

SCIENTIFIC REPORTS



OPEN

Detection of smoking status from retinal images; a Convolutional Neural Network study

Ehsan Vaghefi^{1,2}, Song Yang¹, Sophie Hill³, Gayl Humphrey⁴, Natalie Walker⁴ & David Squirrell³

Cardiovascular diseases are directly linked to smoking habits, which has both physiological and anatomical effects on the systemic and retinal circulations, and these changes can be detected with fundus photographs. Here, we aimed to 1- design a Convolutional Neural Network (CNN), using retinal photographs, to differentiate between smokers and non-smokers; and 2- use the attention maps to better understand the physiological changes that occur in the retina in smokers. 165,104 retinal images were obtained from a diabetes screening programme, labelled with self-reported “smoking” or “non-smoking” status. The images were pre-processed in one of two ways, either “contrast-enhanced” or “skeletonized”. Experiments were run on an Intel Xeon Gold 6128 CPU @ 3.40GHz with 16 GB of RAM memory and a NVIDIA GeForce TiTan VVOLTA 12 GB, for 20 epochs. The dataset was split 80/20 for training and testing sets, respectively. The overall validation outcomes for the contrast-enhanced model were accuracy 88.88%, specificity 93.87%. In contrast, the outcomes of the skeletonized model were accuracy 63.63%, specificity 65.60%. The “attention maps” that were generated of the contrast-enhanced model highlighted the retinal vasculature, perivascular region and the fovea most prominently. We trained a customized CNN to accurately determine smoking status. The retinal vasculature, the perivascular region and the fovea appear to be important predictive features in the determination of smoking status. Despite a high degree of accuracy, the sensitivity of our CNN was low. Further research is required to establish whether the frequency, duration, and dosage (quantity) of smoking would improve the sensitivity of the CNN.

Cardiovascular disease continues to be the leading cause of death globally¹. One of the most important risk factors in the development of cardiovascular disease is cigarette smoking². Smoking has both a physiological and anatomical effect on the systemic and retinal circulation³. Examining the retinal vasculature with fundus photography provides an exclusive opportunity to directly examine blood vessels non-invasively⁴. Fundus photos can thus be used to identify and monitor the progression of those eye diseases that have a systemic involvement.

Analysis of retinal images has revealed that there a number of biomarkers that are associated with increased cardiovascular risk. These include vessel tortuosity and bifurcation⁵ calibre^{6–12}, microvascular changes^{13,14} and vascular fractal dimensions^{15–17}. Smoking has been shown in a number of epidemiological studies to influence the appearance of the retinal vasculature, resulting in wider retinal venular calibre^{18,19}.

The technique of fundus photography has advanced and evolved rapidly over the last century²⁰, and recently there has been a surge in the use of Deep Learning (DL) to analyse the retinal fundus photographs. DL is a subset of machine learning that involves providing a system with series of labelled examples of images of specific quality, so that the system can train itself to identify predictive features without explicit instructions. Convolutional Neural Networks (CNN) are a class of artificial neural networks, which use a variation of multilayer perceptrons and non-linear activation functions^{21,22}.

Early studies used RIGA and SCES datasets and a custom CNN architecture for classification of optic-disk images to diagnose glaucoma²³. This model was developed further to extract features and classify patients into those with or without glaucoma via a random forest classifier using a transfer learning of AlexNet²⁴. Other

¹School of Optometry and Vision Sciences, University of Auckland, Auckland, New Zealand. ²Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand. ³Department of Ophthalmology, University of Auckland, Auckland, New Zealand. ⁴National Institute for Health Innovation, University of Auckland, Auckland, New Zealand. Correspondence and requests for materials should be addressed to E.V. (email: e.vaghefi@auckland.ac.nz)

Received: 3 January 2019
Accepted: 29 April 2019
Published online: 09 May 2019

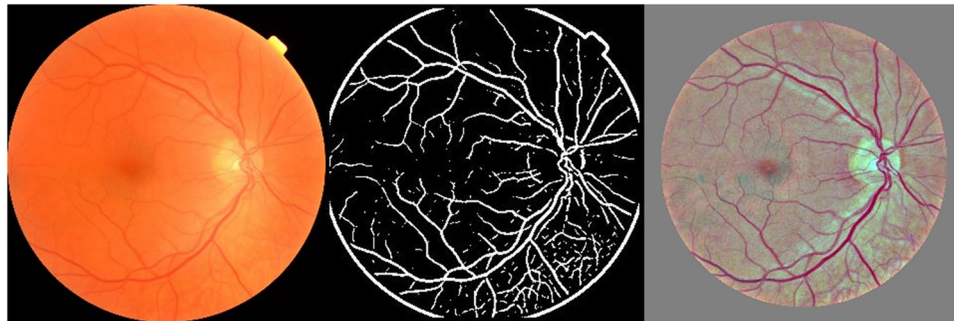


Figure 1. Showing the original (left), skeletonized (centre) and contrast enhanced (right) fundus images.

modifications of AlexNet have been used in the detection of retinal lesions in diabetic retinopathy (DR)^{25,26} or to determine the severity of age-related macular degeneration (AMD) or DR^{27,28}. A more sophisticated network (i.e. Inception-v3) architecture and the EyePACS and Messidor-2 dataset were used to develop and validate the grading of retinal images into normal versus referable DR or referable diabetic macular edema, or both²⁹. Other studies, using a similar approach, have also effectively categorised DR into different grades^{30–32}. Freely available online databases are also generated for the advancement of this field. The freely available STARE database a pretrained VGG19 and AlexNet networks were recently used to classify retinal images into ten different classes of pathologies³³. The efficacy of the freely available DRIVE dataset and a custom CNN and gray-scale thresholding for segmentation of retinal vasculature has also been reported³⁴.

The effectiveness of a CNN to detect cardiovascular risk factors from retinal images has recently been previously demonstrated³⁵. This study used Google Inceptionv3 neural-network architecture to distinguish patient characteristics from retinal images such as age, gender, hypertension, and smoking status³⁵. The last mentioned study has achieved 0.71% (0.70–0.73%) accuracy as measured by the area under the curve (AUC), using not pre-processed fundus photos. Although that study provided ‘attention maps’ to assist with the areas of the training data that were ‘noticed’ by their model, no further conclusion could have been drawn on the potential physiological changes in the ‘noted’ area that had led to CNN’s acquired knowledge.

As the deleterious effects on cardiovascular health in particular are compounded in patients with diabetes, there is a need to develop effective smoking cessation strategies for patients with diabetes who smoke. However in order to test the efficacy of a smoking cessation strategy, one ideally needs an inexpensive and acceptable objective measure of the patients smoking status. In this project we set out to determine whether, using nothing more than the retinal photograph that was obtained when the patient attended for screening, labelled with self-reported smoking status, whether we could build an algorithm capable of detecting whether the individual smoked or not. In this paper, we report the efficacy of our custom-designed CNN for the automated prediction of smoking status, in a self-reported population, using a diabetic retinal screening dataset. Furthermore, by using two pre-processing methods as oppose to common unprocessed fundus photos, we have attempted to create a better understanding of the ‘learned’ knowledge by our CNN and address its ‘black box’ nature.

Methods

The current clinical study was congruent with the ethical principles conveyed in the 2002 version of the Helsinki Declaration and accepted by the Ethical Committee of New Zealand Health and Disability Ethics Committee, reference #18CEN124. The local regulatory authority in New Zealand (National Ethics Advisory Committee) has waived the need for informed consent. After obtaining ethical approval, 165,104 retinal images were obtained from the Auckland Diabetic Eye Screening Database. All patients in this dataset therefore have diabetes, the grade of the diabetic retinopathy being graded to the New Zealand Ministry of Health Diabetic retinopathy standard³⁶. The images had been de-identified and were labelled as “smoking” or “non-smoking” based on the patient’s self-reported smoking status.

The images which were obtained in Auckland Diabetic screening during 2009–2018, were coloured (RGB), in JPG format and resized (320 * 320 pixels) to fit the input criteria of our neural network. Coloured fundus images with one target label: smoking-status (Yes/No) were then split randomly to 60% ‘training set’, 20% ‘validation set’ and 20% ‘test set’. Prior to CNN training, these images were pre-processed using two different filtering methods. Next, the same CNN architecture and hyper-parameters were used for model training (using the ‘training set’) and validation (using the ‘validation set’). Finally, the CNN performance was checked using the not-used-before ‘test set’. The outcome presented in the Results section are from this ‘test set’.

Experiments were run in an Intel Xeon Gold 6128 CPU @ 3.40 GHz with 16 GB of RAM memory and a NVIDIA GeForce TiTan V VOLTA 12 GB, for 20 epochs and training lasted 8 hours. The 20 EPOCHS training-stop criterion was chosen as it was observed that the CNN validation loss (measured as negative log-likelihood and residual sum of squares) has reached a stable minimum over the last 3 EPOCHS of training. Hence, any further training would have led to model ‘over-fitting’, in which the neural network ‘memorizes’ the training examples.

Pre-processing. The images were filtered in two ways (1) “skeletonized”, and (2) “contrast-enhanced” Fig. 1.

The “skeletonization” process was based on a previously published model³⁷. Briefly, after loading the fundus image, the green channel of the image was extracted as it provided the highest contrast between the background and the blood vessels. The resultant grey-scale image was then thresholded to improve the contrast of the blood vessels. Luminance was then inverted so that the blood vessels show up as bright pixels against a dark background in grayscale. A Gaussian filter was then applied to smooth the image. For each pixel in the image, the Hessian and Eigen values of the Hessian (λ_1, λ_2 , where $|\lambda_1| < |\lambda_2|$) were computed. The eigenvalues were used to compute measures:

$$R_B = \frac{\lambda_1}{\lambda_2}$$

and

$$S = \sqrt{\lambda_1^2 + \lambda_2^2}$$

The Vesselness measure for each pixel was then calculated using:

$$V = \begin{cases} 0, & \text{if } \lambda_2 > 0 \\ \exp\left(-\frac{R_B^2}{2\beta^2}\right) \left(1 - \exp\left(-\frac{S^2}{2c^2}\right)\right), & \text{otherwise} \end{cases}$$

where β and c are threshold parameters which control the sensitivity of the Vesselness filter.

The Vesselness measure indicated the probability of a pixel being a vessel. Thresholding was applied and all pixels with a probability higher than the threshold value were assigned as pixels belonging to a vessel Fig. 1B.

“Contrast-enhanced” dataset was obtained using another published method^{38,39}. Here, the following Gaussian filter was applied to the original fundus photo:

$$I_c = \alpha I + \beta G(\rho) * I + \gamma$$

where $*$ denotes the convolution operation, I denotes input image and $G(\rho)$ represents the Gaussian filter with a standard deviation of ρ ¹⁹ Fig. 1C. These images were then normalized to prevent the well-documented CNN “gradient explosion problem”⁴⁰.

CNN model. The CNN architecture that was used in this project is presented below Fig. 2. In short, five convolution layers, five pooling layers and three fully-connected layers composed the main body of our CNN. Batch normalization layers were also added for accelerating converge, dropout and regularization layers were also added to prevent overfitting.

Within the dataset, 85% of the images were self-reported as non-smokers and only 15% as smokers. Imbalanced distribution is common in medical datasets (e.g. healthy vs diseased); and can lead to CNN imbalanced learning⁴¹. In order to address this issue, the smoker-labelled images were augmented and replicated so that each training mini-batch had similar number of smoker and non-smoker images. This data augmentation strategy was not applied during the validation process.

Evaluation metrics. For our model evaluation, we adopted several evaluation parameters. These included accuracy, specificity, sensitivity, and area under receiver operating characteristic curve (AUC).

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TM + FP + FN}$$

FP represents false positive values, where non-smoking images are wrongly classified as smoking and FN represents false negative values where smoking images are wrongly classified as non-smoking. TP represents correctly classified smoking images and TN represents correctly classified non-smoking images.

Attention maps. To better understand how our better-performing (i.e. contrast-enhanced) CNN, attention maps were produced, to identify the areas on the retinal image that had been used as a predictive feature for smoking status. This method is explained in detail elsewhere^{42–44}. Briefly, the output of the first two convolutional layers (layer 1 and layer 5 in Fig. 2) were extracted and averaged for all cases.

Results

165,104 photographs obtained from 81,711 participants were used in this study. The data was drawn from the Auckland diabetes screening program, and all suitable images from screening visits 2008–2018 inclusive were used. Of the cohort of patients from whom these images were drawn from, 7354 (9%) of them identified as being a current smoker. The demographics of the population, and the status of the diabetic retinopathy from which these

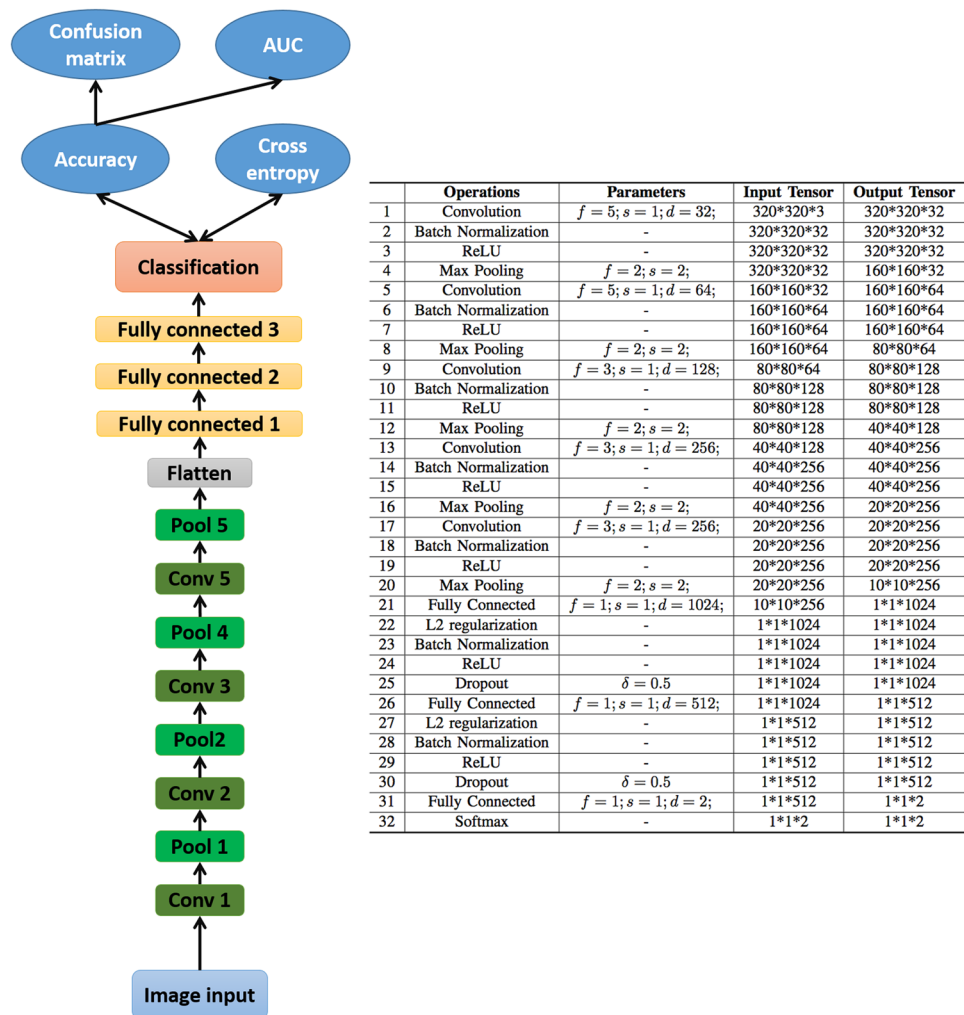


Figure 2. Showing the CNN architecture (left) and each CNN layer property (right).

	Non-smoker (74,357 individuals)	Smoker (7354 individuals)
Age mean (SD)	63.6 ± 16.7	58.8 ± 13.2
Gender (%Male)	52%	66%
HbA1c mmol/mol	64 ± 10.4	64 ± 10.7
Dyslipidaemia on treatment.	61%	63%
Hypertension on treatment	64%	62%
Retinopathy level		
R0	46%	45%
R1	35%	36%
R2	15%	14%
R3	2%	3%
R4-5	2%	2%

Table 1. Demographics of the patients from who the image dataset was derived.

fundus photographs were obtained is detailed in Table 1. 66% of the cohort who identified as smokers were male compared to 52% of the cohort in the cohort who identified as non-smokers, a difference that was statistically significant. Regression analysis revealed that there was also a statistically significant relationship between gender and smoking ($p < 0.001$).

There was no significant difference in the proportion of patients who were being treated for hypertension or dyslipidaemia in the group who self-identified as smokers compared to those who identified as non-smokers. The diabetic control, as assessed by HbA1C taken closest to the date of the screening event, was well matched between

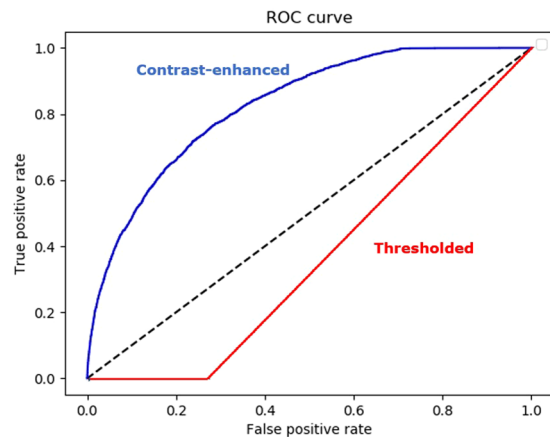


Figure 3. The ROC plots of the contrast-enhanced and thresholded -trained datasets.

groups. Whilst there was a trend for smokers to be younger, this difference was not statistically significant. The majority of patients in both groups had at least mild non proliferative diabetic retinopathy, but there was no difference in the level of retinopathy between the two groups.

Both skeletonized and contrast-enhanced images were used independently, to train and test a CNN model. Using the skeletonized test dataset, our CNN achieved an accuracy of 63.63%, a specificity of 65.60% and a sensitivity of 47.14%. The AUC was 0.58. Using the contrast-enhanced image dataset our CNN produced a superior overall test accuracy 88.88%, and an improved specificity 93.87% and sensitivity of 62.62%. Finally, the AUC was 0.86 Fig. 3.

Representative examples of the attention maps derived from the contrast-enhanced fundus image dataset are shown in Fig. 4. Attention maps were not generated from the skeletonised image dataset as this CNN failed to reach 80% accuracy. Within the attention map the retinal vessels, the perivascular region and the fovea have been highlighted, indicating that the CNN used data from these areas when making its decision on the smoking status of the image under test. The attention maps were similar for all the analysed images and there was no visually identifiable difference between the attention maps derived from the images that were obtained from smokers compared to those from non-smokers.

Discussion

This study has shown that Computational Neural Networks can be utilised to accurately predict smoking status from retinal fundus images.

In highlighting the retinal vasculature, the perivascular region and the fovea, the attention maps derived from the analysis of the contrast-enhanced image dataset, demonstrate that the CNN has identified these regions on the images as being the most important for predicting the smoking status. A wider retinal venular calibre has previously been reported to be linked to smoking^{10,11}, so the finding that the retinal vasculature is highlighted on the attention maps suggests that the CNN is, at least in part, deriving a conclusion based on these structural changes. Similar findings were reported by Poplin, *et al.*³⁵, although a lower accuracy 0.71% (0.70–0.73%) as measured by AUC, and using non pre-processed fundus photos. We believe that the main reason why our network outperformed the previous study has been the use of pre-processed fundus photos, as oppose to unprocessed images. We believe that providing our CNN with ‘contrast enhanced’ features of the perivascular region has assisted with its classification task. The observation that removing this region of fundus photos in our ‘skeletonized’ images led to much poorer performance would further support this hypothesis. Meanwhile, it is interesting to observe that a CNN trained on skeletonised images, a pre-processing method, which reduces the image to a geometric representation of vasculature isolated from the rest of the fundus, is unable to accurately classify the smoking status of the images. It suggests that changes in the architecture of the vasculature, such as vessel calibre^{6–12}, vessel tortuosity and bifurcation⁵, and vascular fractal dimensions^{15–17} alone are not sufficiently strong predictive markers for the accurate detection of the smoking status of an individual who has diabetes.

Whilst it is difficult to identify with any certainty why the skeletonised model failed to accurately predict whether an individual smoked, this model does not make a clear distinction between the retinal arteries and veins and it includes some of the larger choroidal vessels Fig. 1. This lack of clarity between the different components of the retinal vasculature and the inclusion of some of the larger choroidal vessels may have introduced noise into model, which reduced its accuracy. Moreover, the finding that the paravascular area and fovea in the contrast-enhanced images are also important indicates that the CNN is deriving important predictive data from these areas. Both hypertension and dyslipidaemia are associated with well recognised changes within the retinal vasculature^{45–48} and it is therefore possible that the paravascular changes that the algorithm used to classify the images were based, at least in part, on these changes. However, the relationship between chronic cigarette smoking and hypertension is inconclusive with large epidemiological studies concluding that any independent chronic effect of smoking on blood pressure is small^{49,50}. The one exception perhaps being older male smokers had higher systolic BP adjusted for age, BMI, social class, and alcohol intake compared to matched non-smoking peers⁵¹. Moreover, nearly two thirds of patients in our study were on treatment for hypertension, and the proportion of

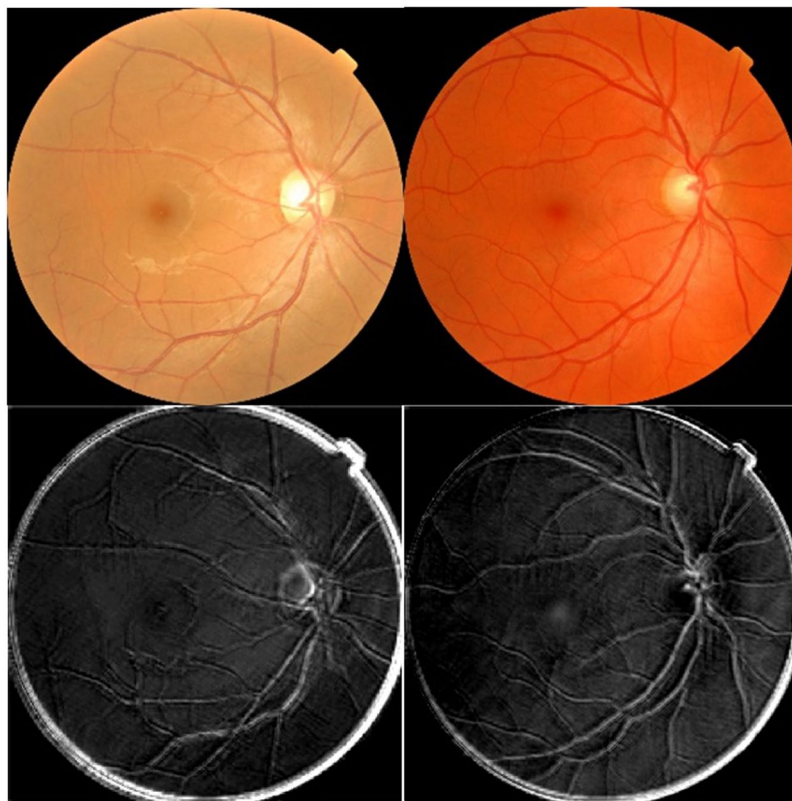


Figure 4. The fundus photos (top row) and attention maps (bottom row) of the enhanced dataset, from a smoker (left) and non-smoker (right) participant, demonstrating the sensitivity of the CNN to the perivascular area.

patients who were being treated for hypertension was similar between those who identified as smokers, compared to those who identified as non-smokers. These data are likely to reflect that patients with diabetes will have their blood pressure checked regularly as part of their regular systemic review and be treated appropriately. One has to acknowledge that an individual being on treatment for hypertension, is not the same as knowing what their blood pressure is. It is therefore still possible that the cohort who identified as smokers had more signs of hypertensive retinopathy compared to non-smokers. However, as the majority of these individuals were regularly being reviewed by their physician, and that the blood pressure targets for each cohort would be similar, it is probably reasonable to assume that the blood pressure control in both cohorts was also similar. The finding that the HbA1C was very similar in each cohort suggests that the management of their diabetes and associated co-morbidities was similar across groups. Resolving this uncertainty would require the retinal images to be labelled with either the blood pressure at the time of image acquisition, and/or whether hypertensive retinopathy was present. However, this data was not available to us so we were unable to test this hypothesis. The possible influence that dyslipidaemia had on the function of our algorithms can be addressed with a very similar set of arguments, but again we did not have the data, which would allow us to interrogate this association any further.

For any cohort of patients with diabetes one potential confounding variable is the proportion of patients with retinopathy of differing levels. Whilst the majority of patients in this study had at least mild non proliferative diabetic retinopathy, the actual proportions of patients who had retinopathy levels R0-R5 were very similar. Moreover, the attention maps suggest that the algorithms were not sensitive to changes within the retina beyond the larger vessels implying that the level of a patients retinopathy had very little influence on the output of the algorithm.

Whilst CNNs are as a rule fairly insensitive to subtle differences in colours, one has to consider the possibility that the colour, or differences in the colour between different components of the fundal image, could be an important discriminator. The oxygen carriage of haemoglobin is reduced in smokers⁵², something that might affect the colour of the blood in the vessels and the subfoveal choroid. The finding that the fovea was also highlighted in the derived attention maps, albeit to a lesser extent than the retinal vasculature and perivascular region was unexpected. Whilst this finding has previously been considered to be a result of the centrality of the fovea in retinal images⁵³, it could also reflect the fact that the CNN was detecting a difference in the colour of the subfoveal choroid in smokers compared to non-smokers. However, in a previous study³⁵, the fovea was highlighted on the attention maps that predicted gender. How the CNN was able to accurately predict gender is unknown, but we know that the central macular thickness, as measured by OCT, is thicker in males compared to females, so it is conceivable that the CNN was able to discern this difference⁵⁴. One therefore also has to consider the possibility that there is an inherent bias in our dataset with smoking being unbalanced between females and males and that our CNN is, at least in part, simply reporting this difference. The finding that a greater proportion of our cohort

who identified as being a current smoker were male is in keeping with data obtained from the 2013 New Zealand census which found that smoking prevalence was higher in men (16.4%) than women (13.9%)⁵⁵. It is however of note that the prevalence of smoking in our cohort of patients with diabetes who smoked (9%) was actually less than the national average (15%).

It seemed in our data that there was a significant relationship between self-reported smoking and gender. It is possible then that the algorithm was sensitive to gender when making a judgement about the smoking status of an individual, particularly when labelling them a non-smoker. However, the observation that the attention maps were not focused solely on the fovea strongly suggests that the algorithm was not making judgements based on gender alone. Repeating the study with a cohort of smokers and non-smokers who are matched for gender would potentially help address the question as to what influence gender has on the algorithm.

These findings demonstrates both the utility of attention maps to assess which factors the algorithm is using to determine its judgement, but they also highlight the need to be cognisant of the fact that there may be unexpected factors that are powerful confounding variables which may influence the algorithms behaviour. At the extreme, the CNN could even be using a different, but strongly related variable, as a surrogate for that factor you are using the CNN to make a judgement on. This phenomenon probably explains why many apparently well functioning algorithms perform poorly when presented with a dataset derived from a different population, in which these co-variables are inadvertently balanced differently.

Despite the possible anatomical correlations between the predictive features identified on the activation maps and the previously described vascular changes identified in the retinal vasculature of smokers, neither these previous results nor our current data could prove causation. Given the significant number of variables likely to have been analysed by our CNN, and that fact that some of these factors may be unknown to us, in reality we can only speculate on the predictive features used by our CNN for determining smoking status identification in this dataset.

The strengths of this study include the large number of images that were analysed and the use of validated pre-processing image methods. Potential limitations include the imbalanced distribution of smoking and non-smoking images within the dataset and the fact that we were not in a position to balance other cofounding variables, such as gender, that could effected the way the algorithm behaved. This could have led to sampling bias. Data augmentation has been used commonly in DL to address the imbalanced data issue and has also been implemented here. Using this technique, similar number of non-smoking and (augmented) smoking fundus images were included in each min-batch of the CNN training process.

Finally, all the images analysed in the study were taken from a diabetic retinopathy screening database. Patients with diabetics are known to have retinal vascular changes related to the duration and severity of disease^{56,57} and these diseases relate alterations may also have confounded our analysis of the images. However, smoking has previously been found to have one of the largest influences on retinal vessel calibre, independent of other factors in large study evaluating retinal vasculature changes in a diabetic population⁵⁸.

While our CNN has high degree of accuracy and specificity, it suffers from low sensitivity. In other words, the model could distinguish non-smokers from smokers with high degree of accuracy but did not perform as well in identifying smokers. Previous studies have reported that the vascular calibre changes are greatest in those who have the highest number of “pack years”⁵⁸. It is therefore likely that the influence of smoking on the vasculature could be both subtle and cumulative, and as such, these changes may not be apparent until the individual’s smoking habit exceeds a certain threshold. Furthermore, confirmation of smoking status in our study was both self-reported and binary; yes/no. Since our data lacked the frequency, duration, or extent (dose) of an individual smoking behaviour, it is very possible that our CNN was not able to detect with any reliability those individuals whose smoking behaviour was either below a given threshold or were ex-smokers. More information regarding duration and dosage of smoking, including ex-smokers, would have allowed a gradient association pattern between the smoking “dose” and the retinal vasculature to have been analysed.

In summary, we have demonstrated that a CNN analysis of image enhanced retinal photographs can determine smoking status with a high degree of accuracy. Further research is required to improve the accuracy and sensitivity of the model by controlling for more potential confounding variables including ex-smoking status, number of cigarettes smoked and frequency. Further exploration of whether this technique can be used to determine other cardiovascular risk factors from retinal images will require access to data-sets that are gathered from both the general population (i.e. arguably largely healthy individuals) as well as those that are derived from health-care based systems.

Data Availability

The data that support the findings of this study are available from Auckland Diabetic Eye Screening Database, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Auckland Diabetic Eye Screening Database.

References

1. Organisation, W. H. *The Top 10 Causes of Death*, <http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (2017).
2. Benowitz, N. L. The role of nicotine in smoking-related cardiovascular disease. *Prev Med* **26**, 412–417, <https://doi.org/10.1006/pmed.1997.0175> (1997).
3. Esen, A. M. *et al.* Effect of smoking on endothelial function and wall thickness of brachial artery. *Circ J* **68**, 1123–1126 (2004).
4. Saine, P. J. & Tyler, M. E. *Ophthalmic photography: retinal photography, angiography, and electronic imaging*. Vol. 132 (Butterworth-Heinemann Boston, 2002).
5. Witt, N. *et al.* Abnormalities of retinal microvascular structure and risk of mortality from ischemic heart disease and stroke. *Hypertension* **47**, 975–981, <https://doi.org/10.1161/01.HYP.0000216717.72048.6c> (2006).

6. Wang, J. J. *et al.* Retinal vascular calibre and the risk of coronary heart disease-related death. *Heart* **92**, 1583–1587, <https://doi.org/10.1136/hrt.2006.090522> (2006).
7. Wong, T. Y. *et al.* Quantitative retinal venular caliber and risk of cardiovascular disease in older persons: the cardiovascular health study. *Arch Intern Med* **166**, 2388–2394, <https://doi.org/10.1001/archinte.166.21.2388> (2006).
8. Seidelmann, S. B. *et al.* Retinal Vessel Calibers in Predicting Long-Term Cardiovascular Outcomes: The Atherosclerosis Risk in Communities Study. *Circulation* **134**, 1328–1338, <https://doi.org/10.1161/CIRCULATIONAHA.116.023425> (2016).
9. Wong, T. Y. *et al.* Retinal vascular caliber, cardiovascular risk factors, and inflammation: the multi-ethnic study of atherosclerosis (MESA). *Invest Ophthalmol Vis Sci* **47**, 2341–2350, <https://doi.org/10.1167/iovs.05-1539> (2006).
10. McGeechan, K. *et al.* Meta-analysis: retinal vessel caliber and risk for coronary heart disease. *Annals of internal medicine* **151**, 404–413 (2009).
11. McGeechan, K. *et al.* Prediction of incident stroke events based on retinal vessel caliber: a systematic review and individual-participant meta-analysis. *Am J Epidemiol* **170**, 1323–1332, <https://doi.org/10.1093/aje/kwp306> (2009).
12. Wong, T. Y. *et al.* Retinal arteriolar narrowing and risk of coronary heart disease in men and women. The Atherosclerosis Risk in Communities Study. *JAMA* **287**, 1153–1159 (2002).
13. Wong, T. Y. *et al.* Retinal microvascular abnormalities and 10-year cardiovascular mortality: a population-based case-control study. *Ophthalmology* **110**, 933–940, [https://doi.org/10.1016/S0161-6420\(03\)00084-8](https://doi.org/10.1016/S0161-6420(03)00084-8) (2003).
14. Cheung, C. Y. *et al.* Retinal microvascular changes and risk of stroke: the Singapore Malay Eye Study. *Stroke* **44**, 2402–2408, <https://doi.org/10.1161/STROKEAHA.113.001738> (2013).
15. Liew, G. *et al.* Fractal analysis of retinal microvasculature and coronary heart disease mortality. *Eur Heart J* **32**, 422–429, <https://doi.org/10.1093/eurheartj/ehq431> (2011).
16. Kawasaki, R. *et al.* Fractal dimension of the retinal vasculature and risk of stroke: a nested case-control study. *Neurology* **76**, 1766–1767, <https://doi.org/10.1212/WNL.0b013e31821a7d7d> (2011).
17. Cheung, C. Y. *et al.* Retinal vascular fractal dimension and its relationship with cardiovascular and ocular risk factors. *Am J Ophthalmol* **154**, 663–674 e661, <https://doi.org/10.1016/j.ajo.2012.04.016> (2012).
18. Ikram, M. K. *et al.* Are retinal arteriolar or venular diameters associated with markers for cardiovascular disorders? The Rotterdam Study. *Invest Ophthalmol Vis Sci* **45**, 2129–2134 (2004).
19. Kifley, A., Wang, J. J., Cugati, S., Wong, T. Y. & Mitchell, P. Retinal vascular caliber, diabetes, and retinopathy. *Am J Ophthalmol* **143**, 1024–1026, <https://doi.org/10.1016/j.ajo.2007.01.034> (2007).
20. Abramoff, M. D., Garvin, M. K. & Sonka, M. Retinal imaging and image analysis. *IEEE reviews in biomedical engineering* **3**, 169–208 (2010).
21. Deng, L. & Yu, D. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing* **7**, 197–387 (2014).
22. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436 (2015).
23. Chen, X., Xu, Y., Wong, D. W. K., Wong, T. Y. & Liu, J. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. 715–718 (IEEE) (2015).
24. Muhammad, H. *et al.* Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *Journal of glaucoma* **26**, 1086–1094 (2017).
25. Abramoff, M. D. *et al.* Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning. *Invest Ophthalmol Vis Sci* **57**, 5200–5206, <https://doi.org/10.1167/iovs.16-19964> (2016).
26. Gargeya, R. & Leng, T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology* **124**, 962–969, <https://doi.org/10.1016/j.ophtha.2017.02.008> (2017).
27. Burlina, P. M. *et al.* Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA ophthalmology* **135**, 1170–1176 (2017).
28. Abbas, Q., Fondon, I., Sarmiento, A., Jimenez, S. & Alemany, P. Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. *Med Biol Eng Comput* **55**, 1959–1974, <https://doi.org/10.1007/s11517-017-1638-6> (2017).
29. Gulshan, V. *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama* **316**, 2402–2410 (2016).
30. Takahashi, H., Tampo, H., Arai, Y., Inoue, Y. & Kawashima, H. Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy. *PLoS One* **12**, e0179790 (2017).
31. Zhou, K. *et al.* In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2724–2727 (IEEE) (2018).
32. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. In *AAAI*. 12 (2017).
33. Choi, J. Y. *et al.* Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database. *PLoS One* **12**, e0187336 (2017).
34. Li, Q. *et al.* A Cross-Modality Learning Approach for Vessel Segmentation in Retinal Images. *IEEE Trans. Med. Imaging* **35**, 109–118 (2016).
35. Poplin, R. *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering* **2**, 158–164, <https://doi.org/10.1038/s41551-018-0195-0> (2018).
36. Ministry of Health, G. o. N. Z. Ministry of Health. 2016. *Diabetic Retinal Screening, Grading, Monitoring and Referral Guidance*. Wellington: Ministry of Health, <https://www.health.govt.nz/system/files/documents/publications/diabetic-retinal-screening-grading-monitoring-referral-guidance-mar16.pdf> (2016).
37. Frangi, A. F., Niessen, W. J., Vincken, K. L. & Viergever, M. A. editors In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (ed Springer) (1998).
38. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2818–2826 (2016).
39. Burlina, P., Pacheco, K. D., Joshi, N., Freund, D. E. & Bressler, N. M. Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis. *Computers in biology and medicine* **82**, 80–86 (2017).
40. Pascanu, R., Mikolov, T. & Bengio, Y. In *International conference on machine learning*. 1310–1318 (2013).
41. Mazurowski, M. A. *et al.* Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. *Neural Netw* **21**, 427–436, <https://doi.org/10.1016/j.neunet.2007.12.031> (2008).
42. Krizhevsky, A., Sutskever, I. & Hinton, G. E. In *25th Conference on Advances in Neural Information Processing Systems*. (ed Pereira, F. *et al.*) 1097–1105 (2012).
43. Cho, K., Courville, A. & Bengio, Y. Describing multimedia content using attention-based encoder–decoder networks. *IEEE Trans. Multimed* **17**, 1875–1886 (2015).
44. Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate, <https://arxiv.org/abs/1409.0473> (2014).
45. Omae, T., Nagaoka, T. & Yoshida, A. Effects of habitual cigarette smoking on retinal circulation in patients with type 2 diabetes. *Investigative ophthalmology & visual science* **57**, 1345–1351 (2016).
46. Mondal, R. N. *et al.* Journal of Hypertension: Open Access (2017).
47. Modi, P. & Arsiwalla, T. In *StatPearls [Internet]* (StatPearls Publishing, 2018).

48. Nagasree, D. & Rachakonda, R. Study of Prevalence of Diabetic Retinopathy and Correlation with Risk Factors. *Alcohol* **210**, 9 (2018).
49. Green, M. S., Jucha, E. & Luz, Y. Blood pressure in smokers and nonsmokers: epidemiologic findings. *American heart journal* **111**, 932–940 (1986).
50. Pankova, A. *et al.* No difference in hypertension prevalence in smokers, former smokers and non-smokers after adjusting for body mass index and age: a cross-sectional study from the Czech Republic, 2010. *Tobacco induced diseases* **13**, 24 (2015).
51. Primates, P., Falaschetti, E., Gupta, S., Marmot, M. G. & Poulter, N. R. Association between smoking and blood pressure: evidence from the health survey for England. *Hypertension* **37**, 187–193 (2001).
52. Sagone, A. L., Lawrence, T. & Balcerzak, S. P. Effect of Smoking on Tissue Oxygen Supply. *Blood* **41**, 845 (1973).
53. Varadarajan, A. V. *et al.* Deep Learning for Predicting Refractive Error From Retinal Fundus Images. *Invest Ophthalmol Vis Sci* **59**, 2861–2868, <https://doi.org/10.1167/jovs.18-23887> (2018).
54. Adhi, M., Aziz, S., Muhammad, K. & Adhi, M. I. Macular thickness by age and gender in healthy eyes using spectral domain optical coherence tomography. *PloS One* **7**, e37638–e37638, <https://doi.org/10.1371/journal.pone.0037638> (2012).
55. Tu, D., Newcombe, R., Edwards, R. & Walton, D. Socio-demographic characteristics of New Zealand adult smokers, ex-smokers and non-smokers: results from the 2013 Census. *NZ Med J* **129**, 43–56 (2016).
56. Kifley, A., Wang, J. J., Cugati, S., Wong, T. Y. & Mitchell, P. Retinal vascular caliber and the long-term risk of diabetes and impaired fasting glucose: the Blue Mountains Eye Study. *Microcirculation* **15**, 373–377, <https://doi.org/10.1080/10739680701812220> (2008).
57. Tsai, A. S. *et al.* Differential association of retinal arteriolar and venular caliber with diabetes and retinopathy. *Diabetes Res Clin Pract* **94**, 291–298, <https://doi.org/10.1016/j.diabres.2011.07.032> (2011).
58. Klein, R., Klein, B. E., Moss, S. E., Wong, T. Y. & Sharrett, A. R. Retinal vascular caliber in persons with type 2 diabetes: the Wisconsin Epidemiological Study of Diabetic Retinopathy: XX. *Ophthalmology* **113**, 1488–1498, <https://doi.org/10.1016/j.ophtha.2006.03.028> (2006).

Acknowledgements

This study was made possible by a grant from the University of Auckland Performance Based Research Fund (PBRF).

Author Contributions

Dr. Ehsan Vaghefi performed the full analysis, supervised Mr. Song, and wrote bulk of the manuscript including the Methods and Results. Mr. Song Yang performed part of the analysis. Dr. Sophie Hill wrote part of the manuscript, including Introduction and Discussion. Dr. Gayl Humphrey proposed the original study and obtained funding support. Dr. Natalie Walker proposed the original study and obtained funding support. Dr. David Squirrell supervised the entire project and revised the final manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019