Expert Review of Molecular Diagnostics

ISSN: 1473-7159 (Print) 1744-8352 (Online) Journal homepage: https://www.tandfonline.com/loi/iero20

# A CNN-aided method to predict glaucoma progression using DARC (Detection of Apoptosing Retinal Cells)

Eduardo M Normando, Tim E Yap, John Maddison, Serge Miodragovic, Paolo Bonetti, Melanie Almonte, Nada Mohammad, Sally Ameen, Laura Crawley, Faisal Ahmed, Philip A Bloom & Maria Francesca Cordeiro

**To cite this article:** Eduardo M Normando, Tim E Yap, John Maddison, Serge Miodragovic, Paolo Bonetti, Melanie Almonte, Nada Mohammad, Sally Ameen, Laura Crawley, Faisal Ahmed, Philip A Bloom & Maria Francesca Cordeiro (2020): A CNN-aided method to predict glaucoma progression using DARC (Detection of Apoptosing Retinal Cells), Expert Review of Molecular Diagnostics, DOI: 10.1080/14737159.2020.1758067

**To link to this article:** https://doi.org/10.1080/14737159.2020.1758067

Accepted author version posted online: 20 Apr 2020.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=iero20

Check for updates

**Publisher:** Taylor & Francis & Informa UK Limited, trading as Taylor & Francis Group

**Journal:** *Expert Review of Molecular Diagnostics*

**DOI:** 10.1080/14737159.2020.1758067

Title

# A CNN-aided method to predict glaucoma progression using DARC (Detection of Apoptosing Retinal Cells)

Authors

Eduardo M Normando MD PhD, [1, 2] Tim E Yap MBBS,[1, 2] John Maddison PhD,[3] Serge Miodragovic,[1] Paolo Bonetti,[1] Melanie Almonte MSc, [1] Nada Mohammad MD, [1] Sally Ameen MBBS,[1] Laura Crawley MBBS, [1] Faisal Ahmed MBBS, [1] Philip A Bloom MBBS, [1,2] and Maria Francesca Cordeiro MBBS PhD[1,2,4]#

[1] ICORG, Imperial College London

[2] Western Eye Hospital Imperial College Healthcare NHS Trust, London

[3] Maddisys Ltd,

[4] UCL Institute of Ophthalmology, London

# Corresponding author: Prof M F Cordeiro, UCL Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL

ACCEPTED MANUSCRIPT

**Original Research**

# A CNN-aided method to predict glaucoma progression using DARC (Detection of Apoptosing Retinal Cells)

ACCEPTED MANUSCRIPT


## Abstract

Objectives: A key objective in glaucoma is to identify those at risk of rapid progression and blindness. Recently, a novel first-in-man method for visualising apoptotic retinal cells called DARC (Detection-of-Apoptosing-Retinal-Cells) was reported. The aim was to develop an automatic CNN-aided method of DARC spot detection to enable prediction of glaucoma progression.

Methods: Anonymised DARC images were acquired from healthy control (n=40) and glaucoma (n=20) Phase 2 clinical trial subjects (ISRCTN10751859) from which 5 observers manually counted spots. The CNN-aided algorithm was trained and validated using manual counts from control subjects, and then tested on glaucoma eyes.

Results: The algorithm had 97.0% accuracy, 91.1% sensitivity and 97.1% specificity to spot detection when compared to manual grading of 50% controls. It was next tested on glaucoma patient eyes defined as progressing or stable based on a significant ($p<0.05$) rate of progression using OCT-retinal nerve fibre layer measurements at 18 months. It demonstrated 85.7% sensitivity, 91.7% specificity with AUC of 0.89, and a significantly ($p=0.0044$) greater DARC count in those patients who later progressed.

Conclusion: This CNN-enabled algorithm provides an automated and objective measure of DARC, promoting its use as an AI-aided biomarker for predicting glaucoma progression and testing new drugs.

ACCEPTED MANUSCRIPT

## 3. Background

Artificial intelligence (AI) is increasingly used in healthcare, especially ophthalmology[1][2]. Machine learning algorithms have become important analytical aids in retinal imaging, being frequently advocated in the management of diabetic retinopathy, age-related macular degeneration and glaucoma, where their utilization is believed to optimise both sensitivity and specificity in diagnosis and monitoring[3][4]–[7]. The use of deep learning in these blinding conditions has been heralded as an advance to reduce their health and socio-economic impact, although their accuracy is confounded by dataset size and deficient reference standards[4].

Glaucoma is a progressive and slowly evolving ocular neurodegenerative disease that it is the leading cause of global irreversible blindness, affecting over 60.5 million people, predicted to double by 2040, as the aging population increases[8], [9]. A key objective in glaucoma research over the last few years is to identify those at risk of rapid progression and blindness. This has included, methods involving multiple levels of data including structural (optical coherence tomography (OCT), disc imaging) and functional (visual fields or standard automated perimetry (SAP)) assessments. However, several studies have demonstrated there is great variability amongst clinicians in agreement over progression using standard assessments including SAP, OCT and optic disc stereo photography. [10][11][3][12] However, clinical grading is regarded as the gold standard in real world practice. In deep learning datasets, manual grading of retinal images is regarded as the "ground truth" and is essential in order to train and test AI strategies in the automated detection of diseases such as glaucoma. [13]–[19] Moreover, it is recognised that both OCT and SAP change only after significant death of a large number of retinal ganglion cells (RGC), [20] and with this the unmet need for earlier markers of disease.

Recently, we reported a novel method to visualise apoptotic retinal cells in the retina in humans called DARC (Detection of Apoptosing Retinal Cells)[21]. The molecular marker used in the

ACCEPTED MANUSCRIPT

technology is fluorescently labelled annexin A5, which has a high affinity for phosphatidylserine exposed on the surface of cells undergoing stress and in the early stages of apoptosis. The published Phase 1 results suggested that the number of DARC positively stained cells seen in a retinal fluorescent image could be used to assess glaucoma disease activity, but also correlated with future glaucoma disease progression, albeit in small patient numbers. DARC has recently been tested in more subjects in a Phase 2 clinical trial (ISRCTN10751859).

The aim of this study was to devise an automatic method of DARC spot detection using a convolutional neural network (CNN). CNNs have shown strong performance in computer vision tasks in medicine, including medical image classification. This CNN-aided algorithm was to be developed by training on a control cohort of subjects and then tested on glaucoma patients in the Phase 2 clinical trial of DARC.

ACCEPTED MANUSCRIPT

# 4. Research design and methods

## 4.1 Study population and setting

The Phase 2 clinical trial of DARC was conducted at The Western Eye Hospital, Imperial College Healthcare NHS Trust, as a single-centre, open-label study with subjects each receiving a single intravenous injection of fluorescent annexin 5 (ANX776, 0.4 mg) between 15th February 2017 and 30th June 2017. Both healthy and progressing glaucoma subjects were recruited to the trial, with informed consent being obtained according to the Declaration of Helsinki after the study was approved by the Brent Research Ethics Committee. (ISRCTN10751859).

All glaucoma subjects were already under the care of the glaucoma department at the Western Eye Hospital. Patients were considered for inclusion in the study if no ocular or systemic disease other than glaucoma was present and they had a minimum of three recent, sequential assessments with retinal optical coherence tomography (Spectralis SD OCT, software version 6.0.0.2; Heidelberg Engineering, Inc., Heidelberg, Germany) and standard automated perimetry (SAP, HFA 640i, Humphrey Field Analyzer; Carl Zeiss Meditec, Dublin, CA) using the Swedish interactive threshold algorithm standard 24-2. Patient eligibility was deemed possible if evidence of progressive disease in at least one eye of any parameter summarised in Table 1 & 2, was found to be present, where progression was defined by a significant (*$p<0.05$; **$p<0.01$) negative slope in the rate of progression (RoP). SAP parameters included the visual field index (VFI) and mean deviation (MD). OCT parameters included retinal nerve fibre layer (RNFL) measurements at three different diameters from the optic disc (3.5, 4.1, and 4.7 mm) and Bruch's membrane opening minimum rim width (MRW). Where it was not possible to use machine in-built software to define the rate of progression, due to the duration of the pre-intervention period of assessment, linear rates of change of each parameter with time were computed using ordinary least squares[22], [23]. Inclusion of these features of progression was to ensure that only patients with active disease were included in the study, especially as the majority of eyes had early disease, with a diagnosis of glaucoma suspect in 23 eyes of 12 patients, as indicated in Table 1.

Healthy volunteers were initially recruited from people escorting patients to clinics and referrals from local optician services who acted as Patient Identification Centres (PICs). Healthy volunteers were also recruited from the Imperial College Healthcare NHS Trust healthy volunteers database. Potential participants were approached and given an invitation letter to participate. Participants at PICs who agreed to be contacted were approached by the research team and booked an appointment to discuss the trial. Enrolment was performed once sequential participants were considered eligible, according to the inclusion and exclusion criteria detailed in Table S1. Briefly, healthy subjects were included if: there was no ocular or systemic disease, as confirmed by their GP;

ACCEPTED MANUSCRIPT

there was no evidence of any glaucomatous process either with optic disc, RNFL (retinal nerve fibre layer) or visual field abnormalities and with normal IOP (intraocular pressure); and they had repeatable and reliable imaging and visual fields.

## 2.2 DARC Image Acquisition and Study Blinding

All participants received a single dose of 0.4mg of ANX776 via intravenous injection following pupillary dilatation (1% tropicamide and 2.5% phenylephrine), and were assessed using a similar protocol to Phase 1[21]. Briefly, retinal images were acquired using a cSLO (HRA+OCT Spectralis, Heidelberg Engineering GmbH, Heidelberg, Germany) with ICGA infrared fluorescence settings (diode laser 786 nm excitation; photodetector with 800-nm barrier filter) in the high resolution mode. Baseline infrared autofluorescent images were acquired prior to ANX776 administration, and then during and after ANX776 injection at 15, 120 and 240 minutes. Averaged images from sequences of 100 frames were recorded at each time point. All images were anonymised before any analysis was performed. For the development of the CNN-algorithm, only baseline and 120 minute images from control and glaucoma subjects were used.

The breakdown of the images analysed are shown in the "Consort" diagram in Figure 1. For the CNN- training, 73 control eyes at baseline and 120 minutes were available for the analysis. Similarly, of the 20 glaucoma patients who received intravenous ANX776, images were available for 27 eyes at baseline and 120 time-points.

## 2.4 DARC Spot Detection

ACCEPTED MANUSCRIPT

### 2.3.1 Manual observer analysis

Anonymised images were randomly displayed on the same computer and under the same lighting conditions, and manual image review was performed by five blinded operators using ImageJ® (National Institutes of Mental Health, USA). [24] The ImageJ 'multi-point' tool was used to identify each structure in the image which observers wished to label as an ANX776 positive spot. Each positive spot was identified by a vector co-ordinate. Manual observer spots for each image were compared: spots from different observers were deemed to be the same spot if they were within 30 pixels of one another. A breakdown of concordance including interobserver agreement is provided in Figure S2, with more than 50% of candidate spots having agreement with more than 2 observers. The criteria for spots used in the automated application to train and compare the systems was when there was concordance of two or more observers.

### 2.3.2 Automated analysis

#### *2.3.2.1 Automated Image Analysis Overview (Figure 2)*

To detect the DARC labelled cells, candidate spots were identified in the retinal images, then classified as "DARC" or "not DARC" using an algorithm trained using the candidates and the spots identified by manual observers. Figure 2 provides an overview of the process.

#### *2.3.2.2 Image Optimisation*

As part of the automated image analysis pipeline, images at 120 minutes were aligned to the baseline image for each eye using an affine transformation followed by a non-rigid

ACCEPTED MANUSCRIPT

transformation. Images were then cropped to remove alignment artefacts. The cropped images then had their intensity standardised by Z-Scoring each image to allow for lighting differences. Finally, the high-frequency noise was removed from the images with a Gaussian blur with a sigma of 5 pixels.

#### *2.3.2.3 Spot Candidate Detection*

Template matching, specifically Zero Normalised Cross-Correlation (ZNCC) is a simple method to find candidate spots. 30x30 pixel images of the spots identified by manual observers were combined using a mean image function to create a spot template. This template was applied to the retinal image producing a correlation map. Local maxima were then selected and filtered with thresholds for the correlation coefficient and intensity standard deviation (corresponding to the brightness of the spot). These thresholds were set low enough to include all spots seen by manual observers. Some of the manual observations were very subtle (arguably not spots at all) and correlation low for quite distinct spots due to their proximity to blood vessels. This means the thresholds needed to be set very low and produce many more spot candidates than manually observed spots (approximately 50-1).

As can be seen from Figure 3, the spot candidates cover much of the retinal image, however this reduces the number of points to classify by a factor of 1500 (compared with looking at every pixel). Using local maxima of the ZNCC, each candidate detection is centred on a spot-like object, typically with the brightest part in the centre. This means the classifier does not have to be tolerant to off-centred spots. It also means that the measured accuracy of the classifier will be more meaningful as it reflects its ability to discern DARC spots from other spot-like objects, not just its ability to discern DARC spots from random parts of the image.

#### *2.3.2.4 Spot Classification*

ACCEPTED MANUSCRIPT

To determine which of the spot candidates were DARC cells, the spots were classified using an established Convolutional Neural Network (CNN) called MobileNet v2[25]–[28]. This CNN enables over 400 spot images to processed in a single batch. This allows it to cope with the 50-1 unbalanced data since each batch should have about 4 DARC spots.

Although the MobileNet v2 architecture was used, the first and last layers were adapted. The first layer became a 64x64x1 input layer to take the 64x64 pixel spot candidate images (this size was chosen to include more of the area around the spot to give the network some context). The last layer was replaced with a dense layer with sigmoid activation to enable a binary classification (DARC spot or not) rather than multiple classification. An alpha value for MobileNet of 0.85 was found to work best, appropriately adjusting the number of filters in each layer

#### *2.3.2.5 Training*

CNN training and validation sets were split by eye then trained at spot level. Training of spot identification was performed only on spots from 50% of control eyes; the remaining 50% were then used for validation/testing. Final testing was done on the whole glaucoma dataset to avoid over-training given the relatively small sample numbers. All training, validation and testing were performed at spot level and not image or eye level, although the selection of train/test was done at an eye level, again to prevent over-training. Briefly, retinal images were randomly selected from baseline and 120 minute images of 50% of the control patients. The CNN was trained using candidate spots, marked as DARC if 2 or more manual observers observed the spot. 58,730 spot candidates were taken from these images (including 985 2-agree manually observed DARC spots). 70% of these spots were used to train, and 30% to validate. The retinal images of the remaining 50% of control patients were used to test the classification accuracy (48610 candidate spots of which 898 were 2-agree manually observed).

ACCEPTED MANUSCRIPT

The data was augmented to increase the tolerance of the network by rotating, reflecting and varying the intensity of the spot images. The DARC spots class weights were set to 50 for spots and 1 for other objects to compensate for the 50-1 unbalanced data.

The training validation accuracy converges, and the matching validation accuracy also shows similar accuracy without signs of over training. As the training curves show (see Figure 4 ) a good accuracy is achieved in 200 epochs, although training was left for 300 epochs to verify stability.

Three training runs were performed, creating three CNN models. For inference, the three models were combined: each spot was classified based on the mean probability given by each of the three models.

#### *2.3.2.6 Testing on Glaucoma DARC images*

Once the CNN-aided algorithm was developed, it was tested on the glaucoma cohort of patients in images captured at baseline and 120 minutes. Spots were identified by manual observers and the algorithm. The DARC count was defined as the number of a ANX776-positive spots seen in the retinal image at 120 minutes after baseline spot subtraction.

## 2.4 Glaucoma progression assessment – the comparator

Rates of progression were computed from serial OCTs on glaucoma patients 18 months after DARC, where progression was defined by a significant (*p<0.05; **p<0.01) negative slope RoP greater than 1 µm/year, based on the 5% lower limit for age-related change of −0.92 µm/year identified by Wu et al [29] to distinguish between normal aging losses[30][31]. All OCTs were

ACCEPTED MANUSCRIPT

assessed for quality at the time of imaging; where motion artefact or other issues were identified, patients were rescanned at the same visit. Those patients with a significant ($p<0.05$) negative slope were defined as progressing compared to those without who were defined as stable. Additionally, assessment was performed by 5 masked expert clinicians using visual field (including VFI, MD, and PSD), OCT (including RNFL and MRW parameters), optic disc clinical observations (cup-disc ratio on biomicroscopy), and documented treatment changes.

## 2.5 Main Outcome Measures & Statistical Analysis

Statistical analysis was performed using GraphPad Prism (version 8.01), SPSS (version 25.0) and Python. Receiver-operating characteristic (ROC) curves were constructed with the area under the curve (AUC), standard errors, maximal sensitivities and specificities generated for CNN training, validation and testing data and comparisons with manual observers counts. Interobserver agreements were calculated for manual observer DARC counts and clinician progression status using Cohen's kappa coefficient. Rates of progression (RoP) were calculated from serial tests of visual field MDs and VFIs, and OCT RNFL and MRW parameters and progression was defined by a significant (*$p<0.05$; **$p<0.01$) negative slope. Where it was not possible to use machine in-built software to define RoP due to the duration of the pre-intervention period of assessment, linear rates of change of each parameter with time were computed using ordinary least squares. Comparisons between stable and progressing glaucoma eyes were made with DARC counts, age, CCT, BP, baseline MD, VFI, RNFL and corresponding abnormal RNFL and BMO-MRW sectors using the unpaired t-test ((*$p<0.05$; **$p<0.01$).

ACCEPTED MANUSCRIPT

# 3 Results

## 3.1 Patient Demographics

Demographics of glaucoma and control subjects are shown in Table supp2. 60 glaucoma patients were screened according to the inclusion/exclusion criteria in Table 1, from which 20 patients with progressing (defined by a significant ($p<0.05$) negative slope in any parameter in at least one eye) glaucoma underwent intravenous DARC. Baseline characteristics of these glaucoma patients are presented in Table 2. 38 eyes were eligible for inclusion, of which 3 did not have images available for manual observer counts, 2 had images captured in low resolution mode and another 2 had intense intrinsic autofluorescence. All patients apart from 2 were followed up in the Eye clinic, with data being available to perform a post hoc assessment of progression.

## 3.2 Testing of Spot Classification

The results in Figure 4 were achieved when testing the CNN-aided algorithm with the 50% of the control eyes that were reserved for test (and so were not used in training). The accuracy was found to be 97%, with 91.1% sensitivity and 97.1% specificity.

The sensitivity and specificity were encouragingly high, especially as the manual observation data that it was trained and tested on had been shown to have high levels of inter-observer variation (see supplementary data). Typical examples of images and manual observer/algorithm spots are shown in Figure 5

## 3.3 Testing in Glaucoma Cohort

ACCEPTED MANUSCRIPT

Follow-up data to 18 months after DARC was available for 18 patients, with a mean number of 6.11 ± 2.27 (SD) tests each with a range of 3 to 11 tests. Using only the OCT global RNFL rates of progression (RoP 3.5 ring) performed at 18 months to define progression, the glaucoma cohort was divided into progressing and stable groups. Clinical agreement was poor between observers, as shown in Figure S2, hence, the use of objective, simple and single OCT parameter. Those patients with a significant ($p<0.05$) negative slope were defined as progressing compared to those without who were defined as stable, and are detailed in Table 3. Of the 29 glaucoma eyes analysed, 8 were found to be progressing and 21 stable, by this definition.

Using this definition of glaucoma progression, a Receiver Operating Characteristic (ROC) curve was constructed for both CNN-aided algorithm and manual observer 2-agree and shown in Figure 6, to investigate if the DARC count was predictive of glaucoma progression at 18 months. Maximal sensitivity (85.7%) and specificity (91.7%) were achieved above a DARC count of 24 with an AUC of 0.88 and likelihood ratio of 8.57 with the CNN algorithm as opposed to the manual observer with maximal sensitivity (71.4%) and specificity (87.5%) above a DARC count of 11, an AUC of 0.79, and likelihood ratios of 4.76, showing the CNN-aided algorithm to be performing superiorly. A comparison of the CNN and all agreement states of the manual observers is shown in Supplementary Figure S3.

## 3.4 DARC counts as a Predictor of Glaucoma Progression

DARC counts in both stable and progressing glaucoma groups with the CNN-aided algorithm are shown in Figure 7a and manual DARC counts (observer 2 agree) in Figure 7b. The DARC count, was found to be significantly higher in patients who were later found to be progressing at 18 months (mean 26.13 ; 95% CI 9.41 to 42.84) compared to those who were stable (mean 9.71; 95% CI 5.68 to 13.75 ) using the CNN-aided algorithm ($p=0.0044$; unpaired t-test). Manual observers (2

ACCEPTED MANUSCRIPT

agree or more) DARC counts, were also significantly higher in those progressing at 18 months (mean 12.25; ; 95% CI 3.67 to 20.83) compared to stable (mean 4.38; ; 95% CI 2.20 to 6.57 ) glaucoma patients (p=0.0084; unpaired t-test). No stable eyes had a CNN DARC count above 30 (dashed line, Figure 7a), highlighting this as a threshold that could be confidently used to separate those at risk of progression.

Analyses of baseline age, CCT, BP, visual field MD, VFI and average RNFL thickness (all at the time of DARC), were performed in a similar manner to that described for DARC counts, but none were found to be significantly predictive of progression. Additionally, we have used a newly described structural OCT parameter being the baseline topographically correspondent abnormal sectors on OCT RNFL and BMO-MRW imaging [32]. This interestingly did show a significant (p=0.045) difference between stable and progressing groups, as shown in Supplement Figure S5.

# 4 Discussion

The main goal of glaucoma management is to prevent vision loss. As the disease progresses slowly over many years, current gold standards of assessing changes not only take a long time to develop, but also after significant structural and functional damage has already occurred[21]. There is an unmet need in glaucoma for reliable measures to assess risk of future progression and effectiveness of treatments[33], [34]. Here, we describe a new CNN-aided algorithm which when combined with DARC - a marker of retinal cell apoptosis, is able to predict glaucoma progression defined by RNFL thinning on OCT, 18 months later. This method when used with DARC was able to provide an automated and objective biomarker.

ACCEPTED MANUSCRIPT

The development of surrogate markers has been predominantly in cancer where they are used as predictors of clinical outcome. In glaucoma, the most common clinical outcome measure is vision loss followed by a decrease in quality of life for assessing treatment efficacy. Surrogates should enable earlier diagnoses, earlier treatment, and also shorter, and therefore more economical clinical trials. However, to be a valid surrogate marker, the measures have to be shown to be accurate. For example, OCT, which is in widespread use has been found to have a sensitivity and specificity of 83% and 88% respectively for detecting significant RNFL abnormalities [35] in addition to good repeatability [36][37]. In comparison, our CNN algorithm had a sensitivity of 85.7% and specificity of 91.7% to glaucoma progression, with an AUC of 0.88. The CNN algorithm appeared to perform better than the manual 2-agree counts in terms of sensitivity, specificity and likelihood ratio of predicting progression; nevertheless, the AUCs were not significantly different which could be related to the small sample numbers. Despite this, both the manual 2-agree and CNN were able to significantly predict progression. The main advantage however, of the CNN lies in its practicality; employing manual observers is labour-intensive, time-consuming and expensive compared to an automated system, with its scalability enabling wider accessibility in the future. [38] Moreover, automated detection being more objective than manual observers, may facilitate the consistency and accuracy of DARC as a biomarker in clinical care.

Although the Phase 1 results suggested there was some level of DARC being predictive, this was done on a very small dataset [21] with different doses of Anx776 of 0.1, 0.2, 0.4 and 0.5 mg, with a maximum of 4 glaucoma eyes per group, of which there were only 3 in the 0.4 mg group. In this present study, all subjects received 0.4 mg Anx776, and 27 eyes were analysed. However, moving forwards, we would hope to obtain even more data on glaucoma patients, as this study had a small sample size, and also establish repeatability and test-retest metrics, as we were limited to a single DARC assessment in this study, as per protocol. Furthermore, we recognise that longitudinal studies will be needed to further validate our findings, especially if we are to investigate the correlation of disease severity and DARC. In this Phase 2 study, the majority of patients had early

ACCEPTED MANUSCRIPT

disease (MD between -1.61 –2.22 dB) so it was not possible to investigate the full spectrum of glaucoma disease severity.

In clinical practice, glaucoma patients are assessed for risk of progression based on establishing the presence of risk factors including: older age, a raised intraocular pressure (IOP, too high for that individual), ethnicity, a positive family history for glaucoma, stage of disease, and high myopia [39]. More advanced disease risks included a vertical cup:disc ratio > 0.7, pattern standard deviation of visual field per 0.2 dB increase, bilateral involvement and disc asymmetry, as also the presence of disc haemorrhages and pseudexfoliation [40]–[44]. In this study, we assessed baseline age, CCT, BP, visual field MD, VFI, average RNFL thickness and topographically correspondent abnormal sectors on OCT RNFL and BMO-MRW imaging [32]. Only the OCT RNFL and BMO-MRW sector parameter was found to be predictive of progression, and emphasizes the importance of this new parameter. Compared to the CNN DARC count however, there was overlap between the stable and progressing groups, with no clear threshold in the number of abnormal sectors that could be used to define those at greatest risk of progression, and therefore difficult to apply in practice as a biomarker. However, the presence of predictive structural changes in the same eyes where the CNN DARC count is higher does provide some confidence in the validity of our results.

Objective assessment is increasingly recognised as being important in glaucoma, as there is variable agreement between clinicians, even with technological aids. Poor agreement has been shown with respect to defining progression in patients using visual fields, OCT and stereophotography [10][3][45][11]. Indeed, for this study, we asked five masked senior glaucoma specialists (co-authors) to grade for progression of patients using their clinical judgement based on optic disc assessment, OCT, visual fields and management history; unfortunately, there was variable agreement between them although three of the more senior clinicians (OBS 1, 2 and 3, Supplementary Figure S4), did appear to agree more than the more junior experts (OBS 4 and 5). For

ACCEPTED MANUSCRIPT

this reason, a single, objective metric [46] [29] of rate of progression was used to define the groups used to test the CNN-aided algorithm.

The analysis of progression was post-hoc, and there was no protocol guiding treating clinicians during the 18 month period of follow-up. Similar to the oral memantine trial, [47] management of patients, especially with regard to IOP lowering, was left to the discretion of the glaucoma specialist, and following normal standard of care. As a result, there was a wide range in the number of OCT scans each patient had in the 18 month period from n=3 to n=11 scans per eye. However, despite this and using the OCT global RNFL 3.5 ring RoP, 8 of 29 eyes were progressing at 18 months. To compensate for ageing losses and high false positive rates of progression, we defined a significant RoP only when the negative slope was greater than 1 µm/year to distinguish between normal aging losses. [29] It would be interesting in the future to compare DARC to conventional risk factors such as IOP using a multivariable analysis, though this will require a larger, prospective, longitudinal trial with a strict protocol regarding patient management.

We recognise that it is debatable to include both right and left eyes of patients as individual study eyes, although this was not a treatment trial [48]. Indeed, there are several published clinical studies that have recorded different rates of glaucoma progression between the right and left eyes of the same individuals [49]–[53]. However, we do see this as a limitation of our own study, and will address this in the future, especially with the opportunity of larger trials.

The poor agreement between clinicians identifying progression has generated great interest in the last few years in the use of artificial intelligence to help aid glaucoma diagnosis and prognosis using AI with optic disc photographs[17][54][19], visual fields[27][14] and OCT [55][56]. A recent study by Medeiros et al described an algorithm to assess fundus photographs based on predictions of estimated RNFL thickness, achieved by training a CNN using OCT RNFL thickness measurements[56]. At specificity of 95%, the predicted measurements had a sensitivity of 76% whereas actual SD OCT measurements had sensitivity of 73%. For specificity at 80%, the predicted

ACCEPTED MANUSCRIPT

measurements had sensitivity of 90% compared to OCT measurements which had sensitivity of 90%. The authors suggest their method could potentially could be used to extract progression information from optic disc photographs, but comment that further validation on longitudinal datasets is needed, in the same way we are suggesting from this work.

Template matching is routinely used for tracking cells in microscopy with similar assessment needed to analyse single cells in vivo longitudinally in this study. For template matching here, a 30x30 pixel template was used, for the CNN a 64x64 pixel image was used. The reason for this size difference is template matching is sensitive to blood vessels and so a small template is beneficial to reduce the likelihood of a blood vessel being included. For the CNN a larger image is useful to give the CNN more context of the area around the spot which may be useful in classification. As mentioned previously, there was some unbalanced data between the template-matching and the training of the CNN algorithm. We compensated for this by setting the DARC spots class weights to 50 for spots and 1 for other objects, but we believe our method for template matching could be improved to reduce the number of candidate spots.

Although the algorithm performs well, providing a viable method to detect progressive glaucoma 18 months ahead of alternative methods, we believe there are areas where it can be optimised, some of which are described below.

Alternative classification algorithms to MobileNetV2 such as Support Vector Machines (SVMs) or Random Forests require “hand-crafted” features which are difficult to produce as they need to account for complexities caused by the image capture such as non-linear intensity variation, optical blur, registration blur and low light noise, as well as biological complexities such as the patterning in the choroidal vasculature, blood vessels, blur due to cataracts etc. The network has some biases to do with the intensity of the original retinal image. We believe we can improve results by looking at the intensity standardisation and augmenting the data by varying the intensity in ways more realistic with a larger dataset. The performance of other networks such as VGG16 were

ACCEPTED MANUSCRIPT

evaluated, at the time of writing MobiNetV2 was found to perform best. We are continuing to evaluate if this network is optimum for this need. In comparison, VGG16, an alternate CNN, would be limited to 64 spots in a batch which could mean a batch has no DARC spots in it which hinders training. We have an alternative method that detects and classifies spots in a single step using the detection and segmentation algorithm, YOLO3. We believe this may be a more efficient and effective method with more data, however at this stage the highest accuracy we have achieved with YOLO is not as good as the method outlined in this document.

# 5 Conclusion

This study describes a CNN-aided algorithm to analyse DARC as a marker of retinal cell apoptosis in retinal images in glaucoma patients. The algorithm enabled a DARC count to be computed which when tested in patients was found to successfully predict OCT RNFL glaucoma progression 18 months later. Further validation with longitudinal studies is needed, but this data supports use of this method to provide an automated and objective biomarker with potentially widespread clinical applications.


## Funding

This paper was funded by the Wellcome Trust (grant no. WT099729)


## Declaration of Interest

MF Cordeiro and J Maddison are named inventors on patents on DARC and the discussed algorithm, owned by UCL. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

## Reviewer Disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

ACCEPTED MANUSCRIPT

## Trial Registration

ISRCTN10751859

## Author contributions

EMN and PAB were PIs of study; EMN, PAB, SA, LC, NM and FA were investigators; EMN, TEY, SM, PB, NM, SA, LC, FA and PAB collected data; EMN, TEY, JM, MA performed analysis; EMN, JM and MFC wrote paper; JM and MFC conceived idea

## Acknowledgements

We would like to thank Dr Benjamin Davis and Jessica Bonetti for their invaluable help for the manual observer assessments, and Dr Roger A'Hern for statistical advice.

ACCEPTED MANUSCRIPT

# References


[1] R. Poplin *et al.*, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nat. Biomed. Eng.*, vol. 2, no. 3, pp. 158–164, Mar. 2018.

** High impact publication of retinal biomarker using AI

[2] D. S. W. Ting *et al.*, "Artificial intelligence and deep learning in ophthalmology," *British Journal of Ophthalmology*, vol. 103, no. 2. BMJ Publishing Group, pp. 167–175, 01-Feb-2019.

[3] S. A. Banegas, A. Antón, A. Morilla-Grasa, M. Bogado, E. M. Ayala, and J. Moreno-Montañes, "Agreement among spectral-domain optical coherence tomography, standard automated perimetry, and stereophotography in the detection of glaucoma progression," *Investig. Ophthalmol. Vis. Sci.*, vol. 56, no. 2, pp. 1253–1260, 2015.

[4] J. I. Orlando *et al.*, "REFUGE Challenge: A Unified Framework for Evaluating Automated Methods for Glaucoma Assessment from Fundus Photographs," *Med. Image Anal.*, p. 101570, Oct. 2019.

* Guidelines for retinal imaging automated analytical methods

[5] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, "Deep image mining for diabetic retinopathy screening.," *Med. Image Anal.*, vol. 39, pp. 178–193, Jul. 2017.

[6] U. Schmidt-Erfurth *et al.*, "Machine Learning to Analyze the Prognostic Value of Current Imaging Biomarkers in Neovascular Age-Related Macular Degeneration.," *Ophthalmol. Retin.*, vol. 2, no. 1, pp. 24–30, 2018.

[7] U. Schmidt-Erfurth *et al.*, "Prediction of Individual Disease Conversion in Early AMD Using Artificial Intelligence.," *Invest. Ophthalmol. Vis. Sci.*, vol. 59, no. 8, pp. 3199–3208, 2018.

[8] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010


ACCEPTED MANUSCRIPT

and 2020," *Br J Ophthalmol*, vol. 90, no. 3, pp. 262–267, 2006.

[9] Y. C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C. Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis," *Ophthalmology*. 2014.

* Epidemeological assesment of glaucoma

[10] A. C. Viswanathan *et al.*, "Interobserver agreement on visual field progression in glaucoma: a comparison of methods," 2003.

[11] J. Moreno-Montañés *et al.*, "Intraobserver and interobserver agreement of structural and functional software programs for measuring glaucoma progression," *JAMA Ophthalmol.*, vol. 135, no. 4, pp. 313–319, Apr. 2017.

[12] M. Iester *et al.*, "Agreement to detect glaucomatous visual field progression by using three different methods: A multicentre study," *Br. J. Ophthalmol.*, 2011.

[13] M. Wang *et al.*, "An Artificial Intelligence Approach to Detect Visual Field Progression in Glaucoma Based on Spatial Pattern Analysis.," *Invest. Ophthalmol. Vis. Sci.*, vol. 60, no. 1, pp. 365–375, 2019.

[14] Ş. S. Kucur, G. Holló, and R. Sznitman, "A deep learning approach to automatic detection of early glaucoma from visual fields.," *PLoS One*, vol. 13, no. 11, p. e0206081, 2018.

[15] F. A. Medeiros, A. A. Jammal, and A. C. Thompson, "From Machine to Machine: An OCT-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs," *Ophthalmology*, vol. 126, no. 4, pp. 513–521, Apr. 2019.

[16] R. Asaoka *et al.*, "Using Deep Learning and Transfer Learning to Accurately Diagnose Early-Onset Glaucoma From Macular Optical Coherence Tomography Images," *Am. J. Ophthalmol.*, vol. 198, pp. 136–145, Feb. 2019.

ACCEPTED MANUSCRIPT

[17] Y. Jiang *et al.*, "Optic Disc and Cup Segmentation with Blood Vessel Removal from Fundus Images for Glaucoma Detection," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018, vol. 2018-July, pp. 862–865.

[18] I. J. C. MacCormick *et al.*, "Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile.," *PLoS One*, vol. 14, no. 1, p. e0209409, 2019.

[19] A. C. Thompson, A. A. Jammal, and F. A. Medeiros, "A Deep Learning Algorithm to Quantify Neuroretinal Rim Loss From Optic Disc Photographs.," *Am. J. Ophthalmol.*, vol. 201, pp. 9–18, May 2019.

[20] R. S. Harwerth, A. S. Vilupuru, N. V Rangaswamy, and E. L. Smith 3rd, "The relationship between nerve fiber layer and perimetry measurements," *Invest Ophthalmol Vis Sci*, vol. 48, no. 2, pp. 763–773, 2007.

[21] M. F. Cordeiro *et al.*, "Real-time imaging of single neuronal cell apoptosis in patients with glaucoma," *Brain*, vol. 140, no. 6, 2017.

** First clincial trial with DARC

[22] M. Pathak, S. Demirel, and S. K. Gardiner, "Nonlinear, multilevel mixed-effects approach for modeling longitudinal standard automated perimetry data in glaucoma.," *Invest. Ophthalmol. Vis. Sci.*, vol. 54, no. 8, pp. 5505–13, Aug. 2013.

[23] Y. X. Wang, N. O'Leary, N. G. Strouthidis, E. T. White, T. A. Ho, and D. F. Garway-Heath, "Comparison of neuroretinal rim area measurements made by the Heidelberg Retina Tomograph I and the Heidelberg Retina Tomograph II.," *J. Glaucoma*, vol. 22, no. 8, pp. 652–8.

[24] "ImageJ." .

ACCEPTED MANUSCRIPT

[25] I. Pan, S. Agarwal, and D. Merck, “Generalizable Inter-Institutional Classification of Abnormal Chest Radiographs Using Efficient Convolutional Neural Networks.,” *J. Digit. Imaging*, vol. 32, no. 5, pp. 888–896, Oct. 2019.

[26] Z. Chen *et al.*, “Feature Selection May Improve Deep Neural Networks For The Bioinformatics Problems.,” *Bioinformatics*, Oct. 2019.

[27] S. Pang, S. Wang, A. Rodríguez-Patón, P. Li, and X. Wang, “An artificial intelligent diagnostic system on mobile Android terminals for cholelithiasis by lightweight convolutional neural network.,” *PLoS One*, vol. 14, no. 9, p. e0221720, 2019.

* interesting use of CNN in medical application.

[28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *arXiv e-prints IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018.

[29] Z. Wu, L. J. Saunders, L. M. Zangwill, F. B. Daga, J. G. Crowston, and F. A. Medeiros, “Impact of Normal Aging and Progression Definitions on the Specificity of Detecting Retinal Nerve Fiber Layer Thinning,” *Am. J. Ophthalmol.*, vol. 181, pp. 106–113, 2017.

[30] C. K. S. Leung *et al.*, “Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: A prospective analysis of age-related loss,” *Ophthalmology*, vol. 119, no. 4, pp. 731–737, 2012.

[31] B. C. Chauhan *et al.*, “Bruch’s Membrane Opening Minimum Rim Width and Retinal Nerve Fiber Layer Thickness in a Normal White Population: A Multicenter Study,” *Ophthalmology*, vol. 122, no. 9, pp. 1786–1794, 2015.

[32] H. Yang *et al.*, “Optical Coherence Tomography Structural Abnormality Detection in Glaucoma Using Topographically Correspondent Rim and Retinal Nerve Fiber Layer Criteria,” *Am. J. Ophthalmol.*, vol. 213, pp. 203–216, 2020.

ACCEPTED MANUSCRIPT

** Newest glaucoma OCT diagnostic parameter

[33] R. N. Weinreb and P. L. Kaufman, “Glaucoma research community and FDA look to the future, II: NEI/FDA Glaucoma Clinical Trial Design and Endpoints Symposium: measures of structural change and visual function.,” *Investigative ophthalmology & visual science*, vol. 52, no. 11. United States, pp. 7842–7851, Oct-2011.

[34] R. N. Weinreb and P. L. Kaufman, “The glaucoma research community and FDA look to the future: a report from the NEI/FDA CDER Glaucoma Clinical Trial Design and Endpoints Symposium,” *Invest Ophthalmol Vis Sci*, vol. 50, no. 4, pp. 1497–1505, 2009.

[35] R. T. Chang, O. J. Knight, W. J. Feuer, and D. L. Budenz, “Sensitivity and specificity of time-domain versus spectral-domain optical coherence tomography in diagnosing early to moderate glaucoma.,” *Ophthalmology*, vol. 116, no. 12, pp. 2294–2299, Dec. 2009.

[36] J. E. DeLeon Ortega *et al.*, “Effect of glaucomatous damage on repeatability of confocal scanning laser ophthalmoscope, scanning laser polarimetry, and optical coherence tomography.,” *Invest. Ophthalmol. Vis. Sci.*, vol. 48, no. 3, pp. 1156–1163, Mar. 2007.

[37] B. B. Tan, M. Natividad, K.-C. Chua, and L. W. Yip, “Comparison of retinal nerve fiber layer measurement between 2 spectral domain OCT instruments.,” *J. Glaucoma*, vol. 21, no. 4, pp. 266–273, 2012.

[38] S. M. Mckinney *et al.*, “International evaluation of an AI system for breast cancer screening,” *Nature*, vol. 577, no. July 2019, 2020.

[39] J. B. Jonas, T. Aung, R. R. Bourne, A. M. Bron, R. Ritch, and S. Panda-Jonas, “Glaucoma,” *The Lancet*. 2017.

** Good review of glaucoma

[40] S. Miglior, V. Torri, T. Zeyen, N. Pfeiffer, J. C. Vaz, and I. Adamsons, “Intercurrent Factors

ACCEPTED MANUSCRIPT

Associated with the Development of Open-Angle Glaucoma in the European Glaucoma Prevention Study," *Am. J. Ophthalmol.*, 2007.

[41] M. O. Gordon, J. A. Beiser, J. D. Brandt, and C. M. Grasso, "Ocular hypertension treatment study: Baseline factors that predict the onset of primary open-angle glaucoma," *Evidence-Based Eye Care*. 2003.

[42] D. L. Budenz *et al.*, "Detection and Prognostic Significance of Optic Disc Hemorrhages during the Ocular Hypertension Treatment Study," *Ophthalmology*, 2006.

[43] R. A. Levine *et al.*, "Asymmetries and visual field summaries as predictors of glaucoma in the ocular hypertension treatment study," *Investig. Ophthalmol. Vis. Sci.*, 2006.

[44] M. O. Gordon *et al.*, "The Ocular Hypertension Treatment Study: Baseline factors that predict the onset of primary open-angle glaucoma," *Arch. Ophthalmol.*, 2002.

[45] D. M. Blumberg *et al.*, "Technology and the glaucoma suspect," *Investig. Ophthalmol. Vis. Sci.*, vol. 57, no. 9, pp. OCT80–OCT85, 2016.

[46] A. J. Tatham and F. A. Medeiros, "Detecting Structural Progression in Glaucoma with Optical Coherence Tomography.," *Ophthalmology*, vol. 124, no. 12S, pp. S57–S65, Dec. 2017.

[47] R. N. Weinreb *et al.*, "Oral Memantine for the Treatment of Glaucoma: Design and Results of 2 Randomized, Placebo-Controlled, Phase 3 Studies," *Ophthalmology*, vol. 125, no. 12, pp. 1874–1885, 2018.

[48] R. A. Armstrong, "Statistical guidelines for the analysis of data obtained from one or both eyes," *Ophthalmic Physiol. Opt.*, vol. 33, no. 1, pp. 7–14, 2013.

[49] T. C. W. Chan, C. Bala, A. Siu, F. Wan, and A. White, "Risk Factors for Rapid Glaucoma Disease Progression," *Am. J. Ophthalmol.*, vol. 180, pp. 151–157, 2017.

[50] A. J. Anderson, "Significant glaucomatous visual field progression in the first two years: What

ACCEPTED MANUSCRIPT

does it mean?," *Transl. Vis. Sci. Technol.*, vol. 5, no. 6, pp. 0–7, 2016.

[51] D. Jeong, K. R. Sung, and J. H. Na, "Comparison of clinical characteristics and progression rates of bilaterally and unilaterally progressing glaucoma," *Korean J. Ophthalmol.*, vol. 29, no. 1, pp. 40–46, 2015.

[52] H. Öhnell, A. Heijl, L. Brenner, H. Anderson, and B. Bengtsson, "Structural and Functional Progression in the Early Manifest Glaucoma Trial," *Ophthalmology*, vol. 123, no. 6, pp. 1173–1180, 2016.

[53] V. Carr, R. M., Oranu, A., & Khungar, "乳鼠心肌提取 HHS Public Access," *Physiol. Behav.*, vol. 176, no. 1, pp. 139–148, 2016.

[54] I. J. C. MacCormick *et al.*, "Correction: Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile (PLoS ONE (2019) 14:1 (e0209409) DOI: 10.1371/journal.pone.0209409)," *PLoS One*, vol. 14, no. 4, pp. 1–20, 2019.

[55] R. Asaoka *et al.*, "Using Deep Learning and Transfer Learning to Accurately Diagnose Early-Onset Glaucoma From Macular Optical Coherence Tomography Images," *Am. J. Ophthalmol.*, vol. 198, pp. 136–145, 2019.

[56] F. A. Medeiros, A. A. Jammal, and A. C. Thompson, "From Machine to Machine: An OCT-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs," *Ophthalmology*, vol. 126, no. 4, pp. 513–521, 2019.

* AI glaucoma study showing extrapolation of OCT assessment to fundus photographs

Papers of special note have been highlighted as:
* of interest
** of considerable interest

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

# Figure Legends

**Figure 1 Consort Diagram Showing Glaucoma and Control Cohorts Subjects and DARC Image Analysis**

***** 3 screen failures, 37 declined ** 24 screen failure, 36 were ineligible from pre-screening GP letter, 50 declined

**Figure 2 CNN-aided Algorithm Flowchart showing Analysis Stages of DARC Images**

**Figure 3 Representative Retinal Image of the Possible Spot Candidates**

Candidate spots were detected using template matching and a correlation map. Local maxima were selected and filtered with thresholds for the correlation coefficient and intensity standard deviation (corresponding to the brightness of the spot). These thresholds were set very low and produce many more spot candidates than manually observed spots (approximately 50-1).

**Figure 4 CNN Training and Validation Stages**

CNN training (A) and validation (B) curves. A good accuracy is achieved in 200 epochs (training cycles) although training was left for 300 epochs to verify stability. The matching validation accuracy also shows similar accuracy without signs of over training. The accuracy was found to be 97%, with 91.1% sensitivity and 97.1% specificity.

**Figure 5 Representative Comparison of Manual Observer and CNN-algorithm DARC Spots**

Spots found by the CNN and spots found by at least 2 manual observers shown on an original retinal image. (A) Patient 6, left eye. Progressive glaucoma (as measured by OCT global RFNL 3.5 ring) (B) Patient 31, left eye. Stable glaucoma. Green circles: manual observers only (False negative); Blue circles: CNN-aided algorithm only, (False Positive); Turquoise circle: Algorithm and manual observers agree (True Positive)

**Figure 6 ROC Curves of Glaucoma Progression of Manual Observer and CNN-algorithm analysis**

Receiver Operating Characteristic (ROC) curves were constructed for both the CNN-aided algorithm (A) and manual observer 2-agree or more (B), to test predictive value of glaucoma progression at 18 months. The rate of progression (RoP) was calculated from the Spectralis OCT global retinal nerve fibre layer (RNFL) measurements at 3.5 mm from the optic disc at 18 months follow up of glaucoma subjects after DARC. Those patients with a significant ($p<0.05$) negative slope were defined as progressing compared to those without who were defined as stable. Maximal sensitivity (90.0%) and specificity (85.71%) were achieved at a DARC count of 23 with the AUC of 0.89 with the CNN

ACCEPTED MANUSCRIPT

algorithm as opposed to the manual observer count with maximal sensitivity (0.85%) and specificity (71.43%) at DARC count of 12, with the AUC of 0.79, showing the CNN-aided algorithm to be performing superiorly.

**Figure 7 CNN DARC counts significantly increased in glaucoma patients who go on to progress compared to those who are stable**

Violin plots illustrating the distribution of data in glaucoma eyes with and without significant RoP as measured by OCT global RFNL 3.5 ring at 18 months follow-up. (A) The CNN DARC count was significantly higher in patients progressing at 18 months (mean 26.13) compared to those who were stable (mean 9.71 ) using the CNN-aided algorithm (p=0.0044). The DARC count was defined as the number of ANX776-positive spots seen in the retinal image at 120 minutes after baseline spot subtraction. No stable eyes had a CNN DARC count above 30 (dashed line), highlighting this as a threshold that could be confidently used to separate those at risk of progression. (B) A similar trend was found with manual observers (2 agree or more) DARC counts, although counts were lower compared to the CNN. Again, a significant difference was found between those progressing at 18 months (mean 12.25) compared to stable (mean 4.38) glaucoma patients (p=0.0084). Asterisks indicate level of significance by unpaired t-test. Horizontal lines indicate medians and interquartile ranges with minimum and maximum points.

ACCEPTED MANUSCRIPT

**Table 1a. Glaucoma Eligibility (Exclusion/Inclusion Criteria Glaucoma)**

| Subject ID | Eligible eye | Diagnosis |
|---|---|---|
| 6 | Both | Primary Open Angle Glaucoma |
| 7 | Both | Glaucoma suspect |
| 9 | Both | Glaucoma suspect |
| 11 | Both | Glaucoma suspect |
| 13 | Both | Glaucoma suspect |
| 17 | Both | Glaucoma suspect |
| 18 | Both | Glaucoma suspect |
| 21 | Both | Primary Open Angle Glaucoma |
| 23 | Both | Primary Open Angle Glaucoma |
| 25 | Both | Glaucoma suspect |
| 31 | Left | Primary Open Angle Glaucoma |
| 32 | Both | Primary Open Angle Glaucoma |
| 38 | Both | Primary Open Angle Glaucoma |
| 39 | Both | Glaucoma suspect |
| 44 | Both | Primary Open Angle Glaucoma |
| 45 | Both | Glaucoma suspect |
| 52 | Both | Glaucoma suspect |
| 61 | Both | Primary Open Angle Glaucoma |
| 72 | Left | Glaucoma suspect |
| 74 | Both | Glaucoma suspect |

**Table1b. Glaucoma characteristics on study entry**

| **Diagnosis** | **n** | **(%)** |
|---|---|---|
| *Glaucoma* | 8 | (40) |
| *Glaucoma suspect* | 12 | (60) |
| *Ocular hypertension* | 0 | (0) |
| **Total** | **20** | |

ACCEPTED MANUSCRIPT

**Table 2. Baseline and Qualification Progression parameters Glaucoma Patients,** where progression was defined by a significant ($p<0.05$) negative slope in the rate of progression.

| Subject | Eye | OCT | | | | SAP | |
|---|---|---|---|---|---|---|---|
| | | **RNFL 3.5** µm/year | **RNFL 4.1** µm/year | **RNFL 4.7** µm/year | **MRW** µm/year | **MD** dB/year | **VFI** %/year |
| **6** | R | | + | + | | | |
| | L | | | | | | |
| **7** | R | | | | | | |
| | L | + | | | + | | |
| **9** | R | + | | | | | |
| | L | + | | | | | |
| **11** | R | | | | | | |
| | L | + | + | | | | |
| **13** | R | | | + | + | | + |
| | L | | | + | | | |
| **17** | R | + | | | | | |
| | L | + | | | | | |
| **18** | R | + | + | | + | | |
| | L | | | | | | |
| **21** | R | + | + | | | | |
| | L | + | | | | | |
| **23** | R | | | | + | | |
| | L | + | | + | | + | |
| **25** | R | | | | | | |
| | L | + | | + | | | |
| **31** | R | + | | | | | |
| | L | | | | | | |
| **32** | R | | + | | | | |
| | L | + | | | | | |
| **38** | R | + | + | + | + | | |
| | L | + | + | + | | | |
| **39** | R | + | | + | | | |
| | L | | | | | | |
| **44** | R | + | | | | | |
| | L | + | | | | | |
| **45** | R | + | | | | | |
| | L | | | | | | |
| **52** | R | + | | | + | | |
| | L | + | | + | | | |
| **61** | R | | | | + | + | + |
| | L | + | + | + | + | | |
| **72** | R | + | | | + | | |
| | L | | + | | + | | |
| **74** | R | + | | + | | | |
| | L | + | + | + | + | | |

ACCEPTED MANUSCRIPT

**Table 3a: Progression classification per glaucoma eye (OCT global RNFL 3.5 ring) 18 months after DARC**

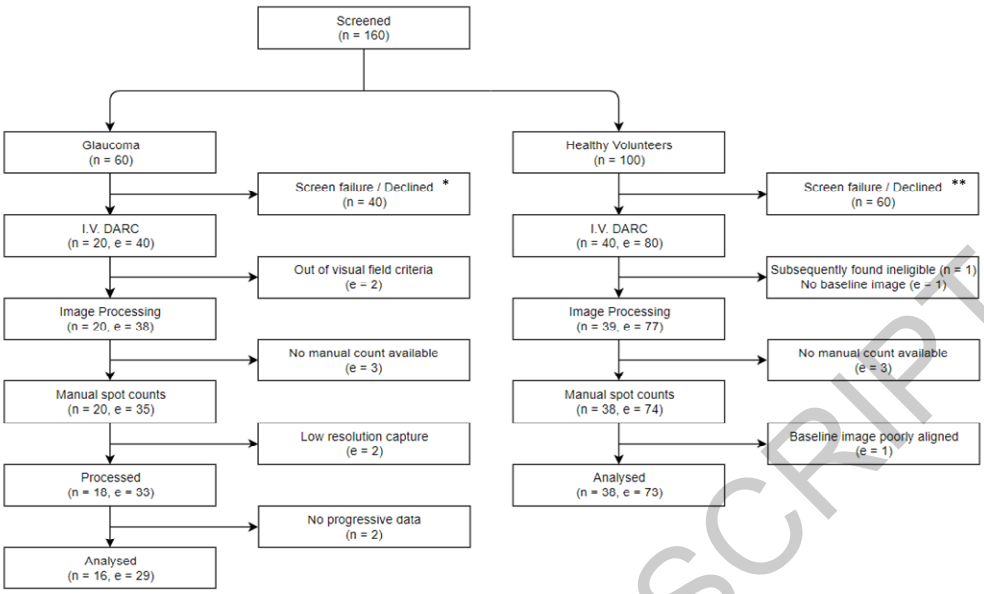
| Category | Number of eyes |
|---|---|
| Progressing | 8 |
| Stable | 21 |
| Unknown | 4 |
| N/A | 5 |
| Total | 38 |

**Table 3b. Clinical findings of eyes meeting the inclusion criteria**

| | **Glaucoma** | | **Healthy volunteer** | | **p** |
|---|---|---|---|---|---|
| | *mean* | *(SD)* | *mean* | *(SD)* | |
| **Age** | 61.1 | (13.7) | 47.6 | 17.1 | <0.005 |
| **Females**<br>**Males** | 6<br>14 | | 21<br>18 | | |
| **Systolic BP** | 140 | (17.0) | 129 | (14.5) | <0.05 |
| **Diastolic BP** | 77.9 | (10.3) | 75.8 | (9.0) | NS |
| **BCVA, logmar** | 0.01 | (0.08) | -0.03 | (0.08) | <0.05 |
| **IOP, mmHg** | 18.90 | (2.61) | 13.63 | (2.50) | <0.005 |
| **Corneal pachimetry (CCT)** | 555.58 | (33.21) | 529.99 | (25.60) | <0.005 |
| **MD (dB)** | -1.7 | (2.1) | -0.3 | (0.1) | <0.005 |
| **OCT RNFL (um)** | 80.0 | (17.4) | 100.1 | (10.7) | <0.005 |

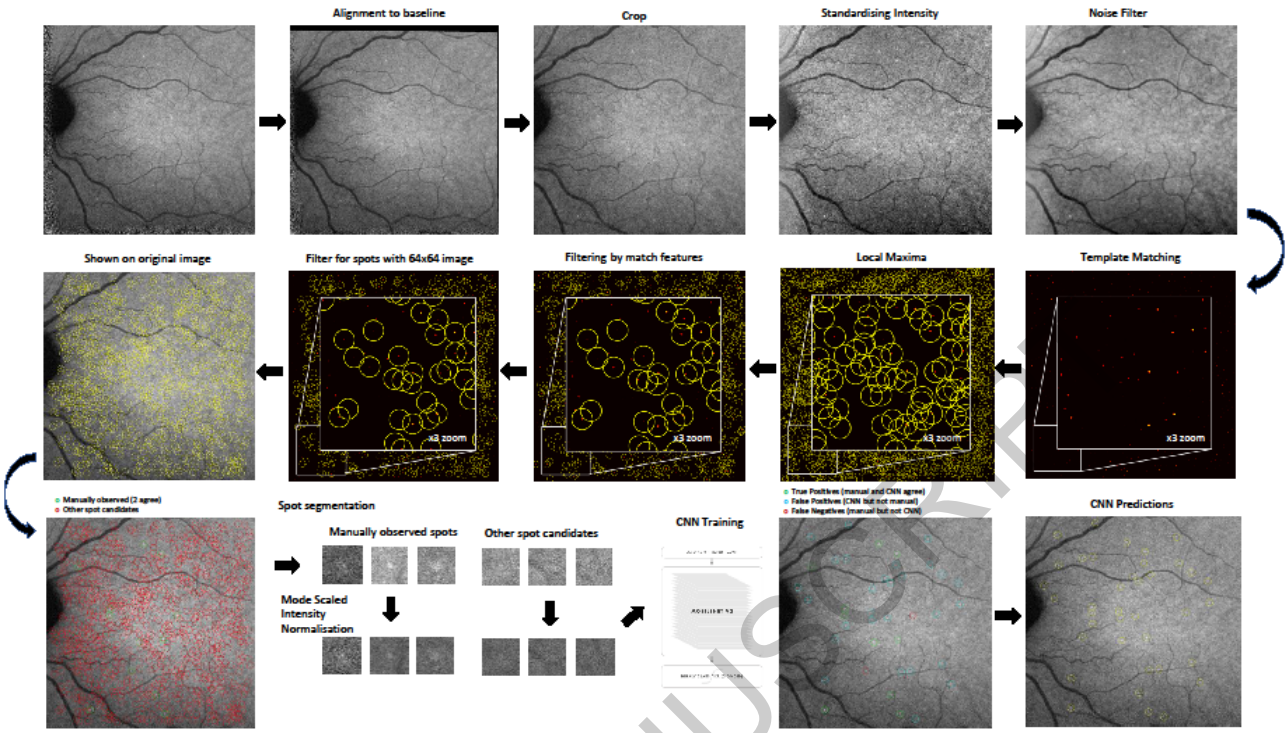
ACCEPTED MANUSCRIPT

Figure 1



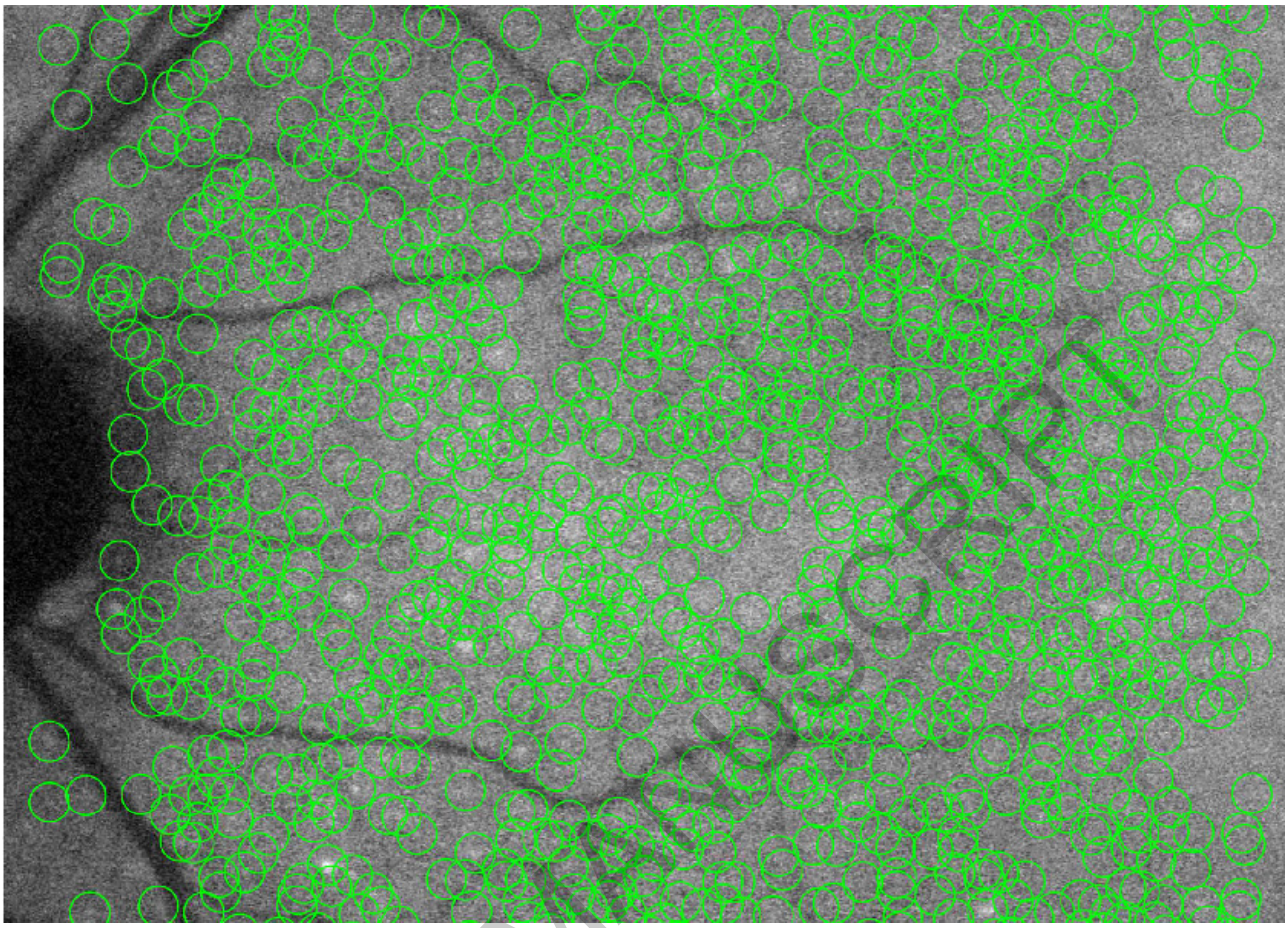

ACCEPTED MANUSCRIPT

Figure 2




ACCEPTED MANUSCRIPT

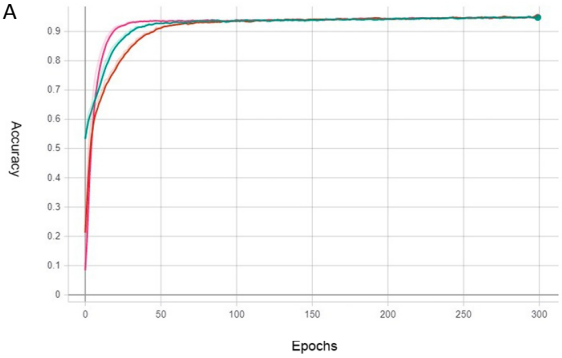
Figure 3



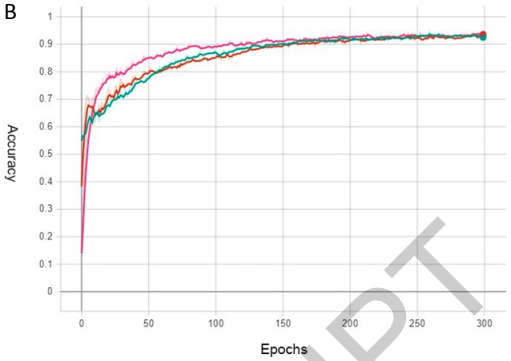
ACCEPTED MA

Figure 4

A




B

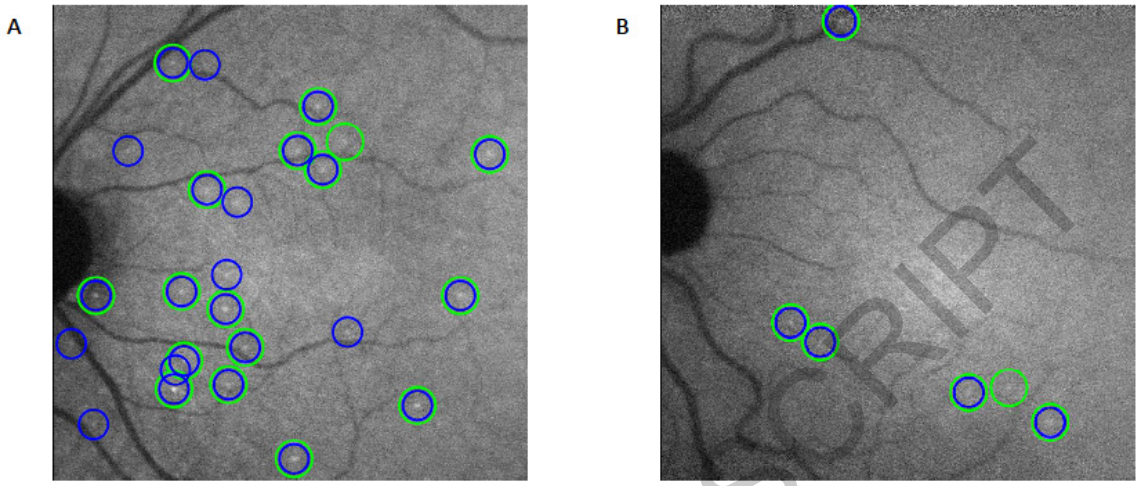



C

| | | | |
|---|---|---|---|
| Accuracy | 97.0% | Total candidate spots | 48610 |
| Sensitivity | 91.1% | Total manual 2-agree spots | 898 |
| Specificity | 97.1% | | |
| | | | |
| True Positives | 818 | False Positives | 1395 |
| False Negatives | 80 | True Negatives | 46317 |

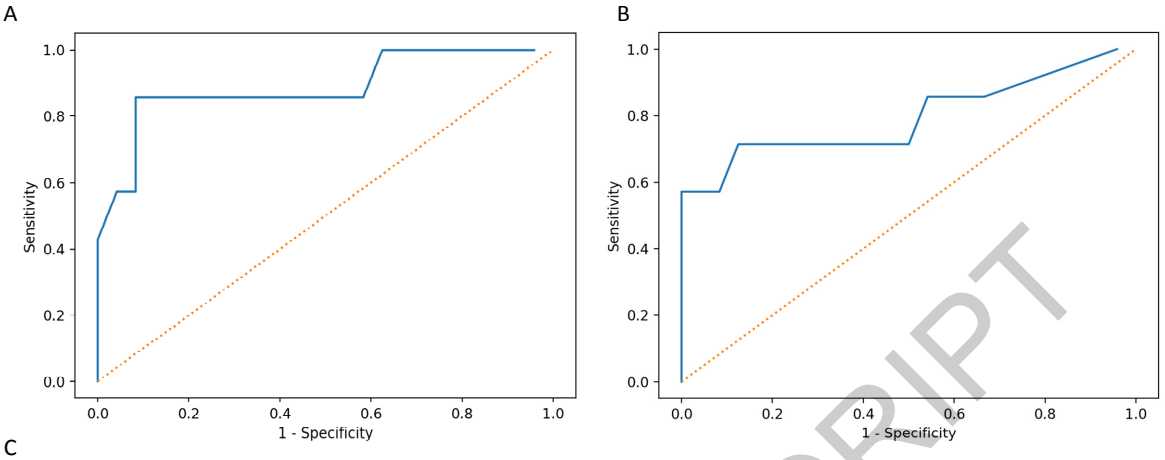
ACCEPTED MANUSCRIPT

Figure 5




ACCEPTED MANUSCRIPT

Figure 6




C

Algorithmic Count

AUC: 0.88 (95% CI 0.67 – 1.00)

Sensitivity: 85.7%, Specificity: 91.7%

Likelihood ratio: 8.57

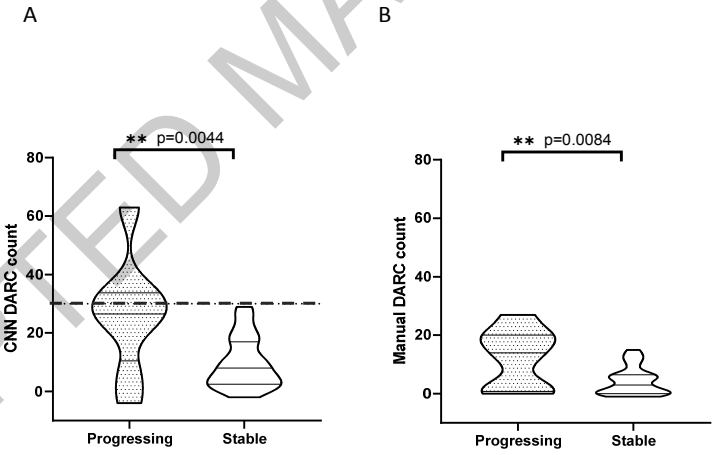
Manual Observers Count (2 agree)

AUC 0.79 (95% CI 0.49 – 1.00)

Sensitivity: 71.4%, Specificity: 87.5%

Likelihood ratio: 4.76

Figure 7




ACCEPTED MANUSCRIPT