

Automatic Classification of Heterogeneous Slit-Illumination Images Using an Ensemble of Cost-Sensitive Convolutional Neural Networks

Jiewei Jiang

School of Electronics Engineering, Xi'an University of Posts and Telecommunications, Xi'an, 710121, China

Liming Wang

School of Computer Science and Technology, Xidian University, Xi'an, 710071, China

Haoran Fu

School of Computer Science and Technology, Xidian University, Xi'an, 710071, China;

Erping Long

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China

Yibin Sun

School of Electronics Engineering, Xi'an University of Posts and Telecommunications, Xi'an, 710121, China

Ruiyang Li

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China

Zhongwen Li

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China

Mingmin Zhu

School of Mathematics and Statistics, Xidian University, Xi'an, 710071, China

Zhenzhen Liu

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China

Jingjing Chen

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China

Zhuoling Lin

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China

Xiaohang Wu

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China

Dongni Wang

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China

Xiyang Liu

School of Computer Science and Technology, Xidian University, Xi'an, 710071, China

Haotian Lin (✉ haot.lin@hotmail.com)

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University

Research

Keywords: cost-sensitive, deep convolutional neural networks, ensemble learning, heterogeneous slit-illumination images, pediatric cataract

Posted Date: September 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-72323/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1 **Automatic classification of heterogeneous slit-illumination**
2 **images using an ensemble of cost-sensitive convolutional**
3 **neural networks**

4 Jiewei Jiang^{1#}, Liming Wang^{2#}, Haoran Fu^{2#}, Erping Long³, Yibin Sun¹, Ruiyang Li³, Zhongwen
5 Li³, Mingmin Zhu⁴, Zhenzhen Liu³, Jingjing Chen³, Zhuoling Lin³, Xiaohang Wu³, Dongni
6 Wang³, Xiyang Liu^{2*}, Haotian Lin^{3*}

7 ¹School of Electronics Engineering, Xi'an University of Posts and Telecommunications, Xi'an,
8 710121, China;

9 ²School of Computer Science and Technology, Xidian University, Xi'an, 710071, China;

10 ³State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen
11 University, Guangzhou, 510060, China;

12 ⁴School of Mathematics and Statistics, Xidian University, Xi'an, 710071, China;

13 [#]These authors contributed equally to this work.

14

15 ***Co-corresponding authors:**

16 Email: haot.lin@hotmail.com (H.T.L) and xyliu@xidian.edu.cn (X.Y.L)

17

18

19

20

21 **Abstract**

22 **Background:** Lens opacity seriously affects the visual development of infants. Slit-illumination
23 images play an irreplaceable role in lens opacity detection; however, these images exhibited
24 varied phenotypes with severe heterogeneity and complexity, particularly among pediatric
25 cataracts. Therefore, it is urgently needed to explore an effective computer-aided method to
26 automatically diagnose heterogeneous lens opacity and to provide appropriate treatment
27 recommendations in a timely manner.

28 **Methods:** We integrated three different deep learning networks and a cost-sensitive method into
29 an ensemble learning architecture, and then proposed an effective model called ECCNN-
30 Ensemble (ensemble of cost-sensitive convolutional neural networks) for automatic lens
31 opacity detection. A total of 470 slit-illumination images of pediatric cataracts were used for
32 training and comparison between the CCNN-Ensemble model and conventional methods.
33 Finally, we used two external datasets (132 independent test images and 79 Internet-based
34 images) to further evaluate the model's generalizability and effectiveness.

35 **Results:** Experimental results and comparative analyses demonstrated that the proposed method
36 was superior to conventional approaches and provided clinically meaningful performance in
37 terms of three grading indices of lens opacity: area (specificity and sensitivity; 92.00% and
38 92.31%), density (93.85% and 91.43%) and opacity location (95.25% and 89.29%).
39 Furthermore, the comparable performance on the independent testing dataset and the internet-
40 based images verified the effectiveness and generalizability of the model. Finally, we developed
41 and implemented a website-based automatic diagnosis software for pediatric cataract grading
42 diagnosis in ophthalmology clinics.

43 **Conclusions:** The CCNN-Ensemble method demonstrates higher specificity and sensitivity
44 than conventional methods on multi-source datasets. This study therefore provides a practical
45 strategy for heterogeneous lens opacity diagnosis and has the potential to be applied to the
46 analysis of other medical images.

47 **Key words:** cost-sensitive; deep convolutional neural networks; ensemble learning;
48 heterogeneous slit-illumination images; pediatric cataract.

49 **Background**

50 Optical imaging technologies play a vital role in clinical diagnosis and treatment of
51 ophthalmology [1, 2]. Computational vision approaches for automatic diagnosis of lens opacity
52 have greatly improved the efficiency of ophthalmologists and the entire treatment chain,
53 providing real benefits for patients [3-6]. In our previous studies, we applied artificial
54 intelligence methods to the classification of diffuse-light ocular images [7-9]. However,
55 diagnosis that is solely dependent on diffuse-light images will inevitably miss a substantial
56 proportion of potential ophthalmology patients [10-12]. The common slit-illumination image
57 offers another effective diagnosis medium and provides an essential supplement to these
58 diffuse-light images [13, 14]. Therefore, development of computer vision techniques for slit-
59 illumination images will move the automatic diagnosis of ophthalmic diseases towards a more
60 comprehensive and intelligent strategy.

61 At present, the existing computer-aided diagnosis methods generally focus on senile cataracts
62 using slit-illumination images [3-5, 15]. Thresholding localization and support vector
63 regression methods were used to grade the nuclear cataract [16]. Recursive convolutional neural
64 networks and support vector regression methods were implemented to enable automatic
65 learning of features for evaluating the severity of nuclear cataracts [17]. However, the
66 phenotypes of senile cataracts are relatively simple and fairly homogeneous. The study of such
67 senile cataracts alone will not be sufficient for the development of a computer-aided diagnosis
68 system for lens opacity in complex clinical scenarios. Practical clinical applications need the
69 ability to diagnose heterogeneous lens opacities with high recognition rates [18-20]. It is

70 therefore essential to develop an efficient, feasible and automatic diagnostic system to address
71 heterogeneous slit-illumination images.

72 The pediatric cataract is a typical lens opacity disease that suffers from severe heterogeneity
73 and complex phenotypes [21-23]. Large-scale slit-illumination images of pediatric cataracts
74 were collected from the long-term Childhood Cataract Program of the Chinese Ministry of
75 Health (CCPMOH) project [24], which covered a wide variety of lens opacities. In addition,
76 imbalance between the categories is an inevitable problem in pediatric cataract diagnosis [21,
77 25], where the number of positive samples is relatively smaller than the number of negative
78 samples. This can easily cause the classifiers to produce a higher false negative rate. Therefore,
79 these datasets represent an ideal medium for exploration of the appropriate computational
80 vision methods required to adapt to complex clinical application scenarios.

81 To develop an effective and efficient computer vision method for analysis of these
82 heterogeneous slit-illumination images, we integrated three deep convolutional neural networks
83 (AlexNet, GoogLeNet and ResNet) [26-28] and a cost-sensitive algorithm [29, 30] into an
84 ensemble learning framework and created the CCNN-Ensemble model (ensemble of cost-
85 sensitive convolutional neural networks). The three convolutional neural networks (CNNs)
86 with their different structures were used to improve both overall recognition rate and stability
87 of the model. The cost-sensitive algorithm was used to address the imbalanced dataset problem
88 and thus significantly reduce the model's false negative rate. We performed detailed
89 experiments to compare performance of the CCNN-Ensemble method with that of conventional
90 methods in three grading indices of lens opacity. We also used two external datasets (an

91 independent testing dataset and an Internet-based dataset) to validate the method's versatility
92 and stability. Finally, potential computer-aided diagnostic software was developed and
93 deployed for use by ophthalmologists and their patients in clinical applications.

94 **Methods**

95 **Dataset**

96 The slit-illumination datasets consist of the following three parts: the training and validation
97 dataset, the independent testing dataset, and the Internet-based dataset. A total of 470 training
98 and validation datasets were derived from the routine examinations of the Zhongshan
99 Ophthalmic Center in Sun Yat-sen University (Fig. 1a) [24]. 132 independent testing images
100 were selected randomly in advance from the Zhongshan Ophthalmic Center; 79 Internet-based
101 images were collected using a keyword search (including words such as congenital cataract,
102 infant and pediatric) of the Baidu and Google search engines.

103 There are no special pixel requirements for the enrolled images provided that the lens area of
104 the image is retained. To ensure grade labeling accuracy, three senior ophthalmologists jointly
105 determine the grade of each image and comprehensively evaluate its severities in terms of three
106 lens lesion indices (opacity area, density and location) [7, 9]. An opacity area that covers more
107 than half of the pupil is defined as extensive; otherwise, it is defined as limited. An opacity
108 density that completely blocks the light is labelled as dense; otherwise, it is defined as
109 transparent. An opacity location that fully covers the visual axis of the pupil is called central;
110 otherwise, it is called peripheral. The collected datasets covered a variety of pediatric cataracts,
111 which were divided into limited and extensive categories for area, dense and transparent

112 categories for density, and central and peripheral categories for location, as shown in Table 1.

113 **Table 1. Distributions of slit-illumination datasets in terms of three grading indices.**

Datasets	Total number	Opacity area		Opacity density		Opacity location	
		limited	extensive	transparent	dense	peripheral	central
Training and validation dataset	470	275	195	260	210	274	196
Independent testing dataset	132	91	41	104	28	100	32
Internet-based dataset	79	19	60	18	61	16	63

114 **Preprocessing and model evaluation**

115 We preprocessed all labeled datasets using twice-applied Canny detection and Hough
116 transformation [31, 32] to acquire the lens region of interest and eliminate surrounding noise
117 zones such as the eyelids and the sclera (Fig. 1a). The localized images were subsequently
118 resized to a size of 256×256 pixels and were then input into the computational vision models.
119 Using these training and validation datasets, we performed a five-fold cross-validation
120 procedure to compare and evaluate the performances of the different models (Fig. 1b). Four
121 representative handcrafted features (WT: wavelet transformation; LBP: local binary pattern;
122 SIFT: scale-invariant feature transform; and COTE: color and texture features) [8, 9, 33-35]
123 were selected and combined with support vector machine (SVM) and adaptive boosting
124 (Adaboost) classifiers for performance comparison. After selection of the optimal CCNN-
125 Ensemble model, we further verified its effectiveness and stability using the two external
126 datasets (the independent testing dataset and the Internet-based dataset).

127 **Evaluation metrics**

128 To provide a full assessment of the superiority of the CCNN-Ensemble method when compared

129 with the conventional methods, we calculated several evaluation metrics, including accuracy,
130 sensitivity, specificity, F1-measure, and G-mean, as follows.

$$131 \quad Accuracy = (TP + TN) / (TP + FN + TN + FP) \quad (1)$$

$$132 \quad Sensitivity(Recall) = TP / (TP + FN) \quad (2)$$

$$133 \quad Specificity = TN / (TN + FP) \quad (3)$$

$$134 \quad Precision = TP / (TP + FP) \quad (4)$$

$$135 \quad F1 - measure = \frac{2 * Recall * Precision}{Recall + Precision} \quad (5)$$

$$136 \quad G - mean = \sqrt{\frac{TP}{TP + FN} * \frac{TN}{TN + FP}} \quad (6)$$

137 where TP , FP , TN and FN denote the numbers of true positives, false positives, true negatives
138 and false negatives, respectively. Accuracy, sensitivity and specificity are the most commonly
139 used evaluation measures. The F1-measure and G-mean [36] indicators simultaneously
140 consider the accuracies of both classes and can thus effectively measure the recognition abilities
141 of models in the case of an imbalanced dataset. Additionally, three more vital objective
142 measures – the receiver operating characteristic curve (ROC), the area under the ROC curve
143 (AUC), and the precision recall curve (PR) – were used for visual comparison and analysis.

144 **Overall framework of CCNN-Ensemble**

145 As shown in Fig. 2, the overall diagnosis framework of the CCNN-Ensemble consists primarily
146 of three deep CNN models (GoogLeNet, AlexNet and ResNet), a cost-sensitive adjustment

147 layer, ensemble learning, dataset augmentation technology and transfer learning. The three
148 heterogeneous CNN models, as classifiers, were employed to construct the ensemble learning
149 framework to enhance the recognition rates of the algorithms. The cost-sensitive adjustment
150 layer was used to manage the imbalanced dataset problem, and the dataset augmentation and
151 transfer learning processes were adopted to overcome the overfitting problem and accelerate
152 model convergence. The technical details are described below.

153 **Ensemble learning of multiple CNNs**

154 We used three heterogeneous CNNs (AlexNet, GoogLeNet and ResNet) to form the ensemble
155 learning framework (Fig. 2). The AlexNet CNN, which was proposed by Krizhevsky [26],
156 performed image classification and won first prize in the ImageNet Large Scale Visual
157 Recognition Challenge (ILSVRC) in 2012, mainly used convolutional layers, overlapping
158 pooling, nonsaturating rectified linear units (ReLU) and three fully-connected layers to
159 construct an eight-layer CNN. Subsequently, a number of variants of CNN method were
160 proposed to enhance its recognition rate and incorporated many emerging technologies. In
161 particular, a 22-layer inception deep network was achieved by Google researchers [27] that was
162 based on the Hebbian principle, intuition of multi-scale processing, filter aggregation, average
163 pooling and auxiliary classifier technologies. Kaiming He then used the residual connection
164 scheme, batch normalization and scale operations to establish a 50-layer ultra-deep residual
165 CNN (ResNet) [28]. Because the above CNNs implemented different principles and techniques,
166 their network structures show distinct heterogeneity, and this can effectively improve the
167 recognition rate of the ensemble learning model.

168 In order to adequately utilize the advantages of the three convolutional neural networks, we
169 implemented a two-stage ensemble learning scheme. Specifically, in the first stage, starting
170 with the initial parameters of models pre-trained on the ImageNet dataset, three CNNs with
171 different structures were trained using transfer learning, respectively. Thus, the optimal
172 parameters of each CNN were obtained. In the second stage, the Softmax functions of the three
173 CNNs were removed, the high-level features of the CNNs were merged into the same cost-
174 sensitive Softmax classification function to construct a unified ensemble CNN. The learning
175 rate of the feature extraction layers was set to one-tenth of the ensemble learning layer. The
176 transfer learning method was adopted to fully train the ensemble learning layer and fine-tune
177 the previous feature extraction layers. Through the above two-stage ensemble learning scheme,
178 three different types of CNNs can complement their shortcomings, which is beneficial to
179 improve the overall performance of intelligent diagnosis for pediatric cataract.

180 **Transfer learning**

181 Because the number of medical images is very small, the fully-trained deep learning system
182 cannot adequately optimize the millions of trainable parameters from scratch and this can easily
183 lead to overfitting. Transfer learning [37, 38] is a critical technology for application to such
184 small datasets that allows the model to be trained from a better starting point and uses the color,
185 texture and shape characteristics that have been learned from natural images. Fine-tuning
186 allowed the final trained CNN model to obtain the unique features of the ophthalmic images
187 and also overcame the overfitting problem. Additionally, data augmentation methods, including
188 transformed images and horizontal reflections [26, 39], were adopted to accelerate the

189 convergence of the models.

190 Cost-sensitive method and optimization process

191 To address the imbalanced dataset problem of the slit-illumination images effectively, the cost-
 192 sensitive approach [29, 30, 40] was adopted to adjust the cost-sensitive weight of the positive
 193 samples in the loss function (Fig. 2). Specifically, we discriminatively determined the cost of
 194 misclassification of the different classes and assigned a larger cost-sensitive weight to the
 195 positive class. For one iterative training stage, n samples were selected at random to form a
 196 training dataset $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(n)}, y^{(n)})\}$, where $x^{(i)} \in R^l$ and $y^{(i)} \in \{1, \dots, k\}$.
 197 Here, $x^{(i)}$ denotes the features of the i -th sample and $y^{(i)}$ is the category label. The cost-
 198 sensitive loss function can be expressed as shown in Eq. 7.

$$199 \quad F(\theta) = -\frac{1}{n} \left[\sum_{i=1}^n \sum_{j=1}^k I\{y^{(i)} = j\} * CS\{y^{(i)} = \text{positive class}\} * \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{s=1}^k e^{\theta_s^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=1}^m \theta_{ij}^2 \quad (7)$$

200 where n , m , k and θ denote the number of training samples, the number of input neurons, the
 201 number of classes, and trainable parameters, respectively. $I\{y^{(i)} = j\}$ represents the indicator
 202 function ($I\{y^{(i)} \text{ is equal to } j\} = 1$ and $I\{y^{(i)} \text{ is not equal to } j\} = 0$) while
 203 $CS\{y^{(i)} = \text{positive class}\}$ is the cost-sensitive weight function
 204 ($CS\{y^{(i)} \text{ is the positive class lable}\} = C$ and $CS\{y^{(i)} \text{ is the negative class label}\} = 1$). Using
 205 a grid-search procedure, we determined that the value of the effective cost-sensitive weight
 206 parameter C was within the interval [4–6]. $\frac{\lambda}{2} \sum_{i=1}^k \sum_{j=1}^m \theta_{ij}^2$ is a weight decay term that is applied
 207 to penalize the larger trainable weights. To obtain the optimal trainable weights θ^* (see Eq. 8),
 208 we needed to minimize $F(\theta)$ using a mini-batch gradient descent (Mini-batch-GD) [41] as

209 shown in Eq. 9.

$$210 \quad \theta^* = \arg \min_{\theta} F(\theta) \quad (8)$$

$$211 \quad \nabla_{\theta_j} F(\theta) = -\frac{1}{n} \sum_{i=1}^n \left[PW \{y^{(i)} = \textit{positive class}\} * x^{(i)} * (I\{y^{(i)} = j\} - p(y^{(i)} = j | x^{(i)}; \theta)) \right] + \lambda \theta_j$$

212 (9)

213 **Experimental environment**

214 In this study, we implemented dataset preprocessing, automatic lens region of interest (ROI)
215 localization, conventional feature extraction, the SVM and Adaboost classifiers, and uniform
216 dataset partitioning for cross-validation using MATLAB R2014a [8, 9]. The CCNN-Ensemble
217 training, validation and testing procedures were all performed in parallel using eight Nvidia
218 Titan X graphics processing units (GPUs) based on the Caffe toolbox [42] in the Ubuntu 16.4
219 OS. The initial learning rate was set at 0.001 and successively reduced by one tenth of the
220 original value after every 500 iterations; a total of 2000 iterations were performed. We set the
221 mini-batch size to 32 on one GPU and used eight GPUs; we thus acquired a total of 256 samples
222 in every iteration and calculated the average value of these samples to update the trainable
223 parameters. Appropriate settings for these parameters can ensure better performance and rapid
224 convergence for the CCNN-Ensemble method.

225 **Results**

226 To achieve an effective solution to assist in the diagnosis of pediatric cataracts using slit-
227 illumination images, we explored three different methods, including four conventional features,
228 four Adaboost ensemble methods, and the CCNN-Ensemble method. First, we trained and

229 compared the performances of these methods on the training and validation datasets to obtain
230 the optimal CCNN-Ensemble method. Then, we used two external datasets to provide further
231 evaluation of the robustness and the clinical effectiveness of the CCNN-Ensemble. Finally, we
232 developed and deployed cloud-based software to serve patients that were located in remote
233 areas.

234 **Performance of CCNN-Ensemble and conventional methods**

235 After application of the five-fold cross-validation [43], we compared the performances of the
236 nine intelligent algorithms for diagnosis of the lens opacity in terms of the three grading indices
237 (opacity area, density and location). We calculated three main indicators – accuracy (ACC),
238 specificity (SPE) and sensitivity (SEN) (Fig. 3) – along with more detailed test results with
239 means and standard deviations (Table 2 and Supplementary Table S1–S2). First, when using
240 the conventional feature methods, both the ACC and SEN indicators are low; for example, the
241 SEN of the LBP method is less than 70% for all grading indices. Second, after application of
242 the Adaboost ensemble learning methods, the SEN indicator is greatly improved, whereas the
243 value of the SPE indicator is reduced. As a result, the ACC is almost equal to the performance
244 of the conventional feature methods (see Fig. 3). Notably, the SEN of the SIFT method
245 increased from 76.41% to 84.62%, whereas the SPE decreased from 76.73% to 65.45% for
246 opacity area grading (Fig. 3 and Table 2); the SEN of the LBP method increased from 68.88%
247 to 81.10%, whereas the SPE again decreased from 80.27% to 73.34% for opacity location
248 grading (Fig. 3 and Supplementary Table S2). The comparison results for the other feature
249 methods and the Adaboost ensemble learning methods are also similar. Third, the CCNN-

250 Ensemble method provided significantly improved recognition rates for all grading indices (Fig.
 251 3). All the average ACCs were maintained at 92% or more, while both the SPE and the SEN
 252 were satisfactory for the grading opacity area (92.00% and 92.31%), the opacity density (93.85%
 253 and 91.43%) and the opacity location (95.25% and 89.29%). Similarly, the F1-measure, G-
 254 mean and AUC indicators also showed values of more than 90% (Table 2 and Supplementary
 255 Table S1-S2).

256 **Table 2. Performance comparison of the different methods in opacity area grading.**

Method	ACC (%)	SPE(%)	SEN (%)	F1_M (%)	G_M (%)	AUC (%)
WT	80.21(3.33) §	87.27(5.45)	70.26(2.92)	74.73(3.50)	78.26(2.86)	87.47(2.87)
WT-Adaboost	81.28(2.77)	83.27(3.25)	78.46(5.90)	77.61(3.59)	80.76(3.13)	89.68(2.54)
LBP	75.11(4.09)	80.73(4.56)	67.18(5.85)	69.11(5.09)	73.59(4.26)	83.45(3.82)
LBP- Adaboost	76.17(4.36)	73.82(5.08)	79.49(5.13)	73.48(4.69)	76.56(4.36)	83.69(3.38)
SIFT	76.60(4.32)	76.73(8.76)	76.41(5.56)	73.12(3.56)	76.35(3.90)	85.66(4.05)
SIFT- Adaboost	73.40(3.98)	65.45(6.03)	84.62(4.80)	72.56(3.67)	74.33(3.94)	85.61(4.15)
COTX	79.79(7.52)	86.18(10.5)	70.77(7.82)	74.62(8.54)	77.93(7.02)	87.22(5.22)
COTX- Adaboost	84.68(4.02)	88.73(6.48)	78.97(7.78)	81.01(4.92)	83.53(4.34)	91.07(2.85)
CCNN-Ensemble	92.13(1.21)	92.00(2.07)	92.31(2.56)	90.68(1.42)	92.14(1.25)	97.76(0.81)

257 Notes: ACC: accuracy; SPE: specificity; SEN: sensitivity; F1_M: F1-measure; G_M: G-mean;
 258 AUC: area under the receiver operating characteristic curve; WT: wavelet transformation; LBP:
 259 local binary pattern; SIFT: scale-invariant feature transform; COTE: color and texture features;
 260 Adaboost: adaptive boosting ensemble learning; CCNN-Ensemble: ensemble learning of cost-
 261 sensitive convolutional neural networks; §Mean (Standard Deviation).

262 Additionally, we used the ROC and PR curves to compare the performances of the above
 263 methods (Fig. 4, Supplementary Fig. S1-S2). The ROC curve of the CCNN-Ensemble is close

264 to the upper-left area of the graph and the PR curve shows a similar performance. All the AUC
 265 indicators were maintained at more than 0.969 for the three grading indices. This indicates that
 266 the CCNN-Ensemble method is superior to the conventional features and Adaboost ensemble
 267 learning methods.

268 **Performance in independent testing dataset**

269 **Table 3. Quantitative evaluation of the CCNN-Ensemble method using two external**
 270 **datasets.**

External Datasets	Grading	ACC (%)	SPE (%)	SEN (%)	F1_M (%)	G_M (%)	AUC (%)
Independent testing dataset	opacity area	94.70	96.70	90.24	91.36	93.42	96.94
	opacity density	93.18	94.23	89.29	84.75	91.72	97.70
	opacity location	93.18	94.00	90.63	86.57	92.30	98.13
Internet-based dataset	opacity area	89.87	89.47	90.00	93.10	89.74	94.65
	opacity density	88.61	88.89	88.52	92.31	88.71	95.63
	opacity location	87.34	87.50	87.30	91.67	87.40	93.06

271 Notes: ACC: accuracy; SPE: specificity; SEN: sensitivity; F1_M: F1-measure; G_M: G-mean;
 272 AUC: area under the receiver operating characteristic curve.

273 To ensure an adequate investigation of the generalizability and the effectiveness of the CCNN-
 274 Ensemble method, we used an independent testing dataset for further validation of the proposed
 275 method. A total of 132 slit-illumination images were selected randomly in advance from the
 276 Zhongshan Ophthalmic Center (details are given in the Methods section). Using the expert
 277 group’s decisions for reference, we presented detailed quantitative evaluation results (as shown
 278 in Table 3) and performance comparisons of the ACC, SPE and SEN indicators (Fig. 5a). We

279 also reported the ROC and PR curves for the three grading indices: opacity area, density and
280 location (Fig. 5a). The experimental results indicated that the performance of the CCNN-
281 Ensemble method on the independent testing dataset is almost equal to that of the validation
282 dataset, with the ACC and the SPE being maintained at more than 93% and 94%, respectively,
283 and the SEN values are 90.24%, 89.29% and 90.63% for the opacity grading area, density and
284 location, respectively.

285 **Performance in Internet-based dataset**

286 In addition, we also collected 79 slit-illumination images from the Internet (details are given in
287 the Methods section). While the quality of these images varied significantly, the CCNN-
288 Ensemble was still able to detect the appropriate cases with a higher recognition rate. In the
289 same manner, we obtained detailed prediction results (given in Table 3), intuitive comparison
290 graphs for the main indicators (ACC, SPE and SEN), the ROC curve, and the PR curve (Fig.
291 5b). Specifically, the CCNN-Ensemble method also offered satisfactory accuracy, specificity
292 and sensitivity in terms of opacity area (89.87%, 89.47%, and 90.00%), opacity density
293 (88.61%, 88.89%, and 88.52%) and opacity location (87.34%, 87.50%, and 87.30%),
294 respectively.

295 **Web-based software**

296 To serve both patients and ophthalmologists located in remote areas, we developed and
297 deployed an automatic diagnosis software based on cloud service ([http://www.cc-
298 cruiser.com:5007/SignIn](http://www.cc-cruiser.com:5007/SignIn)), which included user registration, an image upload module, a
299 prediction module, regular re-examinations, sample downloads, and instructions. Before using

300 the website for diagnosis, the users needed to submit personal information including age,
301 gender and telephone number to complete the registration process. This registration process
302 allowed the doctor to contact patients who were diagnosed with serious conditions, and also
303 prevented the illegal use of our software. After registration, either the patient or the
304 ophthalmologist can upload the slit-illumination images for diagnosis; the software can then
305 perform image preprocessing, make three grading predictions and provide a final treatment
306 recommendation. Our software can diagnose multiple images simultaneously. A total of 30
307 sample images were available for download, and our e-mail address and telephone number were
308 also provided for all registered patients.

309 **Discussion**

310 The inferior performance of conventional feature methods when applied to diagnosis using the
311 slit-illumination images is mainly attributed to the following two causes. First, the conventional
312 feature methods use handcrafted descriptors to represent the original images, which are
313 completely reliant on the designer's experience and operator techniques, and which cannot
314 learn statistical features from the existing large dataset. Second, the conventional feature
315 methods and the SVM classifier do not take the problem of the imbalanced dataset into account,
316 and this results in the final predictions being biased towards the majority class and ignoring the
317 minority class (i.e., the positive samples). Therefore, these methods lead to inferior overall
318 accuracy and lower sensitivity.

319 The Adaboost ensemble learning methods led to moderate improvement of the recognition rates
320 when compared with the conventional feature methods because they train and apply multiple

321 classifiers jointly to determine the final grading results. Simultaneously, an under-sampling
322 method is incorporated into Adaboost to address the imbalanced dataset. Therefore, the
323 sensitivity of the methods is greatly enhanced, but this improvement leads to reduction of the
324 specificity. The overall accuracy rate is almost equal to that obtained when using the
325 conventional feature methods alone.

326 The CCNN-Ensemble method is significantly superior to the above methods in terms of all
327 grading indices, which was attributed to the following four improvements. First, the CCNN-
328 Ensemble method does not need to design any feature descriptor manually because it learns
329 high-level and statistical features directly from the original images. Second, we use three
330 different CNNs for ensemble learning, so that they can learn the different characteristics from
331 three different perspectives to enable joint determination of the final prediction. This ensemble
332 of multiple CNN technologies is beneficial in enhancing the overall performance. Third, the
333 cost-sensitive approach is integrated into the CCNN-Ensemble method and takes greater
334 account of the minority class to ensure that the sensitivity indicator is valid for the imbalanced
335 dataset. In addition, transfer learning is applied to our model to enable fine-tuning of the
336 trainable parameters from a better starting point, thus making it easier to jump out from the
337 local minimum. As a result, the higher accuracy and specificity performances are maintained
338 while the sensitivity is also greatly enhanced.

339 The CCNN-Ensemble method also demonstrated better performance on two external datasets,
340 and their recognition rates were almost equal to that of the validation dataset. This indicates
341 that the proposed approach is insensitive to the different data sources, and that its

342 generalizability and robustness are better than those of the conventional methods. These
343 experimental conclusions provide sufficient evidence to justify the application of the CCNN-
344 Ensemble method in complex clinical scenarios.

345 Based on our proposed method, automated diagnostic software was developed and deployed to
346 serve patients and ophthalmologists remotely in the form of a cloud service, which provided
347 important clinical value for pediatric cataract diagnosis. By accessing our automatic diagnostic
348 software remotely, any patient can upload slit-illumination images and can then quickly obtain
349 prediction results and an appropriate treatment recommendation. Therefore, this remotely-aided
350 diagnosis method freed the doctors from performing tedious examinations and helped patients
351 located in remote areas. In addition, this work can also provide a teaching role for junior doctors.

352 However, several limitations of this study should be mentioned. First, multiple CNNs with
353 different structures are integrated into an architecture. Although the strategy of ensemble
354 learning significantly improves the accuracy, it is slightly less cost-effective due to the high
355 requirement of the computing resource than a single CNN model. Second, our model is solely
356 depended on the slit-illumination image, which is insufficient to identify the lens opacity in
357 occasional situations. Combining the electronic medical records and other optical images may
358 provide valuable supplements for the comprehensive assessment of lens opacity. Third, the
359 robustness and stability of our method are required to be verified before the further
360 generalization of other medical situations. Despite the above limitations, this study provides a
361 practical strategy for heterogeneous lens opacity diagnosis with promising performance
362 validated in multi-source datasets. Further studies with the integration of electronic medical

363 records and more optical images will pave the way for wide-range clinical application of our
364 work.

365 **Conclusions**

366 In this paper, we proposed a feasible and automated CCNN-Ensemble method for effective
367 diagnosis of pediatric cataracts using heterogeneous slit-illumination images. We integrated
368 three deep CNNs and cost-sensitive technology to construct an ensemble learning method that
369 could identify the severity of lens opacity based on three grading indices. The experimental
370 results and comparison analyses verified that the proposed method is superior to other
371 conventional methods. The performance of the CCNN-Ensemble method on two external
372 datasets indicated its improved robustness and generalizability. Finally, a set of cloud-based
373 automatic diagnostic software was produced for use by both patients and ophthalmologists.
374 This research could provide a helpful reference for analysis of other medical images and will
375 help to promote the use of artificial intelligence techniques in clinical applications.

376 **Supplementary file**

377 **Supplementary file 1: Detailed performance comparison of the different methods in terms**
378 **of the opacity density and location grading.** The detailed comparison results of the different
379 methods in terms of the opacity density and location grading are presented in Supplementary
380 files.

381 **Declarations**

382 **Ethics approval and consent to participate**

383 The study was approved by the Ethics Committee of Zhongshan Ophthalmic Center of Sun Yat-
384 sen University. Written informed consent was obtained from all the study participants' parents
385 or legal guardian.

386 **Consent for publication**

387 Not applicable.

388 **Availability of data and materials**

389 The datasets of the current study are available from the corresponding author on reasonable
390 request.

391 **Abbreviations**

392 CNN: convolutional neural network; CCNN-Ensemble: ensemble of cost-sensitive
393 convolutional neural networks; ResNet: residual convolutional neural network; Adaboost:
394 adaptive boosting ensemble learning; SVM: support vector machine; LBP: local binary pattern;
395 WT: wavelet transformation; SIFT: scale-invariant feature transform; COTE: color and texture
396 features; ReLUs: rectified linear units; Mini-batch-GD: mini-batch gradient descent; ACC:
397 accuracy; SPE: specificity; SEN: sensitivity; F1_M: F1-measure; G_M: G-mean; ROC:
398 receiver operating characteristic curve; AUC: area under the ROC curve; PR: precision recall
399 curve; ROI: region of interest; CCPMOH: Childhood Cataract Program of the Chinese Ministry
400 of Health.

401 **Competing interests**

402 The authors declare that they have no competing interests.

403 **Authors' contributions**

404 J.W.J., H.T.L. and X.Y.L. designed the research; J.W.J., E.P.L., L.M.W., Z.W.L., M.M.Z. and
405 R.Y.L. conducted the study; E.P.L., Z.Z.L., Z.L.L., J.J.C. and D.N.W. collected the data; J.W.J.,
406 Y.B.S., H.R.F. and L.M.W. were responsible for coding and computer-aided diagnosis software;
407 M.M.Z. supported the mathematical theory; J.W.J., X.H.W., Z.W.L., J.J.C., H.R.F. and Y.B.S
408 analyzed and completed the experimental results; and J.W.J., E.P.L., H.T.L. and X.Y.L. co-
409 wrote the manuscript. All the authors discussed the results and reviewed the manuscript.

410 **Acknowledgements**

411 We am grateful thanks to Dr. Lin Liu, Shuai Wang, and Fan Liu for their helpful guidance and
412 suggestion with this project.

413 **Funding**

414 This study was funded by the National Key R&D Program of China (No. 2018YFC0116500),
415 the National Natural Science Foundation of China (No. 81770967), the National Natural
416 Science Fund for Distinguished Young Scholars (No. 81822010), the Science and Technology
417 Planning Projects of Guangdong Province (No. 2018B010109008), the Science and Technology
418 Planning Projects of Guangdong Province (No. 2019B030316012), and the Fundamental
419 Research Funds for the Central Universities (No. JBX180704). The sponsor or funding
420 organization had no role in the design or conduct of this research.

421 **Authors' information**

422 Correspondence and requests for materials should be addressed to H.T.L.
423 (haot.lin@hotmail.com) or X.Y.L. (xyliu@xidian.edu.cn).

424

425 **References**

- 426 1. Bernardes R, Serranho P, Lobo C. Digital ocular fundus imaging: a review.
427 *Ophthalmologica*. 2011;226(4):161-81.
- 428 2. Ng EY, Acharya UR, Suri JS, Campilho A. *Image Analysis and Modeling in*
429 *Ophthalmology*: CRC Press; 2014.
- 430 3. Zhang Z, Srivastava R, Liu H, Chen X, Duan L, Kee Wong DW, et al. A survey on computer
431 aided diagnosis for ocular diseases. *BMC Med Inform Decis Mak*. 2014;14:80.
- 432 4. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial
433 intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*.
434 2019;103(2):167-75.
- 435 5. Armstrong GW, Lorch AC. A (eye): A review of current applications of artificial
436 intelligence and machine learning in ophthalmology. *International Ophthalmology Clinics*.
437 2020;60(1):57-71.
- 438 6. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial
439 intelligence in ophthalmology: a review. *Clinical & experimental ophthalmology*.
440 2019;47(1):128-39.
- 441 7. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for
442 the multihospital collaborative management of congenital cataracts. *Nature Biomedical*
443 *Engineering*. 2017;1:0024.
- 444 8. Wang L, Zhang K, Liu X, Long E, Jiang J, An Y, et al. Comparative analysis of image
445 classification methods for automatic diagnosis of ophthalmic images. *Scientific Reports*.
446 2017;7.
- 447 9. Liu X, Jiang J, Zhang K, Long E, Cui J, Zhu M, et al. Localization and diagnosis framework
448 for pediatric cataracts based on slit-lamp images using deep features of a convolutional
449 neural network. *PLOS ONE*. 2017;12(3):e0168606.
- 450 10. Klein BEK, Klein R, Linton KLP, Magli YL, Neider MW. Assessment of cataracts from
451 photographs in the Beaver Dam Eye Study. *Ophthalmology*. 1990;97(11):1428-33.
- 452 11. Reid JE, Eaton E. Artificial intelligence for pediatric ophthalmology. *Current opinion in*
453 *ophthalmology*. 2019;30(5):337-46.

- 454 12. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic efficacy and therapeutic
455 decision-making capacity of an artificial intelligence platform for childhood cataracts in
456 eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine*. 2019;9:52-9.
- 457 13. Chylack LT, Jr, Wolfe JK, Singer DM, et al. The lens opacities classification system iii.
458 *Archives of Ophthalmology*. 1993;111(6):831-6.
- 459 14. Kumar S, Yogesan K, Constable I. Telemedical diagnosis of anterior segment eye diseases:
460 validation of digital slit-lamp still images. *Eye*. 2009;23(3):652-60.
- 461 15. Kolhe S, Guru MSK. *Cataract Classification and Grading: A Survey*. 2015.
- 462 16. Li H, Lim JH, Liu J, Mitchell P, Tan AG, Wang JJ, et al. A Computer-Aided Diagnosis
463 System of Nuclear Cataract. *IEEE Transactions on Biomedical Engineering*.
464 2010;57(7):1690-8.
- 465 17. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on
466 deep learning. *IEEE Transactions on Biomedical Engineering*. 2015;62(11):2693-701.
- 467 18. Amaya L, Taylor D, Russell-Eggitt I, Nischal KK, Lengyel D. The morphology and natural
468 history of childhood cataracts. *Survey of ophthalmology*. 2003;48(2):125-44.
- 469 19. Wu X, Long E, Lin H, Liu Y. Prevalence and epidemiological characteristics of congenital
470 cataract: a systematic review and meta-analysis. *Scientific Reports*. 2016;6:28564.
- 471 20. Medsinghe A, Nischal KK. Pediatric cataract: challenges and future directions. *Clinical
472 ophthalmology (Auckland, NZ)*. 2015;9:77.
- 473 21. Lenhart PD, Courtright P, Wilson ME, Lewallen S, Taylor DS, Ventura MC, et al. Global
474 challenges in the management of congenital cataract: proceedings of the 4th International
475 Congenital Cataract Symposium held on March 7, 2014, New York, New York. *Journal of
476 American Association for Pediatric Ophthalmology and Strabismus*. 2015;19(2):e1-e8.
- 477 22. Ellis FJ. Management of pediatric cataract and lens opacities. *Current opinion in
478 ophthalmology*. 2002;13(1):33-7.
- 479 23. Wilson ME, Trivedi RH, Pandey SK. *Pediatric cataract surgery: techniques, complications,
480 and management: Lippincott Williams & Wilkins; 2005*.
- 481 24. Lin H, Long E, Chen W, Liu Y. Documenting rare disease data in China. *Science*.
482 2015;349(6252).
- 483 25. Chen W, Long E, Chen J, Liu Z, Lin Z, Cao Q, et al. Timing and approaches in congenital

- 484 cataract surgery: a randomised controlled trial. *The Lancet*. 2016;388:S52.
- 485 26. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional
486 neural networks. *Advances in Neural Information Processing Systems*. 2012;25(2):2012.
- 487 27. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with
488 convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern
489 Recognition*. 2015:1-9.
- 490 28. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *arXiv preprint
491 arXiv:151203385*. 2015.
- 492 29. Ali S, Majid A, Javed SG, Sattar M. Can-CSC-GBE: Developing cost-sensitive classifier
493 with gentleboost ensemble for breast cancer classification using protein amino acids and
494 imbalanced data. *Computers in biology and medicine*. 2016;73:38-46.
- 495 30. Zhou Z-H, Liu X-Y. Training cost-sensitive neural networks with methods addressing the
496 class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering*.
497 2006;18(1):63-77.
- 498 31. Daugman J. New methods in iris recognition. *IEEE Transactions on Systems, Man, and
499 Cybernetics, Part B*. 2007;37(5):1167-75.
- 500 32. Masek L. Recognition of human iris patterns for biometric identification. The University
501 of Western Australia. 2003;2.
- 502 33. Yang J-J, Li J, Shen R, Zeng Y, He J, Bi J, et al. Exploiting ensemble learning for automatic
503 cataract detection and grading. *Computer methods and programs in biomedicine*.
504 2016;124:45-57.
- 505 34. Guo L, Yang J-J, Peng L, Li J, Liang Q. A computer-aided healthcare system for cataract
506 classification and grading based on fundus image analysis. *Computers in Industry*.
507 2015;69:72-80.
- 508 35. Huang W, Chan KL, Li H, Lim JH, Liu J, Wong TY. A computer assisted method for nuclear
509 cataract grading from slit-lamp images using ranking. *IEEE Transactions on Medical
510 Imaging*. 2011;30(1):94-107.
- 511 36. Tang Y, Zhang Y-Q, Chawla NV, Krasser S. SVMs modeling for highly imbalanced
512 classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*.
513 2009;39(1):281-8.

- 514 37. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural
515 networks for computer-aided detection: CNN architectures, dataset characteristics and
516 transfer learning. *IEEE Trans Med Imaging*. 2016.
- 517 38. Ravishankar H, Sudhakar P, Venkataramani R, Thiruvankadam S, Annangi P, Babu N, et al.
518 Understanding the mechanisms of deep transfer learning for medical images. *arXiv preprint*
519 *arXiv:170406040*. 2017.
- 520 39. Ciresan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image
521 classification. *Computer Vision and Pattern Recognition (CVPR)*. 2012:3642-9.
- 522 40. Krawczyk B, Schaefer G, Woźniak M. A hybrid cost-sensitive ensemble for imbalanced
523 breast thermogram classification. *Artificial intelligence in medicine*. 2015;65(3):219-27.
- 524 41. Bottou L. Large-scale machine learning with stochastic gradient descent. *Proceedings of*
525 *COMPSTAT'2010*. 2010:177-86.
- 526 42. Jia YaS, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick,
527 Ross and Guadarrama, Sergio and Darrell, Trevor. Caffe: Convolutional architecture for
528 fast feature embedding. *arXiv preprint arXiv:14085093*. 2014.
- 529 43. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model
530 selection. *International Joint Conference on Artificial Intelligence*. 1995;14(2):1137-45.
- 531
- 532

533 **Figure legends**

534 **Fig. 1. Dataset preparation and performance evaluation of multiple methods. (a) Dataset**
535 **labelling and preprocessing.** 470 training and validation samples and 132 independent test
536 samples were derived from samples provided by the Zhongshan Ophthalmic Center of Sun Yat-
537 sen University; 79 Internet-based samples were collected using the Baidu and Google search
538 engines. Each image was independently graded and labeled by three senior ophthalmologists;
539 subsequently, the images were cropped automatically using twice-applied Canny detection and
540 Hough transformation. **(b) Model comparison and evaluation.** The training and validation
541 dataset was used to train and evaluate the performances of the different methods and select the
542 best model. Independent testing and Internet-based datasets were also used to evaluate the
543 stability and the generalizability of the CCNN-Ensemble method. Notes: WT: wavelet
544 transformation; LBP: local binary pattern; SIFT: scale-invariant feature transform; COTE:
545 color and texture features; Adaboost: adaptive boosting ensemble learning; CCNN-Ensemble:
546 ensemble learning of cost-sensitive convolutional neural networks.

547 **Fig. 2. Framework of the CCNN-Ensemble method.** The preprocessed images were input
548 into three parallel deep learning CNNs (AlexNet, GoogLeNet, and ResNet) with different
549 network structures for feature extraction and classification; a unified ensemble learning of
550 CNNs was then used to improve the recognition rate of the classifier. The cost-sensitive layer
551 was used to adjust the costs of the positive and negative samples in the loss function to address
552 the imbalanced dataset problem. Notes: CNN: convolutional neural network; AlexNet: eight-
553 layer Alex CNN; GoogLeNet: 22-layer inception CNN developed by Google researchers;
554 ResNet: 50-layer residual CNN.

555 **Fig. 3. Performance comparisons of the different methods for the three grading indices.**

556 Images **(a)–(c)** show performance comparisons of conventional features, adaboost ensemble

557 learning and CCNN-Ensemble methods for the lens opacity area, opacity density, and opacity

558 location, respectively. The sensitivity of adaboost ensemble learning methods is greatly

559 improved over the conventional feature methods, whereas their specificity indicator is reduced

560 and the accuracy has no significant improvement. The CCNN-Ensemble method outperforms

561 other conventional features and adaboost ensemble approaches and offers exceptional accuracy,

562 specificity and sensitivity in terms of three grading indices of lens opacity: area (92.13%,

563 92.00%, and 92.31%), density (92.77%, 93.85%, and 91.43%) and location (92.76%, 95.25%,

564 and 89.29%). Notes: ACC: accuracy; SPE: specificity; SEN: sensitivity; WT: wavelet

565 transformation; LBP: local binary pattern; SIFT: scale-invariant feature transform; COTE:

566 color and texture features; Ada: adaptive boosting ensemble learning; WT-Ada: adaptive

567 boosting ensemble learning with wavelet transformation feature; CCNN-Ensemble: ensemble

568 learning of cost-sensitive convolutional neural networks.

569 **Fig. 4. ROC and PR curves for the different methods in opacity area grading. (a)** ROC

570 curves and AUC values for the CCNN-Ensemble method and four comparison methods: WT-

571 Ada, SIFT-Ada, LBP-Ada, and COTE-Ada. **(b)** PR curves for the CCNN-Ensemble method

572 and the four comparison methods. Notes: WT: wavelet transformation; LBP: local binary

573 pattern; SIFT: scale-invariant feature transform; COTE: color and texture features; Ada:

574 adaptive boosting ensemble learning; WT-Ada: adaptive boosting ensemble learning with

575 wavelet transformation feature; CCNN-Ensemble: ensemble learning of cost-sensitive

576 convolutional neural networks; ROC: receiver operating characteristic curve; AUC: area under

577 the ROC curve; PR: precision recall curve.

578 **Fig. 5. Performance analysis results for the CCNN-Ensemble on two external datasets. (a)**

579 The performance comparison, ROC curves and PR curves of the CCNN-Ensemble method for

580 lens opacity area, density and location grading on independent testing dataset. **(b)** The

581 performance comparison, ROC curves and PR curves for lens opacity area, density and

582 location grading on Internet-based dataset. The model performances are satisfactory when

583 applied to the two external datasets, independent test images: area (94.70%, 96.70%, and

584 90.24%), density (93.18%, 94.23%, and 89.29%) and location (93.18%, 94.00%, and 90.63%);

585 internet-based images: area (89.87%, 89.47%, and 90.00%), density (88.61%, 88.89%, and

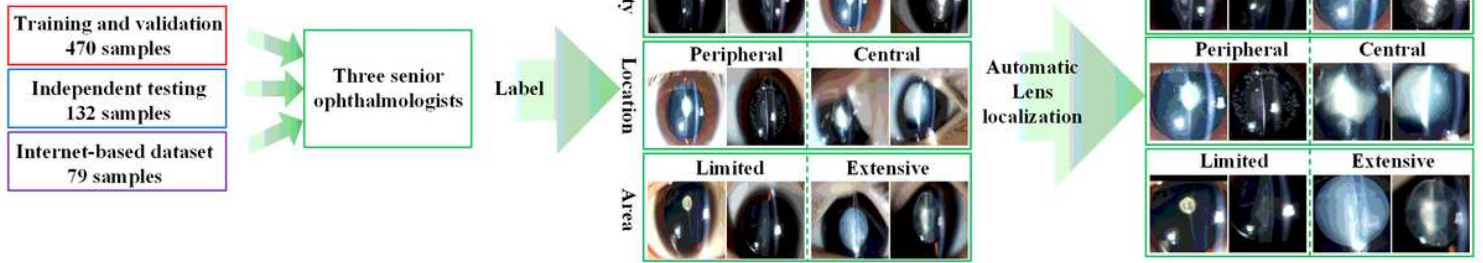
586 88.52%) and location (87.34%, 87.50%, and 87.30%), indicating that the model is universal

587 and effective. Notes: ACC: accuracy; SPE: specificity; SEN: sensitivity; ROC: receiver

588 operating characteristic curve; AUC: area under the ROC curve; PR: precision recall curve.

Figures

a. Dataset labelling and preprocessing



b. Models comparison and evaluation

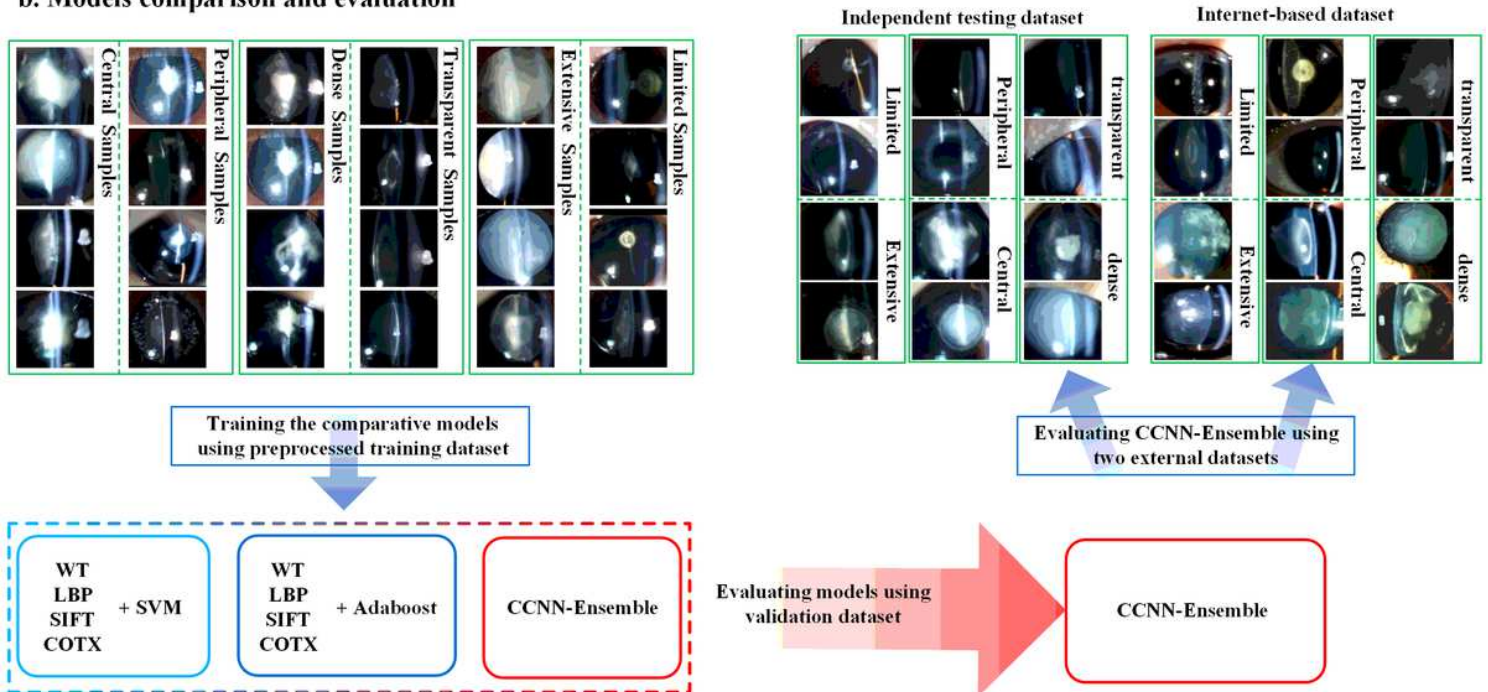


Figure 1

Dataset preparation and performance evaluation of multiple methods. (a) Dataset labelling and preprocessing. 470 training and validation samples and 132 independent test samples were derived from samples provided by the Zhongshan Ophthalmic Center of Sun Yat-sen University; 79 Internet-based samples were collected using the Baidu and Google search engines. Each image was independently graded and labeled by three senior ophthalmologists; subsequently, the images were cropped automatically using twice-applied Canny detection and Hough transformation. (b) Model comparison and evaluation. The training and validation dataset was used to train and evaluate the performances of the different methods and select the best model. Independent testing and Internet-based datasets were also used to evaluate the stability and the generalizability of the CCNN-Ensemble method. Notes: WT: wavelet transformation; LBP: local binary pattern; SIFT: scale-invariant feature transform; COTE: color and texture features; Adaboost: adaptive boosting ensemble learning; CCNN-Ensemble: ensemble learning of cost-sensitive convolutional neural networks.

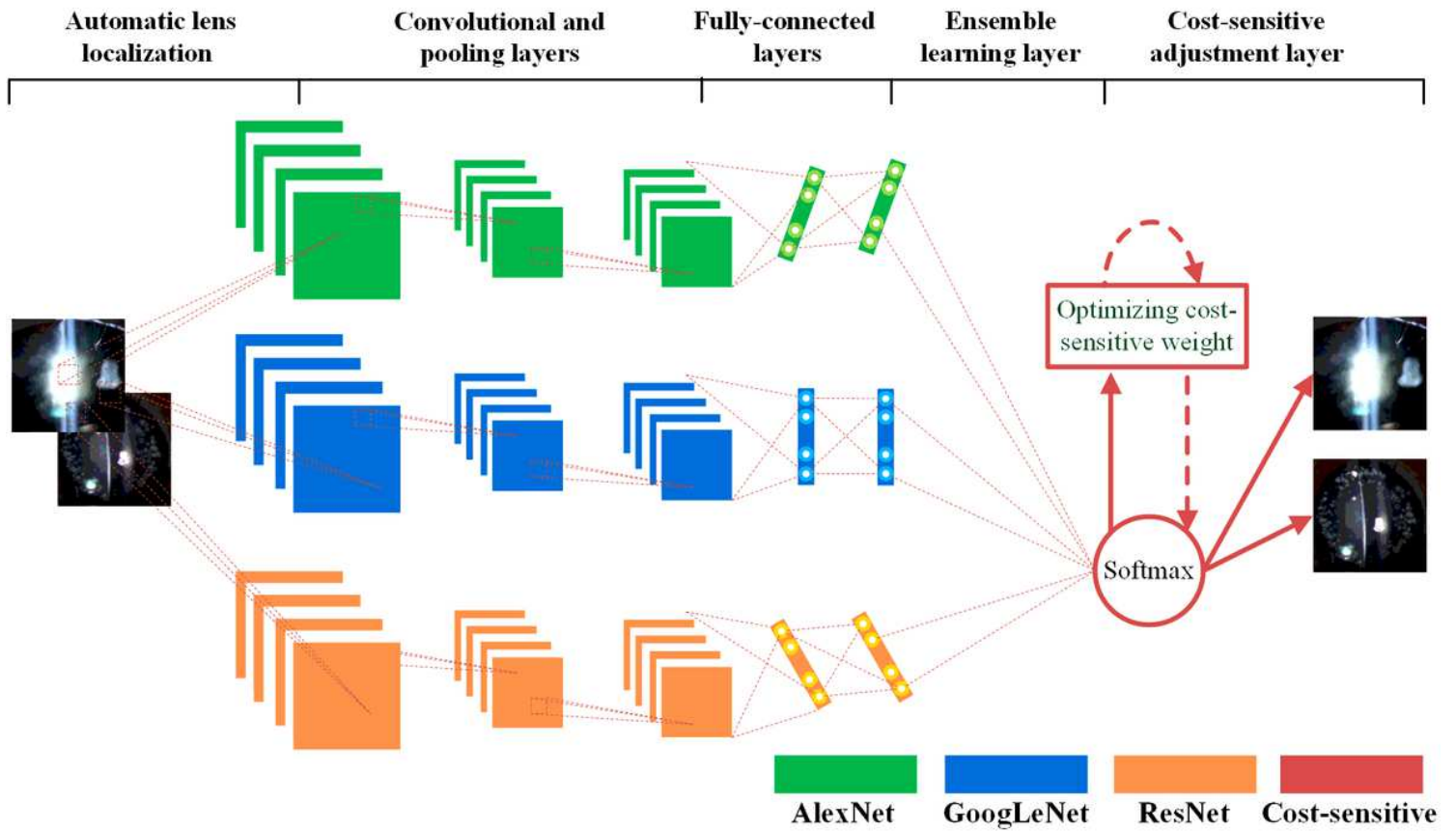


Figure 2

Framework of the CCNN-Ensemble method. The preprocessed images were input into three parallel deep learning CNNs (AlexNet, GoogLeNet, and ResNet) with different network structures for feature extraction and classification; a unified ensemble learning of CNNs was then used to improve the recognition rate of the classifier. The cost-sensitive layer was used to adjust the costs of the positive and negative samples in the loss function to address the imbalanced dataset problem. Notes: CNN: convolutional neural network; AlexNet: eight-layer Alex CNN; GoogLeNet: 22-layer inception CNN developed by Google researchers; ResNet: 50-layer residual CNN.

WT
 WT-Ada
 SIFT
 SIFT-Ada
 LBP
 LBP-Ada
 COTX
 COTX-Ada
 CCNN- Ensemble

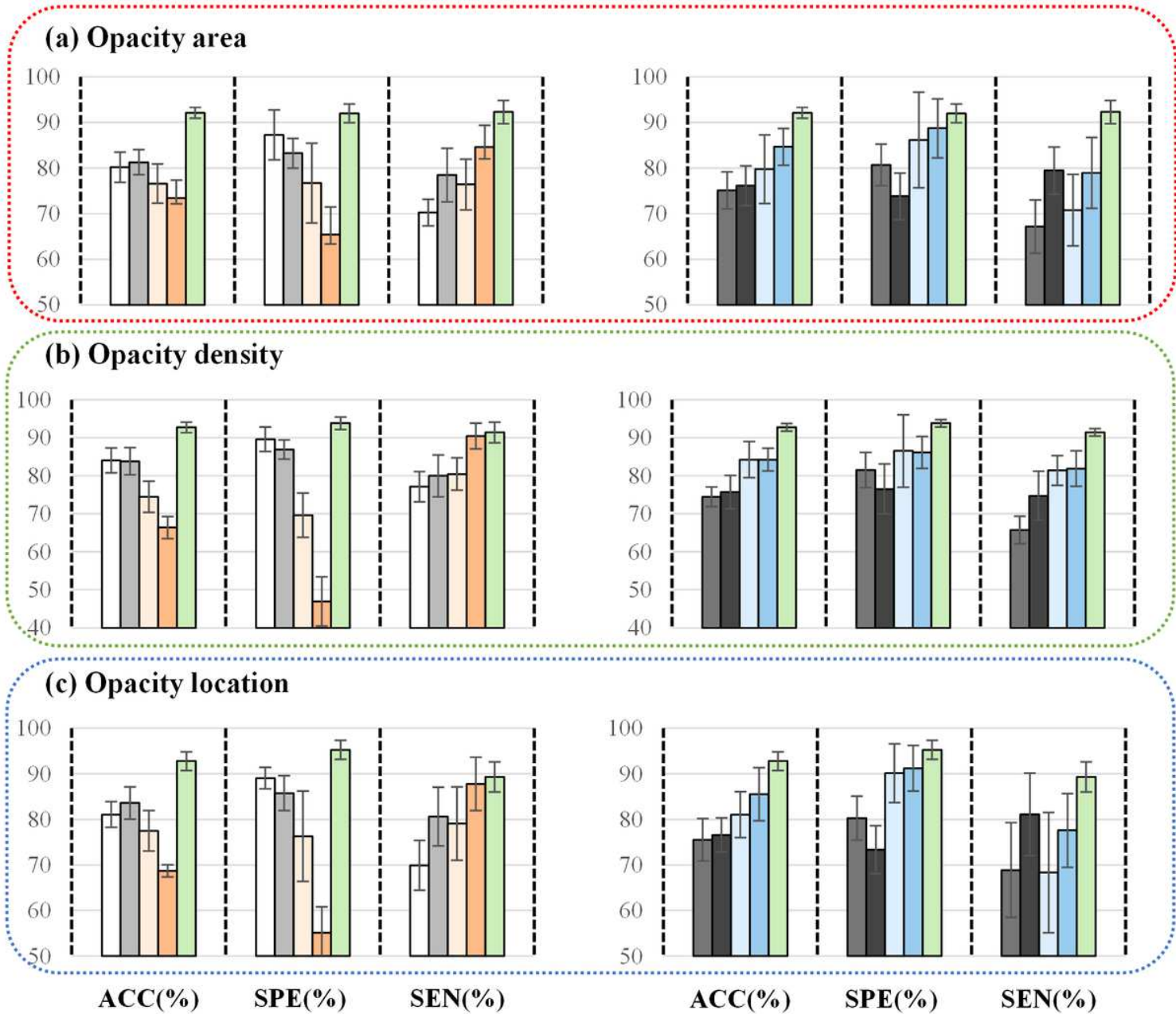


Figure 3

Performance comparisons of the different methods for the three grading indices. Images (a)–(c) show performance comparisons of conventional features, adaboost ensemble learning and CCNN-Ensemble methods for the lens opacity area, opacity density, and opacity location, respectively. The sensitivity of adaboost ensemble learning methods is greatly improved over the conventional feature methods, whereas their specificity indicator is reduced and the accuracy has no significant improvement. The CCNN-Ensemble method outperforms other conventional features and adaboost ensemble approaches and offers exceptional accuracy, specificity and sensitivity in terms of three grading indices of lens opacity: area (92.13%, 92.00%, and 92.31%), density (92.77%, 93.85%, and 91.43%) and location (92.76%, 95.25%, and 89.29%). Notes: ACC: accuracy; SPE: specificity; SEN: sensitivity; WT: wavelet

transformation; LBP: local binary pattern; SIFT: scale-invariant feature transform; COTE: color and texture features; Ada: adaptive boosting ensemble learning; WT-Ada: adaptive boosting ensemble learning with wavelet transformation feature; CCNN-Ensemble: ensemble learning of cost-sensitive convolutional neural networks.

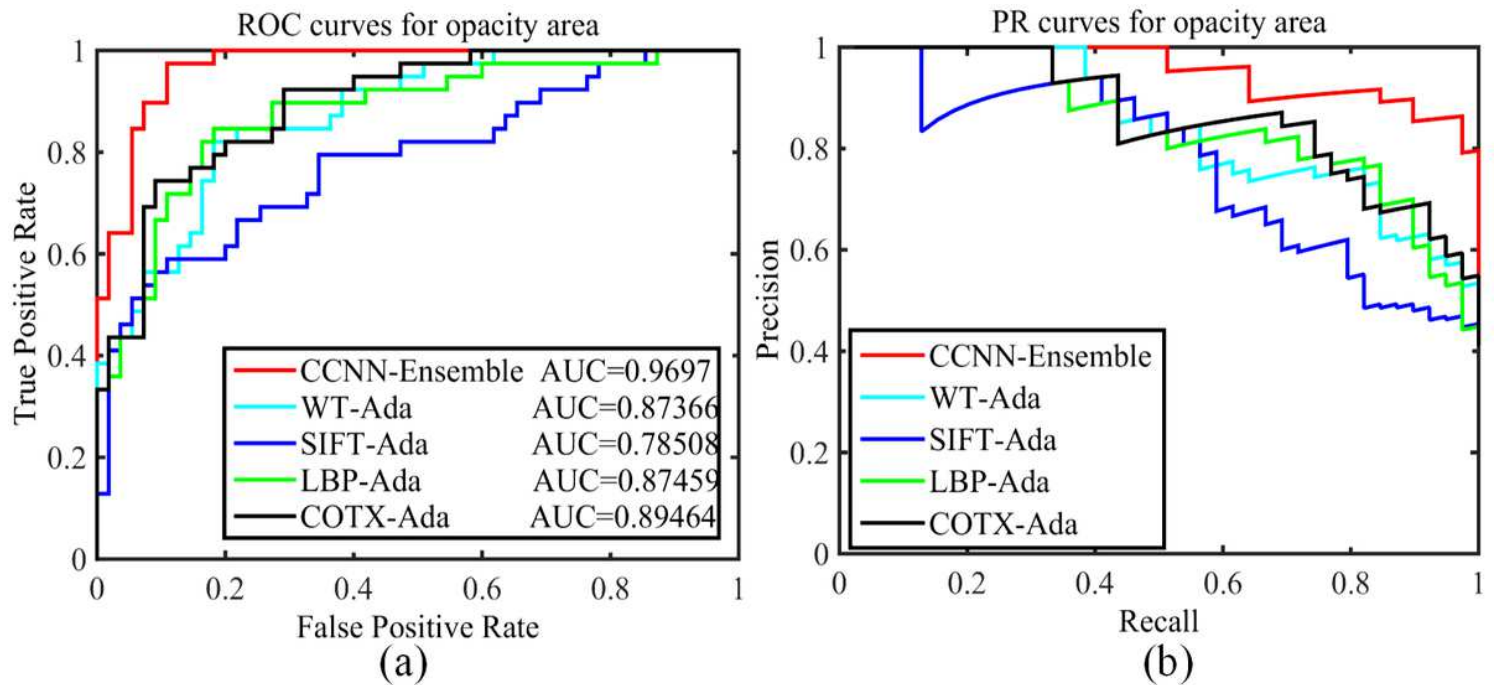


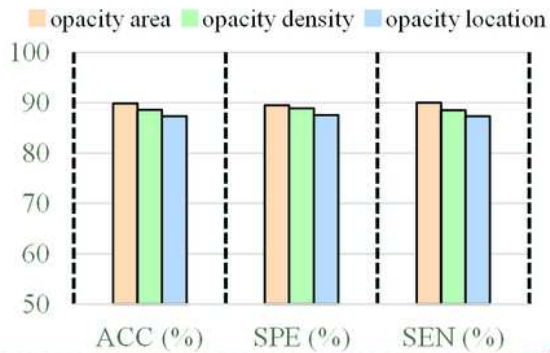
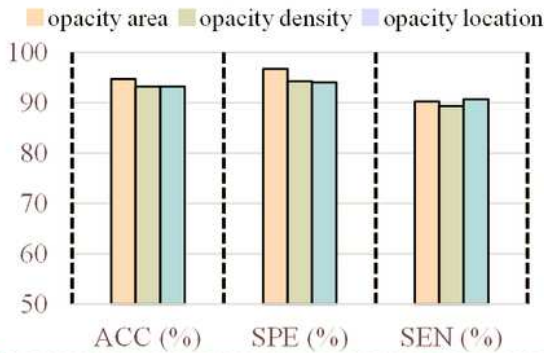
Figure 4

ROC and PR curves for the different methods in opacity area grading. (a) ROC curves and AUC values for the CCNN-Ensemble method and four comparison methods: WT-Ada, SIFT-Ada, LBP-Ada, and COTE-Ada. (b) PR curves for the CCNN-Ensemble method and the four comparison methods. Notes: WT: wavelet transformation; LBP: local binary pattern; SIFT: scale-invariant feature transform; COTE: color and texture features; Ada: adaptive boosting ensemble learning; WT-Ada: adaptive boosting ensemble learning with wavelet transformation feature; CCNN-Ensemble: ensemble learning of cost-sensitive convolutional neural networks; ROC: receiver operating characteristic curve; AUC: area under the ROC curve; PR: precision recall curve.

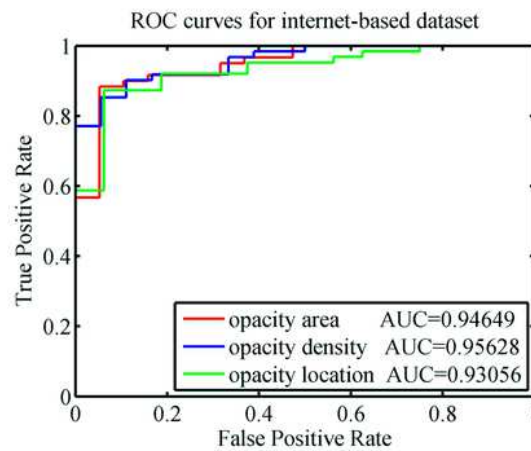
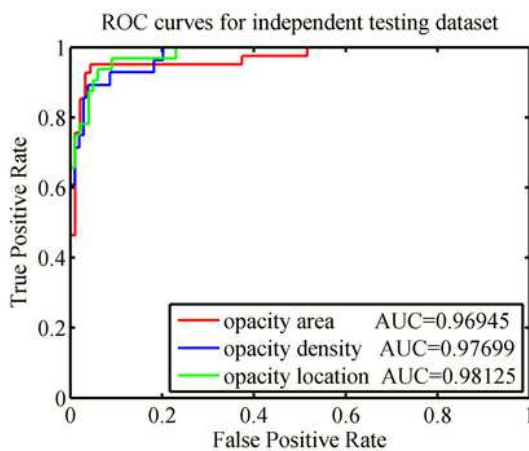
(a) Independent testing dataset

(b) Internet-based dataset

Performance comparison



ROC curves



PR curves

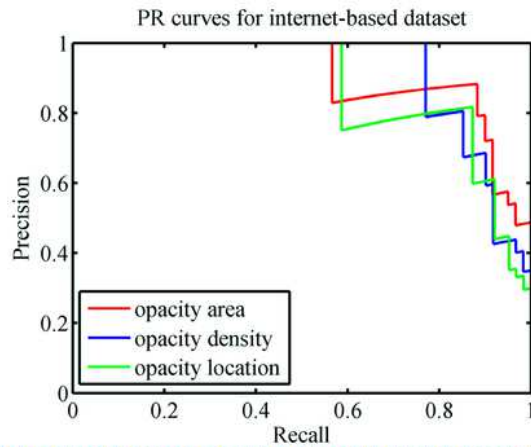
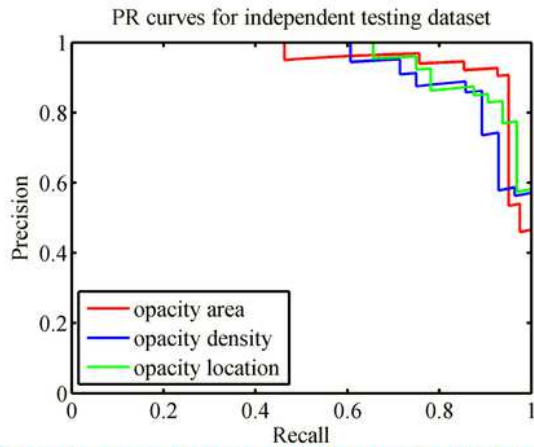


Figure 5

Performance analysis results for the CCNN-Ensemble on two external datasets. (a) The performance comparison, ROC curves and PR curves of the CCNN-Ensemble method for lens opacity area, density and location grading on independent testing dataset. (b) The performance comparison, ROC curves and PR curves for lens opacity area, density and location grading on Internet-based dataset. The model performances are satisfactory when applied to the two external datasets, independent test images: area

(94.70%, 96.70%, and 90.24%), density (93.18%, 94.23%, and 89.29%) and location (93.18%, 94.00%, and 90.63%); internet-based images: area (89.87%, 89.47%, and 90.00%), density (88.61%, 88.89%, and 88.52%) and location (87.34%, 87.50%, and 87.30%), indicating that the model is universal and effective. Notes: ACC: accuracy; SPE: specificity; SEN: sensitivity; ROC: receiver operating characteristic curve; AUC: area under the ROC curve; PR: precision recall curve.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile.docx](#)