Automated Retinal Image Quality Assessment on the UK Biobank Dataset for Epidemiological Studies

R.A. Welikala[a,*], M.M. Fraz[b], P.J. Foster[c], P.H. Whincup[d], A.R. Rudnicka[d], C.G. Owen[d], D.P. Strachan[d], and S.A. Barman[a] on behalf of the UK Biobank Eye and Vision Consortium[e]

[a] School of Computing and Information Systems, Kingston University, Surrey, KT1 2EE, United Kingdom

[b] School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad 44000, Pakistan

[c] NIHR Biomedical Research Centre, Moorfields Eye Hospital, London, EC1V 2PD and UCL Institute of Ophthalmology, London, EC1V 9EL, United Kingdom

[d] Population Health Research Institute, St. George's, University of London, London, SW17 0RE, United Kingdom

[e] Members of the UK Biobank Eye and Vision Consortium are listed before References

[*] Corresponding author. Email address: R.Welikala@kingston.ac.uk

Abstract:

Morphological changes in the retinal vascular network are associated with future risk of many systemic and vascular diseases. However, uncertainty over the presence and nature of some of these associations exists. Analysis of data from large population based studies will help to resolve these uncertainties. The QUARTZ (QUantitative Analysis of Retinal vessel Topology and siZe) retinal image analysis system allows automated processing of large numbers of retinal images. However, an image quality assessment module is needed to achieve full automation. In this paper, we propose such an algorithm, which uses the segmented vessel map to determine the suitability of retinal images for use in the creation of vessel morphometric data suitable for epidemiological studies. This includes an effective 3-dimensional feature set and support vector machine classification. A random subset of 800 retinal images from UK Biobank (a large prospective study of 500,000 middle aged adults; where 68,151 underwent retinal imaging) was used to examine the performance of the image quality algorithm. The algorithm achieved a sensitivity of 95.33% and a specificity of 91.13% for the detection of inadequate images. The strong performance of this image quality algorithm will make rapid automated analysis of vascular morphometry feasible on the entire UK Biobank dataset (and other large retinal datasets), with minimal operator involvement, and at low cost.

Keywords:

Retinal image

Image quality

Vessel segmentation

Large retinal datasets

UK Biobank

Epidemiological studies

1. Introduction:

Examination of the blood vessel structure in retinal images offers an opportunity to directly and non-invasively observe the blood circulatory system. The morphological characteristics of retinal vessels (e.g. vessel calibre, tortuosity) have been prospectively associated with cardiovascular and systemic disease [1-4]. Approximately 20 million people in the UK have a long term health condition and its increasing prevalence is a major challenge for the healthcare system [5]. Cardiovascular disease alone accounts for nearly 200,000 deaths in the UK per year [6], with coronary heart disease, stroke and heart failure accounting for most of these deaths. Early detection and prevention of disease outcome is key. Accurate assessment of retinal vessel morphology may be an important biomarker of vascular health, which might predict those at high risk of disease [7].

UK Biobank includes probably the world's largest retinal image repository (nearly 136,000 retinal images) in a middle-aged population-based cohort study. The depth and breadth of health data recorded make it a powerful research resource for improving the prevention, diagnosis and treatment of a wide range of serious illnesses [8]. Considerable potential exists in using this retinal dataset to discover biomarkers for the identification of high risk patients. Retinal vessel morphology may provide such a measure, which could be used to identify those at high risk, particularly of vascular related disease. However, the extraction of quantitative measures from the vessel morphology in large datasets are needed to establish the presence (or absence) of

associations, and this has been prohibitive to date given the considerable amount of manual operator involvement required. Therefore, automation is needed to process and analyse the large amount of images and extract useful, objective and quantitative information from vessel morphology [9].

Our research group have developed a retinal image analysis system called QUARTZ (QUantitative Analysis of Retinal vessel Topology and siZe), which is used for the automated processing of large numbers of retinal images and obtains quantitative measures of vessel morphology to be used in epidemiological studies [10]. QUARTZ was designed to be fully automated and includes vessel segmentation, measurements from retinal vessels, arteriole/venule (a/v) classification and optic disc localization. The software also derives information from the whole retina and not simply concentric areas centred on the optic disc or manually selected vessels. The vessel analysis screen of the QUARTZ software is illustrated in figure 1.

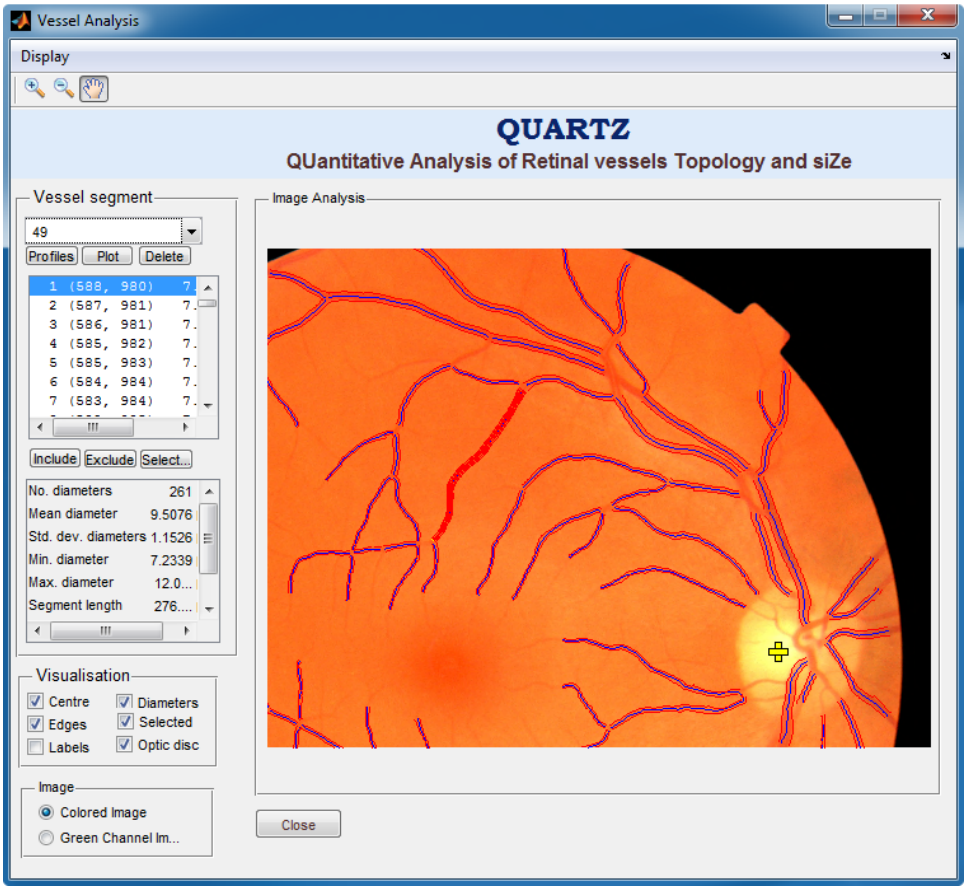


Figure 1: QUARTZ vessel analysis screen.

QUARTZ is considered fully automated in respect to its ability to run on large datasets uninterrupted. However, if quantitative results are to be of high precision, then it is vital to identify and remove images of inadequate quality. Currently, this step of image quality assessment is performed manually. This takes 15 seconds per image, which totalled to approximately 67 hours of manual processing time when QUARTZ was previously used to process 16,000 images [11] from another large population based study (the European Prospective Investigation on Cancer study in Norfolk; EPIC-Norfolk) [12]. VAMPIRE [13,14], SIRIUS [15], ARIA [16] and other notable vessel analysis software, also rely on manual image quality processing. Hence, the

development of automated image quality assessment would be extremely beneficial in speeding up the process.

Retinal images can be poorly and unevenly illuminated, blurred and obstructed. This is caused by factors including camera exposure, focal plane error, poorly dilated/small pupils, eye lashes/blinking, lens artefacts, media opacity (e.g. cataracts, vitreous haemorrhages, asteroid hyalosis) and head/eye movement.

The criteria for an image to be classed as adequate differs for retinal images used to make a conventional diagnosis (e.g. detection of diabetic retinopathy) to those which are useful for extracting vessel morphometric data suitable for epidemiological studies. For the former, there must be good clarity of the entire image to ensure any signs of possible pathology are not missed. For the latter, the criteria are that image clarity must be sufficient to allow for the accurate vessel segmentation for at least a portion of the image. Useful information can still be extracted from well segmented sections of the vasculature, even if this only represents a portion of the vascular tree. From our experience most large retinal datasets used in epidemiological studies (EPIC-Norfolk, UK Biobank) contain large amounts of images of poor quality. Poorly/unevenly illuminated images are a particular problem which is expected when images are captured without the use of pharmacological mydriasis. Therefore this approach ensures that there is little wastage by making use of many of the poor images, extracting as much information as possible from these retinal datasets.

In this paper, an automated image quality assessment methodology is proposed which is designed to classify retinal images as inadequate or adequate for use in the creation of vessel morphometric data suitable for epidemiological studies. As far as we are aware, this paper presents the first automated methodology designed to tackle this important issue. The methodology is based on the assessment of the segmented

vasculature, which involves the extraction of three global features (area, fragmentation and complexity) measured from the segmented vessel map and support vector machine (SVM) classification. This step completely removes the need for any manual processing, allowing QUARTZ to analyse the entire UK Biobank dataset (and other large retinal datasets) without manual intervention.

2. Quality Assessment Algorithms:

To the best of our knowledge no literature exists on automated retinal image quality assessment for the purpose of epidemiological studies. However, there are several publications on automated retinal image quality assessment to determine diagnostic suitability. What constitutes an adequate image for these publications (clarity across the entire image) differs to our requirements. A brief overview of these methodologies shall be given as relevance still exists. These methodologies can be divided into two major categories: (i) classifiers based on generic image quality parameters such as illumination, sharpness and contrast and (ii) classifiers based on structural image quality parameters such as field definitions and the identification of the vascular structure. The latter category holds more relevance to the objective of this paper.

Methods based on generic image quality parameters generally avoid eye structure (e.g. vessel) segmentation. Lee [17] proposed a method to evaluate illumination and contrast based on a measure of similarity between a template intensity histogram and the intensity histogram of the image to be classified. Lalonde [18] adapted this approach by defining a set of local intensity histogram templates, as well as analysing the global edge magnitude histogram. Dias [19] evaluated the generic image attributes of colour, focus, contrast and illumination with features derived from colour indexing and the Sobel operator.

Usher [20] was the first paper to consider image quality evaluation based on structural parameters. Image clarity was calculated based on the area of segmented vessels over the entire image. Fleming [21] judged image clarity by specifically looking at the area of segmented vasculature within the macula. The authors also performed evaluation based on field definition, which were constraints to ensure the optic disc, macula and arcades were all correctly positioned within the image. Hunter [22] proposed a method based on measuring both the contrast and quantity of segmented vessels within the macula, and on the contrast between the fovea region and the retinal background. Giancardo [23] described a technique that extracted local measures of vessel density by measuring the segmented vessel area over different regions of the image.

Paulus [24] combined generic and structural image quality parameters, independent from segmentation methods. This included the use of Haralick texture metrics to describe the generic criteria and Image Structure Clustering [25] that has the ability to cluster the image pixels into the contained anatomical structures. The cluster sizes and inter-cluster differences were used to evaluate the structural recognisability.

As stated above, what constitutes an adequate image for these publications differs to our requirements. These methodologies have the objective of detecting retinal images with good clarity across the entire image, our requirements are to detect to retinal images that have sufficient clarity to allow for the accurate segmentation for at least a portion of the image. Therefore, the proposed methodology takes the logical approach of assessing only the structural attributes of the segmented vasculature.

3. Methodology:

A simple and highly effective method is proposed to detect retinal images of inadequate quality for epidemiological studies by the assessment of the segmented vasculature. Once detected, the inadequate images are rejected and thereby not

included in the vessel morphometric dataset to be used by the epidemiologists. We start this section with an overview of the segmentation process. This is followed by a description of the image quality assessment algorithm, which is based on three global features measured from the segmented vessel map and support vector machine (SVM) classification. The classifier labels images as inadequate or adequate.

3.1 Vessel Segmentation

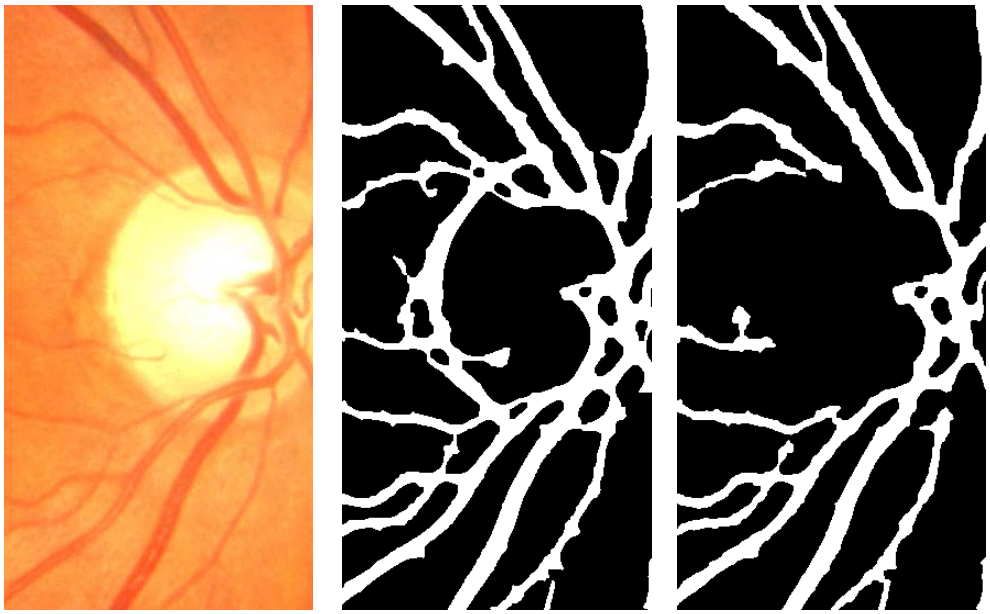
Retinal vessels can be approximated as being piecewise linear. In QUARTZ, we have implemented an unsupervised vessel segmentation approach based on a multi-scale line detector [26]. Using the inverted green channel, the average grey-level of the pixels along a line passing through the target pixel was calculated for multiple orientations. The line-strength of the target pixel was obtained from the orientation with the largest value subtracted by the average grey-level of a square sub-window centred at the target pixel. This was a multi-scale approach, so lines of different lengths (15,17,19,21 pixels) were applied and a linear combination of responses resulted in a final line-strength image. The application of a hysteresis thresholding based morphological reconstruction then followed to produce the segmented vasculature. A full description on this methodology is provided in [10].

Additional steps, to those documented in [10], of pre-processing and post-processing were performed to further improve the segmentation. For epidemiological studies, the need to keep the segmentation of non-vessel objects to a minimum is more important than the segmentation of the complete vasculature.

Pre-processing: pixels surrounding areas of bright intensity, such as the optic disc and lens artefacts, are often falsely segmented (see figures 2-3). Lens artefacts are very

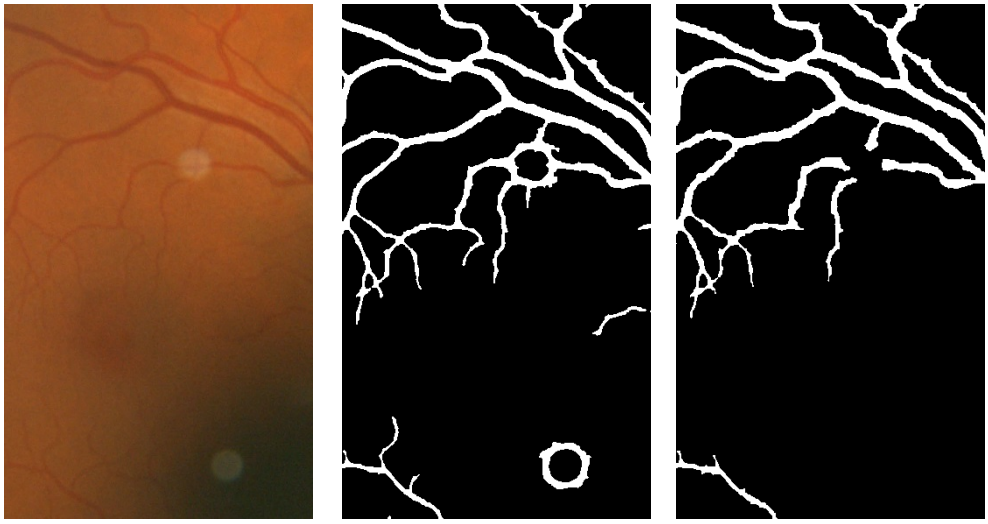
common in the UK Biobank retinal dataset, appearing in over 30% of the images (measured from a subset of 800 images). To prevent this, bright intensities were removed as follows. An image estimating the background was produced by using a median filter of size 105 x 105 pixels. This background image was subtracted from the original image. The resultant has negative values for all pixels which have intensity lower than the background. All pixels with a positive value were set to zero.

Post-processing: The fovea can be falsely segmented; this was resolved using a circular mask of radius 80 pixels to remove the centre of the image. Also all objects with an area of less than 1200 pixels were removed to further eliminate spurious objects; this was at the cost of removing some smaller unconnected vessels.



(a) (b) (c)

Figure 2: Zoom-in region of retinal image: (a) optic disc, (b) segmentation of (a) without pre/post processing, (c) segmentation of (a) with pre/post-processing. © UK Biobank.



(a) (b) (c)

Figure 3: Zoom-in region of retinal image: (a) lens artefacts, (b) segmentation of (a) without pre/post processing, (c) segmentation of (a) with pre/post-processing. © UK Biobank.

3.2 Image Quality

Before proceeding with a description of the automated algorithm, it is important to detail the criteria used for manual image quality assessment. Using the segmented vessel map, images were manually rejected if:

(i) Less than half of the vasculature was segmented. This is often caused by poor illumination. Effects of poor illumination on segmentation are shown in figure 4.

(ii) Segmentation was considerably fragmented/unconnected.

(iii) Multiple non-vessel objects were segmented. The last category includes the false segmentation caused by eye lashes, lens artefacts, choroidal vessels, exudates,

haemorrhages, the fovea, the optic disc, retinal scars, retinitis pigmentosa, asteroid hyalosis etc. Figure 5 provides examples of falsely segmented non-vessel objects causing images to be manually rejected.

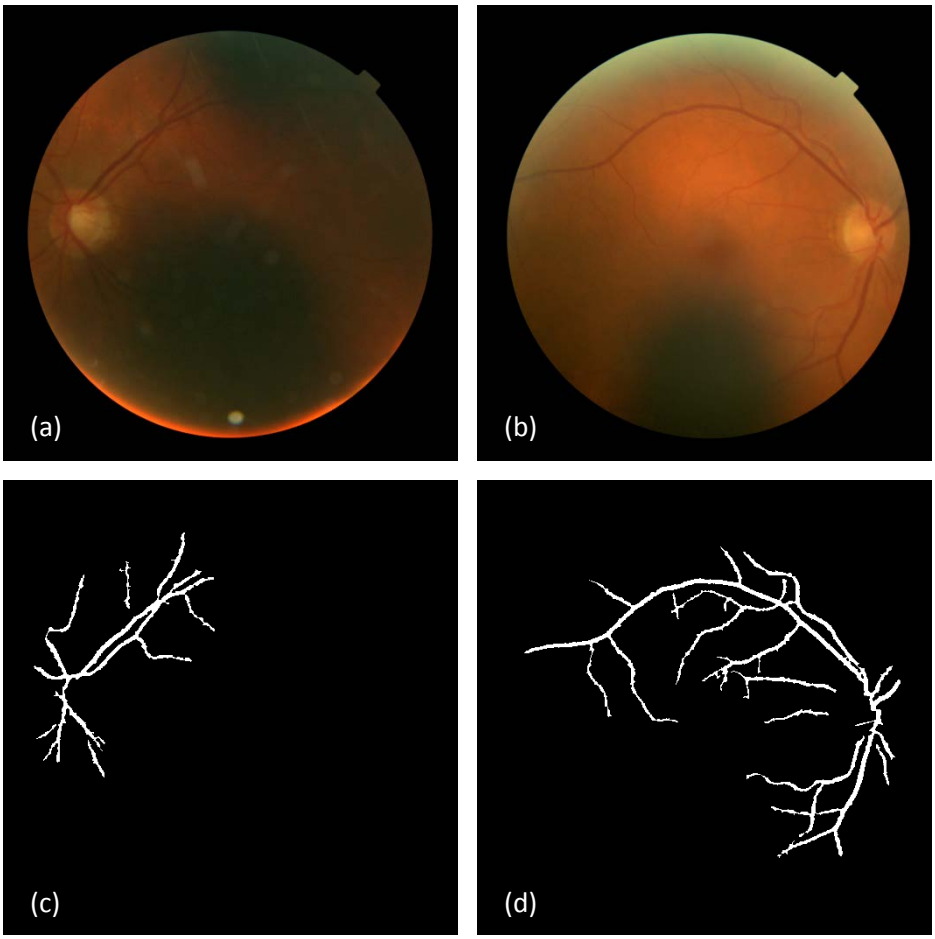


Figure 4: (a)-(b) Poorly illuminated retinal images, (c)-(d) segmentation of (a)-(b). The manual operator has labelled (c) inadequate and (d) adequate. © UK Biobank.

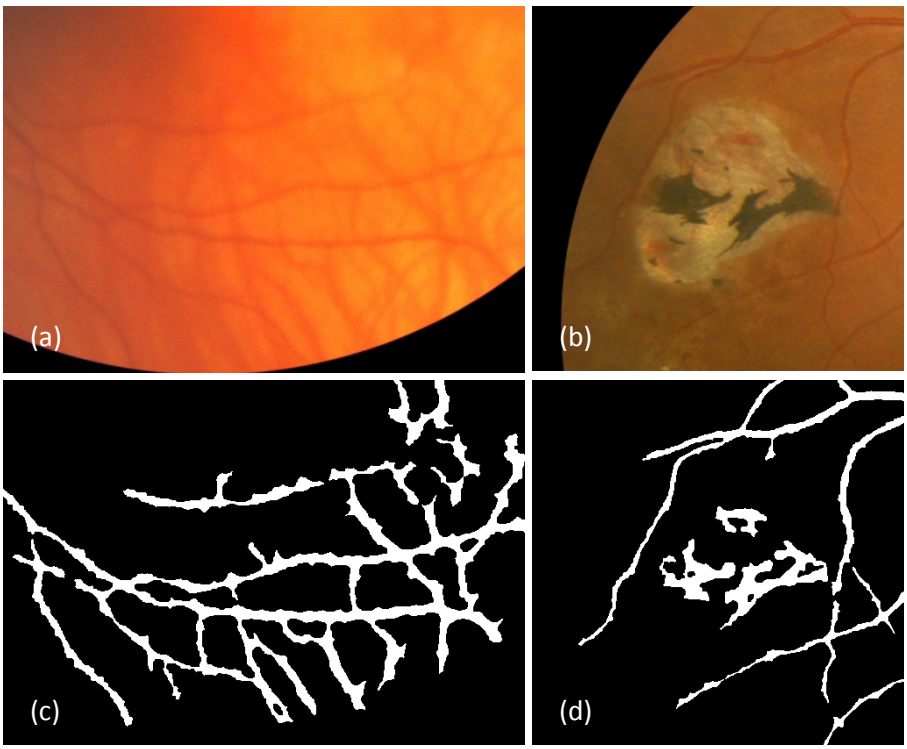


Figure 5: Zoom-in region of retinal images manually labelled as inadequate: (a) choroidal vessels, (b) retinal scar tissue, (c)-(d) segmentation of (a)-(b) showing falsely segmented non-vessel objects. © UK Biobank.

3.2.1 Feature Measurements

The segmented vessel map was skeletonised by means of morphological thinning [27] prior to the measurements of the features. The features were designed to mimic the criteria followed by the manual operator (stated above). The features are as follows:

(i) Area: The sum of all segmented pixels within the image.

(ii) Fragmentation: The number of connected components (8-connected) divided by the sum of all segmented pixels within the image.

(iii) Complexity: Spurs were removed (≤10 pixels). Bifurcation and crossover points were removed which were pixels with more than two eight-way neighbours, thus dividing the skeletonised vessel map into segments. Short segments were removed (≤10 pixels). Complexity is defined as the number of segments divided by the sum of all segmented pixels within the image.

The first feature was put in place to assess if a sufficient amount of vasculature had been segmented. The second feature was put in place to penalise fragmented segmentations. The perfect segmentation of a vascular tree would theoretically have only one connected component. The last feature was put in place to detect segmentations whose shape had deviated from that of a vessel map.

3.2.2 Classification

All features were normalised so that each feature had zero mean and unit standard deviation. Figure 6 provides a visualization of the distribution of two classes, plotted using the 3-dimensional feature set. Classification of retinal images into classes of inadequate or adequate was performed using a SVM classifier [28,29], chosen because it is a state-of-the-art method which is reported to possess a good classification performance. SVMs search for a linear decision surface (hyperplane) that can separate classes of objects and has the largest gap/margin between border-line objects. If the classes are not linearly separable, then a mathematical construction known as the kernel trick can be used to map the data in a higher dimensional space known as the feature space, where the separating linear decision exists and can be determined. There are several different kernel functions, which include the linear kernel, the Gaussian radial basis function kernel (RBF) and the polynomial kernel. These kernels often have parameters whose values need to be selected. Another

parameter associated with SVMs is the soft margin parameter C which is tuned to deal with noisy measurements and outliers.

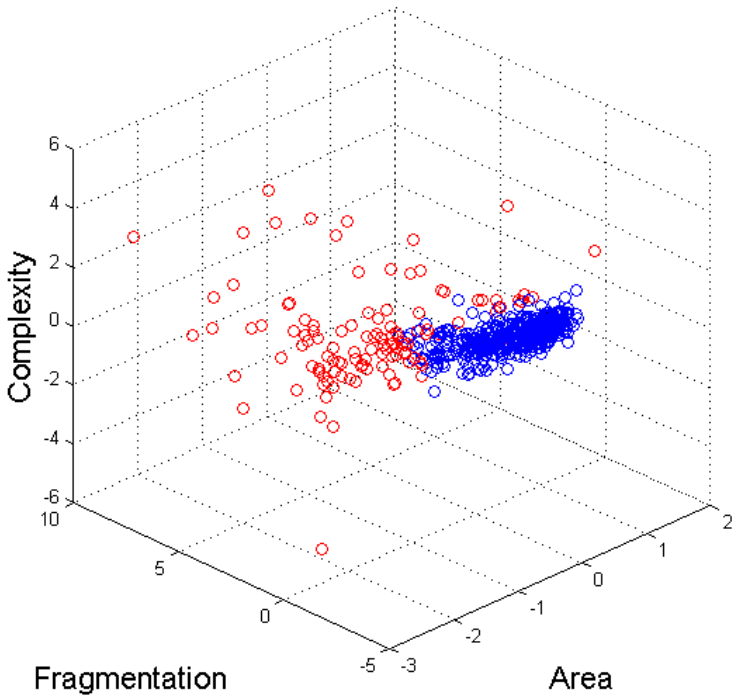


Figure 6: Distribution of the two classes (red = inadequate, blue = adequate). All features were normalised.

Due to its use in other modules of QUARTZ (a/v classification [30]), an ensemble classifier of bagged decision trees was also applied for classification. The classifier used 30 decision trees which was determined by using the out-of-bag classification error.

4. Experimental Evaluation:

4.1 Materials:

UK Biobank recruited 500,000 people aged between 40-69 years in 2006-2010 from across the country. They have undergone measures, provided blood, urine and saliva samples, provided detailed information about themselves and agreed to have their health followed. A subset of 68,151 participants had retinal images captured. This amounts to over 130,000 macular centred retinal images (captured from both eyes). Images were captured with a non-mydriatic fundus camera (Topcon 3DOCT-1000 Mk 2) with a 45 degree field-of-view and saved in PNG format with a resolution of 2048 x 1536 pixels. The UK Biobank study was approved by the Northwest Region NHS research ethics committee.

For the purpose of this evaluation, 800 retinal images were used from 400 random participants. The quality of images of this subset were assumed to represent the whole of the UK Biobank dataset. Manual image quality assessment was performed on these images using the criteria stated at the start of section 3.2. Out of the 800 images, 213 (26.62%) were manually labelled as inadequate and 587 (73.38%) were labelled as adequate. This equated to 327 (81.75%) participants with at least one image being adequate. The manually labelled images were randomly divided into two sets S_1 and S_2, each containing 400 images.

Two human observers were used for manual labelling. The labelling stated above was performed by the first observer and was used as the reference standard. The second observer labelled a random subset of 100 images out of the 800. A high agreement of 99% was achieved between the two observers, which aided the validation of the reliability of the manual labels for use in training and testing.

A SVM model selection procedure [31] was used to determine the optimal SVM kernel and parameters. This involved a grid search with the training and validation sets created from set S_1 using 10-fold cross validation. The feature value normalization was recalculated each time, leaving out the validation data. The performance criteria for the grid search was to find the kernel and parameters whose receiver operating characteristic curve got closest to the (0,1) point which represents a prefect classification. From inspection of figure 6, it's apparent that a linear kernel would perform reasonably well at separating the two classes. However, the best outcome was achieved with the use of the RBF kernel with the scaling factor (γ) of 2^{-1} and the soft margin parameter (C) of 2^{-4}.

With model selection complete, the SVM classifier was trained on the whole of set S^1 to generate the final classifier. Testing was performed on set S^2.

4.2 Performance Measures

Images were classified as either inadequate or adequate. Consequently there are four outcomes, two classifications and two misclassifications which are defined in table 1. The algorithm was evaluated in terms of sensitivity (SN) and specificity (SP). These are often used in machine learning and are measures of the quality of binary classification. These metrics are defined in table 2.

The receiver operating characteristic (ROC) curve allows for the visualization of the performance of the binary classifier system, expressing the trade-off between increased detection and false alarms. It is created by plotting the true positive rate (SN) versus the false positive rate (1-SP) at various thresholds of the probability score of the classifier (various operating points). The area under the curve (AUC) is an effective

measure of the system's performance. Finding the point on the curve closest to (0,1) is one particular approach to find the optimal operating point of the system.

	Inadequate image	Adequate image
Inadequate image detected	True positive (TP)	False positive (FP)
Inadequate image not detected	False negative (FN)	True negative (TN)

Table 1: Outcomes of image quality classification

Measure	Description
SN	TP/(TP+FN)
SP	TN/(TN+FP)

Table 2: Performance measures for image quality classification

4.2 Results

The ROC curve representing the performance for the detection of images of inadequate quality by the SVM based proposed algorithm on the test data (set S^2) is depicted in figure 7. The AUC value is 0.9828. The closest point on the curve to (0,1) is a distance away of 0.0937 and gives a sensitivity of 94.39% and a specificity of 92.49%. An alternate/preferred operating point gives a sensitivity of 95.33% and a specificity of 91.13%. A high sensitivity is preferred over a high specificity as it is important to detect and therefore avoid inadequate images from being included in the morphometric dataset for epidemiological analysis. The operating point achieved from

the maximum accuracy (TP+TN/TP+FP+TN+FN) was avoided as this would favour a high specificity considering a large portion of the images are negative (adequate quality).

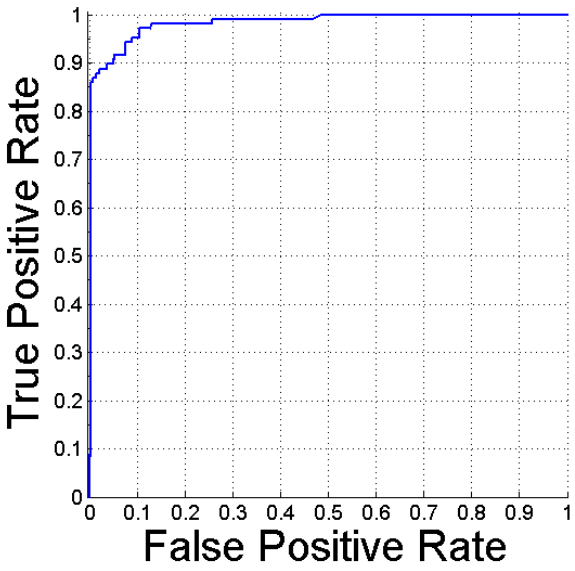


Figure 7: ROC curve for the proposed algorithm (SVM based).

The ROC curve representing the performance of the proposed system with the classifier switched to an ensemble system of bagged decision trees is shown in figure 8. The AUC value is 0.9726. The closest point on the curve to (0,1) is a distance away of 0.1003 and gives a sensitivity of 91.59% and a specificity of 94.54%. An alternate operating point gives a sensitivity of 95.33% and a specificity of 89.08%.

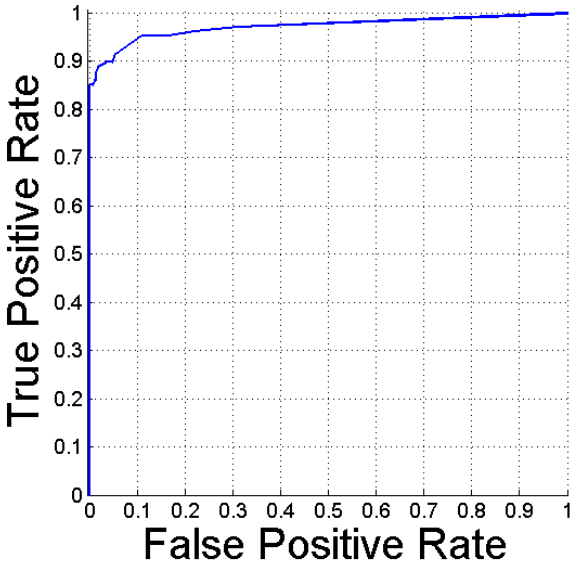


Figure 8: ROC curve for the proposed algorithm (ensemble based).

The results of the proposed method under these two classifiers are summarized in table 3. Although not directly relatable, it is useful to have an insight into the reported results of systems designed to detect retinal images of inadequate quality for conventional diagnostic procedures and these are provided in table 4.

Examples of classified images, for the SVM based proposed method at the stated alternate operating point, are given in figures 9-11. The Matlab code took 0.57 seconds on an Intel(R) core(TM) i7-4700MQ at 2.40 GHz to process each image. The segmentation which was separately timed took 13 seconds to process each image.

Method	AUC	Distance to (0,1)	SN	SP	Alternate	
					SN	SP
Proposed (SVM)	0.9828	0.0937	94.39%	92.49%	95.33%	91.13%
Proposed (ensemble)	0.9726	0.1003	91.59%	94.54%	95.33%	89.08%

Table 3: Performance of the proposed system designed for epidemiological studies.

Method	Year	AUC	SN	SP
Dias [19]	2014	0.9987	99.76%	99.49%
Usher [20]	2003	-	84.30%	95.00%
Fleming [21]	2006	-	99.10%	89.40%
Hunter [22]	2011	-	100.00%	93.00%
Paulus [24]	2010	0.9530	96.90%	80.00%

Table 4: Reported results of image quality assessment systems designed for conventional diagnostic procedures (e.g. detection of diabetic retinopathy).

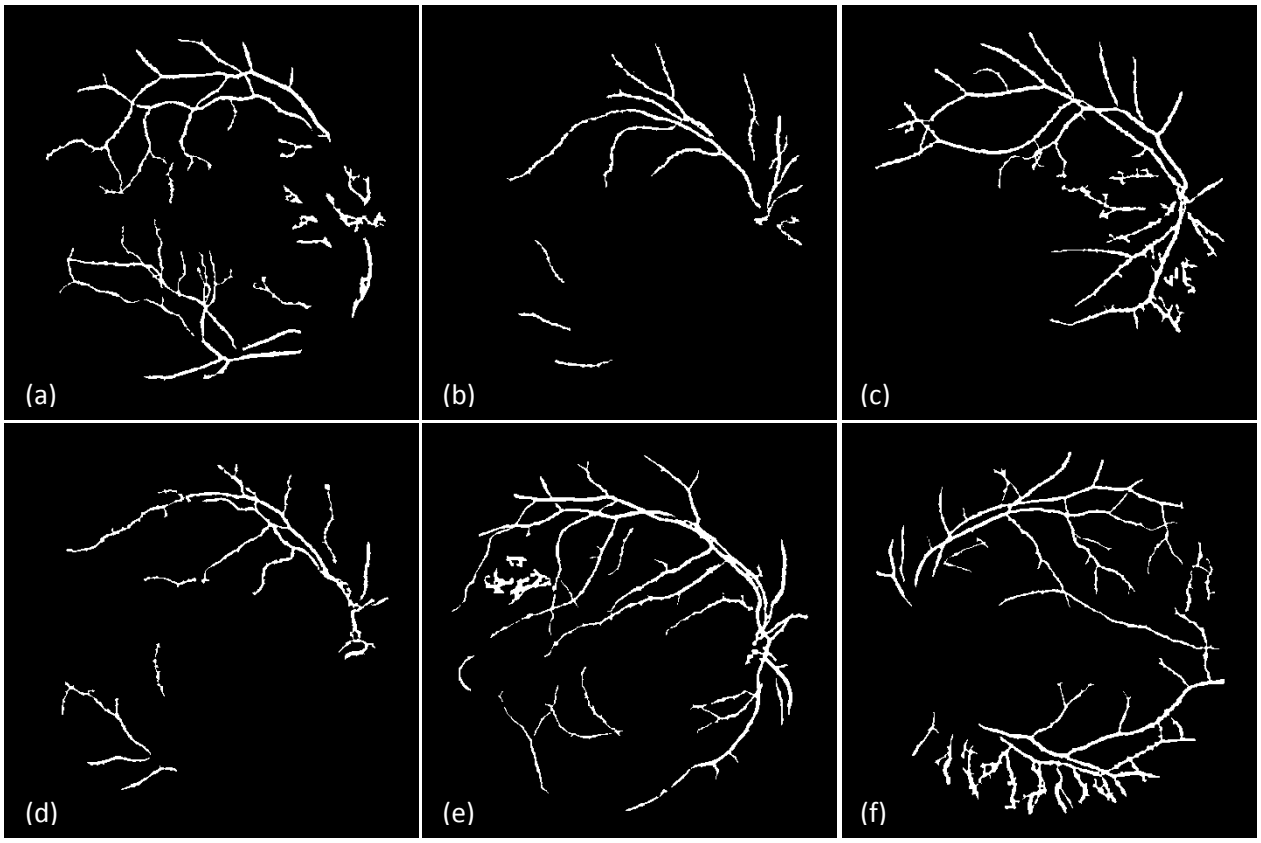


Figure 9: True positive outputs of the proposed system. © UK Biobank.

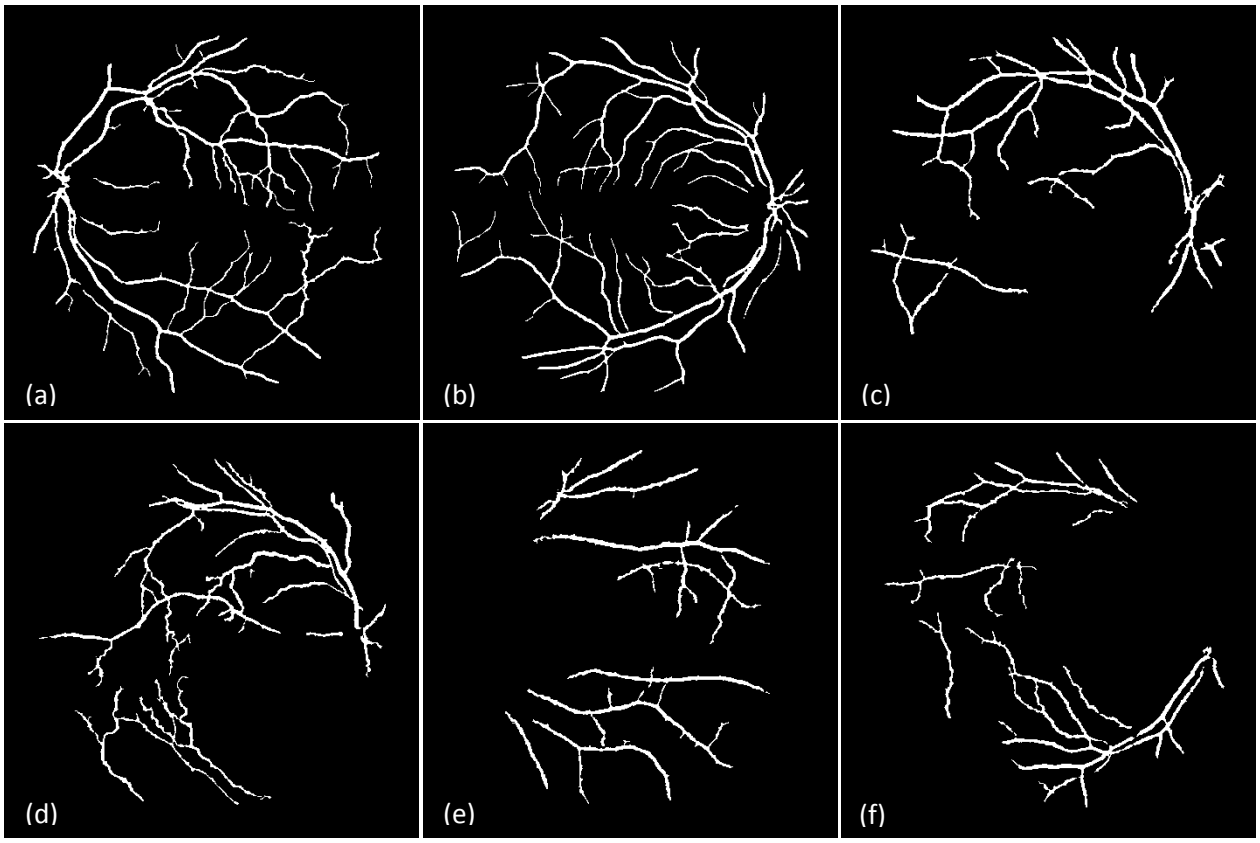


Figure 10: True negative outputs of the proposed system. © UK Biobank.

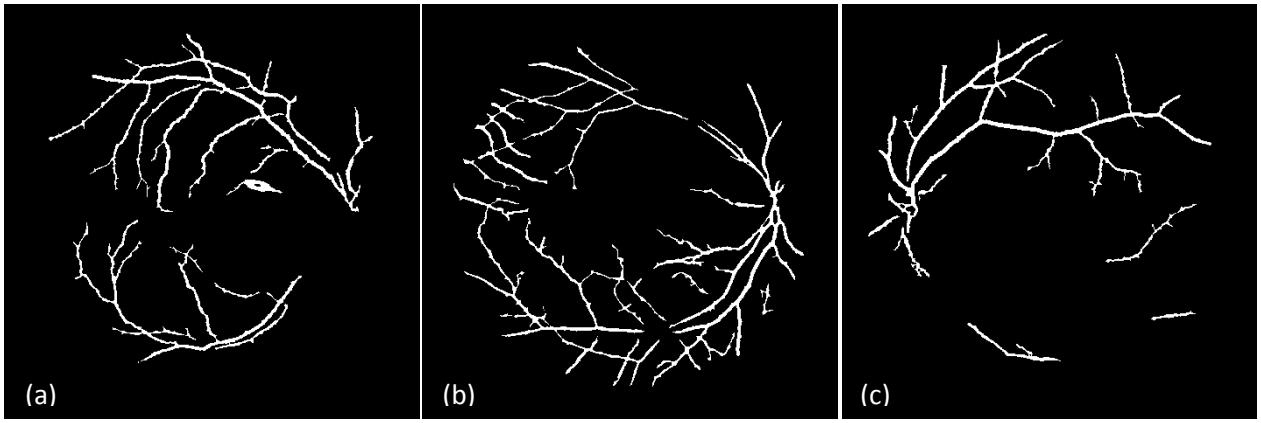


Figure 11: (a)-(b) False negative and (c) false positive outputs of the proposed system.

© UK Biobank.

5. Discussion and Conclusion:

In this paper, we have presented an effective image quality assessment method designed to detect retinal images that are of inadequate quality for retinal vascular morphometry in large-scale epidemiological studies. The method is based on the classification of images using the 3 simple global features that were measured from the segmented vasculature.

The performance of two different state-of-the-art classifiers was explored. Both provided impressive results; however, the SVM classifier is preferred as it slightly outperforms the ensemble classifier of bagged decision trees. The ease of implementing feature selection is a smart feature of the bagged ensemble, therefore this classifier would be advantageous if a larger feature set was involved. The proposed method using the SVM classifier produces a ROC curve with an AUC of 0.9828, containing an operating point with a sensitivity of 95.33% and a specificity of 91.13%. Once detected, inadequate images are then removed from further processing/analysis and therefore are not included in morphometric dataset to be used by the epidemiologists. Therefore, this operating point equates to 68.00% of all images being used, in which 98.16% of them will correctly be of adequate quality.

Image quality assessment is application specific. The application of our proposed method for epidemiological studies differs to those methods whose reported results are stated in table 4 for conventional diagnostic procedures, thus no real comparisons can be made in terms of performance. Assessing image quality in large retinal datasets for epidemiological studies has its own criteria and problems. Images from such large population datasets, such as UK Biobank and EPIC-Norfolk, have a high portion of images of poor quality, especially poorly/unevenly illuminated images. This is mainly caused by the capture of images without the use of pharmacological mydriasis. In

addition to this camera operators may only be provided with basic training/instructions and are not expert retinal photographers. Despite this there is a wealth of useful information that can be extracted from such datasets, and it is important to also make use of the images of poor quality where possible. Thus, when image quality was manually performed, poor images could be retained if they contained a good level of segmentation for at least half of the vasculature. Our proposed method was designed to mimic the manual operator.

Upon manual image quality labelling, 73.38% of this subset of UK Biobank dataset was deemed as having images with segmentations of adequate quality. This is higher than the 64.73% of images from the EPIC-Norfolk dataset reported in our prior work [11]. The higher percentage for UK Biobank was due to improvements in the pre-processing and post-processing steps of segmentation rather than any significant difference in the quality of images between the two datasets.

From the examples of classified images shown in figure 9 it is evident that our algorithm is capable of detecting inadequate images from a variety of causes. This includes images with the false segmentation of choroidal vessels and retinal scars and images with insufficient segmentation of the vasculature. Figure 10 illustrates the images correctly labelled as adequate. The majority of images contain a segmented vasculature with a full appearance as shown in figure 10(a-b). But also classified as adequate are poor images with incomplete vasculature segmentations figure 10(c-f), which is common due to poor illumination. As stated previously, it is important to extract information from such images. The results are impressive when considering the simplicity of the feature set containing global measures of area, fragmentation and complexity. Clearly the feature set is descriptive enough to allow the SVM to produce effective decision boundaries.

Figure 11 demonstrates when the proposed system can fail. In the first two examples, images are incorrectly labelled as adequate which contain the false segmentation of an elongated haemorrhage (figure 11(a)) and choroidal vessels (figure 11(b)). Choroid vessels are from a layer below the retina and hence, they are not normally visible. When visible they often contain little contrast to background, particularly in the green channel, and consequently are not segmented. On the occasions when they are falsely segmented, they usually cause dense and irregular regions in the segmentation and thus get a straight forward label from the proposed system of inadequate. However, in this occasion the segmentation of the choroidal vessels are dispersed throughout the image, and therefore the segmentation falsely appears to be adequate. Figure 11(c) illustrates an image incorrectly labelled as inadequate.

In the proceeding vessel analysis performed by QUARTZ, the vessel segmentation is skeletonized to only provide the initial vessel centrelines, which then have splines and Gaussian functions fitted, in order to perform quantitative analysis. However the segmentation must still perform to a high level. Whilst the algorithm proposed in this paper is assessing the quality of the segmentation, the standard method of evaluating vessel segmentation against a pixel labelled ground truth can be insightful. The performance of vessel segmentation and all of the other QUARTZ modules (a/v classification, optic disc localization, vessel width measures) have been previously evaluated [10,11] against several datasets including the EPIC-Norfolk dataset. The performances of these modules on the UK Biobank dataset, which now operate in conjunction with the automated image quality module, have shown impressive preliminary results. Upon complete evaluation, a full publication of these performances will be produced.

In conclusion, this paper has demonstrated an automated system that is capable of detecting and thus removing retinal images of inadequate quality for the creation of

vessel morphometric data suitable for epidemiological studies. This will be incorporated in to our QUARTZ software and as a result the need for any manual processing will be completely removed. For the UK Biobank retinal dataset, the image quality assessment stage could be completed in approximately 22 hours on a single computer as opposed to 567 hours of manual processing time. This will make it practical to run the QUARTZ software on the 136,000 images of the UK Biobank retinal dataset and other large retinal datasets. Once processed the UK Biobank dataset will be used in epidemiological studies, examining the association of vessel morphology to disease risk factors and outcomes.

Acknowledgements:

This research has been conducted using the UK Biobank resource. The UK Biobank Eye and Vision Consortium is supported by funding from The Special Trustees of Moorfields Eye Hospital NHS Foundation Trust, and at the NIHR Biomedical Research Centre at Moorfields Eye Hospital and UCL Institute of Ophthalmology. The image analysis work is supported by grants from the Medical Research Council Population and Systems Medicine Board (Grant no. MR/L02005X/1) and Fight for Sight.

Conflict of Interest Statement:

None declared.

Members of the UK Biobank Eye and Vision Consortium:

Prof Tariq Aslam, Dr Sarah Barman, Prof Paul Bishop, Mr Peter Blows, Dr Catey Bunce, Dr Roxana Carare, Prof Usha Chakravarthy, Miss Michelle Chan, Mrs

Antonietta Chianca, Dr Valentina Cipriani, Prof David Crabb, Mrs Philippa Cumberland, Dr Alexander Day, Miss Parul Desai, Prof Bal Dhillon, Prof Andrew Dick, Prof Paul Foster, Dr John Gallacher, Prof David (Ted) Garway-Heath, Mr Srini Goverdhan, Prof Jeremy Guggenheim, Mrs Priyal Gupta, Prof Chris Hammond, Dr Ruth Hogg, Prof Anne Hughes, Mr Pearse Keane, Prof Sir Peng Tee Khaw, Mr Anthony Khawaja, Mr Gerassimos Lascaratos, Prof Andrew Lotery, Prof Phil Luthert, Dr Tom MacGillivray, Dr Sarah Mackie, Prof Keith Martin, Ms Michelle McGaughey, Dr Bernadette McGuinness, Dr Gareth McKay, Mr Martin McKibbin, Dr Danny Mitry, Prof Tony Moore, Prof James Morgan, Ms Zaynah Muthy, Mr Eoin O'Sullivan, Dr Chris Owen, Mr Praveen Patel, Dr Tunde Peto, Prof Jugnoo Rahi, Dr Alicja Rudnicka, Miss Carlota Grossi Sampedro, Mr David Steel, Mrs Irene Stratton, Mr Nicholas Strouthidis, Prof Cathie Sudlow, Dr Caroline Thaung, Miss Dhanes Thomas, Prof Emanuele Trucco, Mr Adnan Tufail, Prof Stephen Vernon, Mr Ananth Viswanathan, Miss Cathy Williams, Dr Katie Williams, Prof John Yates, Dr Max Yates, Dr Jennifer Yip, Dr Haogang Zhu.

References:

[1] T.Y. Wong, R. Klein, A.R. Sharrett, B.B. Duncan, D.J. Couper, J.M. Tielsch, et al., Retinal arteriolar narrowing and risk of coronary heart disease in men and women: the Atherosclerosis Risk in Communities Study, JAMA: The Journal of the American Medical Association 287 (2002) 1153-1159.

[2] N. Cheung, S.M. Saw, F.M. Islam, S.L. Rogers, A. Shankar, K. Haseth, et al., BMI and retinal vascular caliber in children, Obesity 15 (2007) 209-209.

[3] T.Y. Wong, F.M. Islam, R. Klein, B.E. Klein, M.F. Cotch, C. Castro, et al., Retinal vascular caliber, cardiovascular risk factors, and inflammation: the multi-ethnic study of

atherosclerosis (MESA), Investigative Ophthalmology & Visual Science 47 (2006) 2341-2350.

[4] C.G. Owen, A.R. Rudnicka, C.M. Nightingale, R. Mullen, S.A. Barman, N. Sattar, et al., Retinal arteriolar tortuosity and cardiovascular risk factors in a multi-ethnic population study of 10-year-old children; the Child Heart and Health Study in England (CHASE), Arteriosclerosis, Thrombosis, and Vascular Biology 31 (2011) 1933-1938.

[5] Department of Health, UK Government. Long Term Conditions Compendium of Information (Third Edition), https://www.gov.uk/ (2012).

[6] Deaths registered in England and Wales in 2013, by cause, Office for National Statistics (2014).

[7] M.D. Abràmoff, M.K. Garvin, M. Sonka, Retinal imaging and image analysis, Biomedical Engineering, IEEE Reviews in 3 (2010) 169-208.

[8] UK Biobank, http://www.ukbiobankeyeconsortium.org.uk/ (2013).

[9] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J.P. Hubschman, et al., Validating retinal fundus image analysis algorithms: issues and a proposal, Investigative Ophthalmology & Visual Science 54 (2013) 3546-3559.

[10] M.M. Fraz, R.A. Welikala, A.R. Rudnicka, C.G. Owen, D.P. Strachan, S.A. Barman, QUARTZ: Quantitative Analysis of Retinal Vessel Topology and size - An automated system for quantification of retinal vessels morphology, Expert Systems with Applications 42 (2015) 7221-7234.

[11] R.A. Welikala, M.M. Fraz, S. Hayat, A.R. Rudnicka, P.J. Foster, P.H. Whincup, et al., Automated retinal vessel recognition and measurements on large datasets,

Engineering in Medicine and Biology Society, EMBC, Annual International Conference of the IEEE (2015) 5239-5242.

[12] EPIC-Norfolk, European Prospective Investigation of Cancer (EPIC). http://www.Srl.Cam.Ac.uk/epic/ (2013).

[13] A. Perez-Rovira, T. MacGillivray, E. Trucco, K.S. Chin, K. Zutis, C. Lupascu, et al., VAMPIRE: Vessel Assessment and Measurement Platform for Images of the REtina, Engineering in Medicine and Biology Society, EMBC, Annual International Conference of the IEEE (2011) 3391-3394.

[14] T.J. MacGillivray, J.R. Cameron, Q. Zhang, A. El-Medany, C. Mulholland, Z. Sheng, et al., Suitability of UK Biobank retinal images for automatic analysis of morphometric properties of the vasculature, PLoS ONE 10 (2015) e0127914.

[15] M. Ortega, N. Barreira, J. Novo, M.G. Penedo, A. Pose-Reino, F. Gómez-Ulla, Sirius: A web-based system for retinal image analysis, International Journal of Medical Informatics 79 (2010) 722-732.

[16] P. Bankhead, C.N. Scholfield, J.G. McGeown, T.M. Curtis, Fast retinal vessel detection and measurement using wavelets and edge location refinement, PLoS ONE 7 (2012) e32435.

[17] S.C. Lee and Y. Wang, Automatic retinal image quality assessment and enhancement, Proceedings of SPIE Medical Imaging: Image Processing (1999) 1581-1590.

[18] M. Lalonde, L. Gagnon, M.C. Boucher, Automatic visual quality assessment in optical fundus images, Proceedings of Vision Interface 32 (2001) 259-264.

[19] J.M.P. Dias, C.M. Oliveira, L.A. da Silva Cruz, Retinal image quality assessment using generic image quality indicators, Information Fusion 19 (2014) 73-90.

[20] D.B. Usher, M. Himaga, M.J. Dumskyj, J.F. Boyce, Automated assessment of digital fundus image quality using detected vessel area, Proceedings of Medical Image Understanding and Analysis (2003) 81-84.

[21] A.D. Fleming, S. Philip, K.A. Goatman, J.A. Olson, P.F. Sharp, Automated assessment of diabetic retinal image quality based on clarity and field definition, Investigative Ophthalmology & Visual Science 47 (2006) 1120-1125.

[22] A. Hunter, J. Lowell, M. Habib, B. Ryder, A. Basu, D. Steel, An automated retinal image quality grading algorithm, Engineering in Medicine and Biology Society, EMBC, Annual International Conference of the IEEE (2011) 5955-5958.

[23] L. Giancardo, M.D. Abràmoff, E. Chaum, T.P. Karnowski, F. Meriaudeau, K.W. Tobin Jr, Elliptical local vessel density: a fast and robust quality metric for retinal images, Engineering in Medicine and Biology Society. EMBS. 30th Annual International Conference of the IEEE (2008) 3534-3537.

[24] J. Paulus, J. Meier, R. Bock, J. Hornegger, G. Michelson, Automated quality assessment of retinal fundus photos, International Journal of Computer Assisted Radiology and Surgery 5 (2010) 557-564.

[25] M. Niemeijer, M.D. Abramoff, B. van Ginneken, Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening, Medical Image Analysis 10 (2006) 888-898.

[26] U.T.V. Nguyen, A. Bhuiyan, L.A.F. Park, K. Ramamohanarao, An effective retinal blood vessel segmentation method using multi-scale line detection, Pattern Recognition 46 (2013) 703-715.

[27] Z. Guo and R.W. Hall, Parallel thinning with two-subiteration algorithms, Communications of the ACM 32 (1989) 359-373.

[28] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, Proceedings of the Fifth Annual Workshop on Computational Learning Theory (1992) 144-152.

[29] N. Cristianini and J. Shawe-Taylor, An introduction to support vector machines and other kernel-based learning methods, (Cambridge University Press, 2000).

[30] M.M. Fraz, A.R. Rudincka, C.G. Owen, D.P. Strachan, S.A. Barman, Automated arteriole and venule recognition in retinal images using ensemble classification, 9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP) (2014).

[31] C.W. Hsu, C.C. Chang, C.J. Lin, A practical guide to support vector classification, Department of Computer Science, National Taiwan University (2003).