

Deep Learning-Based Cataract Detection and Grading from Slit-Lamp and Retro-Illumination Photographs

Model Development and Validation Study

Ki Young Son, MD,^{1,*} Jongwoo Ko, MSc,^{2,*} Eunseok Kim, BSc,^{3,4} Si Young Lee, BSc,⁴ Min-Ji Kim, MD,¹ Jisang Han, MD,⁵ Eunhae Shin, MD,¹ Tae-Young Chung, MD, PhD,^{1,‡} Dong Hui Lim, MD, PhD^{1,6,‡}

Purpose: To develop and validate an automated deep learning (DL)-based artificial intelligence (AI) platform for diagnosing and grading cataracts using slit-lamp and retroillumination lens photographs based on the Lens Opacities Classification System (LOCS) III.

Design: Cross-sectional study in which a convolutional neural network was trained and tested using photographs of slit-lamp and retroillumination lens photographs.

Participants: One thousand three hundred thirty-five slit-lamp images and 637 retroillumination lens images from 596 patients.

Methods: Slit-lamp and retroillumination lens photographs were graded by 2 trained graders using LOCS III. Image datasets were labeled and divided into training, validation, and test datasets. We trained and validated AI platforms with 4 key strategies in the AI domain: (1) region detection network for redundant information inside data, (2) data augmentation and transfer learning for the small dataset size problem, (3) generalized cross-entropy loss for dataset bias, and (4) class balanced loss for class imbalance problems. The performance of the AI platform was reinforced with an ensemble of 3 AI algorithms: ResNet18, WideResNet50-2, and ResNext50.

Main Outcome Measures: Diagnostic and LOCS III-based grading prediction performance of AI platforms.

Results: The AI platform showed robust diagnostic performance (area under the receiver operating characteristic curve [AUC], 0.9992 [95% confidence interval (CI), 0.9986–0.9998] and 0.9994 [95% CI, 0.9989–0.9998]; accuracy, 98.82% [95% CI, 97.7%–99.9%] and 98.51% [95% CI, 97.4%–99.6%]) and LOCS III-based grading prediction performance (AUC, 0.9567 [95% CI, 0.9501–0.9633] and 0.9650 [95% CI, 0.9509–0.9792]; accuracy, 91.22% [95% CI, 89.4%–93.0%] and 90.26% [95% CI, 88.6%–91.9%]) for nuclear opalescence (NO) and nuclear color (NC) using slit-lamp photographs, respectively. For cortical opacity (CO) and posterior subcapsular opacity (PSC), the system achieved high diagnostic performance (AUC, 0.9680 [95% CI, 0.9579–0.9781] and 0.9465 [95% CI, 0.9348–0.9582]; accuracy, 96.21% [95% CI, 94.4%–98.0%] and 92.17% [95% CI, 88.6%–95.8%]) and good LOCS III-based grading prediction performance (AUC, 0.9044 [95% CI, 0.8958–0.9129] and 0.9174 [95% CI, 0.9055–0.9295]; accuracy, 91.33% [95% CI, 89.7%–93.0%] and 87.89% [95% CI, 85.6%–90.2%]) using retroillumination images.

Conclusions: Our DL-based AI platform successfully yielded accurate and precise detection and grading of NO and NC in 7-level classification and CO and PSC in 6-level classification, overcoming the limitations of medical databases such as few training data or biased label distribution. *Ophthalmology Science* 2022;2:100147 © 2022 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Supplemental material available at www.ophtalmologyscience.org.

Cataract was the leading cause of blindness among the global population who were blind in 2015.¹ As the population ages, the prevalence and incidence of cataracts are expected to increase.² Surgical removal of the lens and implantation of intraocular lens are the only effective treatments of a visually significant cataract.

Decision-making for cataract surgery is a major challenge for clinicians. Cataract classification systems have

been developed to assess the extent of cataracts, and the Lens Opacities Classification System (LOCS) III is a widely used subjective cataract classification system for measuring cataract severity.³ Ophthalmologists determine the grade of cataract by observing the lens image under the slit lamp compared with the LOCS III standard lens images. Studies have reported that LOCS III classification shows good interobserver agreement compared with other methods of

grading cataracts.^{4,5} Although several studies of the LOCS III showed that practice and meeting to discuss interpretations improved interobserver agreement, it is obvious that the experience of the observer may affect the reliability of the evaluations.^{6,7} To overcome this matter, additional detailed clinical histories and formal questionnaires are used to assess the visual functional status of patients.⁸

Recently, artificial intelligence (AI)-assisted cataract diagnosis has attracted many researchers due to its feasibility and potential. In the early stages of AI applications in adults with cataract, many research groups have validated the deep learning (DL) algorithm for cataract diagnosis and grading using slit-lamp photographs. However, the results were tentative for application to clinical practice, considering the limitations of technology at the time. Subsequently, researchers shifted the focus to fundus photographs as an imaging method. Recently, an AI algorithm using slit-lamp images with a large amount of data was introduced and showed improved outcomes.⁹ However, the researchers established and validated a DL algorithm for cataract diagnosis and grading based on LOCS II and trained datasets of slit-lamp or diffuse-beam images, which have disadvantages in detecting cortical opacity-type or posterior capsular opacity-type cataracts.

We developed and validated an automated DL-based AI platform for diagnosing and grading cataracts using 2 types of lens images—slit-lamp photographs and retroillumination photographs—of patients based on LOCS III. Our deep convolutional neural network (CNN) was trained with various slit-lamp and retroillumination images to identify the presence and severity of cataract and recommend a proper treatment plan for the patients.

Methods

Study Approval

This study followed all guidelines for experimental investigations in human subjects. The study was approved by the Samsung Medical Center Institutional Review Board (IRB file number, 2020-08-035) and adhered to the tenets of the Declaration of Helsinki. The research protocol was approved by the Samsung Medical Center Institutional Review Board and ethic committees and all human participants gave written informed consent.

Participants

This cross-sectional retrospective study included patients 14 to 94 years of age with different types of cataract (nuclear sclerotic, cortical, and posterior subcapsular) who visited the outpatient clinic between January 2017 and December 2020 and had available anterior segment photograph data. Patients with pathologic features of the cornea, anterior chamber, lens, or iris that interfere with the detection of lens images (e.g., corneal opacity or edema, uveitis, and iris defects including aniridia, coloboma, and iridocorneal endothelial syndrome) and a medical history of previous ophthalmic surgery (e.g., keratoplasty, implantable Collamer lens, and cataract surgery) were excluded. The patients with retinal and vitreal diseases involving visual pathways that could interfere with visual acuity and final management plan were also excluded.

Data Collection

The authors reviewed the medical charts of all patients. The following data were collected: demographic data (age and sex), visual acuity, anterior segment evaluation, and fundus evaluation. Two types of lens images were obtained using slit-lamp digital cameras (SL-D7 [Topcon Medical System, Inc]; and D850 [Nikon, Inc]) according to the LOCS III protocols. A slit-lamp color photograph was obtained for each eye at a magnification of $\times 10$, a slit width of 0.2 mm, and a slit-beam orientation of 30° . The retroillumination photograph was obtained with a camera and a flash brightness of 100 in a dark room (< 15 lux). Two independent observers (K.Y.S. and S.Y.L.) specifically trained in the use of LOCS III and masked to the study participants' clinical information reviewed images and performed clinical lens grading of nuclear color (NC), nuclear opalescence (NO), cortical opacity (CO), and posterior subcapsular opacity (PSC) based on standard lens photographs of LOCS III. Additional lens-grading levels were established for severe nuclear cataract over LOCS III grade 6 (brunescens/white cataract) to 6 plus grade and for normal slit-lamp, retroillumination photograph to grade 0 (Supplemental Figs S1 and S2). To obtain 4 class variables, we divided the lens grading into 4 categories: normal for grade 0, mild for grades 1 and 2, moderate for grades 3 and 4, and severe for grade 5 or more in all types of cataract, which was defined as severity-based grading. In the event of discrepancies between the 2 graders, the senior expert on cataract (D.H.L.) made the final grading diagnosis.

Development of Deep Learning Algorithm

In this study, the DL algorithm predicted cataract grades from slit-lamp and retroillumination images (NO and NC grades for slit-lamp images and CO and PSC grades for retroillumination images) by performing severity prediction as a multiclass classification task. Previous state-of-the-art studies⁹ have demonstrated the effectiveness of the DL algorithm in this direction, but with relatively unrealistic settings (e.g., large training data, balanced label distribution, etc.). However, most medical images consist of a small amount of data or skewed label distribution. Therefore, our system aimed to achieve data-efficient and robust learning of the DL algorithm by focusing on 4 problems that exist in cataract images:

1. Redundant information inside data: Because only limited constituent parts of cataract images have discriminative information for diagnosing the disease, other unimportant constituent parts may hinder the algorithm from correctly classifying images.
2. Dataset size: Because the amount of data is small, the network is easy to overfit to the training dataset of specific patients and shows poor performance during the test procedure.
3. Dataset bias: Because different experts can diagnose the same image as different grades, we cannot expect an absolutely objective diagnosis of the trained network.
4. Class imbalance: The number of images of normal patients is much larger than that of patients with diseases.

To remedy the problems in grading cataract images, our framework adopted 2 networks each for constraining salient regions, classifying images, and learning methods for empowering generalization ability on unseen data. In contrast to previous methods⁹ that require a large amount of data to be trained, our algorithm achieved robust and data-efficient learning by adapting the following strategy. Specifically, our system chose the following recent DL techniques to overcome the problems: (1) a region

detection network (RDN) was used to classify the core region of an image by cropping unnecessary parts out (Supplemental Fig S3); (2) data augmentation was used to provide generalization ability to DL algorithms and transfer learning and to accelerate the training procedure and avoid algorithm overfitting in the training data (for specific techniques and their hyperparameters, see Supplemental Table 1); (3) generalized cross-entropy (GCE) loss was used to enable the DL algorithm to be less prone to the dataset bias problem; and (4) class balanced (CB) loss was used to prevent the DL algorithm from yielding skewed predictions when trained on data with a long tail. We describe the details of our strategies in the Appendixes 1 and 2. Furthermore, we applied the model ensemble technique, which combines multiple models to reduce the high variance and bias of a single model and to obtain a higher test performance. In our work, we combined 3 AI algorithms for the model ensemble—ResNet18, WideResNet50-2, and ResNext50—and compared the accuracy of each of the 3 single networks and their ensembles.

Logic Flow Chart for Cataract Diagnosis and Management

The DL algorithm was designed to perform the following steps. In step 1, image data from slit-lamp or retroillumination photography and visual acuity data are collected from the instrument and medical chart. In step 2, the DL algorithm agent analyzes the lens images as showing a normal lens or showing the presence of any type of cataract. In step 3, the DL algorithm agent determines the severity of cataract based on the lens-grading system. The visual acuity of the participants was not used in step 3 of the DL algorithm, but rather in step 4, where the visual acuity was considered to suggest an optimal management plan for each participant. If the participant showed a normal lens and best-corrected visual acuity (BCVA) of better than 20/32, the DL algorithm agent automatically suggested routine observation for the management plan. For participants with a normal lens with decreased BCVA (worse than 20/32) or mild grade of any type of cataract or moderate grade of any type of cataract with BCVA of better than 20/32, the DL algorithm agent automatically refers the patient to an ophthalmologist. If participants show severe or moderate grade of any type of cataract with decreased BCVA (worse than 20/32), the participants are informed that consultation with an ophthalmologist for consideration for cataract surgery will be recommended on the screening report (Fig 1).

Statistical Analysis

To evaluate our system, we presented the receiver operating characteristic curve and calculated the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy for the cases of slit-lamp and retroillumination images in 2 ways: (1) LOCS III-based grading, which is classified into 7 grades for NO and NC and into 6 grades for CO and PSC and (2) severity-based grading, which classifies all NO, NC, CO, and PSC into normal, mild, moderate, or severe. We used the Python scikit-learn package (https://scikit-learn.org/stable/_sources/about.rst.txt) to compute the performance metrics (AUC, sensitivity, specificity, and accuracy) and Python matplotlib package (<https://matplotlib.org/stable/users/license.html>) to generate plots. The performance metrics—true positive (TP), true negative (TN), false positive (FP), and false negative (FN)—were calculated using the following formulas: sensitivity = $TP / (TP + FN)$, specificity = $TN / (TN + FP)$, and Accuracy = $(TP + TN) / (TP + TN + FP + FN)$. Bootstrapping was used to estimate the 95% confidence intervals (CIs) of the performance metrics.

Visualization

Heatmaps are provided for visual explanations of predictions using gradient-weighted class activation mapping (Grad-CAM).^{10,11} This visualization technique, and Grad-CAM++, are designed for interpreting the output of black-box neural networks; the heatmaps visualize important regions in cataract images where the network is considered a critical aspect for grading. Supplemental Figure S4 shows the heatmaps from the fourth layer of classification networks; the semantically important regions for cataract classification are highlighted in the heatmaps.

Ablation Study

We conducted ablation studies to confirm that each component of the algorithm independently contributed to the success of the model. The 4 components of the algorithm were designed to resolve key problems in cataract image classification: (1) redundant information inside data, (2) dataset size, (3) dataset bias, and (4) class imbalance. Supplemental Figure S5 depicts the ablation experimental results, where each row represents grading criterion (NO, NC, CO, and PSC from top to bottom) and the column refers to the components of the algorithm (region detection, data augmentation and transfer learning, GCE loss, and CB loss from left to right). Therefore, 16 instances were tested by applying 4 components, 1 for each problem. Specifically, for each component, we obtained 8 instances in which the component was turned on or off. For brevity, we averaged the results of the 8 instances for each metric and grade and illustrated them using a bar plot. Because our focus is on the behaviors of each component of our system, we used ResNet18 for the ablation study on behalf of the algorithms.

To address point 1, we hypothesized that if the CNN is trained using the original images without any preprocessing, the network might reflect unnecessary visual information, consequently leading to lower classification ability. Therefore, we used the RDN to crop the image first, resulting in a cropped image that contains salient visual information only. To address point 2, we used data augmentation and transfer learning techniques to enhance the generalization ability of our algorithm (Supplemental Fig S6). To address points 3 and 4, we designed a CB and GCE loss by fusing the GCE and CB losses to train the classification network of our algorithm.

Results

Image Dataset Demographics

A total of 1335 slit-lamp photograph images and 637 retroillumination images were obtained from 596 patients (887 eyes). Among the 1335 slit-lamp and 637 retroillumination images, 918 (68.8%) and 435 (68.3%) images were in the training datasets, 152 (11.4%) and 71 (11.1%) images were in the validation datasets, and 265 (19.9%) and 131 (20.6%) images were in the test datasets (Table 1).

Classification Performance in the Test Dataset

The overall diagnostic performance showed an AUC of 0.9992 and 0.9994, sensitivity of 98.82% and 98.51%, specificity of 96.02% and 92.31%, and accuracies of 98.82% and 98.51% for the experiments on slit-lamp images (NO and NC), respectively. The system showed high diagnostic performance for the retroillumination images of CO

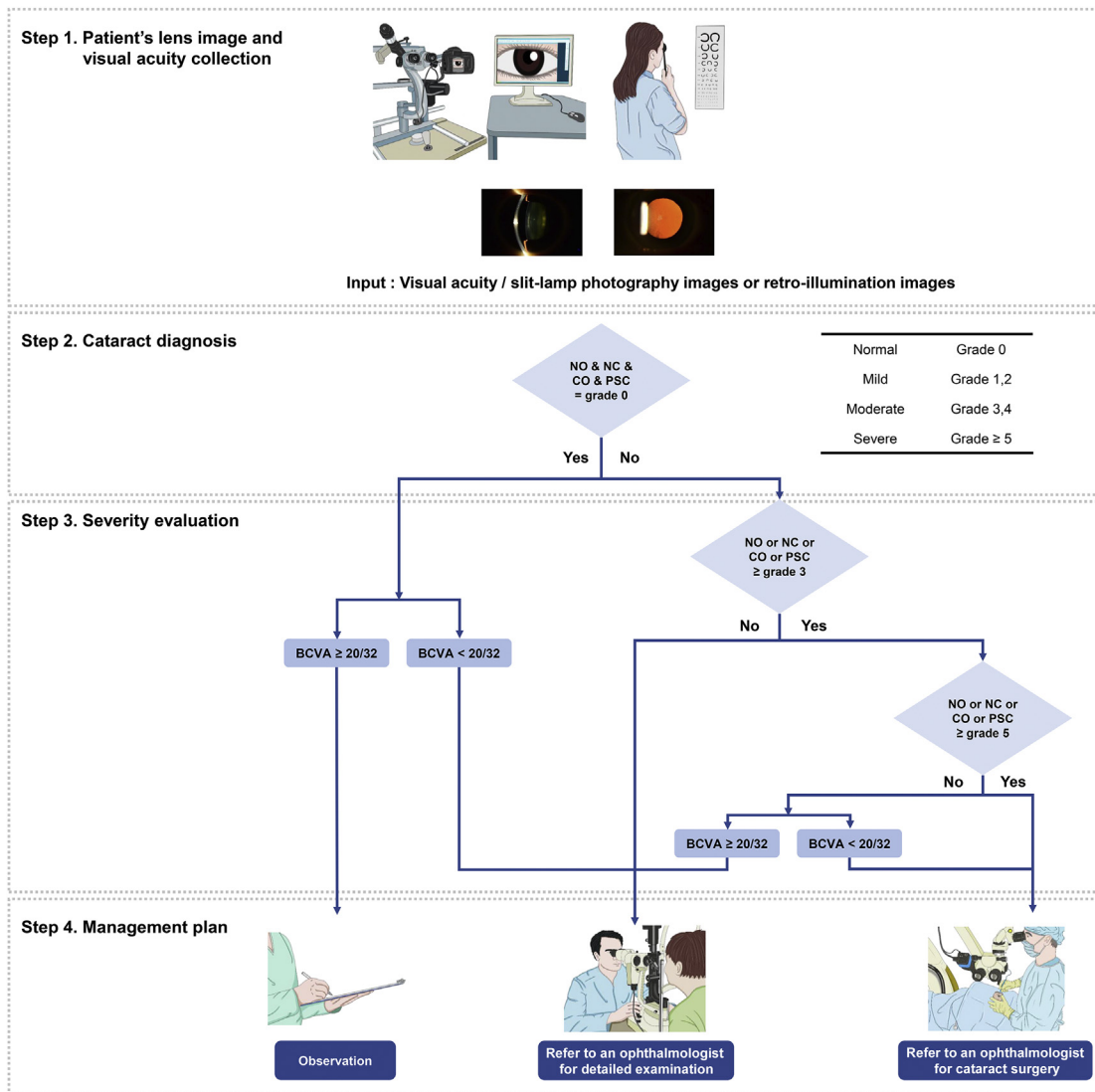


Figure 1. Logic flowchart for cataract diagnosis and management. The deep learning (DL) algorithm agent was designed to perform the following steps. In step 1, the patient's lens slit-lamp and retroillumination photograph images and visual acuity are collected. In step 2, the DL algorithm agent analyzes the lens images to determine whether they are normal or if any type of cataract is present. In step 3, the DL algorithm agent determines the patient's cataract severity based on grading by the network. The visual acuity of the subjects was not used in step 3 of the DL algorithm but rather in step 4, where the visual acuity was considered to suggest an optimal management plan for each subject. BCVA = best-corrected visual acuity; CO = cortical opacity; NC = nuclear color; NO = nuclear opalescence; PSC = posterior subcapsular opacity.

and PSC (AUC, 0.9680 and 0.9465; sensitivity, 96.94% and 92.13%; specificity, 96.78% and 89.36%; and accuracy, 96.21% and 92.17% for CO and PSC, respectively; [Table 2](#)). The model ensemble of the 3 AI algorithms enabled an increase in the diagnostic performance of the system. The detailed diagnostic performances of each algorithm are listed in [Supplemental Tables S2, S3, S4, and S5](#).

The system achieved excellent LOCS III-based grading prediction performance: AUC of 0.9567 and 0.9650; sensitivity of 81.71% and 75.69%; specificity of 93.10% and 93.71%; and accuracy of 91.22% and 90.26% for NO and NC, respectively. Additionally, for the experiments on retroillumination images (CO and PSC), the system

achieved good LOCS III-based grading prediction performance (AUC, 0.9044 and 0.9174; sensitivity, 78.02% and 71.49%; specificity, 94.05% and 91.64%; and accuracy, 91.33% and 87.89% for CO and PSC, respectively). [Figure 2](#) (top) and [Supplemental Tables S6, S7, S8, and S9](#) show the LOCS III-based grading prediction performance of the algorithm for NO, NC, CO, and PSC in detail.

Additionally, we confirmed that the system achieved excellent severity-based grading prediction performance in the 4-level classification (normal, mild, moderate, and severe): AUC of 0.9789 and 0.9842; sensitivity of 91.10% and 88.44%; specificity of 94.69% and 93.21%; and accuracy of 93.51% and 91.70% for NO and NC, respectively. Additionally, for experiments on retroillumination images (CO

Table 1. Summary of Training, Validation, and Test Datasets

Lens Type	Grade							Total
	0	1	2	3	4	5	6	
Training datasets: slit-Lamp (NO, NC) and retroillumination (CO, PSC) images								
NO	144	98	261	255	72	49	39	918
NC	140	123	226	194	113	76	46	918
CO	193	71	72	58	30	11	—	435
PSC	296	53	31	29	24	2	—	435
Validation datasets: slit-Lamp (NO, NC) and retroillumination (CO, PSC) images								
NO	24	16	43	43	12	8	6	152
NC	23	21	37	32	18	13	8	152
CO	32	12	11	9	5	2	—	71
PSC	49	9	4	4	4	1	—	71
Test datasets: slit-Lamp (NO, NC) and retroillumination (CO, PSC) images								
NO	41	29	65	77	35	13	5	265
NC	41	33	69	63	35	10	14	265
CO	53	16	22	23	14	3	—	131
PSC	83	17	12	11	7	1	—	131

CO = cortical opacity; NC = nuclear color; NO = nuclear opalescence; PSC = posterior subcapsular opacity; — = not available.

and PSC), the system achieved high performance (AUC, 0.9395 and 0.9102; sensitivity, 88.06% and 81.32%; specificity, 94.26% and 92.30%; and accuracy, 93.30% and 88.82% for CO and PSC, respectively). Figure 2 (bottom) and Supplemental Tables S10, S11, S12, and S13 show the severity-based grading prediction performance of the algorithm for grading cataract in detail.

Furthermore, we found that the model ensemble of the 3 AI algorithms is an effective technique to achieve higher performance. The ensemble model showed the best performance metrics for NO, NC, CO, and PSC compared with each of the 3 single networks in terms of accuracy (see Supplemental Tables S2–S13 for the summary statistics of each grade).

Results of Ablation Study

Regardless of the dataset and grading criteria, we found that all instances equipped with the algorithm components outperform those without them. That is, the network with RDN, data augmentation and transfer learning, and CB and GCE loss consistently improved model performance across all metrics. For example, all components of the systems enhanced performance on NO grading—AUC of 0.8458 and

0.9182; sensitivity of 46.68% and 71.23%; specificity of 86.36% and 89.77%; and accuracy of 79.36% and 85.70%—without and with RDN, respectively. Similar trends were observed when comparing the effects of other components (RDN, data augmentation and transfer learning, and CB and GCE loss) on other domains (NO, NC, and PSC). The detailed results are shown in Supplemental Figure S5.

Discussion

This study was designed to establish and validate an automated cataract diagnosis and grading system based on LOCS III using DL algorithms. The key factors that contributed to the system were 4 strategies in the AI domain: (1) RDN for redundant information within the data, (2) data augmentation and transfer learning for the small dataset size problem, (3) GCE loss for dataset bias, and (4) CB loss for a class imbalance problem. To demonstrate that the 4 components of our system independently contribute to the successful classification of cataract images, an ablation study was conducted, and all 4 components independently enhanced the performance, especially RDN, which

Table 2. Summary Statistics for the Diagnostic Performance of the Deep Learning System on Slit-Lamp Images (Nuclear Color and Nuclear Opalescence) and Retroillumination images (Cortical Opacity and Posterior Subcapsular Opacity) on the Test Dataset

	Area under the Receiver Operating Characteristic Curve	Sensitivity	Specificity	Accuracy
Slit-lamp image				
NO	0.9992 (0.9986–0.9998)	98.82% (97.69%–99.94%)	96.02% (91.36%–100.0%)	98.82% (97.69%–99.94%)
NC	0.9994 (0.9989–0.9998)	98.51% (97.39%–99.62%)	92.31% (86.27%–98.35%)	98.51% (97.39%–99.62%)
Retroillumination image				
CO	0.9680 (0.9579–0.9781)	96.94% (95.53%–98.35%)	96.78% (94.83%–98.74%)	96.21% (94.43%–97.99%)
PSC	0.9465 (0.9348–0.9582)	92.13% (88.65%–95.61%)	89.36% (84.44%–94.29%)	92.17% (88.56%–95.78%)

CO = cortical opacity; NC = nuclear color; NO = nuclear opalescence; PSC = posterior subcapsular opacity. Microaveraged area under the receiver operating characteristic curve, sensitivity, specificity, and accuracy are reported.

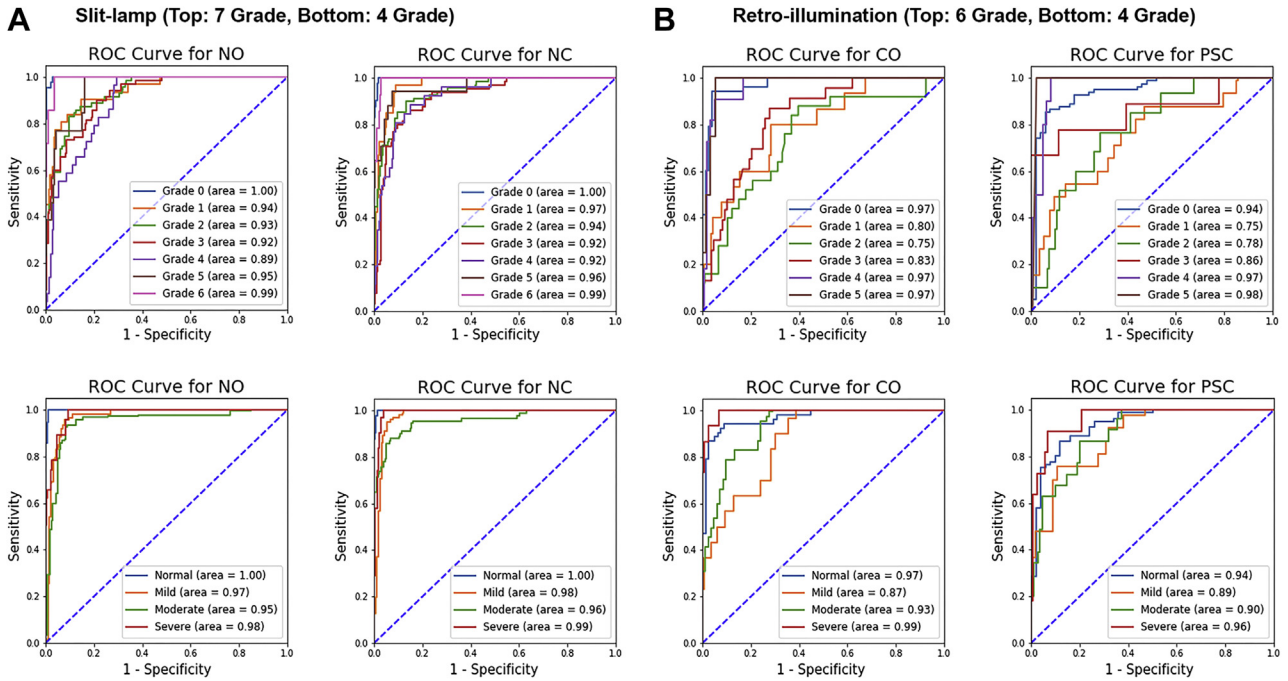


Figure 2. Receiver operating characteristic (ROC) curves and areas under the ROC curve for the grading prediction performance of the deep learning system. **A, B,** Seven class grades for nuclear opalescence (NO) and nuclear color (NC) and 6 class grades for cortical opacity (CO) and posterior subcapsular opacity (PSC) based on Lens Opacities Classification System III grading (top), and 4 class grades based on severity (bottom), evaluated on (A) slit-lamp and (B) retroillumination images on the test dataset.

contributed the most to the enhancement of the algorithm. Furthermore, because our algorithm was highly robust and data efficient under the images obtained from slit-lamp digital camera and augmentation technique conditions, the ensemble of 3 AI algorithms (ResNet18, WideResNet50-2, and ResNext50) showed outstanding performance in diagnosing and differentiating clinical cataract grading for all types of cataract.

In our study, the Grad-CAM and Grad-CAM++ results reflected the regions that most affected CNN model predictions of cataracts. In our CNN model, we inferred that the transition zone between dense and minimal cataracts may be the key region that most affected CNN prediction of cataracts. However, although our CNN model generated a heatmap on the transition zone, the heatmap in the trained CNN model with a large sample size could be different. Further studies with large sample sizes will be needed to elucidate the key regions in the CNN model of cataract grading.

For nuclear sclerotic cataracts, the diagnostic performance of NO and NC showed excellent results (AUC of 0.9992 and 0.9994 and accuracy of 98.82% and 98.51%, respectively). Performance on NO and NC 7-level classification showed excellent results in our study (AUC of 0.9567 and 0.9650 and accuracy of 91.22% and 90.26%, respectively). The AUC ranged from 0.89 to 1.00 for all grades, proving the effectiveness of the system. Previous studies reported that the accuracy of cataract diagnosis, primarily

nuclear cataract using AI agents, ranged from 88.8% to 93.3%.^{9,12–15} However, most studies reporting the outcomes of cataract detection used fundus photography as an imaging method for training AI algorithms. Using fundus photography images to diagnose and grade cataract is simple; however, the obtained fundus image is not a direct lens image. The blurriness of fundus images was used to assess cataract severity, which is not a standard classification in clinical practice. It could be influenced by any opacity along the visual axis, resulting in an incorrect diagnosis and grading of cataracts. In our study, the DL network was trained using various slit-lamp images used in a real clinical setting. These datasets are suitable for accurate evaluation of nuclear sclerotic cataracts because slit-lamp images reflect actual lens configurations.

However, 3 imaging techniques using the slit lamp, including slit-beam photography, broad-beam photography, and retroillumination, make it difficult to obtain uniform data, to process images, and to train AI agents. Thus, the performance of the previous AI algorithms was not of the highest quality. Xu et al¹⁶ and Gao et al¹⁷ reported that the accuracy of cataract grading with an AI agent using slit-lamp images as imaging methods was 69.0% and 70.7%, respectively. Recently, Wu et al⁹ reported the accuracy of the AI algorithm (ResNet) for robust diagnostic performance in 3-step tasks: capture mode recognition (AUC, > 0.99), cataract diagnosis (AUC, > 0.99), and detection of referable cataracts (AUC, > 0.91) using

slit-lamp and diffuse-beam photography based on large datasets. However, cataracts were divided into 3 classes—normal, mild, and severe—and were not graded by the standard LOCS III grading system. In our study, by applying the standard LOCS III image and adopting the 4 strategies, especially RDN with an ensemble of 3 AI algorithms (ResNet18, WideResNet50-2, and ResNext50), our results showed a more precise diagnosis and grading of cataracts, even in the 7-level classification of NO and NC with AUCs of 0.9567 and 0.9650, respectively.

The diagnostic performance of CO and PSC showed excellent results (AUC of 0.9680 and 0.9465 and accuracy of 96.21% and 92.17%, respectively). Performance on CO and PSC 6-level classification showed good results in our study (AUC of 0.9044 and 0.9174 and accuracy of 91.33% and 87.89%, respectively). Methods for detecting CO and PSC have evolved for many years. In 1991, Nidek developed a global thresholding method to detect opacity on retroillumination images.¹⁸ Since then, various improved methods such as contrast-based thresholding, local thresholding, texture analysis, watershed, and Markov random fields were introduced.^{19–23} Li et al²⁰ reported automatic detection and grading outcomes of CO with 85.6% accuracy. The researchers also reported 82.6% sensitivity and 80.0% specificity for automatic detection of PSC opacity.²¹ Gao et al²⁴ proposed the enhanced texture features method on 2 retroillumination lens images (anterior and posterior images) for CO and PSC grading and reported a good diagnostic performance (accuracy of 84.8%, sensitivity of 78.5%, and specificity of 87.8%) based on the Wisconsin grading system (3-class classification for CO and PSC, respectively). Recently, Zhang and Li²³ reported improved outcomes with 91.2% sensitivity and 90.1% specificity of PSC screening on retroillumination images using a combination of Markov random fields and mean gradient comparison methods based on the Wisconsin grading system. In contrast to previous studies, we adopted a 6-level image classification (grades 0–5) for CO and PSC severity evaluation based on the LOCS III and achieved good diagnostic and grading prediction performance. The performance of detecting and grading CO and PSC using the ensemble method was comparable with that of nuclear cataracts. Interestingly, the AUC of CO and PSC tended to be of a higher grade (grades 3, 4, and 5), which means that the DL system predicted CO and PSC grading more precisely when opacities covered large areas in retroillumination images. We postulate that Grad-CAM and Grad-CAM++ may regard key regions as noise and could not generate a heatmap correctly on early CO and PSC images. It is difficult to differentiate between early PSC and CO, even for trained ophthalmologists. Additionally, these results are consistent with previous reports. Gao et al²⁴ reported that the proposed computer-aided cataract detection with enhanced texture extraction method based on the Wisconsin grading system achieved good diagnostic performance for severe CO, PSC, or noncataract images. However, the diagnostic performance of early CO

and PSC was unsatisfactory. Zhang and Li²³ reported that the watershed and Markov random fields methods showed that the accuracy of PSC grading based on the Wisconsin grading system was 91.7% and 91.3% for grades 1 and 3, respectively, but the accuracy of grading for PSC grade 2 was 83.8% using the proposed method. Furthermore, previous studies did not report the AUC and sensitivity of the DL system in early CO and PSC. In our results, the accuracy of detecting and grading early CO and PSC was superior to that of previous reports. Considering data imbalance in medical data, the AUC may reflect the actual performance of DL algorithms, and the performance of DL systems on early CO and PSC previously reported might have been overestimated. Therefore, further studies must develop methods for detecting and discriminating early CO and PSC from noise on retroillumination images to improve the diagnostic and grading prediction performance.

The application of the slit-lamp and retroillumination image classification for automated cataract grading enables accurate objective evaluation of 3 types of cataracts with only lens images. Furthermore, the AI algorithm can detect referable cases by combining information on cataract diagnosis with visual acuity. The incidence of cataracts has increased in the elderly, most of whom live in medically vulnerable areas. This AI-based cataract diagnostic platform can be used in rural areas where ophthalmologists are scarce, and it can improve health care services in underdeveloped poverty areas.

This study has several limitations, the first of which stems from its relatively small dataset size compared with other studies. However, our DL algorithm achieved higher accuracy in classifying the 3 types of cataracts (nuclear sclerotic, cortical, and posterior subcapsular) for 6 to 7 grades with limited data. The second is the result of varying quality of lens images obtained from various ophthalmic examination instruments; external validation is required for use in factual clinical practice. In this study, our model showed robust performance under specific conditions, such as images obtained from the slit-lamp digital camera and augmentation technique. As a preliminary study, further studies with large sample sizes and external validation are planned to evaluate these AI algorithms.

We have proposed an ensemble DL algorithm for nuclear sclerotic cataract, CO, and PSC diagnosis and grading based on LOCS III. This DL-based system successfully yielded accurate and precise cataract detection and grading measurements. This preliminary research showed the potential for accurate detection and grading in 6- and 7-level classifications for all types of cataracts. The future direction of this study will be to provide more robust diagnostic and grading prediction performance and to evaluate the efficiency in factual clinical practice.

Acknowledgments

www.editage.co.kr

Footnotes and Disclosures

Originally received: October 2, 2021.

Final revision: February 23, 2022.

Accepted: March 14, 2022.

Available online: March 18, 2022. Manuscript no. XOPS-D-21-00177.

¹ Department of Ophthalmology, Sungkyunkwan University School of Medicine, Samsung Medical Center, Seoul, Republic of Korea.

² Graduate School of Artificial Intelligence, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea.

³ Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea.

⁴ Samsung Biomedical Research Institute, Seoul, Republic of Korea.

⁵ Department of Ophthalmology, Sungkyunkwan University School of Medicine, Kangbuk Samsung Hospital, Seoul, Korea.

⁶ Department of Digital Health, Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Seoul, Republic of Korea.

*Both authors contributed equally as first authors.

‡Both authors contributed equally as senior authors.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have no proprietary or commercial interest in any materials discussed in this article.

Supported by the National Research Foundation of Korea, funded by the Korean government's Ministry of Education, Seoul, Korea (grant no.: NRF-2021R1C1C1007795 [D.H.L.]).

HUMAN SUBJECTS: Human subjects were included in this study. The study was approved by the Samsung Medical Center Institutional Review Board and adhered to the tenets of the Declaration of Helsinki. All human participants gave written informed consent.

No animal subjects were included in this study.

Author Contributions:

Conception and design: Chung, Lim

Analysis and interpretation: Son, Ko, E.Kim, Lee, M.-J.Kim, Han, Shin, Chung, Lim

Data collection: Son, Ko

Obtained funding: N/A

Overall responsibility: Son, Ko, E.Kim, Chung, Lim

Abbreviations and Acronyms:

AI = artificial intelligence; **AUC** = area under the receiver operating characteristic curve; **BCVA** = best-corrected visual acuity; **CB** = class-balanced; **CI** = confidence interval; **CNN** = convolutional neural network; **CO** = cortical opacity; **DL** = deep learning; **FN** = false negative; **FP** = false positive; **GCE** = generalized cross-entropy; **Grad-CAM** = gradient-weighted class activation mapping; **LOCS** = Lens Opacities Classification System; **NC** = nuclear color; **NO** = nuclear opalescence; **PSC** = posterior subcapsular opacity; **RDN** = region detection network; **TN** = true negative; **TP** = true positive.

Keywords:

Artificial intelligence, Cataract, Deep learning, Lens Opacities Classification System III.

Correspondence:

Dong Hui Lim, MD, PhD, Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, South Korea. E-mail: ldhlse@gmail.com; and Tae-Young Chung, MD, PhD, Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, South Korea. E-mail: tychung@skku.edu.

References

1. Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health*. 2017;5(12):e1221–e1234.
2. Liu YC, Wilkins M, Kim T, et al. Cataracts. *Lancet*. 2017;390(10094):600–612.
3. Chylack Jr LT, Leske MC, Sperduto R, et al. Lens Opacities Classification System. *Arch Ophthalmol*. 1988;106(3):330–334.
4. Chylack Jr LT, Wolfe JK, Singer DM, et al. The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. *Arch Ophthalmol*. 1993;111(6):831–836.
5. Hall AB, Thompson JR, Deane JS, Rosenthal AR. LOCS III versus the Oxford Clinical Cataract Classification and Grading System for the assessment of nuclear, cortical and posterior subcapsular cataract. *Ophthalmic Epidemiol*. 1997;4(4):179–194.
6. Karbassi M, Khu PM, Singer DM, Chylack Jr LT. Evaluation of lens opacities classification system III applied at the slitlamp. *Optom Vis Sci*. 1993;70(11):923–928.
7. Tan AC, Loon SC, Choi H, Thean L. Lens Opacities Classification System III: cataract grading variability between junior and senior staff at a Singapore hospital. *J Cataract Refract Surg*. 2008;34(11):1948–1952.
8. McAlinden C, Gothwal VK, Khadka J, et al. A head-to-head comparison of 16 cataract surgery outcome questionnaires. *Ophthalmology*. 2011;118(12):2374–2381.
9. Wu X, Huang Y, Liu Z, et al. Universal artificial intelligence platform for collaborative management of cataracts. *Br J Ophthalmol*. 2019;103(11):1553–1560.
10. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). 2018:839–847.
11. Selvaraju RR, Cogswell M, Das A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization. *Proc IEEE Int Conf Comput Vis*. 2017:618–626.
12. Acharya RU, Yu W, Zhu K, et al. Identification of cataract and post-cataract surgery optical images using artificial intelligence techniques. *J Med Syst*. 2010;34(4):619–628.
13. Guo L, Yang J-J, Peng L, et al. A computer-aided healthcare system for cataract classification and grading based on fundus image analysis. *Comput Ind*. 2015;69:72–80.

14. Xiong L, Li H, Xu L. An approach to evaluate blurriness in retinal images with vitreous opacity for cataract diagnosis. *J Healthc Eng.* 2017;2017:1–16.
15. Yang JJ, Li J, Shen R, et al. Exploiting ensemble learning for automatic cataract detection and grading. *Comput Methods Programs Biomed.* 2016;124:45–57.
16. Xu Y, Gao X, Lin S, et al. Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression. *Med Image Comput Comput Assist Interv.* 2013;16(Pt 2):468–475.
17. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Trans Biomed Eng.* 2015;62(11):2693–2701.
18. Nidek Co. Ltd. *Anterior Eye Segment Analysis System: Eas-1000 Operator's Manual.* Tokyo: Nidek, Japan; 1991.
19. Gershenzon A, Robman LD. New software for lens retro-illumination digital image analysis. *Aust N Z J Ophthalmol.* 1999;27(3–4):170–172.
20. Li H, Ko L, Lim JH, et al. Image based diagnosis of cortical cataract. *Annu Int Conf IEEE Eng Med Biol Soc.* 2008;2008:3904–3907.
21. Li H, Lim JH, Liu J, et al. Automatic detection of posterior subcapsular cataract opacity for cataract screening. *Annu Int Conf IEEE Eng Med Biol Soc.* 2010;2010:5359–5362.
22. Chow YC, Gao X, Li H, et al. Automatic detection of cortical and PSC cataracts using texture and intensity analysis on retro-illumination lens images. *Annu Int Conf IEEE Eng Med Biol Soc.* 2011;2011:5044–5047.
23. Zhang W, Li H. Lens opacity detection for serious posterior subcapsular cataract. *Med Biol Eng Comput.* 2017;55(5):769–779.
24. Gao X, Li H, Lim JH, Wong TY. Computer-aided cataract detection using enhanced texture features on retro-illumination lens images. 2011 18th IEEE International Conference on Image Processing. 2011:1565–1568.