Accurate and Explainable Cataract Detection Using Eye Images Taken by Hand-held Slit-lamp Cameras

Daniel Kai Xiang Fung[1*], Di Wang[2*], Hao Wang[2], Yongwei Wang[2], Pengcheng Wu[2], Yan Yee Hah[3],
Chee Chew Yip[3], Wee Jin Heng[4], Tock Han Lim[4], Cyril Leung[2], Chunyan Miao[1,2**]

[1]School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore
[2]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly, NTU, Singapore
[3]Khoo Teck Puat Hospital (KTPH), Singapore   [4]Tan Tock Seng Hospital (TTSH), Singapore
*equal contribution   **corresponding author (ascymiao@ntu.edu.sg)

Abstract—Cataract is a leading cause of visual impairment in the elderly. With a greying population globally, there is a pressing need to improve the accessibility of cataract screening. Hand-held Slit-lamp Cameras (HSCs) are often preferred in community eye screening due to their great portability and accessibility. However, the image quality from HSCs is generally inferior to that from conventional bulky fundus cameras. In this paper, we extend the pre-trained ResNet-18 neural network to analyze a limited number of HSC images (n=187) for cataract detection. Model accuracy is improved through augmenting training data samples and complementing the visual features with patients' vision measurements. Explainability (of focal model areas for decision making) is attained via extracting the saliency maps using the Grad-CAM method. Our model achieves a high accuracy of 0.96, on par with state-of-the-art results reported in the literature. Our approach demonstrates the potential of large-scale community-based cataract detection using HSCs and our highly accurate and explainable AI-assisted model.

Index Terms—cataract detection, eye images, hand-held slit-lamp cameras, AI-assisted community health screening

I. INTRODUCTION

Cataract is the clouding of the human crystalline lens. It is the most prevalent and treatable cause of visual impairment and blindness around the globe [1]. Cataract is commonly caused by aging, disease, trauma, medication, and genetic predisposition [2]. Thus, it poses a great challenge to all aging populations including Singapore [3]. Furthermore, because the cataract progresses slowly and painlessly, patients are often unaware of its presence, e.g., most Singaporeans above 40 who have significant cataracts in either eye are unaware of their condition [4]. Generally, the patient is referred to an ophthalmologist when the cataract has progressed significantly. This delay in cataract diagnosis may result in diminished visual quality. Hence, accessible large-scale cataract screening is essential to detect a cataract in its early stage to stop or reverse the progression of visual impairment.
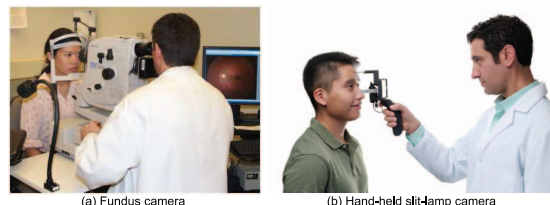
Fig. 1. Two types of cameras often used for cataract evaluation. (a) Fundus cameras (picture excerpted from [8]) are often used in clinical settings. (b) Hand-held slit-lamp cameras (picture excerpted from [9]) can be used in non-clinical settings due to their greater portability and accessibility.

In a typical clinical setting, an ophthalmologist often assesses the patients' optic nerve and disk using the fundus camera. When the ophthalmologist further checks the existence of cataract, he/she also evaluates the patients' visual acuity (e.g., Logarithm of the Minimum Angle of Resolution (LogMAR) Chart [5]) and contrast sensitivity (e.g., Pelli-Robson Contrast Sensitivity Chart [6]) using eye charts [2]. Expertise is needed to operate the fundus camera and to compare the patients' crystal images with the standard graded images such as the Lens Opacity Classification System III (LOCS III) [7].

Fundus cameras are expensive, bulky and requires expertise to operate (see Fig. 1(a)). A few prior studies were carried out to use Hand-held Slit-lamp Cameras (HSCs) as an alternative (see Fig. 1(b)) to capture images of the patients' eyes and subsequently apply various machine learning algorithms to evaluate cataract (see Section II). The use of HSC enables more healthcare providers to perform cataract screening in community settings. However, HSC image quality is generally much inferior to those taken by fundus cameras (see Fig. 2). Thus, the accuracy of cataract diagnosis using HSC images is generally lower than using fundus cameras (see Table IV).

To elevate the accuracy of cataract evaluation using HSC images towards large-scale screening in community settings, in this paper, we incorporate two additional procedures into the general framework, namely (i) data augmentation to alleviate the problem of having a small dataset and an imbalanced class distribution, and (ii) fusion of patients' vision measurements, i.e., visual acuity and contrast sensitivity, as the peripheral information with the deep learning computer vision model.
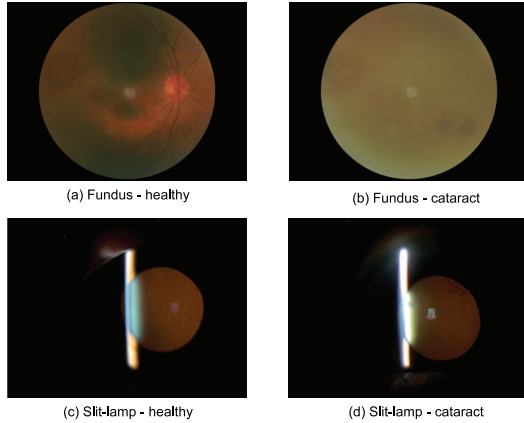
Fig. 2. The quality of the eye images taken by fundus cameras ((a)&(b), taken from [10]) is relatively much higher than those taken by slit-lamp cameras ((c)&(d), collected in this study). In general, the slit-lamp images contains the overexposure of the lamp glare, lack of details, are low in clarity and contrast, and have inconsistent position, periphery and size of the iris.

Specifically, we select the lightweight ResNet-18 [11] pre-trained on ImageNet in this work based on preliminary results. We conduct experiments using own-collected dataset comprising 187 eye images (quite small in size comparing to prior studies). The results show that our model achieves an accuracy of 0.96, which is on par with state-of-the-art results reported in the literature. To further validate the model, we extract saliency maps to visualize the areas of attention where the model focuses on while deriving the prediction. The assessment on these saliency maps provide much transparency to the model's dynamics, which is essential to encourage healthcare providers to adopt our AI-assisted cataract detection model for large-scale screening in community settings.

II. RELATED WORK

Because the images taken by fundus cameras are of high quality (see Figs. 2(a) and 2(b)), many pioneer AI-assisted cataract diagnosis models were developed using fundus images. Especially after the release of the Ocular Disease Intelligent Recognition (ODIR) dataset in 2019 [10], which comprises 5,000 patients' information and their fundus images of both eyes (other than cataract, disease labels such as diabetes, glaucoma, hypertension, etc., are also provided), a series of studies were conducted using ODIR. Among which, Sudarsono et al. [12] fine-tuned AlexNet with the diffGrad optimizer to achieve an accuracy of 0.975 for cataract detection. Hasan et al. [13] used the pre-trained InceptionResnetV2 to achieve an even higher accuracy of 0.982. These studies showcased the high efficacy of the AI-assisted cataract diagnosis models trained using fundus images.

Nonetheless, as afore-introduced, it is impractical to use bulky fundus cameras for cataract detection in non-clinical settings. Therefore, HSCs were also often used in cataract diagnosis studies. Liu et al. [14] constructed their own CNN model to perform cataract detection and grading. It is worth noting that they focused on pediatric cataracts while all the other studies mentioned in this paper focus on cataracts in adults. Their model achieved an accuracy of 0.971 for cataract detection. Xu et al. [15] first applied Faster R-CNN to locate the nuclear region and subsequently applied ResNet-101 for cataract grading. Their model achieved an accuracy of 0.847 for 5-class cataract grading.

Instead of using commercial-off-the-shelf (COTS) HSCs, a group of researchers attempted to make the hand-held devices more portable and accessible by developing prototypes with COTS smartphones integrated. Zhang et al. [16] used Inception-v3 for cataract detection and achieved an accuracy of 0.948. It is worth noting that they also collected a dataset using COTS HSCs and achieved a higher accuracy of 0.959. Hu et al. [17] employed a combination of YOLOv3, ShuffleNet and SVM to grade cataracts and achieved an accuracy of 0.935 for three classes. Askarian et al. [18] extracted luminance and RGB features to train an SVM, achieving an accuracy of 0.966 for cataract detection. Different from all the other studies mentioned in this paper, Askarian et al. [18] used images of eye models rather than human eyes.

As aforementioned, due to the inferior quality of HSC images, models trained using these images generally achieve a lower accuracy than those trained using fundus images (see Table IV). Nonetheless, we still focus on HSC images in this paper because they represent the real-world quality of eye images acquired in most non-clinical settings.

Surprisingly, these prior studies on cataract diagnosis *did not apply any explainable AI (XAI) techniques to visualize or explain the model dynamics*. We deem such explanations as critically necessary to provide 'assurance' to the medical professionals. Hence, we analyse the explainability of our model (see Section IV-D).

III. METHODOLOGY

Same as a number of prior studies [12]–[14], [16], [18], the key objective of this work is to develop an AI-assisted cataract detection model that is capable of determining whether a given eye image has cataract or not. Nonetheless, comparing to prior studies, our model has the following **three innovative aspects**, all **novel** in the field of image-based cataract detection:

1) Augmentation of limited data samples (from our infrared-based HSCs). This alleviates the problems of small data size and imbalanced data distribution in the dataset.

2) Feature complement with patient's vision measurements, namely visual acuity measured by the LogMar Chart [5]) and contrast sensitivity measured by the Pelli-Robson Contrast Sensitivity Chart [6], respectively.

3) Visualisation of the model's decision process by extracting the saliency maps using the Grad-CAM method [19].

We deem these three innovative aspects **should be adopted by subsequent studies** in this field. Specifically, data augmentation should be applied because this process would save cost in data collection; vision measurements including visual acuity and contrast sensitivity should be used because they are always collected during a standard cataract screening process;
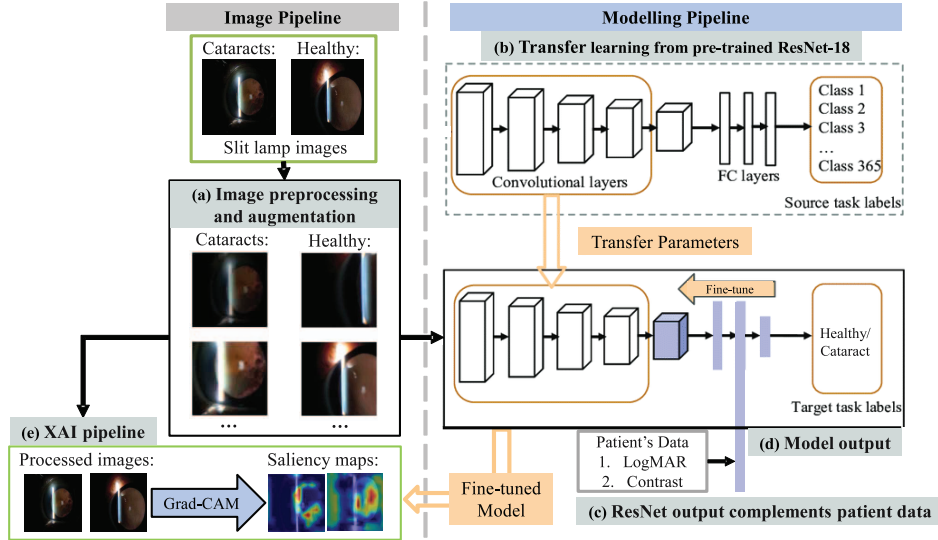
Fig. 3. The architecture of our cataract detection model.

and the model's dynamics should be visualized to assess the rightfulness of the decisions made. In summary, adopting these three procedures **will not lead to any increase in the amount of effort** required for data collection, while at the same time, they would **help to elevate the model performance and provide the necessary transparency of the model**.

The architecture of our cataract detection model is shown in Fig. 3. As shown, there are five key procedures in our model, namely (a) image preprocessing and data augmentation, (b) parameter adoption of ResNet-18 pre-trained on ImageNet, (c) complements with patient's vision measurements, (d) model fine-tuning, and (e) visualization of the decision process. We introduce these five key procedures in the following subsections, respectively.

A. Image Preprocessing and Data Augmentation

We preprocess all the available HSC images as follows:

1) The white space on the image border, if any, is cropped.

2) The contrast of each cropped image is enhanced by applying the Contrast-Limited Adaptive Histogram Equalization (CLAHE) [20] technique.

3) The RGB pixel values for all images are normalized across the dataset towards zero-mean. This step is usually applied to improve the deep learning model's convergence.

Thereafter, we generate synthetic samples to deal with the following two common problems exist in this research field:

Problem 1: The sample size of the collected images is generally small, which may lead to model underfitting.

Problem 2: The number of healthy eye samples (without cataract) is generally much smaller than that of cataract samples, because it is exceptionally uncommon for people with healthy vision have their eye(s) imaged. Nonetheless, this imbalanced class distribution may pose challenges to the classification model.

Specifically, we apply random scaling, random cropping, random rotation, and random brightness scaling to generate synthetic samples based on the original ones. In addition, we purposely apply another round of random rotation (ranging in [-180°, +180°]) to all images in the training dataset. This is to let the model learn that the visual artifacts (glares and their reflections) are irrelevant to the presence of cataracts. In the end, all images are center cropped to the size of 224×224 pixels to fit with the input size of ResNet-18 [11].

B. Adoption of Pre-trained ResNet-18

Because the size of the HSC image datasets is normally small, we choose not to train a deep learning model from scratch. In this work, we select ResNet-18 [11] (see Fig. 4) pre-trained on the ImageNet dataset. As aforementioned, this selection is based on preliminary results that ResNet-18 achieves competitive performance comparing to larger models in the ResNet family. In addition, we deem the small model size will lead to a faster processing speed in future real-world deployments for cataract screening.

C. Complements with Patient's Vision Measurements

Patient's visual acuity and contrast sensitivity are always measured as part of the cataract detection and grading procedures. Such measurements can be easily and are often conducted in non-clinical settings, such as in the spectacles shops. In our work, we complement the HSC images with such vision measurements as the model's inputs, aiming to elevate the model performance by utilizing the highly relevant, existing information.

Specifically, we extend the Fully Connected (FC) layer in the original ResNet-18 (see Fig. 4) with two FC layers (see Fig. 3). The inputs and outputs of the two FC layers are described in Table I. Specifically, the patients' visual acuity
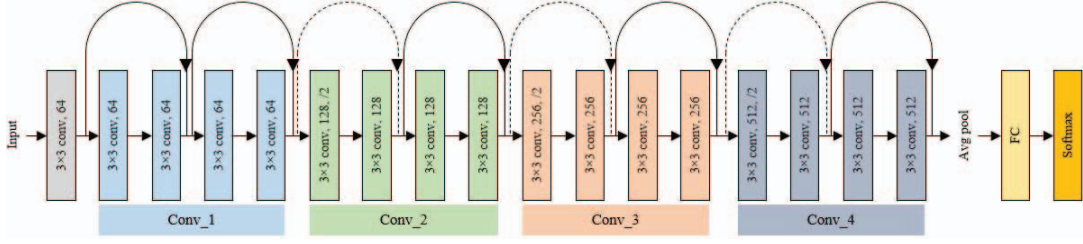
Fig. 4. The architecture of ResNet-18. As shown, multiple convolutional layers are grouped into convolutional blocks: Conv_1, Conv_2 Conv_3 and Conv_4.

TABLE I
THE TWO FC LAYERS IN OUR MODEL

Layer	Input(s)	Shape	Output
FC_1	Latent visual features from ResNet-18	1×128	1×128
	Patient's visual acuity score	1×128	
	Patient's contrast sensitivity score	1×128	
FC_2	Output of FC_1	1×128	1×2

TABLE II
DATASETS USED IN OUR EXPERIMENTS

Dataset	Image Type	Train	Test	Total
Baseline	Healthy	14	5	19
	Cataract	163	5	168
	Total	177	10	187
Balanced-small	Healthy	168	5	173
	Cataract	163	5	168
	Total	331	10	341
Balanced-large	Healthy	490	5	495
	Cataract	489	5	494
	Total	979	10	989

(measured by LogMAR [5]) and contrast sensitivity (measured by Pelli-Robson Contrast Sensitivity [6]) are mapped to vectors with the size of 1×128, respectively. Such mapping is learned (with batch normalization) by the model (see the following subsection).

It is worth mentioning that for the augmented images, the associated vision measurements are adopted from the corresponding original image without any random value alteration.

D. Model Fine-tuning

Because the key objective of our model is to determine whether cataract exists in the input HSC image with the complement of patient's vision measurements, the model essentially performs binary classifications that 0 represents the absence of cataract and 1 represents presence. Therefore, we use the following cross-entropy loss ($J(\theta)$) to train all learnable parameters:

$$J(\theta) = -\frac{1}{N}[\sum^{N} y\log(g(z)) + (1-y)\log(1-g(z))], \quad (1)$$

where θ denotes the set of all learnable parameters, N denotes the total number of data samples used for training, y denotes the ground-truth label, z denotes the model inputs, and $g(z)$ denotes the model output. To speed up the learning process, we optimize θ using the stochastic gradient descent algorithm.

E. Saliency Maps for Model Dynamics Visualization

To provide transparency on how our model detects the presence of cataract in an input image, we apply the Grad-CAM algorithm [19]. Specifically, for each input HSC image, the gradients flowing into the final convolutional layer of each convolutional block (see Fig. 4) are used to produce a coarse saliency map to visualize and examine the attention of the model (see Section IV-D).

IV. EXPERIMENTS

In this section, we first introduce the own-collected dataset. We then describe how we define the various model configurations and set up the experiments. After presenting and discussing the experimental results, we visualize the saliency maps extracted from both healthy and cataract images.

A. Own-collected HSC Image Dataset

The study data were collected from Khoo Teck Puat Hospital (KTPH) and Tan Tock Seng Hospital (TTSH) in Singapore over the period of Q2 2022 to Q2 2023. After removing samples with missing patient data, a total number of 187 samples comprising both the HSC images and patient vision measurements were collated. This dataset is highly imbalanced where only 19 samples (10.16%) are healthy, while the other 168 samples all have cataracts in varying grades.

B. Model Configurations and Experimental Set-ups

Because the key objective of this study is to develop an AI-assisted cataract detection model for future deployment in community settings, we focus on the binary classification problem to determine whether a sample has the presence of cataract or not. Nonetheless, to alleviate the problem of having a highly imbalanced class distribution, we generate two augmented datasets following the steps introduced in Section III-A. As shown in Table II, the Baseline dataset only comprises the original 187 samples. For the Balanced-small dataset, we only augment the healthy class to make the class distribution (almost) equal, while for the Balanced-large dataset, we augment both classes to have a larger dataset. It is worth mentioning that the augmented samples are generated independently for the latter two datasets.

TABLE III
PERFORMANCE OF OUR CATARACT DETECTION MODEL

Configuration	Dataset	Accuracy	F1-score	Sensitivity	Specificity
image-only	Baseline	0.740 ± 0.058	0.795 ± 0.041	0.960 ± 0.036	0.520 ± 0.122
	Balanced-small	0.890 ± 0.035	0.892 ± 0.032	0.900 ± 0.053	0.880 ± 0.077
	Balanced-large	0.925 ± 0.042	0.928 ± 0.040	0.940 ± 0.050	0.910 ± 0.072
image+metadata	Baseline	0.845 ± 0.048	0.863 ± 0.040	0.950 ± 0.056	0.740 ± 0.103
	Balanced-small	0.905 ± 0.039	0.905 ± 0.037	0.900 ± 0.053	0.910 ± 0.072
	Balanced-large	**0.960 ± 0.039**	**0.959 ± 0.041**	**0.960 ± 0.046**	**0.960 ± 0.036**

TABLE IV
PERFORMANCE COMPARISONS WITH PRIOR STUDIES (MODELS RANKED BY ACCURACY)

Study	Year	Image type	Sample size	Model	Accuracy	Remarks
Hasan et al. [13]	2021	fundus	1,088	InceptionResNetV2	0.982	Using images selected from ODIR [10]
Sudarsono et al. [12]	2020	fundus	200	AlexNet	0.975	Using images selected from ODIR [10]
Liu et al. [14]	2017	HSC	886	CNN	0.971	Focused on pediatric cataract
Askarian et al. [18]	2021	HSC	100	SVM	0.966	Using own HSC prototype and eye models; Using extracted luminance and RGB features
Ours	2024	HSC	187	ResNet-18	0.960	Image augmentation + vision measurements
Zhang et al. [16]	2020	HSC	2,516	Inception-v3	0.948	Using own HSC prototype

The size of the Baseline dataset is small, especially because it only has 19 healthy samples. Therefore, instead of splitting the samples with a consistent ratio into the train and test subsets, we maintain a consistent size of the test subset. Specifically, for all experiments, we always keep five samples from each class in the test subset while use all the other samples for training. It is worth highlighting that the samples in the test subset are all the original ones without augmentation and for all samples selected in the test subset, their correspondingly augmented samples, if any, are excluded from the train subset. These precautions are implemented to ensure there is no data leakage between the train and test subsets.

To remove the random effects, for each experiment setting, we report the averaged performance with standard deviations (std) across 20 independent runs.

In addition, to evaluate the effectiveness of having the complements of patient's vision measurements, we conduct an ablation study to examine the performance of our model without taking in the vision measurements as inputs. We refer to the former configuration as image+metadata and the latter as image-only. Specifically, for the image-only configuration, instead of the two FC layers (see Table I), we adopt the original single FC layer of ResNet-18 (see Fig. 4).

C. Experimental Results

The experimental results are presented in Table III. It is not surprising to see that the image+metadata model trained on the Balanced-large dataset achieves the best performance across all performance metrics. Moreover, the image+metadata model is shown as outperforming the image-only model, showcasing the effectiveness and importance of the complement of patient's vision measurements. Furthermore, it is observed that with the increase of the dataset size, the performance of both models generally improves, showcasing the effectiveness and importance of image augmentation. Specifically, with the

augmentation of more and more healthy samples, both models' specificity keeps improving.

Because all HSC image datasets used and the models constructed in prior studies were not made publicly available, it is not viable to make fair comparisons using the same dataset. Alternatively, we compare the performance of our model against the prior studies' results on cataract detection (binary classification on the presence of cataract) simply using their reported model performance while listing the contextual information in Table IV. Although our model comes in second last in terms of accuracy among the six compared in Table IV, its performance is still competitive especially among the studies using HSC images. The top-two models [13], [12] used high-quality fundus images (see Fig. 2) that let them outperform the rest. The third model [14] used a larger dataset with a balanced class distribution (476 healthy and 410 cataract). The fourth model [18] used eye models to take high-quality images with an equal class distribution. As such, considering all these factors that greatly affect the model performance, we deem ours achieves the same level of high performance as these state-of-the-art prior studies.

D. Using Saliency Maps to Visualize Attention

Using the best-performing model trained on the Balanced-large dataset, we extract saliency maps of all the original images for visualization of the model's attention (see Section IV-A). In Fig. 5, we present the saliency maps extracted from two randomly selected images: one with cataract (patient ID: A031, right eye) and one healthy sample (patient ID: K009, right eye). These two sets of saliency maps well demonstrate the model's attention at different depth. Specifically, at the relatively shallower layers (i.e., Conv_1 and Conv_2, see Fig. 4), the model pays more attention to the specific and scattered regions, e.g., the small regions brightened by the HSC and around the peripheral; while at the deeper layers
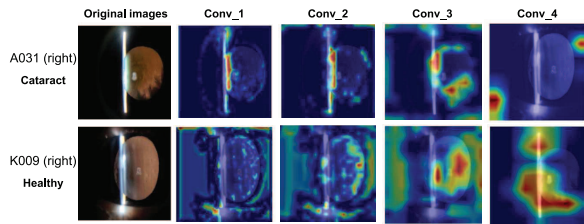
Fig. 5. Saliency maps extracted from two randomly selected images. The **red** regions indicate where the model's attention lies the most, the **black** regions indicate where the model's attention lies the least, while the other colors interpolate the level of attention in between.
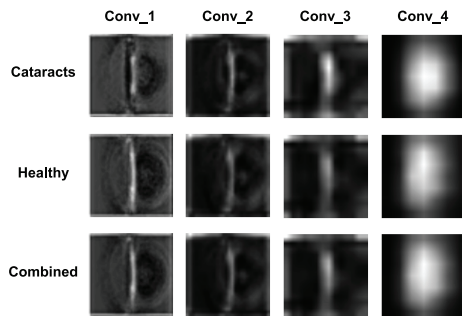

Fig. 6. The saliency maps averaged from all the cataract, healthy and combined (both classes) images in the original dataset, respectively. These saliency maps are rescaled to elevate the contrast for better visual assessment on the areas with the most attention.

(i.e., Conv_3 and Conv_4), the model pays more attention to the abstract and connected regions, e.g., the large regions around the centerline of the iris. Such observations are consistent across most saliency maps that cataracts are generally identified at either Conv_3 or Conv_4 (for A031, the most noticeable part of the cataract is accurately identified in the bottom-right quadrant at Conv_3), while for healthy eyes, the attention is generally spread out at Conv_3 and Conv_4 (see Conv_4 of K009).

Furthermore, to provide a visualized overview of all the extracted saliency maps, we average them at each convolutional block according to the images' class labels and present the averaged results in Fig. 6. As shown, there is little difference between the three sets of averaged saliency maps, suggesting our model has a consistent focus. The HSC glare and the outline of the iris are clearly visible at Conv_1, while they get blurred at Conv_2 and Conv_3. The area of attention at Conv_4 is enlarged and centers around the centerline of the iris with no specific focus.

Observations from both Figs. 5 and 6 suggest that our cataract detection model successfully emulates how an ophthalmologist checks the various regions of an eye step-by-step to arrive at the final decision.

V. Conclusion

This study shows that our AI-assisted cataract detection model using HSC images has high accuracy and explainability (for transparency of the decision process). It may potentially be deployed for large-scale cataract detection in communities.

Going forward, we plan to 1) develop own hand-held cameras to capture high-quality images and 2) automate the screening procedures towards real-world deployments.

References

[1] D. Lam, S. K. Rao, V. Ratra, Y. Liu, P. Mitchell, J. King, M.-J. Tassignon, J. Jonas, C. P. Pang, and D. F. Chang, "Cataract," *Nature Reviews Disease Primers*, vol. 1, p. (article no.) 15014, 2015.

[2] J. Thompson and N. Lakhani, "Cataracts," *Primary Care: Clinics in Office Practice*, vol. 42, pp. 409–423, 2015.

[3] K. W. Wong, W. L. Wong, and H. Feng, "Demographic profile of seniors in Singapore," *Singapore Department of Statistics*, vol. 1, pp. 6–9, 2022.

[4] J. Chua, B. Lim, E. K. Fenwick, A. T. L. Gan, A. G. Tan, E. Lamoureux, P. Mitchell, J. J. Wang, T. Y. Wong, and C.-Y. Cheng, "Prevalence, risk factors, and impact of undiagnosed visually significant cataract: The Singapore epidemiology of eye diseases study," *PLoS One*, vol. 12, no. 1, p. (article no.) e0170804, 2017.

[5] I. L. Bailey and J. E. Lovie-Kitchin, "Visual acuity testing. From the laboratory to the clinic," *Vision Research*, vol. 90, pp. 2–9, 2013.

[6] K. Kaur and B. Gurnani, "Contrast sensitivity," In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2023.

[7] L. T. Chylack Jr, J. K. Wolfe, D. M. Singer, M. C. Leske, M. A. Bullimore, I. L. Bailey, J. Friend, D. McCarthy, and S.-Y. Wu, "The Lens Opacities Classification System III," *Archives of Ophthalmology*, vol. 111, no. 6, pp. 831–836, 1993.

[8] Faculty of Medicine, The University of British Columbia, Canada, "Color fundus photography," https://ophthalmology.med.ubc.ca/patient-care/ophthalmic-photography/color-fundus-photography/.

[9] remidio, "Portable slit Lamp: PSL D20," https://www.remidio.com/products/psl/.

[10] "Peking University International Competition onOcular Disease Intelligent Recognition (ODIR-2019)," https://odir2019.grand-challenge.org/.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[12] E. Sudarsono, A. Bustamam, and P. P. Tampubolon, "An optimized convolutional neural network using diffgrad for cataract image classification," in *ICSAS*, 2020, p. (article no.) 020090.

[13] K. Hasan, T. Tanha, R. Amin, O. Faruk, M. M. Khan, S. Aljahdali, and M. Masud, "Cataract disease detection by using transfer learning-based intelligent methods," *Computational and Mathematical Methods in Medicine*, p. (article no.) 7666365, 2021.

[14] X. Liu, J. Jiang, K. Zhang, E. Long, J. Cui, M. Zhu, Y. An, J. Zhang, Z. Liu, Z. Lin, X. Li, J. Chen, Q. Cao, J. Li, X. Wu, D. Wang, and H. Lin, "Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network," *PLoS ONE*, p. (article no.) 0168606, 2017.

[15] C. Xu, X. Zhu, W. He, Y. Lu, X. He, Z. Shang, J. Wu, K. Zhang, Y. Zhang, X. Rong, Z. Zhao, L. Cai, D. Ding, and X. Li, "Fully deep learning for slit-lamp photo based nuclear cataract grading," in *MICCAI*, 2019, pp. 513–521.

[16] Z. Zhang, W. Shou, D. Xing, W. Ma, Q. Xu, W. Wang, L.-Q. Xu, Z. Liu, and L. Xu, "Automated cataracts screening from slit-lamp images employing deep learning," in *IoTaaS*, 2020, pp. 282–291.

[17] S. Hu, X. Wang, H. Wu, X. Luan, P. Qi, Y. Lin, X. He, and W. He, "Unified diagnosis framework for automatednuclear cataract grading based onsmartphone slit-lamp images," *IEEE Access*, vol. 8, pp. 174 169–174 178, 2020.

[18] B. Askarian, P. Ho, and J. W. Chong, "Detecting cataract using smartphones," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, p. (article no.) 3800110, 2021.

[19] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *ICCV*, 2017, pp. 618–626.

[20] O. Deperlioglu, U. Kose, D. Gupta, A. Khanna, F. Giampaolo, and G. Fortino, "Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation," *Future Generation Computer Systems*, vol. 129, pp. 152–169, 2022.