INTERNATIONAL JOURNAL OF CIRCUMPOLAR HEALTH
2024, VOL. 83, 2314802
https://doi.org/10.1080/22423982.2024.2314802

Taylor & Francis
Taylor & Francis Group

SHORT COMMUNICATION

The use of artificial intelligence to assess diabetic eye disease among the Greenlandic population

Trine Jul Larsen [a], Maria Bråthen Pettersen[b], Helena Nygaard Jensen[b], Michael Lynge Pedersen[a,c], Henrik Lund-Andersen[b,c], Marit Eika Jørgensen[d] and Stine Byberg[b]

[a]Greenland Center of Health Research, Institute of Nursing and Health Science, University of Greenland, Nuuk, Greenland; [b]Clinical Epidemiology, Steno Diabetes Center Copenhagen, Copenhagen, Denmark; [c]Rigshospitalet-Glostrup University Hospital, Glostrup, Denmark; [d]Steno Diabetes Center Greenland, Nuuk, Greenland

ABSTRACT

Background: Retina fundus images conducted in Greenland are telemedically assessed for diabetic retinopathy by ophthalmological nurses in Denmark. Applying an AI grading solution, in a Greenlandic setting, could potentially improve the efficiency and cost-effectiveness of DR screening.

Method: We developed an AI model using retina fundus photos, performed on persons registered with diabetes in Greenland and Denmark, using Optos® ultra wide-field scanning laser ophthalmoscope, graded according to ICDR.

Using the ResNet50 network we compared the model's ability to distinguish between different images of ICDR severity levels in a confusion matrix.

Results: Comparing images with ICDR level 0 to images of ICDR level 4 resulted in an accuracy of 0.9655, AUC of 0.9905, sensitivity and specificity of 96.6%.

Comparing ICDR levels 0,1,2 with ICDR levels 3,4, we achieved a performance with an accuracy of 0.8077, an AUC of 0.8728, a sensitivity of 84.6% and a specificity of 78.8%. For the other comparisons, we achieved a modest performance.

Conclusion: We developed an AI model using Greenlandic data, to automatically detect DR on Optos retina fundus images. The sensitivity and specificity were too low for our model to be applied directly in a clinical setting, thus optimising the model should be prioritised.

ARTICLE HISTORY
Received 2 August 2023
Revised 23 January 2024
Accepted 1 February 2024

KEYWORDS
Diabetic retinopathy; artificial intelligence; screening; ultra wide-field; ICDR-scale

Introduction

Diabetic retinopathy (DR) is a microvascular complication of diabetes, leading to vision loss if not detected and treated. Thus, regular screening of the retina is essential for early detection and treatment, in order to prevent vision loss [1–3].

Greenland is populated by 56.000 inhabitants living around 17 towns and 56 settlements spread along the 44.000 km ice-free coastline, where each destination only can be reached by plane or boat, depending on the weather. Today, the majority (87.5%) of the population lives in towns and a minority (12.5%) of the population lives in smaller settlements with 3–500 inhabitants [4].

In Greenland, nine out of 17 towns are equipped with Optos® ultra wide-field fundus cameras and persons registered with diabetes are invited for regular screenings in the town closest to their home. Images and other screening data are uploaded through a server, and telemedically assessed for DR, by two specialist ophthalmologist nurses at Steno Diabetes Center Copenhagen

(SDCC)/Rigshospitalet-Glostrup University Hospital in Denmark, according to the International Clinical Diabetes Retinopathy disease severity scale (ICDR) [5].

The distances in Greenland are however so vast, travelling to the nearest eye screening station can take days, and at least one week's delay can be expected, when awaiting the result of the eye examination. Advanced diagnostics and treatment are done either by visiting ophthalmologists or in Denmark, as there are no regular ophthalmologists in Greenland.

Through time, artificial intelligence (AI) has emerged as a major frontier in computer science research [6]. Several AI solutions to detect microvascular lesions in conventional retina fundus photos have been developed, using deep learning techniques [7–9]. Deep learning is a machine learning technique where a neural network is trained to detect diabetic retinopathy from the intensities of the pixels in the

CONTACT Trine Jul Larsen Tjul@uni.gl Greenland Center of Health Research, Institute of Nursing and Health Science, University of Greenland, Qupaloraarsuk 40, Nuuk, Ilulissat 3952, Greenland

images to form fundus images already graded by ophthalmologists. The model then "learns" how an image with a specific grading appears [7].

The technique has the potential to be an important screening tool, reducing the workload of healthcare professionals and ensuring timely diagnosis [10].

In Singapore, an integrated DR screening programme using AI has resulted in a better accuracy of the screenings (90% sensitivity and specificity), and faster response of the screening [11,12].

Further, algorithms for automatic grading in detection of DR have been developed in the

Netherlands, and found to be valid for use in the primary care [13].

Applying an AI grading solution, in a Greenlandic setting for automatically distinguishing images with and without microvascular lesions, would mean that only patients in need of treatment and/or further diagnostics would need their images sent to Denmark and/or to be seen by the specialist ophthalmologist nurses and doctors. This would save resources, previously used for image grading and provide immediate diagnoses to the population, reducing the need for patient travels. Studies have shown, directly applying an AI algorithm developed on, e.g., a Western population to other ethnic populations may reduce the performance of the algorithm [14,15], hence the algorithm must be trained on data, representative of the population it is aimed for.

Thus, our aim was to develop an AI model, which automatically can detect DR on retina images, specific for the Greenlandic population and the Optos fundus camera, used in Greenland.

Material and methods

We developed an AI model, using retina fundus photos, gradings and diagnoses from routine DR screenings on non-dilated pupils, performed on persons registered with diabetes in Greenland.

Since, the prevalence of severe DR is rare in Greenland [16], we had to supplement our data, with Optos retina fundus photos from a Danish population, to have sufficient amounts of images with all DR gradings. In order for the model to learn how to distinguish between all ICDR gradings, the model has to be fed with sufficient amounts of images representing all ICDR categories.

Study population and data collection

Pseudo-anonymised data for the Greenlandic population, registered with type 1 diabetes (T1D) or type 2 diabetes (T2D), with screenings for DR performed between 2015 and 2020, was extracted from the electronic medical record systems, Cosmic and Æskulap, for East Greenland [17]. All retina fundus photos were manually exported from a server, to a hard disk and stored in an encrypted server, labelled with their corresponding ICDR grading one by one. We extracted retina fundus photos, from a Danish clinical study, OPTIMISM (unpublished), where persons with T1D and T2D were imaged, using Optos® ultra wide-field scanning laser ophthalmoscope at SDCC and at an eye clinic in a small town in Northern Jutland during 2019 and 2021. All images from the

OPTIMISM study were graded by the same ophthalmological nurses, grading the images from

Greenland. All images were graded according to the ICDR scale [5], consisting of four severity levels.

- Level 0: No DR – No abnormalities
- Level 1: Mild non-proliferative diabetic retinopathy (NPDR)
- Level 2: Moderate non-proliferative DR
- Level 3: Severe non-proliferative DR
- Level 4: Proliferative DR (PDR)

All data from the OPTIMISM study were registered in a RedCap database, with remote access [18].

Data processing

All data consisted of colour images with a pixel range of 4000×4000.

The classes of the DR gradings were divided into folders, with their corresponding grading. Images were excluded in the downloading process, if they were ungradable or if a laser treatment had been conducted. Laser treatments were common with images that had PDR and were excluded due to the nature of the grading system. If laser tracks are present, then the image will always be categorised as an ICDR 4 grading, regardless of whether or not PDR is present. As we wanted the AI model to be able to detect PDR and not laser treatment, we excluded images with laser treatment. Figures 1 and 2 illustrate images in which laser treatment had been conducted and images which were ungradable.

If the presence of cataract and vitreous opacity disturbed the image to an extent that made it nongradable the images were also excluded. If the images were gradable, they were included and no measures were taken to account for cataract.

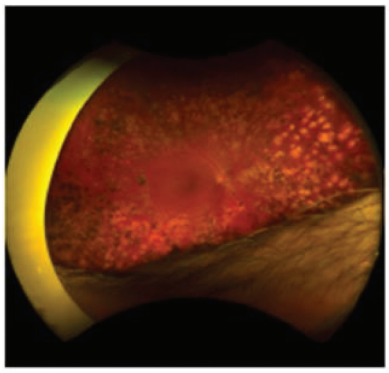

Figure 1. Image inwhich laser treatment has been conducted.

Deep learning model

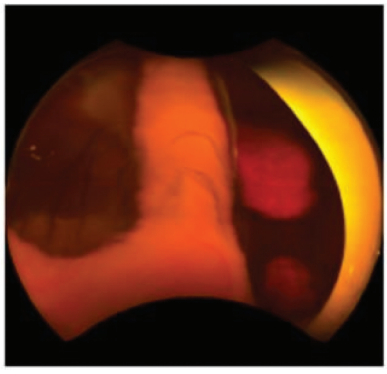

Figure 2. Ungradable image.

We developed our model using the ResNet50 network, a pre-trained network with over a million images from the ImageNet database [19].

In the experimental phase of our model, we developed and compared models using the two networks ResNet50 and VGG16. In addition, we compared different preprocessing techniques. From our tests, we achieved the best results when using cropping, green channel extraction, and brightness and contrast for preprocessing when both using VGG16 and ResNet50. This preprocessing method was used, in our final models.

By using ResNet50 as the pretrained network for transfer learning, it outperforms VGG16 in all of our models. The networks ResNet50 and VGG16 are quite similar, as they both are trained on the same millions of images from ImageNet, have a similar architecture consisting of convolutional layers, pooling layers and a fully connected layer, and are commonly used in medical image analysis. The most significant difference between the two networks is the depth of the layers [20–22]. The network is 50 layer deep and consists of convolutional layers, pooling layers and a fully connected layer, classifying the images based on their label.

Using K-fold cross-validation, we extracted 20% of our data to create a test set. Thus, we used 80% of the remaining data as a training set and 20% as validation data. By using K-fold cross-validation, all of the training and validation data are used, by splitting the data into K non-overlapping subsets [23]. The most common is to have a K between 5 or 10. Afterwards, the model is trained with K-1 fold and then validated with the Kth fold. This is repeated until all the data has functioned as both training and validation [24]. After all the folds are trained, it is possible to estimate the mean test error, which is the average from each fold [23]. By training a model 5 folds, this will increase the computational cost of the model, because it is similar to training a "normal" model five times consequently.

Since we wanted to transfer the network onto our own images, we had to replace the final layers of each network. Thus, we extracted every layer of the architecture, except the last three, when building the model for ResNet50. We replaced and defined the new last three layers with a new fully connected layer, a softmax layer and a classification layer. The classification layer was replaced with our classes defined by the labels of our images as the new classes for the new classification layer.

We sat the fully connected layer to have a WeightLearnRateFactor and a BiasLearnRateFactor at 20 [25]. In the preprocessing function, we extracted the green channel of the image [26]. We used a cropping function, resizing the images to 2473×2980pixels, centred on the macula and adjusted the brightness and contrast level, followed by concatenating the arrays to match the 3 colour channels. We used data augmentation to resize all images to the network's size, 224×224pixels, and to create more images by applying random rotation, reflection and shear to the dataset.

For our model, we used an Adam optimising algorithm, with an initial learning rate of 0.0001.

Measuring performance

We assessed the performance of our model in four sets; images with ICDR grading 0 vs. 4, images with ICDR grading 0 vs. 3 & 4, images with ICDR grading 0–2 vs. 3 & 4 (also classified as "nonreferable" vs. "referable" DR) [27], and lastly images with ICDR grading 0 vs. 1–2 vs. 3–4.

To visualise and describe the performance of our model, we used a confusion matrix.

The confusion matrix gave multiple metrics, such as the classes true positive (TP), false positive (FP), true negative (TN) and false negative (FN).

From these values, we calculated the Sensitivity: TP/(TP+FN), Specificity: TN/(FP+TN) and the Accuracy (TP+TN)/(TP+TN+FP+FN) [28].

We modelled the confusion matrix of our labels, comparing the predicted labels from the model with the actual labels and calculated the accuracy, a good predictor of how well the model predicts the sensitivity, measuring how many of the positive images classified as correct by the model, the specificity, measuring how many of the negative images are correctly predicted as negative by the model, and the Area Under the Curve (AUC) [28], calculated by the Receiver Operator Characteristic (ROC) curve [29], constructed by plotting the true positive rate against the false positive rate, illustrating the predictive accuracy well.

Ethics

Ethical approval was granted by the ethical review committee for Greenland (KVUG 2017–10) and by the Greenland Health Authorities. All participants of the Danish OPTIMISM study had given their informed consent, and ethical approval was granted by the National Committee on Health Research Ethics (journal no. H19044875).

Results

A total of 1700 images from the Greenlandic population were manually extracted from the Greenlandic health server to a hard disk. From the Danish OPTIMISM study, 5000 images were available.

Images which were not "single colour image" and images with no DR were excluded to attempt to achieve balanced dataset including all ICDR gradings. Thus, a total of 248 images from the Greenlandic population and 551 images from the OPTIMISM study were included in the present study. Our final dataset consisted of 799 retina images from a Greenlandic and Danish study population, assessed according to ICDR (Table 1).

We compared the model's ability of distinguishing between different image ICDR severity levels (see Table 2); firstly, we compared images graded 0, with images graded 4. We found a good model performance with an accuracy of 0.9655, AUC of 0.9905, sensitivity and specificity of 96.6%. Secondly, we compared images graded 0 with images graded 3–4. The model performance was modest with an accuracy of 0.8171, an AUC of 0.8543, a sensitivity at 78% and a specificity at 87.8%. Thirdly, we compared images graded 0, 1 and 2, with images graded 3 and 4. The performance was similar to the comparisons above with an accuracy of 0.8077 and an AUC of 0.8728. The sensitivity was higher than above (84.6%) however, the specificity was correspondingly lower (78.8%). Fourthly, we assessed the performance of the model in distinguishing between images grade 0 versus images graded 1 and 2 versus images graded 3 and 4. The models ability to distinguish these three gradings from each other, was lower than in the other comparisons, with an accuracy of 0.6583, an AUC of 0.8063, a sensitivity of 60% and a specificity of 74.6%.

Table 1. Overview of the final dataset.

ICDR scale	Danish	Greenlandic	Total
0 (No DR)	124	83	207
1 (Mild NPDR*)	78	63	141
2 (Moderate NPDR)	130	54	184
3 (Severe NPDR)	95	29	124
4 (PDR**)	124	19	143
Total	551	248	799

*Non-proliferative diabetic retinopathy ** Proliferative diabetic retinopathy.

Table 2. Performance of the model for different images gradings.

ICDR level	Accuracy/ Performance	AUC*	Sensitivity	Specificity
0 vs. 4	0.9655	0.9905	96.6%	96.6%
0 vs. 3+4	0.8171	0.8543	78.0%	87.8%
0+1+2 vs. 3+4	0.8077	0.8728	84.6%	78.8%
0 vs. 1+2 vs. 3+4	0.6583	0.8063	60.0%	74.6%

**Area Under the Curve* .

Discussion

In this study, we aimed to develop an AI model, to automatically detect DR on retina fundus images, specific for the Greenlandic population and the Optos® ultra widefield scanning laser ophthalmoscope, used for DR screening in Greenland. We developed a model with the ability to distinguish between images of different ICDR severity levels and achieved a very good performance and specificity, when we compared images of ICDR level 0 to images of ICDR level 4. However, the performance was suboptimal when comparing no/mild vs. severe DR with sensitivity and specificity for the models of around 80%.

An autonomous approach, independently grades images, without human expert reading of the images, and provides an immediate detection of the DR level. The approach should have a high sensitivity and specificity as there is no human check involved, in order to be applied in a clinical setting [30].

In China, a study aimed to assess the accuracy of AI-based screening for DR, and explore the feasibility of applying the technique to a community hospital [31].

The study enrolled 889 subjects, in which nonmydriatic fundus photos were taken using Topcon

TRC-NW400 camera. Similarly to our study, all the fundus photos were graded to by two independently ophthalmologist, using the ICDR scale. The study managed to develop a model for referable DR, defined as more than mild NPDR and/or macula oedema, with a sensitivity of 91,18% and a specificity of 98,7%, which is much higher compared to the performance of our model. The study finds the results to be feasible in a clinical setting, however they recommend further research to assess the effectiveness of DR detection. Likewise, a clinical trial in the US involving 900 subjects compared Optical Coherence Tomography (OCT) images from the Wisconsin Fundus Photograph Reading Center, to an autonomous AI diagnostic system to detect more than mild DR and diabetes macula oedema. The AI system performed very well with a sensitivity of 87.2% and a specificity of 90.7%, and was the first autonomous AI-based diagnostic system for detection of DR in primary care, approved by the US Food and Drug Administration [30]. The sensitivity and specificity of our model lie well below the sensitivities and specificities in the abovementioned studies, and therefore non-applicable for use in its current state in Greenland. Had the performance been better, especially the sensitivity in our model, we could have initiated a feasibility study on future screenings in Greenland where we could assess the accuracy of the AI grading, the acceptance of AI grading among patients and health care workers, and potential cost-savings.

Application of our model

In 1995, a consensus view was put forward by clinicians at a meeting of the British Diabetic Association in Exeter, that screening for DR should have a minimum sensitivity of 80% and a specificity of 95% [32].

In Greenland, we would need an autonomous model with high specificity, sensitivity as well as overall good performance, to distinguish referable from non-referable images, as the imaging health care personnel in Greenland are not trained in image grading.

Our model achieved a high performance and specificity/sensitivity, when we compared no DR to ICDR level 4. However, detecting DR before it has advanced to a very severe stage and visual symptoms is crucial for adequate treatment outcomes [33], it would be too risky to directly implement a model only capable of recognising very severe forms of DR, for autonomous grading in Greenland. Thus, despite a very good performance, the model would not be applicable in a Greenlandic setting. Though, the model could be applied as a potential decision tool, to immediately identify proliferative changes at the screening station and thus, avoid delays in referral and treatment considering that advanced diagnosis and treatment must be done telemedically or in Denmark.

In China, a semi-automated deep learning algorithm-assisted approach has been developed on Topcon retina fundus photos, to detect vision-threatening referable DR [34]. The approach combined both AI and human grading procedures. The model detected vision-threatening referable DR, of which the high-risk cases, detected by the model, were manually graded by a senior ophthalmologist. Applying the model, in a clinical setting, presented advantages in time and economic savings for grading the images, enabling accurate and efficient diagnoses. However, this would not be feasible in Greenland, where the photographers are not trained in image grading.

In line with the study above, our model achieved a good performance, in distinguishing images of ICDR level 0–2 (non-referable DR) compared with ICDR level 3–4 (referable DR).

The model could be utilised at the point of screening in Greenland, with a higher sensitivity and specificity, thus only images labelled “referable DR”, would be telemedically assessed by the Danish ophthalmological staff. However, the sensitivity of 84%, would lead to too many false negatives to exclusively rely on the model for DR screening. In order to develop an AI model, that could function more or less autonomous in a clinical setting, we would need a model with high specificity and sensitivity as well as overall good performance, to distinguish referable from nonreferable images autonomously, as the imaging healthcare professionals in Greenland are not trained in image grading.

Applying an autonomous AI system for referable detection of DR, in a Greenlandic setting, with a high prevalence of diabetes [35,36], could potentially improve the efficiency and cost-effectiveness of DR screening, by minimising the cost for travelling and the workload of healthcare professionals and finally ensuring timely diagnosis. However, a recent study, conducted in Greenland, found an overall prevalence of DR of 13.6% and less than 2.5% had severe-none proliferative DR or PDR [16]. Thus, specialist examination and treatment, could be reserved for the small group of diabetic patients with more severe changes. In order for our model, to be used as an autonomous AI model in Greenland, the model should be optimised, by development of a large training and evaluation dataset [6].

Optimizing the assessment of the images with more ophthalmological nurses grading the same images, could reduce the potential human errors and increase the quality of the training data [37]. Finally, our model would have to be strictly validated in pre-registered studies for safety, efficacy and equity, involving real-work [38]. This could be done by testing the model in the capital, Nuuk, and at the same time get the ophthalmological nurses at SDCC, to give their assessment of the same image, and we could hereby get an indication of what the model does not recognise.

Strengths

This is the first study to develop an AI model, which automatically can detect DR on retina images, specific for the Greenlandic population, conducted on the Optos fundus camera.

All retina fundus photos from both the Greenlandic and the Danish study populations included in this study were manually graded by the same two specialist ophthalmological nurses.

Limitations

Due to the low prevalence of severe DR in Greenland [16], we had to include images from a Danish population in order to achieve an acceptable performance. This could mean that the model could perform poorer when exclusively applied to Greenlandic data.

However, during recent centuries, the Greenlandic population have intermixed with Europeans, leading to a relatively high proportion of genetic European ancestry in modern Greenlanders, where the average estimated degree of European ancestry admixture in present day Inuit in Greenland is 25%, varying with the degree of geographical isolation [39]. Thus, many of the people screened for DR in Greenland today will share many common features with an ethnic Danish population, suggesting that the model might perform well in Greenland despite its training on partially Danish data. Our model was only developed to detect DR, thus other ocular diseases are not detected when using the model. Currently, the ophthalmological nurses assess the fundus images and supplementary information, for signs of other ocular diseases and refer persons to ophthalmologists if deemed necessary. Nonetheless, the assessment of other ocular diseases can be seen as an extra service for persons with diabetes, as the primary aim of DR screening is to assess images for DR.

In Greenland, a range of other eye data including ocular pressure, autorefraction, visual acuity and optical coherence tomography (OCT) are also performed during DR screening. While the screening technician could refer a patient with high values of ocular pressure and declining visual acuity for telemedical assessment, the results of the OCT still needs to be interpreted by a specialist.

We excluded images in the downloading process, if they were ungradable or if a laser treatment had been conducted, however blurry or partial images could disturb the accuracy of the model. Preprocessing the images and splitting the images of their respective grading, according to quality, and train the model with high quality images first and during the training add blurry or partial images of less quality, could further improve the model [40].

Conclusion

In this study, we developed an AI model, to automatically detect DR on different retina fundus images, based on ICDR levels, specific for a Greenlandic population and the Optos fundus camera. We developed a model, with the potential to be applied in the clinic, for autonomous detection of DR. To apply the model more or less autonomous in a clinical setting, we would need a model with a higher specificity and sensitivity as well as overall good performance, to distinguish referable from non-referable images autonomously.

For safety reasons, sensitivity should be prioritised over specificity, to ensure a true detection of DR and minimise the false-positive referrals.

Future perspectives

Improving the current model should be prioritised, both in terms of adding more Greenlandic or other Inuit population images, as well as improving the model predictions with more images of severe DR. At present the model is not suitable for autonomous grading, however to be of advantage in the Greenlandic health system where telemedicine is an integrated part of the health care delivery, it is important that it can autonomously grade images with high accuracy. After the development of a well performing model, it is important to assess the feasibility of implementing such a model in a Greenlandic setting, and compare the costs and effects to other potential scenarios.

In addition, qualitative studies should be undertaken to assess the attitudes and uptake of an AI model for DR screening, both among patients and healthcare professionals. Furthermore, it could be assessed how the model could be adapted to screening programmes in other arctic populations.

Acknowledgments

This paper have been presented at the NUNAMED conference the 2nd October 2022.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Pinngortitaleriffik [80.28].

ORCID

Trine Jul Larsen http://orcid.org/0000-0001-5070-4231

References

[1] Lee R, Wong TY, Sabanayagam C. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. Eye Vis. 2015;2(1):17. doi: 10.1186/s40662-015-0026-2

[2] Ansari P, Tabasumma N, Snigdha NN, et al. Diabetic retinopathy: an overview on mechanisms, pathophysiology and pharmacotherapy. Diabetology. 2022;3(1):159–175. doi: 10.3390/diabetology3010011

[3] Fairless E, Nwanyanwu K. Barriers to and facilitators of diabetic retinopathy screening utilization in a high-risk population. J Racial Ethn Health Disparities. 2019;6(6):1244–1249. doi: 10.1007/s40615-01900627-3

[4] Grønlands Statistik [cited 2022 Jun 1]. Available from: https://stat.gl/dialog/main.asp?lang=en&version=2021&sc=GF&subthemecode=P1&colcode=P

[5] Wu L. Classification of diabetic retinopathy and diabetic macular edema. World J Diabetes. 2013;4(6):290. doi: 10.4239/wjd.v4.i6.290

[6] Vatathanavaro S, Tungjitnob S, Pasupa K. White Blood Cell Classification: A Comparison between VGG-16 and ResNet-50 Models. 2018.

[7] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402. doi: 10.1001/jama.2016.17216

[8] Grzybowski A, Brona P. Analysis and comparison of two artificial intelligence diabetic retinopathy screening algorithms in a Pilot study: IDx-DR and Retinalyze. J Clin Med. 2021;10(11):2352. doi: 10.3390/jcm10112352

[9] Goh JKH, Cheung CY, Sim SS, et al. Retinal imaging techniques for diabetic retinopathy screening. J Diabetes Sci Technol. 2016;10(2):282–294. doi: 10.1177/1932296816629491

[10] Valverde C, Garcia M, Hornero R, et al. Automated detection of diabetic retinopathy in retinal images. Indian J Ophthalmol. 2016;64(1):26. doi: 10.4103/0301-4738.178140

[11] Xie Y, Nguyen QD, Hamzah H, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. Lancet Digit Health. 2020;2(5):e240–e249.

[12] Bhargava M, Cheung CYL, Sabanayagam C, et al. Accuracy of diabetic retinopathy screening by trained non-physician graders using non-mydriatic fundus camera. Singapore Med J. 2012;53(11):715–719.

[13] van der Heijden AA, Abramoff MD, Verbraak F, et al. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the hoorn diabetes care system. Acta Ophthalmol (Copenh). 2018;96(1):63–68. doi: 10.1111/aos.13613

[14] Burlina P, Joshi N, Paul W, et al. Addressing artificial intelligence bias in retinal diagnostics. Transl Vis Sci Technol. 2021;10(2):13. doi: 10.1167/tvst.10.2.13

[15] Raman R, Dasgupta D, Ramasamy K, et al. Using artificial intelligence for diabetic retinopathy screening: policy implications. Indian J Ophthalmol. 2021;69(11):2993. doi: 10.4103/ijo.IJO_1420_21

[16] Larsen TLJ, Jørgensen ME, Pedersen ML, et al. Low prevalence of retinopathy among Greenland inuit. Int J Circumpolar Health. 2021;80(1):1938420. doi: 10.1080/22423982.2021.1938420

[17] Pedersen ML. Diabetes care in the dispersed population of Greenland. A new model based on continued monitoring, analysis and adjustment of initiatives taken. Int J Circumpolar Health. 2019;78(sup1):1709257. doi: 10.1080/22423982.2019.1709257

[18] Kragelund SH, Kjærsgaard M, Jensen-Fangel S, et al. Research Electronic Data Capture (REDCap®) used as an audit tool with a built-in database. J Biomed Inform. 2018;81:112–118. doi: 10.1016/j.jbi.2018.04.005

[19] ResNet-50 convolutional neural network - MATLAB-resnet50 [Published online 2022 Jun 1]. Available from: https://www.mathworks.com/help/deeplearning/ref/resnet50.html;jsessionid=0258cff6eb076f58e2c243382acb

[20] Gheisari S, Shariflou S, Phu J, et al. A combined convolutional and recurrent neural network for enhanced glaucoma detection. Sci Rep. 2021;11(1):1945. doi: 10.1038/s41598-021-81554-4

[21] Supawit V, Suchat T, Kitsuchart P. White blood cell classification: a comparison between CGG16 and ResNet50 models. Published online December 2018.

[22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Published online 2015 Apr 10. cited 2024 Jan 5. http://arxiv.org/abs/1409.1556.

[23] Goodfellow I. Machine learning basics. In: Deep learning. Chapter 5. MIT Press. p. 110130. http://www.deeplearningbook.org

[24] Sanjay M Why and how to cross validate a model? [cited 2021 Apr 5]. Available from: https://towardsdatascience.com/why-and-how-to-cross-validate-a-model-d6424b45261f.

[25] Fully connected layer - MATLAB [cited 2022 Jun 1]. Available from: https://www.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.fullyconnectedlayer.html

[26] Diagnostic-Atlas-Book-Us.Pdf [cited 2022 Jun 1]. Available from: https://www.dewimed.com.mx/assets/diagnostic-atlas-book-us.pdf

[27] Dwivedi P. Understanding and coding a ResNet in Keras. Medium [Published online 2019 Mar 27; cited 2022 Jun 1]. Available from: https://towardsdatascience.com/confusion-matrix-for-your-multiclass-machine-learning-model-ff9aa3bf7826

[28] Mohajon J. Confusion matrix for your multi-class machine learning model. Medium [Published online 2021 Jul 24; cited 2022 Jun 1]. Available from: https://towardsdatascience.com/confusionmatrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826
[29] What is a ROC Curve and how to interpret it [Published online 2018 Jul 5; cited 2022 Jun 1]. Available from: https://www.displayr.com/what-is-a-roc-curve-how-to-interpret-it/
[30] Abràmoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2018;1(1):39. doi: 10.1038/s41746-018-0040-6
[31] He J, Cao T, Xu F, et al. Artificial intelligence-based screening for diabetic retinopathy at community hospital. Eye Lond Engl. 2020;34(3):572–576. doi: 10.1038/s41433-019-0562-4
[32] Scanlon PH. The English national screening programme for diabetic retinopathy 2003–2016. Acta Diabetol. 2017;54(6):515–525. doi: 10.1007/s00592-017-0974-1
[33] Stefánsson E, Bek T, Porta M, et al. Screening and prevention of diabetic blindness. Acta Ophthalmol Scand. 2000;78(4):374–385. doi: 10.1034/j.16000420.2000.078004374.x
[34] Wang Y, Shi D, Tan Z, et al. Screening referable diabetic retinopathy using a semiautomated deep learning algorithm assisted approach. Front Med. 2021;8:740987. doi: 10.3389/fmed.2021.740987
[35] Koch A. Diabetes in Greenland – how to deliver diabetes care in a country with a geographically dispersed population. Int J Circumpolar Health. 2019;78(sup1):1668592. doi: 10.1080/22423982.2019.1668592
[36] Jørgensen ME, Borch-Johnsen K, Witte DR, et al. Diabetes in Greenland and its relationship with urbanization. Diabet Med. 2012;29(6):755–760. doi: 10.1111/j.1464-5491.2011.03527.x
[37] Wang K, Jayadev C, Nittala MG, et al. Automated detection of diabetic retinopathy lesions on ultrawidefield pseudocolour images. Acta Ophthalmol (Copenh). 2018;96(2). doi: 10.1111/aos.13528
[38] Abràmoff MD, Tobey D, Char DS. Lessons learned about autonomous AI: finding a safe, efficacious, and ethical path through the development process. Am J Ophthalmol. 2020;214:134–142. doi: 10.1016/j.ajo.2020.02.022
[39] Moltke I, Fumagalli M, Korneliussen TS, et al. Uncovering the genetic history of the PresentDay Greenlandic population. Am J Hum Genet. 2015;96(1):54–69.
[40] Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. Prog Retin Eye Res. 2019;72:100759. doi: 10.1016/j.preteyeres.2019.04.003