

ARTIFICIAL INTELLIGENCE

Development and validation of a pixel wise deep learning model to detect cataract on swept-source optical coherence tomography images

Pierre Zéboulon*, Christophe Panthier, Hélène Rouger, Jacques Bijon, Wassim Ghazal, Damien Gatinel

Department of Ophthalmology, Rothschild Foundation, 25 Rue Manin, Paris 75019, France

KEYWORDS

Cataract;
Deep learning;
Artificial intelligence;
Optical coherence tomography

Abstract

Purpose: The diagnosis of cataract is mostly clinical and there is a lack of objective and specific tool to detect and grade it automatically. The goal of this study was to develop and validate a deep learning model to detect and localize cataract on Swept Source Optical Coherence Tomography (SS-OCT) images.

Methods: We trained a convolutional network to detect cataract at the pixel level from 504 SS-OCT images of clear lens and cataract patients. The model was then validated on 1326 different images of 114 patients. The output of the model is a map representing the probability of cataract for each pixel of the image. We calculated the Cataract Fraction (CF), defined as the number of pixel classified as “cataract” divided by the number of pixel representing the lens for each image. Receiver Operating Characteristic Curves were plotted. Area Under the Curve (ROC AUC) sensitivity and specificity to detect cataract were calculated.

Results: In the validation set, mean CF was 0.024 ± 0.077 and 0.479 ± 0.230 ($p < 0.001$). ROC AUC was 0.98 with an optimal CF threshold of 0.14. Using that threshold, sensitivity and specificity to detect cataract were 94.4% and 94.7%, respectively.

Conclusion: We developed an automatic detection tool for cataract on SS-OCT images. Probability maps of cataract on the images provide an additional tool to help the physician in its diagnosis and surgical planning.

© 2022 Spanish General Council of Optometry. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Cataract surgery is one of the most performed surgeries globally annually. Its diagnosis is still mostly clinical today based on slit-lamp examination using the Lens Opacities

Classification System III¹ and visual acuity. Objective grading could be of great use in clinical practice. Indeed, slit lamp assessment and quantification of the lens opacity is subjective and lacks reproducibility. Visual acuity is often a poor measurement of optical quality and does not fully describe the impact of cataract on the patient’s vision.

In many cases, subtle cataracts can produce a significant visual impairment despite relatively preserved visual acuity.

* Corresponding author.

E-mail address: pzeboulon@for.paris (P. Zéboulon).

This dissociation between the clinical aspect of the lens and the patient's symptoms might postpone the surgery or might trigger more tests in search of other causes of decreased vision. Automatic cataract detection would also be interesting in remote settings with limited access to medical care to screen patients and refer them when surgery is needed. Very few objective and automatic tools exist to detect and grade cataract severity. Scheimpflug images have been used to measure the nucleus density.² Despite interesting results, it does not quantify the anterior cortex or posterior subcapsular cataracts, which are often responsible for higher visual discomfort. Double pass aberrometry is an interesting technology that measures light scattering in the ocular media. It produces a sensitive and quantitative measurement well correlated with visual acuity and quality of vision.^{3,4} However, it is incapable of discriminating between scattering due to the lens or to other media such as the cornea or the vitreous. It also requires the spherocylindrical correction of the patient and is unusable in case of high myopia. Several studies have tackled the problem of cataract detection using Deep Learning (DL), either using slit lamp⁵ or fundus photographs⁶ with promising results. These approaches are, however, potentially flawed by the imaging technique used themselves. Indeed, slit lamp photograph requires some training and is operator-dependent. Fundus photographs can be affected by any ocular media opacity, which could produce false-positive cataract cases. Recently, Swept-Source Optical Coherence Tomography (SS-OCT) has been used to quantify cataract severity using the lens average pixel density.⁷⁻⁹ Given its longer wavelengths, SS-OCT is reproducible and is less affected by corneal opacities than other technologies. Recent SS-OCT devices produce high-resolution lens images with an unmatched level of detail. We recently showed how DL could detect corneal edema at the pixel level on OCT images.^{10,11} This study aims to develop and validate a DL model to detect cataract at the pixel level on SS-OCT images following a somewhat similar methodology.

Methods

This retrospective study was conducted at the Rothschild Foundation Hospital and was authorized by our Institutional Review Board. It agrees with the tenets of the declaration of Helsinki. Informed consent was obtained from all patients.

Patients and images

All images were exported from the Anterior (Heidelberg, Heidelberg, Germany) SS-OCT. The Metrics app was used, which produces six radial scans of length 16 mm. The image resolution was 2150×1824 pixels. Axial and transverse resolutions are under $10 \mu\text{m}$ and $30 \mu\text{m}$, respectively.

Included patients belonged to either one of the following clinical categories: clear lens or cataract. The presence of any corneal disease was an exclusion criterion. Clear lens patients were preoperative of refractive surgery and had a Best Corrected Visual Acuity (BCVA) of at least 20/20. Cataract patients had a clinical cataract diagnosis responsible for significant visual discomfort and were scheduled for surgery. All kind of cataracts were

included, and four experienced cataract and refractive surgeons of the department performed the patient inclusions (PZ, CP, WG and DG).

The first set of patients constituted the development set which was randomly split into a training set and a test set with an 80%/20% ratio at the patient level, not at the image level. Both eyes of each patient were included in the same set. The development set was used to train the model and test its performance during training. All images of the development set were manually segmented. All lens pixels of a given image were labeled "Normal" for the clear lens patients or "Cataract" for the cataract patients. All other pixels were labeled as "Background".

A separate set of patients constituted the validation set which was used to validate the model on unseen data.

For each patient, we also collected the age and logMAR BCVA. Clear lens patients' BCVA was reported as 20/20 even if it was measured higher as higher acuity values might not have been tested for all patients.

Preprocessing

All images were cropped to a 1100×1100 pixel square around the lens using fixed coordinates for each image. The resulting cropped image was then rescaled to 0.5 times its original size for memory purposes. No other preprocessing technique was applied.

Deep learning methodology

A U-Net^{12,13} model was trained with a Stochastic Gradient Descent optimizer, a fixed learning rate of 0.001, and a cross-entropy loss function. Data augmentation was performed for each image and epoch with a random rotation between -15° and 15° , a random horizontal flip.

During training, the model's performance was evaluated with individual Dice coefficients for each pixel class. The training was stopped as soon as both the training and the test set reached satisfactory Dice values to prevent overfitting.

Model training was performed using Python 3.7 and PyTorch library.

Metrics and statistics

"The Cataract Fraction (CF) defined as the number of pixels classified as 'Cataract' 'Cataract' divided by the total number of pixels representing the lens was calculated for each image. CF was averaged over all radial scans available for each eye.

The comparability of the development set and the validation set regarding age and logMAR BCVA was tested using Mann Whitney U test as data were not normally distributed according to the 'D'Agostino Pearson test. Mean CF was compared between "Normal" and "Cataract" groups using Mann Whitney U test for the same reason. Receiver Operating Characteristic Curves were plotted and Area Under the Curve (ROC AUC), optimal CF threshold, sensitivity, and specificity to detect cataract were calculated for each set.

The pixel-wise output probabilities of the network for the "Cataract" class are represented color coded overlaid on the original uncropped images for several examples. Hotter colors mean a higher probability of cataract.

A p value below 0.05 was considered as significant.

Statistic tests were performed using Python 3.7 and SciPy libraries.

Figures were created using Matplotlib library.

Results

504 images of 84 eyes of 43 patients were included in the development set, and 1326 images of 221 eyes of 114 patients in the validation set. The main characteristics of patients of each group are presented in Table 1.

LogMAR BCVA were comparable for clear lens patients ($p > 0.05$) and cataracts patients ($p = 0.10$) between both set. Age was comparable for cataract patients of both sets ($p = 0.07$). However, a significant age difference was observed in clear lens patients of both sets ($p = 0.01$).

Training was conducted for 21 epochs. Training curves of the Dice coefficients during training for each set are presented in Fig. 1.

In the development set, the mean Cataract Fraction (CF) was 0.002 ± 0.003 and 0.625 ± 0.236 ($p < 0.001$) for the clear lens and cataract patients, respectively. It was 0.024 ± 0.077 and 0.479 ± 0.230 ($p < 0.001$), respectively, in the validation set.

In the development set, ROC AUC was calculated at 1.00, with an optimal CF threshold value of 0.19. Using that threshold, sensitivity and specificity to detect cataract were both 100%.

In the validation set, ROC AUC was 0.98 with an optimal CF threshold of 0.14. Using that threshold, sensitivity and specificity to detect cataract were 94.4% and 94.7%, respectively (Fig. 2).

Examples of the 'model's result on different nuclear and cortical cataract cases are provided in Fig. 3.

Discussion

We developed a deep learning model to detect cataract on SS-OCT images with good diagnostic performances.

Cataract detection and grading is classically performed clinically using the Lens Opacities Classification System III.¹ However, reliable automatization of the process would be very useful both for telemedicine and in routine practice for cataract surgeons. Indeed, a precise and repeatable assessment of the 'lens' opacity would allow for a better understanding of the 'patient's discomfort in early stages, an objective measurement of 'cataract's progression and could help surgical planning. Also, in some cases of associated

corneal disease or vitreous opacity, it can be hard to determine the impact of the lens opacity on the 'patient's vision. A cataract diagnosis tool unaffected by such conditions would be precious to our practice.

SS-OCT longer wavelengths make it more robust to corneal diseases. Its combination with deep learning may constitute the best option for such a tool so far.

Few other methods have been described to assess lens opacity objectively. Table 2 shows compares the theoretical advantages of each described or existing technique along the 6 following criteria: grading capability, ability to detect all three kinds of cataract, specificity to the diagnosis of cataract, interpretability of the result, robustness to corneal disease and finally inter-operator variability. We selected those criteria because we believe they are the desired requirements for an objective cataract detection and grading tool. We develop those elements for each tool in the following section. It should be noted that some points are hypothesized based on the physical properties of each device.

Scheimpflug tomography was the first objective tool used for that matter with the PNS (Pentacam Nucleus Staging). It provides a single numerical value, based on the pixel intensity in the nucleus region. However, it is limited to the nucleus density and does not consider the anterior and posterior cortex, which are often responsible for more visual discomfort. It is also limited in cases of severe nuclear cataract.⁹ Despite those limitations, it is well correlated to clinical nuclear opalescence, visual acuity, and contrast sensitivity.^{2,3,14} It should be noted that results are probably more affected than SS-OCT in cases of corneal diseases given the shorter wavelength used.

Ocular Scattering Index measured from the double pass aberrometer HD-Analyzer, (Visometrics, Spain) provides a single metric to quantify light scattering from ocular media.¹⁵ It is well correlated to visual acuity, clinical lens opacity grading and contrast sensitivity in cataract eyes.^{3,14} However, it is not reliable in cases of high myopia and is affected by any ocular media opacity.^{16,17} Interpretability is limited as it is not possible to know from which ocular structure the scattering arises.

SS-OCT have also been used to quantify cataract by using lens pixels intensity. It is well correlated to clinical nuclear opalescence and visual acuity.⁷⁻⁹ In a previous study, we compared the diagnostic performances of the OSI, Average Lens Density (ALD) measured on IOL-700 (Carl Zeiss Meditec AG) SS-OCT images, and Scheimpflug based nuclear staging. In a cohort of 285 eyes, ALD had the best diagnostic performance with a Receiver operating characteristic curves (ROC) area under the curve (AUC) of 0.97. In the same cohort, OSI

Table 1 Main patients' characteristics in the different groups.

	Development set		Validation set	
	Clear lens	Cataract	Clear lens	Cataract
Images, N	263	241	792	534
Eyes, N	43	41	132	89
Patients, N	21	22	66	48
Age (mean \pm std)	$38.0 \pm 10.0^*$	70.4 ± 8.4	$34.6 \pm 10.0^*$	73.4 ± 11.5
logMAR BCVA (mean \pm std)	0.00 ± 0.00	0.31 ± 0.20	0.00 ± 0.00	0.27 ± 0.30

* $p < 0.05$

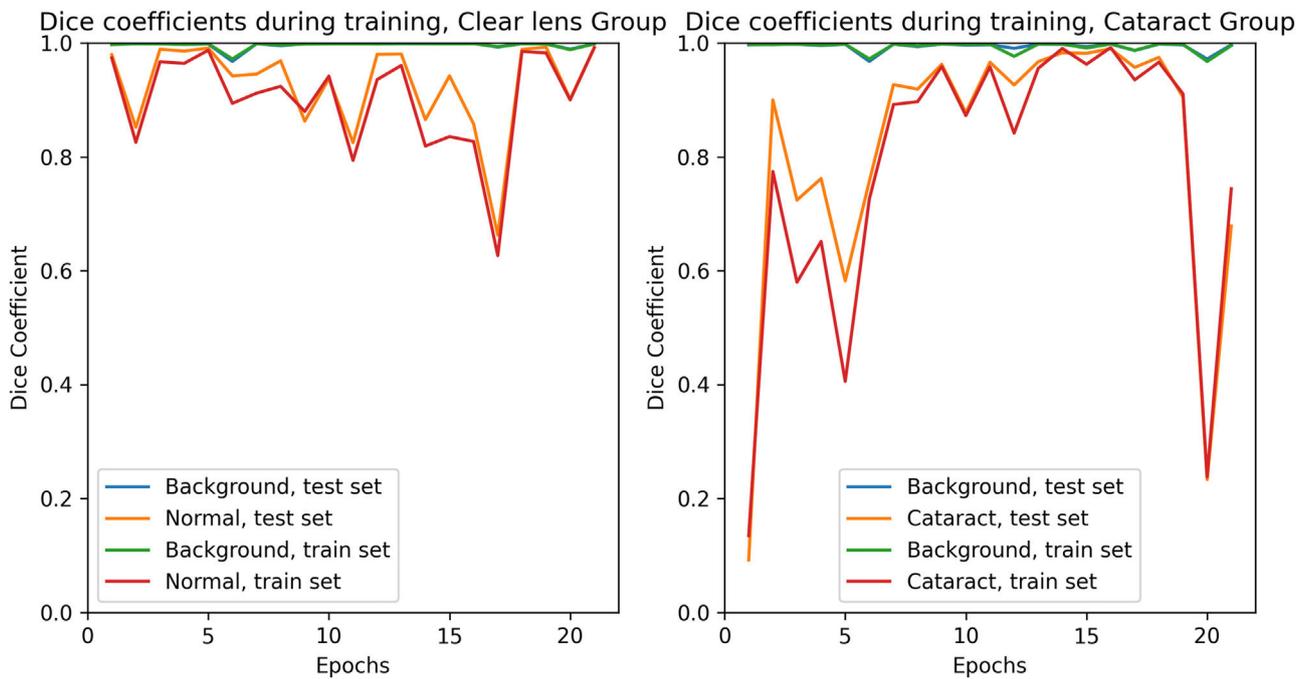


Fig. 1 Training curves for the training and test set for clear lens patients and cataract patients.

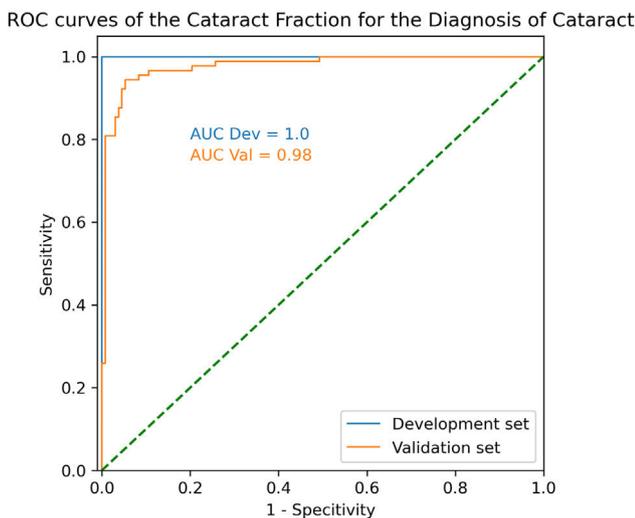


Fig. 2 Receiver operating characteristic curves for the development and validation sets. AUC: Area under the curve; Dev: Development set; Val: Validation set

and Scheimpflug-based nuclear staging had ROC AUCs of 0.96 and 0.90 respectively. This technique can not, however, distinguish the different kinds of cataracts.

In depth analysis of SS-OCT volumetric data has also been described with promising results using a prototype instrument.¹⁸

Two main approaches have been used recently to detect and grade cataract using Deep Learning. The first uses slit lamp photographs, while the second uses fundus photographs. Using slit lamp photographs, Keenan et al. achieved similar or better performances than human readers in grading all types of cataracts.⁵ Lu et al. also describe a model to detect and

grade cataract on slit lamp photographs. It achieved good performances for cortical and nuclear cataract.¹⁹ Slit-lamp photographs are, however, subject to inter-examiner variability and require a certain amount of training to achieve good quality images. Using fundus photographs and a global-local attention network, Xu et al. achieved a 90.64% classification accuracy in detecting cataract.⁶ Fundus images have the advantage of being widely accessible. However, they are affected by any corneal or vitreous opacities and the pupil diameter. In all cases, interpretability is limited to visualization techniques calculated secondarily highlighting with region of the image allowed the classification by the model. In the case of fundus imaging, the clinical relevance of such technique is questionable.

Only two studies^{20,21} used deep learning combined with SS-OCT images to detect and grade cataract. Both describe models to classify OCT scans in different stages of nuclear cataract with good performance. However, in addition to assessing nuclear cataracts only and disregarding cortical cataracts, it should be noted that clinical grading is subjective, and models trained with such data will be biased and produce the same kind of errors as humans. As previously, interpretability is limited to the same visualisation techniques common for deep neural networks. In the case of nuclear cataract staging, this information is of limited interest.

We voluntarily did not use clinical grading as ground truth categories for training as we believe it is highly subjective. Instead, we only used the clinical diagnosis of cataract to label all pixels of a given lens with the same label. Even though some subjectivity exists in this labeling method, it is limited as most ophthalmologists would agree on the definition of a significant cataract based on the association of the patient's symptoms, visual acuity and slit lamp examination. We also used a segmentation model instead of a classification model. This method ensures that interpretability is built

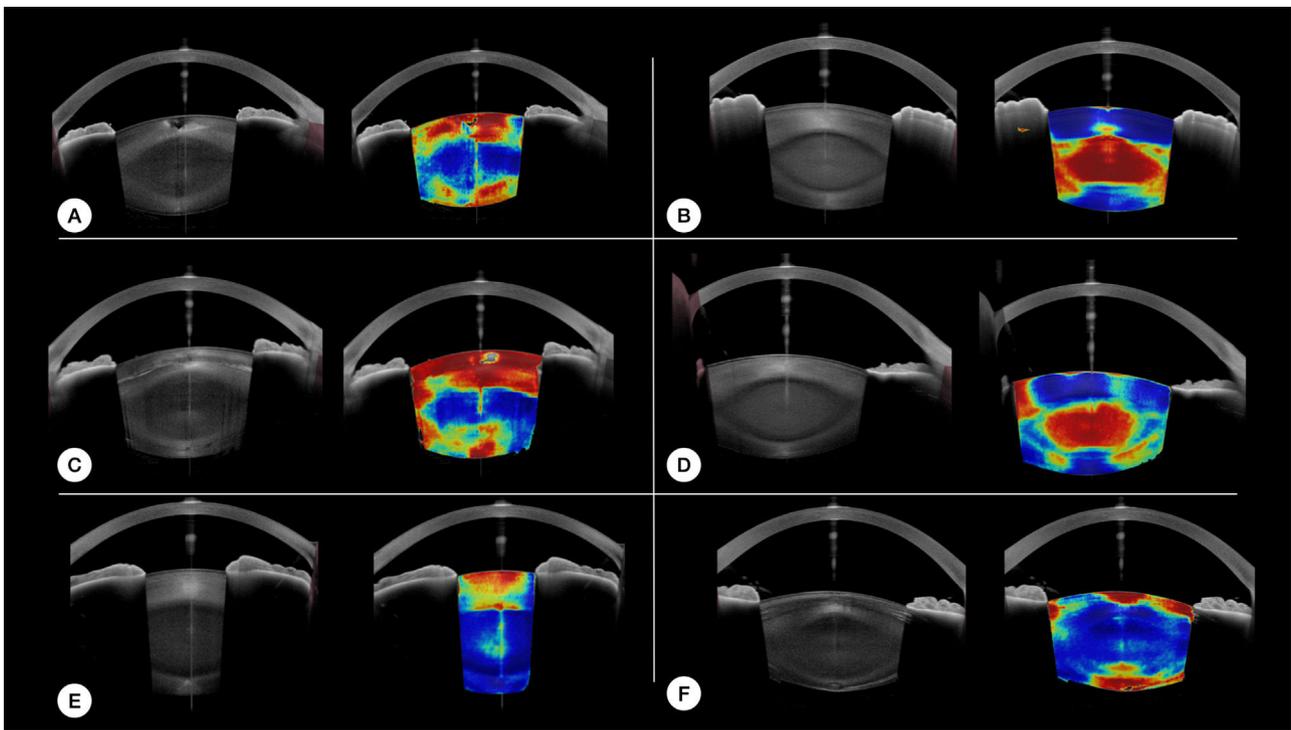


Fig. 3 Examples of our model’s results. For each case, the original image is on the left, and the model results overlaid on the original image is on the right. Hot colors indicate a high probability of cataract. A and C are cases of cortical cataracts. B and D of nuclear cataracts. E is a case of anterior cortical cataract and F is a case of cortical and posterior subcapsular cataract.

Table 2 Comparison of the theoretical advantages of objective cataract detection systems.

	Our approach	HD Analyzer	PNS	SS-OCT ALD	AI Slit Lamp	AI Fundus photograph	AI SS-OCT classification
Grading	✗	✓	✓	✓	✓	✓	✓
Three types of cataracts	✓	✓	✗	✗	✓	✗	✗
Specific to cataract	✓	✗	✓	✓	✓	✗	✓
Interpretability	✓	✗	✓	✓	✓	✗	✗
Robust to corneal diseases	✓	✗	✗	✓	✗	✗	✓
Independent from user’s experience	✓	✓	✓	✓	✗	✓	✓

PNS: Pentacam Nucleus Staging

SS-OCT ALD: Average Lens Density measured on IOL700 images.

AI Slit Lamp: Deep learning classification of Slit lamp photographs

AI Fundus photograph: Deep learning classification of Fundus photographs

AI SS-OCT classification: Deep learning nuclear cataract classification using SS-OCT images.

in the model by providing colormaps of the probability of cataract without using any additional visualization technique. As cataracts might be heterogenous, our labeling process could be characterized as noisy at the pixel levels. The clear lens labeling, however, is rather pure. The combination of our labelling process and a segmentation approach allows the model to learn which part of the image has a higher probability of presence of localized cataract or clear lens.

Eventhough the goal of the study was not to assess specifically the model’s performance to detect each type of cataract, the image outputs of the model seems promising for that matter. Fig. 3 provides examples of local detection of cataract in different cases of cataracts. Despite having no

prior knowledge of the different layers of the lens, the model seems to perform remarkably well in cases of isolated nuclear or cortical cataract. Moreover, Fig. 3 F, shows the correct detection of subcapsular cataract. A more thorough evaluation will be included in a subsequent study.

We found a significant age difference between the development and validation set clear lens patients. Although this difference can be explained by sampling fluctuation given the small size of the development set, it could constitute a bias for the validation set’s result.

Despite having found an AUC of 1 for the development set, we do not believe the model is overfitted. This is certainly due to less variance in the development set. The

validation 'set's results are in-line with clinical findings, and no unexplainable results were observed.

Through error analysis, we found that the false-negative cases were very localized cataracts in young patients, such as polar cataracts. Some false-positive cases were from unusual presentations of clear lens or rare artifacts. These errors could certainly be reduced by increasing the training set size to include more uncommon cases of both clear lens and cataract patients. Other false-positive cases were probably due to age-related modifications of the lens. Indeed, in some older clear lens patients, the model detected weak signals of cataract in the nuclear region. Nuclear opacification is a progressive process, and two 60-year-old patients with 20/20 BCVA and similar nuclear opacity might not exhibit the same subjective discomfort. For that matter, it is impossible to define a clear cut-off between nuclear cataract and clear lens.

The current version of the model is not built for cataract grading, but rather for cataract detection and localization. We aim to address the problem of cataract grading in a subsequent study. Also, the model's robustness to artifacts produced by corneal diseases will be tested in future works. Finally, the model's repeatability should also be evaluated.

We believe that the model's built-in interpretability, helps the clinician reading the images without providing a definite categorical diagnosis and leaves the final diagnosis to the physician. This could certainly aid younger surgeon during their training but might be of limited use in it's current state for experienced surgeons.

In conclusion, we developed a deep learning model to detect cataract at the pixel level on SS-OCT images. It is the only OCT-based tool so far seemingly capable of detecting and differentiating both nuclear and cortical cataract. In that sense, it helps the physician reading the images and could also be used in remote settings with telemedicine for automated diagnosis.

Conflicts of interest

The authors have no conflict of interest.

Acknowledgments

None.

References

- Chylack LT, Wolfe JK, Singer DM, Leske MC, Bullimore MA, Bailey IL, et al. The lens opacities classification system III. The longitudinal study of cataract study group. *Arch Ophthalmol Chic*. 1960;111:831–836. <https://doi.org/10.1001/archophth.1960.01090060119035>.
- Grewal DS, Brar GS, Grewal SPS. Correlation of nuclear cataract lens density using scheinpluf images with lens opacities classification system III and visual function. *Ophthalmology*. 2009;116:1436–1443. <https://doi.org/10.1016/j.ophtha.2009.03.002>.
- Pan AP, Wang QM, Huang F, Huang JH, Bao FJ, Yu AY. Correlation among lens opacities classification system III grading, visual function index-14, pentacam nucleus staging, and objective scatter index for cataract assessment. *Am J Ophthalmol*. 2015;159:241–247.e2. <https://doi.org/10.1016/j.ajo.2014.10.025>.
- Cabot F, Saad A, McAlinden C, Haddad NM, Grise-Dulac A, Gatinel D. Objective assessment of crystalline lens opacity level by measuring ocular light scattering with a double-pass system. *Am J Ophthalmol*. 2013;155:629–635. e1-2; <https://doi.org/10.1016/j.ajo.2012.11.005>.
- Keenan TDL, Chen Q, Agrón E, Tham YC, Goh JHL, Lei X, et al. DeepLensNet: deep learning automated diagnosis and quantitative classification of cataract type and severity. *Ophthalmology*. 2022;129:571–584. <https://doi.org/10.1016/j.ophtha.2021.12.017>.
- Xu X, Li J, Guan Y, Zhao L, Zhao Q, Zhang L, et al. GLA-Net: a global-local attention network for automatic cataract classification. *J Biomed Inform*. 2021;124: 103939. <https://doi.org/10.1016/j.jbi.2021.103939>.
- Panthier C, de Wazieres A, Rouger H, Moran S, Saad A, Gatinel D. Average lens density quantification with swept-source optical coherence tomography: optimized, automated cataract grading technique. *J Cataract Refract Surg*. 2019;45:1746–1752. <https://doi.org/10.1016/j.jcrs.2019.07.033>.
- Panthier C, Burgos J, Rouger H, Saad A, Gatinel D. New objective lens density quantification method using swept-source optical coherence tomography technology: comparison with existing methods. *J Cataract Refract Surg*. 2017;43:1575–1581. <https://doi.org/10.1016/j.jcrs.2017.09.028>.
- Chen D, Li Z, Huang J, Yu L, Liu S, Zhao YE. Lens nuclear opacity quantitation with long-range swept-source optical coherence tomography: correlation to LOCS III and a Scheimpflug imaging-based grading system. *Br J Ophthalmol*. 2019;103:1048–1053. <https://doi.org/10.1136/bjophthalmol-2018-312661>.
- Zéboulon P, Ghazal W, Gatinel D. Corneal edema visualization with optical coherence tomography using deep learning: proof of concept. *Cornea*. 2020. <https://doi.org/10.1097/ICO.0000000000002640>. Publish Ahead of Print.
- Zéboulon P, Ghazal W, Bitton K, Gatinel D. Separate detection of stromal and epithelial corneal edema on optical coherence tomography using a deep learning pipeline and transfer learning. *Photonics*. 2021;8:483. <https://doi.org/10.3390/photonics8110483>.
- Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. arXiv; 2015. <https://doi.org/10.48550/arXiv.1505.04597>.
- milesial. U-Net: semantic segmentation with PyTorch. 2022. <https://github.com/milesial/Pytorch-UNet.git>.
- Lim SA, Hwang J, Hwang KY, Chung SH. Objective assessment of nuclear cataract: comparison of double-pass and scheinpluf systems. *J Cataract Refract Surg*. 2014;40:716–721. <https://doi.org/10.1016/j.jcrs.2013.10.032>.
- Vilaseca M, Peris E, Pujol J, Borrás R, Arjona M. Intra- and inter-session repeatability of a double-pass instrument. *Optom Vis Sci Off Publ Am Acad Optom*. 2010;87:675–681. <https://doi.org/10.1097/OPX.0b013e3181ea1ad3>.
- Ren Z, Xu L, Fan Q, Yang K, Ren S, Zhao D. Assessment of visual quality in eyes with forme fruste keratoconus and mild and moderate keratoconus based on optical quality analysis system II parameters. *J Ophthalmol*. 2020;2020: 7505016. <https://doi.org/10.1155/2020/7505016>.
- Garrido C, Cardona G, Güell JL, Pujol J. Visual outcome of penetrating keratoplasty, deep anterior lamellar keratoplasty and descemet membrane endothelial keratoplasty. *J Optom*. 2018;11:174–181. <https://doi.org/10.1016/j.optom.2017.08.003>.
- Grulkowski I, Manzanera S, Cwiklinski L, Mompeán J, Castro A, Marin JM, et al. Volumetric macro- and micro-scale assessment of crystalline lens opacities in cataract patients using long-

- depth-range swept source optical coherence tomography. *Biomed Opt Express*. 2018;9:3821–3833. <https://doi.org/10.1364/BOE.9.003821>.
19. Lu Q, Wei L, He W, Zhang K, Wang J, Zhang Y, et al. Lens opacities classification system III-based artificial intelligence program for automatic cataract grading. *J Cataract Refract Surg*. 2021. <https://doi.org/10.1097/j.jcrs.0000000000000790>.
 20. Zhang X, Xiao Z, Li X, Wu X, Sun H, Yuan J, et al. Mixed pyramid attention network for nuclear cataract classification based on anterior segment OCT images. *Health Inf Sci Syst*. 2022;10:3. <https://doi.org/10.1007/s13755-022-00170-2>.
 21. Huang X, Mori H, Lukmanto D, Tran TH, Fukuda M, Oshika T, et al. Objective cataract detection and grading with deep learning based on OCT densitometry. *Invest Ophthalmol Vis Sci*. 2021;62:67.