[1]$ are 416, so the obtained positioning coordinates ((left, top), (right, bottom)) in proportion k_1, k_2 are converted into:

$$top_2 = floor(k_1 * top) \tag{5}$$

$$left_2 = floor(k_2 * left) \tag{6}$$

$$bottom_2 = floor(k_1 * bottom) \tag{7}$$

$$right_2 = floor(k_2 * right) \tag{8}$$

C. GRADING

When classifying the nuclear lens regions, we adopt two classification methods, namely ShuffleNet network based on deep learning, and SVM classifier model based on gray conjugate features.

1) ShuffleNet V1

ShuffleNet v1 [33] achieves higher accuracy than other lightweight models in ImageNet classification and MS COCO target detection tasks, such as MobileNet v1. On ARM devices, ShuffleNet is 13 times faster than AlexNet. The architecture designed in this paper adopts ShuffleNet network in one of the cataract classification recognition algorithms, which has the advantages of higher accuracy and faster speed.

This section analyzes the pre-training and migration learning based on ShuffleNet network, and uses ImageNet as the dataset for pre-training. At the same time, the transfer learning technology, Leaky ReLU is used to activate the function. Subsequently, its own dataset is added for re-training. Our models are largely evaluated on the Marked Slit Lamp Picture Project(MLSPP) classification dataset. We follow most of the training settings and hyper-parameters used in most papers, but we set the weight decay to 4e-5 instead of 1e-4. Use linear-decay learning rate policy. It takes 1 or 2 days to train a model on 1 GPU, whose batch size is set to 32. Leaky ReLU is used to activate the function. The input image size is 224 × 224, the learning rate is set to 0.001. We use exactly the same settings for all models to ensure fair comparisons.

2) CONVENTIONAL METHOD FOR GRADING FEATURE EXTRACTION

Statistical texture analysis is divided into two types: the first -order of statistical texture method and the second-order of statistical texture method. The first-order of the statistical texture method is based on the characteristics of histogram image to calculate features. In some cases, the first-order statistical texture method cannot be used to identify the differences between images. In this study, we used Gray Level Co-occurrence Matrix (GLCM). This method distinguishes the normal and cataract. GLCM was proved to be popular statistical method of extracting textural feature from images. GLCM reflects the comprehensive information of image gray level about direction, adjacent interval, and variation amplitude. GLCM can be used to analyze the local features and arrangement rules of images. To describe texture conditions more intuitively with GLCM, the obtained co-occurrence matrix was not directly applied. Quadratic statistics were obtained on the basis of the obtained co-occurrence

matrix. Reference [34], [35] defined 14 feature parameters of GLCM for texture analysis: second moment, contrast, correlation, difference moment, deficit moment (homogeneity), sum average, sum variance, sum entropy, entropy, variance, difference entropy, correlation measure 1, correlation measure 2, and maximum correlation coefficient.

Ulaby *et al.* found that among the 14 texture features based on GLCM, only 4 feature parameters were irrelevant. These 4 features are convenient to calculate and can achieve high classification accuracy. Generally, second-order moment (energy), contrast, correlation and entropy are used. The four most commonly used features are used to extract texture features of images. SVM is used to train the classifier for the gray conjugate matrix generated by the above method, and the model model_svm is generated.

3) MODEL FUSION

With the aim of improving the accuracy of the whole stacking method, we proposed a new fusion process of combining YOLO v3 model, ShuffleNet model, and the support vector machine based on gray level Gray Level Co-occurrence Matrix. Stacking [36] method uses the model output values that need to be merged as the input values in the next stage, and the final classification label is used as the output value, so that the coefficient can be obtained by the logistic regression method.

The logistic regression can use the equation (9):

$$\text{logit P} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m \tag{9}$$

P is the final category. $x_1, x_2, \cdots, x_m$ are deterministic variables associated with probability. α is a constant. $\beta 1, \beta 2, \ldots \beta$m are the logistic regression coefficient, which indicates the coefficients at the possibility of P after the input values of the logistic regression $x_1, x_2, \cdots, x_m$ are confirmed.

In this method, three different models are used for integration. So m=3, the output values of three different models are $x_1 \sim x_3$, as the inputs of logistic regression. P is the known label of the final classification. The $\beta 1$-$\beta 3$ trinomial coefficients can be obtained by fitting calculation.

The detailed process of fusion is as follows: At first, We performed 5-fold cross validation on the training dataset of 500 images, and obtained the results t1∼t5, also used 100 images as the test dataset. Next, we would utilize three machine learning methods to categorize the data: method C, D, and E. (Method C represents YOLOv3; D represents ShuffleNet; E represents method B in P.3. The complete process of stacking is: Towards each method C, D, and E, it would use dataset t2 to t5 to train, and test on t1. The result would be saved as d1, and the results on the testing dataset would be saved as td1. Subsequently, t1, t3, t4, t5 would be used to train the algorithm and be tested on t2. The result would be saved as d2, and the results on the testing dataset would be saved as td2. Under the same process, d1, d2, d3, d4, d5 and td1, td2, td3, td4, td5 would be obtained. The results of d1 to d5 would be combined as d, then d would be fed into model C, D, and E, so the trained results would be considered DC, DD, DE respectively. Meanwhile, we would calculate the average value of td1 to td5 to obtain TC, TD, TE. Eventually, DC, DD, and DE would be considered input values. A logistic regression would be carried out to train further to address the weight of every algorithm before obtaining the complete model of model_final.

IV. EXPERIMENTS

A. DATASETS

The study data was obtained from the Marked Slit Lamp Picture Project (MSLPP) dataset created by He Eye Specialist Hospital (HESH) from 2015 to 2018. The dataset consists of 16,103 anterior ocular images. Among the images, 4,738 images were captured from eyes with pronounced cataracts, 5,346 images from early cataracts, and 6,019 images from non-cataracts.

The images were collected by the community screening team (including physicians, nurses, and consultants) of HESH. The dataset grows as more screenings are completed. During the screening, the method of image capturing is adapted to the environment of the screening in order to maintain the validity and the completeness of the images captured.

For instance, when the images were captured in a bright environment, the individuals being captured would put on eye cover box to minimize the exposure to the ambient light. The reflection from the light was also minimized according to the brightness of the environment. All of the board-certified ophthalmologists with more than 5 years of clinical experience graded the images and checked whether the image showed signs of cataracts or not.

When there were disagreements among the ophthalmologists, they would vote to make a final judgement on the cataract severity, including pronounced cataract, early cataract, and non-cataract. Fig.3 represents exemplary images captured by smartphone-based slit-lamps. The image dataset was randomly separated into the training group and the validation group using the ratio of 7:3 in order to develop the algorithm. The testing set, which is composed of 819 images, was captured by smartphone-based slit-lamps.

B. EVALUATION METRICS

The dimensions used to evaluate the performance of the proposed algorithm included: image entropy, Brenner gradient function, gray variance function SMD, gray variance product function SMD2, mAP, Acc (accuracy), kappa coefficient and F1 value. The suggested quantitative metrics were compared between different methods proposed in other papers using the same dataset.

1) IMAGE ENTROPY

$$E = -\sum_{i=0}^{n} P_i log_2 P_i \tag{10}$$

$P_{(i)}$ is the probability of a certain pixel value i appearing in the image, and n is the gray value range(generally 0-255). The larger the information entropy E value of the image,

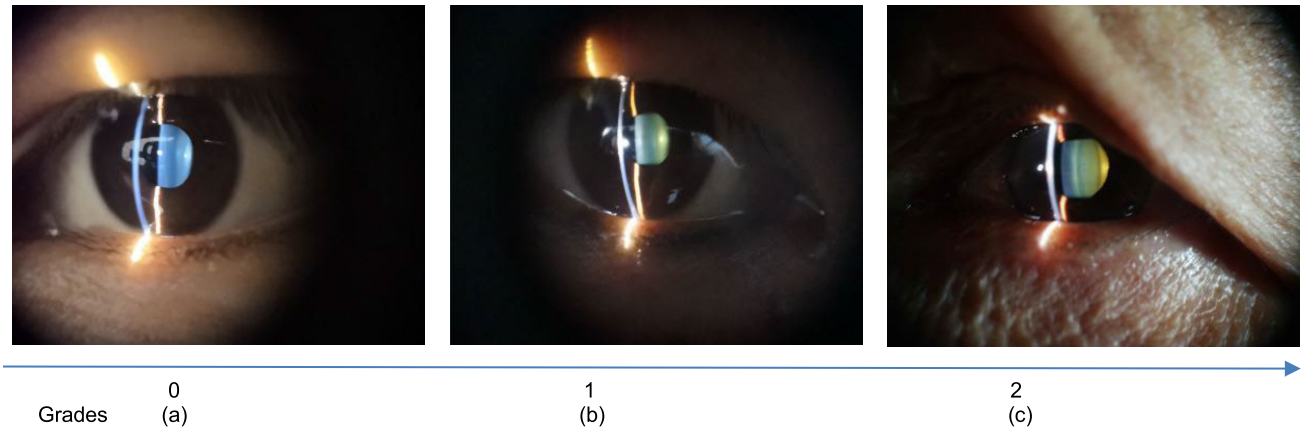


FIGURE 3. The reference standard comes from the board-certified ophthalmologists who grade the images into three categories: (a) normal with a grade of 0. (b) early cataract with a grade of 1. (c) pronounced cataract with a grade of 2.

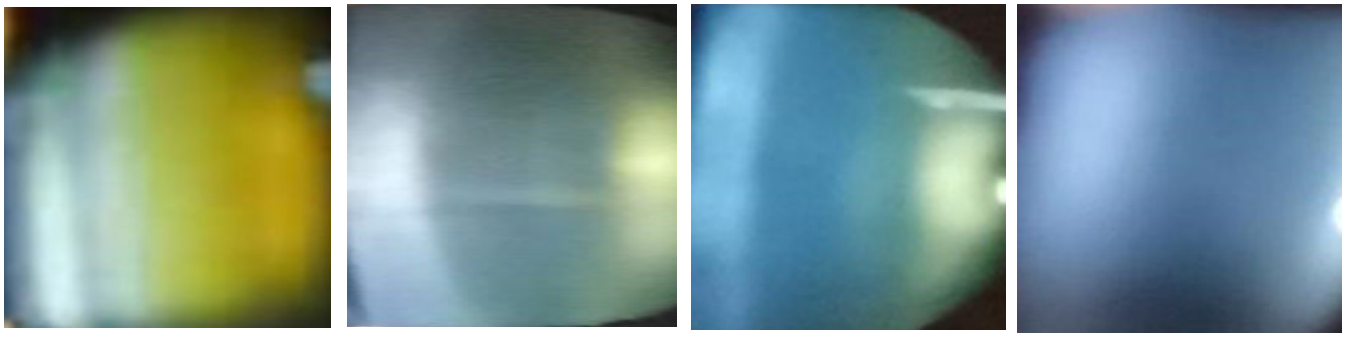


FIGURE 4. The ocular images were randomly selected from the MSLPP dataset. The YOLO v3 algorithm identifies the nuclear section of ocular images and crops the target sections.

the larger the size of the gray area deviating from the peak of the image histogram in the image. The probability of all gray values appearing tends to be equal. And the larger the amount of information carried by the image, the richer the information.

$$D(f) = \sum_y \sum_x |f(x+2, y) - f(x, y)|^2 \tag{11}$$

Among them, $f(x, y)$ represents the gray value of the corresponding pixel points of the image f and $D(f)$ is the result of image definition.

2) GRAY VARIANCE METHOD FUNCTION

$$D(f) = \sum_y \sum_x |f(x, y) - f(x+1, y)| + |f(x, y) - f(x, y+1)| \tag{12}$$

3) GRAY VARIANCE PRODUCT FUNCTION

The equation of the gray variance product function is shown below:

$$D(f) = \sum_y \sum_x |f(x, y) - f(x+1, y)| * |f(x, y) - f(x, y+1)| \tag{13}$$

4) CALCULATION EQUATION OF ACCURACY AND F1 VALUE

The calculation equation of accuracy and F1 value is shown below:

$$Acc = \frac{(TP + TN)}{(TP + FN + TN + FP)} \tag{14}$$

$$Specificity = TNR = \frac{TN}{FP + TN} \tag{15}$$

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

5) CALCULATION EQUATION OF ACCURACY AND F1 VALUE

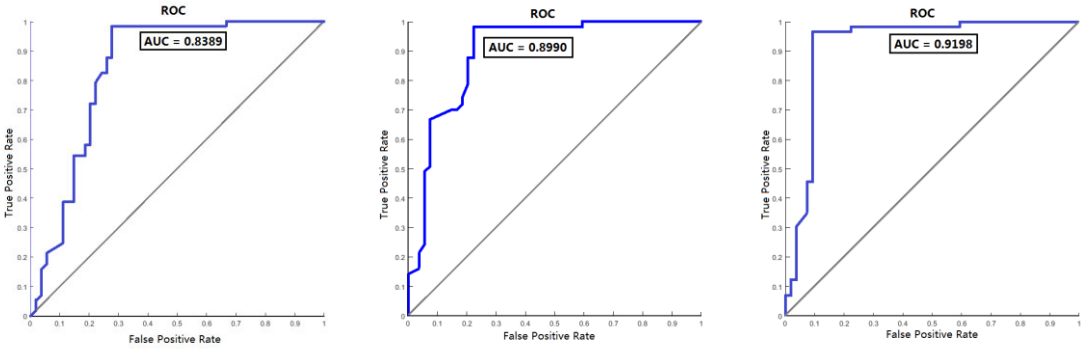
The calculation equation of accuracy and F1 value is shown below:

$$Sensitivity = TPR = Recall = \frac{TP}{TP + FN} \tag{17}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{18}$$

$$Youden\ index = Sensitivity + Specificity - 1 = TPR - FPR \tag{19}$$

TP, FP, TN and FN stand for the number of true positives, false positives, true negatives and false negatives in the detection results respectively.



(a) (b) (c)

FIGURE 5. ROC curves and AUC values of the three models. Figure A shows the ROC curve and AUC value of GoogleNet. Figure b is the ROC curve and AUC value of Resnet101. Figure C is the ROC curve and AUC value of UDFA.

TABLE 1. Diagnosis framework for automated nuclear cataract grading based on slit-lamp images. Use the tenengrad to determine the blur of the eye lens image. Use the YOLOV3 to detect and cut crystal images of high-definition eyes. The cut image data is trained by ShuffleNet and SVM method based on GLCM. Finally, the three models are fused and output.

Inputs	
I1.	Slit-lamp images for Yolov3 training: $\{x \in X\}$
I2.	Use the trained model to perform classification. Import different images from the same eye captured by slit-lamps to image blurriness evaluation algorithm to obtain the maximum value of $max(T_i), i \in I.$
I3.	Use Yolov3 to evaluate images, and follow formula (3) to (8) to crop the images. The cropped images are $\{c_1 \in C\}$.
I4.	Number of Iterations: 126
I5.	Learning rate: 0.0001
Training	
T1.	Use Yolov3 to train the algorithm with more than 10,000 images. When loss if converged, export model_v and carry out step in I.3 to obtain the cropped dataset C.
T2.	Feed dataset C to ShuffleNet algorithm to train cataract classification algorithm. The trained model is named as model_sn.
T3.	Extract features of images using the dataset C following the precess discussed in section B of P.3 to generate feature set of F. Train an SVM using the dataset F with labels.
T4.	Using the fusion process mentioned in P.4, combine the models discussed in T1-T3, and obtain the combined model as model_final.
Validation	
V1.	Feed different images from the same eye to the image blurriness evaluation algorithm to assess the image clarity, and the image with the highest value of $max(T_i), i \in I$ would be exported with the mark of Img.
V2.	Use model_v mentioned in T1 to classify Img. Details explained in P.2.
Outputs	Take the results of V2 to combine with the results obtained in T4 to fuse the model to obtain the final cataract classification output.

According to the description in chapter 3B, we use the box obtained by yolo3 algorithm to intercept the cataract nuclear region of the original size image input into the network according to equation (3) ∼ (8) (the original image size is: 2000 * 3000) as shown in Fig.5. For TABLE 3, the Entropy, SMD2, SMD and Brenner values calculated from the original image (resize 256*256) are higher than those from the box image (resize 256*256) obtained directly by the object detection algorithm. In the table TABLE 4, we further compared the images cut by the method proposed in this paper with the images cut by the method proposed in [21] without scaling, and found that the Entropy, SMD2, SMD and Brenner values obtained by the method proposed in this paper are still higher than those obtained by the method proposed in [21].

TABLE 2. The structure of the image dataset (MSLPP) obtained by the eye care screening team in he eye specialist hospital (HESH) using smartphone-based slit-lamps.

	TRAINING SET	Validation set	Testing set
Pronounced cataract	3316	1422	281
Early cataract	3742	1604	267
Non-cataract	4213	1806	271
Total	11271	4832	819

TABLE 3. The entropy, SMD2, SMD, Brenner values calculated from the uniformly contracted images (256 * 256) of ocular images using the YOLOv3 and FRCNN algorithms after cropping the images.

	ENTROPY	SMD2	SMD	Brenner
YOLOv3 cuts and outputs based on the original image	7.64	2.95	7.27	2.66
Faster-RCNN cuts and outputs	7.55	1.98	6.39	2.09

TABLE 4. The entropy, SMD2, SMD, Brenner values calculated from the images (without contraction) of ocular images using the YOLOv3 and FRCNN algorithms after cropping the images.

	ENTROPY	SMD2	SMD	Brenner
YOLOv3 cuts and outputs based on the original image	7.71	2.73	5.13	1.89
Faster-RCNN cuts and outputs	7.55	1.80	0.83	0.71

This shows that the image screenshot method in this paper can retain more image information and definition and is more suitable for cutting cataract nuclear regions. The results of TABLE 3 and TABLE 4 were obtained by randomly selecting 50 nuclear regions shear maps and calculating the mean value.

In order to compare the effect of nuclear region detection in this paper with that of faster-rcnn used in [21], we use the evaluation indexes in Table TABLE 5 to explain. By comparison, we found that the detection method used in this paper is nearly 3 times faster than faster-rcnn in single graph detection. The results above Mean Average Precision (mAP) are all calculated in the verification set in TABLE 2. In the final step of nuclear cataract classification recognition, we used the test set of TABLE 2 for statistics. The test set data of TABLE 2 were the clinical data from HESH. UDFA uses YOLOv3 to locate the nuclear region of the eye lens image, and intercepts the nuclear region of the original image to obtain a nuclear region dataset, and uses ShuffelNet and SVM classifiers for grading

training. Fuse the grading results probabilities x_1-x_3, then get the final classification model. At the same time, Resnet-101 and GoogleNet are used to perform classification training on the nuclear region datasets obtained by UDFA. Finally, the classification effects obtained by the three methods are compared. The comparision results are shown in TABLE 6. In order to assess how well the model predicts the outcome,

TABLE 5. YOLOv3 and faster-RCNN use the validation set in table 2 to experiment, and calculate the map value. At the same time, detect and calculate the time in a single picture, and judge the detection speed by the obtained map and time.

	MAP	Time
YoLov3	**52.36%**	29ms
Faster-rcnn	**50.13%**	200ms

we also use the idea of receiver operator characteristics (ROC) curves for the test datasets including the area under the curve (AUC) as our criteria. Figure 5. shows the ROC curves and the AUC value of three algorithm in GoogleNet, Resnet101 and the UDFA (Unified Diagnosis Framework for Automated Nuclear Cataract Grading) methods. We can see that the AUC's values from GoogleNet, Resnet101 and the UDFA methods are 0.8389, 0.8990 and 0.9198 respectively. Note that a larger AUC is better predictability, measured by sensitivity and specificity of test datasets. It means that for the MSLPP datasets, the UDFA method performs better than the other two. methods. TABLE 6 represents the results and shows that our method achieved an accuracy of 93.48% when it was detecting cataract, higher than 87.6% of ResNet101 and 83.53% of GoogleNet. In terms of Kappa

TABLE 6. The accuracy, sensitivity, specificity, Youden, F1, and kappa values among the algorithm proposed in this study, GoogleNet, and ResNet when they were trained from MSLPP dataset and classified testing set images.

Method	ACC	SENSITIVITY	Specificity	Youden	F1	Kappa
UDFA	**93.48%**	**89.2%**	97.37%	**0.846**	**92.3%**	**0.954**
ResNet-101	87.6%	80.39%	97.1%	0.775	88.2%	0.905
GoogleNet	83.53%	75%	**99.5%**	0.75	85.71%	0.892

value in classifying cataract severities (pronounced cataract, early cataract, and non-cataract), the Kappa value of our proposed method differs from GoogleNet's Kappa coefficient by 0.062, slightly higher than ResNet101's Kappa coefficient. In terms of F1 value and Youden index, our results are much higher than the other two methods. Through the above experiments, it is shown that the method proposed by us not only has a good classification effect, but also has a good performance in consistency compared with other methods.

V. CONCLUSION

This article first introduces the background and significance of cataract diagnosis. The automatic diagnosis of cataract can improve the accessibility of cataract examination and provide an important reference for underdeveloped areas with scarce medical resources to prevent blindness caused by cataract.

This study proposed a unified framework to perform automated nuclear cataract severity classification using smartphone-based slit-lamp photos. Motivated by Grafting, our framework combines deep learning and traditional feature extraction methods. Compared with the popular deep learning target detection method Faster-RCNN, this method can achieve real-time detection, and the Mean Average Precision (mAP) is higher. In order to make this method applicable to medical circumstances, this paper proposes to evaluate the image blur degree before grading, so as to ensure the quality of the picture uploaded to UDFA. Experimental results show that the framework is competitive among many existing cataract grading methods.

This method simplifies the complicated operation in cataract screening process, reduces the difficulty of screening by artificial intelligence technology, improves the accuracy rate, reduces the misdiagnosis rate and greatly improves the accessibility of medical treatment. The methods involved in this study have been widely used in cataract screening in Liaoning Province, China, and 94246 people have been screened so far. This method is actively involved in China's medical anti-poverty work, currently involved in Pingtang County and Luodian County, Guizhou Province. Internationally, this method is also being used in cataract screening in Nigeria, Africa, which has so far screened 1079 people.

ACKNOWLEDGMENT

The authors would like to thank Xingru He and Chunhong Han of the Social Responsibility Department of He Eye Specislists Hosipital for data acquisition, Zhuoshi Wang of He Eye Specislists Hospital for suggestions during the experiment, Mary Adu of He University for advice and language corrections, Xin Zhou and Xinyuan Zhang of He University for organizing the acquisition of screening data, and Chen Li of College of Medicine and Biological Information Engineering of Northeastern University for structural design given in full text.

REFERENCES

[1] R. R. A. Bourne, S. R. Flaxman, T. Braithwaite, M. V. A. C. Das, J. B. Jonas, and Y. Zheng, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis," *Lancet Global Health*, vol. 5, no. 9, pp. e888–e897, 2017.

[2] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010," *Brit. J. Ophthalmology*, vol. 96, no. 5, pp. 614–618, May 2012.

[3] B. H. Feldman and S. Heersink. (2017). *Cataract. Retrieved March 29, 2018, From American Academy of Ophthalmology Website*. [Online]. Available: http://eyewiki.aao.org/Cataract#Ophthalmic_Examination

[4] P. Desai, A. Reidy, D. C. Minassian, G. Vafidis, and J. Bolger, "Gains from cataract surgery: Visual function and quality of life," *Brit. J. Ophthalmology*, vol. 80, no. 10, pp. 868–873, Oct. 1996.

[5] G. C. Brown, M. M. Brown, A. Menezes, B. G. Busbee, H. B. Lieske, and P. A. Lieske, "Cataract surgery cost utility revisited in 2012," *Ophthalmology*, vol. 120, no. 12, pp. 2367–2376, Dec. 2013.

[6] P. J. Foster, T. Y. Wong, D. Machin, G. J. Johnson, and S. K. L. Seah, "Risk factors for nuclear, cortical and posterior subcapsular cataracts in the Chinese population of Singapore: The Tanjong Pagar survey," *Brit. J. Ophthalmology*, vol. 87, no. 9, pp. 112–1120, 2003.

[7] L. Chylack, J. Wolfe, D. Singer, M. C. Leske, M. A. Bullimore, I. L. Bailey, J. Friend, D. McCarthy, and S. Y. Wu, "The lens opacities classification system III," *Arch Ophthalmology*, vol. 111, no. 6, pp. 831–836, 1993.

[8] B. E. K. Klein, R. Klein, K. L. P. Linton, Y. L. Magli, and M. W. Neider, "Assessment of cataracts from photographs in the beaver dam eye study," *Ophthalmology*, vol. 97, no. 11, pp. 1428–1433, Nov. 1990.

[9] B. Thylefors, L. T. Chylack Jr., K. Konyama, K. Sasaki, R. Sperduto, H. R. Taylor, and S. West4, "A simplified cataract grading system the WHO cataract grading group," *Ophthalmic Epidemiology*, vol. 9, no. 2, pp. 83–95, Apr. 2002.

[10] Z. Zhang, R. Srivastava, H. Liu, X. Chen, L. Duan, D. W. K. Wong, C. K. Kwoh, T. Y. Wong, and J. Liu, "A survey on computer aided diagnosis for ocular diseases," *BMC Med. Informat. Decis. Making*, vol. 14, no. 1, pp. 1–29, Dec. 2014.

[11] W. Huang, H. Li, K. L. Chan, J. H. Lim, J. Liu, and T. Y. Wong, "A computer-aided diagnosis system of nuclear cataract via ranking," in *Proc. 12th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Berlin, Germany: Springer, 2009, pp. 803–810.

[12] W. Huang, K. L. Chan, H. Li, J. H. Lim, J. Liu, and T. Y. Wong, "A computer assisted method for nuclear cataract grading from slit-lamp images using ranking," *IEEE Trans. Med. Imag.*, vol. 30, no. 1, pp. 94–107, Jan. 2011.

[13] H. Li, J. H. Lim, J. Liu, D. Wing, K. Wong, and T. Y. Wong, "Feature analysis in slit-lamp image for nuclear cataract diagnosis," in *Proc. 3rd Int. Conf. Biomed. Eng. Informat.*, Oct. 2010, pp. 253–256.

[14] S. Fan, C. R. Dyer, L. Hubbard, and B. Klein, "An automatic system for classification of nuclear sclerosis from slit-lamp photographs," in *Medical Image Computing and Computer-Assisted Intervention*. Berlin, Germany: Springer, 2003.

[15] H. Li, J. H. Lim, J. Liu, T. Y. Wong, A. Tan, J. J. Wang, and P. Mitchell, "Image based grading of nuclear cataract by SVM regression," *Proc. SPIE*, vol. 6915, Mar. 2008, Art. no. 691536.
[16] H. Li, J. H. Lim, J. Liu, D. W. K. Wong, N. M. Tan, S. Lu, Z. Zhang, and T. Y. Wong, "Computerized systems for cataract grading," in *Proc. 2nd Int. Conf. Biomed. Eng. Informat.*, Oct. 2009, pp. 1–4.
[17] H. Li, J. Hwee Lim, J. Liu, P. Mitchell, A. Grace Tan, J. Jin Wang, and T. Yin Wong, "A computer-aided diagnosis system of nuclear cataract," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1690–1698, Jul. 2010.
[18] J. Nayak, "Automated classification of normal, cataract and post cataract optical eye images using SVM classifier," in *Proc. World Congr. Eng. Comput. Sci.*, vol. 1, 2013, pp. 23–25.
[19] X. Gao, D. W. K. Wong, T.-T. Ng, C. Y. L. Cheung, C.-Y. Cheng, and T. Y. Wong, "Automatic grading of cortical and PSC cataracts using retroillumination lens images," in *Proc. Asian Conf. Comput. Vis.*, Berlin, Germany: Springer, 2012, pp. 256–267.
[20] R. Srivastava, X. Gao, F. Yin, D. W. K. Wong, J. Liu, C. Y. Cheung, and T. Y. Wong, "Automatic nuclear cataract grading using image gradients," *J. Med. Imag.*, vol. 1, no. 1, Jun. 2014, Art. no. 014502.
[21] X. Gao, S. Lin, and T. Y. Wong, "Automatic feature learning to grade nuclear cataracts based on deep learning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2693–2701, Nov. 2015.
[22] J.-J. Yang, J. Li, R. Shen, Y. Zeng, J. He, J. Bi, Y. Li, Q. Zhang, L. Peng, and Q. Wang, "Exploiting ensemble learning for automatic cataract detection and grading," *Comput. Methods Programs Biomed.*, vol. 124, pp. 45–57, Feb. 2016.
[23] Y. Xu, X. Gao, S. Lin, D. W. K. Wong, J. Liu, D. Xu, C. Y. Cheng, C. Y. Cheung, and T. Y. Wong, "Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Berlin, Germany: Springer, 2013, pp. 468–475.
[24] X. Gao, H. Li, J. H. Lim, and T. Y. Wong, "Computer-aided cataract detection using enhanced texture features on retro-illumination lens images," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 1565–1568.
[25] L. Guo, J.-J. Yang, L. Peng, J. Li, and Q. Liang, "A computer-aided healthcare system for cataract classification and grading based on fundus image analysis," *Comput. Ind.*, vol. 69, pp. 72–80, May 2015.
[26] H. Li, L. Ko, J. Hwee Lim, J. Liu, D. Wing Kee Wong, and T. Yin Wong, "Image based diagnosis of cortical cataract," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 3904–3907.
[27] H. Li, L. Ko, J. H. Lim, J. Liu, D. W. K. Wong, T. Y. Wong, and Y. Sun, "Automatic opacity detection in retro-illumination images for cortical cataract diagnosis," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2008, pp. 553–556.
[28] X. Liu, J. Jiang, K. Zhang, E. Long, J. Cui, M. Zhu, Y. An, J. Zhang, Z. Liu, Z. Lin, X. Li, J. Chen, Q. Cao, J. Li, X. Wu, D. Wang, and H. Lin, "Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0168606.
[29] L. Zhang, J. Li, i. Zhang, H. Han, B. Liu, J. Yang, and Q. Wang, "Automatic cataract detection and grading using deep convolutional neural network," in *Proc. IEEE 14th Int. Conf. Netw., Sens. Control (ICNSC)*, May 2017, pp. 60–65.
[30] C. Xu, X. Zhu, W. He, Y. Lu, X. He, Z. Shang, J. Wu, K. Zhang, Y. Zhang, X. Rong, and Z. Zhao, "Fully deep learning for slit-lamp photo based nuclear cataract grading," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Cham, Switzerland: Springer, 2019, pp. 513–521.
[31] J. M. Tenenbaum, "Accommodation in computer vision," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, 1970.
[32] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018. [Online]. Available: https://arxiv.org/abs/1804.02767
[33] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 6848–6856.
[34] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using GLCM approach," *Int. J. Sci. Res. Pulications*, vol. 3, no. 5, p. 1, May 2013.
[35] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features of image classification," *IEEE Trans. Syst., Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[36] KazAnova, "Stacking made easy: An introduction to StackNet by competitons grand-660 master Marios Michailidis," *AnalyticsWeek*, 2018.

SHENMING HU received the M.S. degree from the Faculty of Computer Science, Tsukukba University, Japan, in 2012. He is currently pursuing the Ph.D. degree with the College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China.

XIAOTING WANG is currently pursuing the M.S. degree with the School of Statistics, Renmin University of China. She is also a Lecturer with the Artificial Intelligence and Big Data College, she University, Shenyang, China.

HONG WU is currently an Algorithm Engineer with Shenyang Eyerobo Company Ltd., Shenyang, China.

XINZE LUAN is currently a Lecturer with the Artificial Intelligence and Big Data College, He University, Shenyang, China.

PENG QI is currently with the Department of Ophthalmic Optometry, He University. He is also with the Social Responsibility Department, He Eye Specialist Hospital, Shenyang, China.

YI LIN is currently with the Department of Ophthalmic Optometry, He University, Shenyang, China.

XIANGDONG HE received the Ph.D. degree from the Medical School, Fukuoka University, Fukuoka, Japan, in 1993. He is currently the Director of He's Eye Corporation, Shenyang, China.

WEI HE received the Ph.D. degree from the Medical School, Fukuoka University, Fukuoka, Japan, in 1995. He is currently the Chairman and the President of He's Eye Corporation and ICO Board of Trustees Member, China.