

An Artificial Intelligence-Based Deep Learning Algorithm for the Diagnosis of Diabetic Neuropathy Using Corneal Confocal Microscopy: A Development and Validation Study

Dr Bryan M. Williams^{1,2,3}, Dr Davide Borroni^{2,4}, Dr Rongjun Liu⁵, Dr Yitian Zhao^{2,7}, Dr Jiong Zhang⁸, Dr Jonathan Lim⁹, Mr Baikai Ma⁵, Dr Vito Romano^{1,2}, Prof Qi Hong⁵, Dr Maryam Ferdousi¹⁰, Dr Ioannis N. Petropoulos⁶, Mr Georgios Ponirakis⁶, Prof Stephen Kaye^{1,2}, Prof Rayaz A. Malik⁶, Dr Uazman Alam^{*,1,9}, Dr Yalin Zheng^{*,1,2}

¹Department of Eye and Vision Science, University of Liverpool, Liverpool L7 8TX, UK

²St. Paul's Eye Unit, Royal Liverpool University Hospital, Liverpool, L7 8XP, UK

³Data Science Institute, Lancaster University, Lancaster, LA1 4YW, UK

⁴Department of Ophthalmology, Riga Stradins University, Riga, Latvia

⁵Department of Ophthalmology, Peking University Third Hospital, Beijing, China

⁶Weill Cornell Medicine – Qatar, Doha, Qatar.

⁷Cixi Institute of Biomedical Engineering, Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo, China

⁸Laboratory of Neuro Imaging, Institute for Neuroimaging and Informatics, Keck School of Medicine, University of Southern California, Los Angeles, USA

⁹Institute of Cardiovascular Sciences, University of Manchester, Manchester, UK.

¹⁰Department of Endocrinology and Diabetes, University Hospital Aintree, Longmoor Lane, Liverpool L9 7AL, United Kingdom

*Donates joint senior and corresponding authors:

Corresponding Authors

Yalin Zheng, Department of Eye and Vision Science, University of Liverpool, William Henry Duncan Building, West Derby Street, Liverpool, L7 8TX, 0151 794 9055, yalin.zheng@liverpool.ac.uk, <https://orcid.org/0000-0002-7873-0922>.

Uazman Alam, Department of Eye and Vision Science, University of Liverpool, William Henry Duncan Building, West Derby Street, Liverpool, L7 8TX, 0151 529 5920, uazman.alam@liverpool.ac.uk, <https://orcid.org/0000-0002-3190-1122>.

Word count of main text

3931 words.

Tweet

Check out our latest study, developing an artificial intelligence approach for analysing the corneal nerve fibres and diagnosing diabetic neuropathy. Top results in quantifying nerve fibre length, branch points, tail points, nerve segments and fractals. Excellent classification of healthy controls and participants with neuropathy. Results are fast and accurate. Huge potential for diabetic neuropathy screening.

Abstract

Aims

Corneal confocal microscopy is a rapid non-invasive ophthalmic imaging technique that identifies peripheral and central neurodegenerative disease. Quantification of the corneal subbasal nerve plexus morphology, however, currently requires either time-consuming manual annotation or a less sensitive automated image analysis approach. The aim of this study was to describe the development and validation of an artificial intelligence–based, deep learning algorithm for the quantification of nerve fibre properties relevant to the diagnosis of diabetic neuropathy in comparison to a validated automated analysis program, ACCMetrics.

Methods

Our deep learning algorithm, which employs a convolutional neural network with data augmentation, was developed for the automated quantification of the corneal subbasal nerve plexus for the diagnosis of diabetic neuropathy. The algorithm was trained using a high-end graphics processor unit on 1,698 corneal confocal microscopy images; for external validation, it was further tested on 2137 images. The algorithm was developed to identify total nerve fibre length, branch points, tail points, number and length of nerve segments, and fractals. Sensitivity analyses were undertaken to determine the area under the curve for ACCMetrics and our algorithm for the diagnosis of diabetic neuropathy.

Results

The intra-class correlation coefficient was superior with our algorithm compared to ACCMetrics for total corneal nerve fibre length (0.933 vs 0.825), mean length per segment (0.656 vs 0.325), number of branch points (0.891 vs 0.570), number of tail points (0.623 vs 0.257), number of nerve segments (0.878 vs 0.504) and fractals (0.927 vs 0.758). In addition, our proposed algorithm achieved an area under the curve of 0.83, specificity of 0.87 and sensitivity of 0.68 for the classification of healthy controls (n=90) and participants with neuropathy (n=132) (defined by the Toronto Criteria).

Conclusions

These results have demonstrated that our deep learning algorithm provides rapid and excellent localisation performance for the quantification of corneal nerve parameters. This model has potential for adoption into clinical screening programmes for diabetic neuropathy.

Keywords

Corneal Confocal Microscopy

Corneal nerve

Deep Learning

Image Segmentation

Diabetic Neuropathy

Ophthalmic Imaging

Image Processing and Analysis

Small nerve fibres

Research in Context

What is already known about this subject?

Diabetic neuropathy can be identified with corneal confocal microscopy, a rapid non-invasive ophthalmic imaging technique.

Corneal confocal microscopy provides an objective quantitative method to diagnose diabetic neuropathy which is currently not done in routine clinical practice

Quantification of the corneal subbasal nerve plexus morphology requires either time-consuming manual annotation or a less sensitive automated image analysis approach.

What is the key question?

Can a more accurate, fully automated algorithm be developed to quantify corneal nerve fibre parameters for the diagnosis of diabetic neuropathy?

What are the new findings?

This newly developed algorithm is capable of segmenting the fibres of the subbasal nerve plexus.

The performance of this algorithm is superior to that of the existing state-of-the-art, ACCMetrics in quantifying nerve morphology.

On the back of this algorithm, good results are achieved in the automated diagnosis of diabetic neuropathy.

How might this impact on clinical practice in the foreseeable future?

The speed and accuracy of this algorithm give it potential to be adopted into clinical screening programmes for diabetic neuropathy.

Abbreviations

ACCM:	ACCMetrics Model
ANOVA:	Analysis of Variance
AUC:	Area under Curve
CCM:	Corneal Confocal Microscopy
CNN:	Corneal Nerve Fibre
CNN:	Convolutional Neural Network
DLA:	Deep Learning Algorithm
DPN:	Diabetic Peripheral Neuropathy
DSC:	Dice Similarity Coefficient
ENA:	Early Neuropathy Assessment
ICC:	Intra-class Coefficient
IENF:	Intra-epidermal Nerve Fibres
IENFD:	Intra-epidermal Nerve Fibres Density
LCNN:	Our Preliminary Model
LDLA:	Our Refined Ensemble Deep Learning Algorithm
LoA:	Limits of Agreement
MA:	Manual Annotation
RMSE:	Root Mean Squared Error
ROC:	Receiver Operating Characteristic
SBP:	Sub-basal Nerve Plexus
SD:	Standard Deviation
SFN:	Small Fibres Neuropathy
SVM:	Support Vector Machines

Introduction

The prevalence of diabetic peripheral neuropathy (DPN) can be as high as 50% in an unselected population. Currently, screening for DPN most commonly relies on the 10g monofilament test, which identifies those at risk of foot ulceration but is poor for identifying patients with early neuropathy [1]. While other methods may be employed for screening i.e. clinical examination, questionnaires, vibration perception threshold, they do not directly quantify small nerve fibres directly, which are the earliest site of injury. Skin biopsy enables direct visualisation of thinly myelinated and unmyelinated nerve fibres, which are the earliest affected in DPN. Skin biopsy can be used to diagnose small fibres neuropathy (SFN) [2]. The assessment of intra-epidermal nerve fibres (IENF) and their densities (IENFD) are currently advocated in clinical practice in the US [3] and recommended as an endpoint in clinical trials [4]. However, skin biopsy is invasive and requires specialised laboratory facilities for analysis. The cornea is the most densely innervated tissue of the human body, containing a network of unmyelinated axons (small nerve fibres) called the sub-basal nerve plexus (SBP).

Corneal confocal microscopy (CCM) has been utilised to image the SBP, which has been shown to be remarkably stable in healthy corneas over three years [5] but demonstrates early and progressive pathology in a range of peripheral and central neurodegenerative conditions [6-11]. Fig. 1 (a, b) and (c, d) shows examples from control and diabetic neuropathy participants respectively. Previous studies have demonstrated analytical validation by showing that CCM reliably quantifies early axonal damage in DPN [12, 13] with high sensitivity and specificity [14, 15] and closely correlates to the loss of intra-epidermal nerve fibres [15, 16]. CCM also predicts incident diabetic neuropathy [17] and can detect corneal nerve regeneration in people with DPN [18]. CCM may also detect early nerve fibre

loss before intra-epidermal nerve fibre loss in skin biopsy [19]. In some patients, corneal nerve fibre (CNF) loss may be the first evidence of subclinical DPN [19]; Brines et al. [20] have shown that CNF area and width distribution area may improve the diagnostic and predictive ability of CCM.

To accurately quantify CNF morphology, nerves must be distinguished from background and other cell structures accurately. The major limitation for wider clinical utilisation is the need for manual image analyses, which is highly labour-intensive and requires considerable expertise to quantify nerve pathology [21]. The development of methods for the objective, reliable and rapid analysis of corneal nerves is vital if CCM is to be adopted for screening and large clinical trial programmes. Furthermore, to be used as a diagnostic tool, it is essential to extract the measurements automatically with high reliability [21]. Dabbah et al. [22] presented a dual-model automated detection method of CNFs using CCM, showing excellent correlation with manual ground-truth analysis ($r=0.92$). They further refined this method utilising the dual-model property in a multi-scale framework to generate feature vectors from localised information at every pixel and showed an even stronger correlation with the ground-truth ($r=0.95$) [21]. This study, however, used neural networks without convolution layers [21], which necessitates pre-processing and encourages overfitting.

Kim et al. [23] have also developed a nerve segmentation technique to delineate corneal nerve morphology in CCM. These range from filtering methods with rapid implementation but low-contrast and imprecise focus, to more complex support vector machine approaches, which rely on features defined by the user. Chen et al [24] presented a method based on feature engineering, achieving state-of-the-art results; but the reliance on hand-crafted features increases complexity to the user and can introduce user-bias, returning sub-optimal results [25].

Recently, machine learning based approaches have achieved excellent performance in computer vision and medical image analysis tasks. Deep learning and particularly Convolutional Neural Networks (CNN), a class of deep neural networks, have emerged as a highly effective branch of machine learning for image classification [25]. This approach allows for “end-to-end” classification results to be achieved, allowing the end user to obtain a result without the need for specifying or designing features or setting example-specific parameters. CNN design follows vision processing in living organisms [26] with the connectivity pattern between neurons resembling visual cortex organisation. Based on training with pre-annotated data, CNNs combine the traditionally separate machine learning tasks of (i) feature designing, (ii) learning and (iii) image classification in a single model, relieving the traditional machine-learning burden of designing handcrafted features. More recently, this has extended beyond image-wise classification to efficient pixel-wise classification, allowing image segmentation to be achieved, i.e. pixels may be classed as belonging to an object of interest or not. There has been a significant increase in recent years in the development of deep learning algorithms (DLAs) with CNNs achieving excellent performance in many computer vision applications and showing significant clinical utility in healthcare [27]. Accurate automated detection of corneal nerves from CCM has many potential benefits over manual detection such as objectivity, increased efficiency and reproducibility. This can enhance early disease diagnostics and improve patient outcomes. In addition, artificial intelligence-based DLAs have the added advantage of continual learning and refinement alongside concurrent analysis.

The aim of this study was to develop and validate a DLA for corneal nerve segmentation in CCM images and compare this with a widely used and validated automated image analysis software, ACCMetrics (Early Neuropathy Assessment (ENA) group, University of Manchester, Manchester, UK) [24].

Research Design and Methods

Participants

All participants gave informed consent at the respective institutions and the studies were conducted in accordance with the Declaration of Helsinki. Relevant ethical and institutional approvals were gained prior to the imaging of all participants.

Image Datasets

In this study, 3835 confocal images of the sub-basal corneal nerve plexus were utilised from healthy volunteers and people with diabetes from Padova, Italy (n=120), Beijing, China (n=1578) and Manchester, UK (n=2137). Fig. 1(e) shows an example CCM image, (f) is the manual annotation of (e) with branching and tail points highlighted in red and green respectively in (g).

Dataset 1 (BioImLab, University of Padova, Italy)

120 images were obtained from Ruggeri's BioImLab at the Department of Information Engineering, University of Padova, Italy. Of these, the first 30 images are from 30 volunteers (1 image per volunteer) who are either healthy or showing different pathologies (diabetes, pseudoexfoliation syndrome, keratoconus) [28]. The images were captured in tiff format at size 384x384 pixels with a Heidelberg Retina Tomograph II using the Rostock Corneal Module (HRTII32-RCM) confocal laser microscope (Heidelberg Engineering GmbH, Heidelberg, Germany). The remaining 90 images are of the corneal sub-basal epithelium from normal and pathological subjects with one image per subject [29] using a ConfoScan 4 CCM at 40X magnification (Nidek Technologies, Padova, Italy). An area of 460x350µm was captured in 768x576 pixels and stored in monochrome jpg compressed format.

Dataset 2 (Peking University Third Hospital, Beijing, China)

1578 images (384x384 pixels in tiff format) were acquired from healthy volunteers (n=90) and corneas with various corneal conditions (n=105 including 52 participants with diabetes) using the Heidelberg Retina Tomograph 3/Rostock Cornea Module (Heidelberg Engineering, Heidelberg, Germany). Six images per eye were obtained where possible from the corneal apex using the same methodology developed and utilised by ENA group, University of Manchester, Manchester, UK.

Dataset 3 (Early Neuropathy Assessment (ENA) Group, University of Manchester, UK)

2137 images were analysed from healthy-volunteers and participants with diabetes (n=444). All CCM images were obtained using the same methodology utilising the standard, internationally accepted protocol developed by the ENA group [13]. The images were captured using the Rostock corneal module set at +12 objective lens. The image size obtained was 400x400 micrometres (384x384 pixels). The images were exported in bmp format, which is compatible with the image analysis software. Images were from the following cohorts:

Group 1: Healthy-volunteer controls (n=90).

Group 2: Participants with impaired glucose tolerance (n=53 including 26 with definite neuropathy).

Group 3: Participants with type 1 diabetes with definite neuropathy (n=37)

Group 4: Participants with type 1 diabetes without neuropathy (n=53).

Group 5: Participants with type 2 diabetes without (n=101) and definite with neuropathy (n=49).

Group 6: Participants with type 2 diabetes with mild neuropathy (n=41) and definite neuropathy (n=20).

Definite neuropathy was defined as the presence of an abnormality of nerve conduction studies (age-matched) and a symptom or symptoms or a sign or signs of neuropathy as defined by the Toronto consensus statement by the American Diabetes Association on DPN [30]. In total, 132 participants had definite neuropathy across the groups. Note that the depth of images are only marginally different for each participant and depends on their corneal thicknesses so there is no definitive depth. However, the subbasal nerve plexus occurs at $\sim 50\mu\text{m}$ depth in most people, with no difference in people with diabetes and healthy volunteer controls [31].

Image Annotation

In order to obtain a ground-truth for each image in the BioImLab and Beijing datasets, the corneal nerves in each image were manually traced by a clinical ophthalmologist (DB) using an in-house program written in Matlab (Mathworks R2017, Natwick). Our previous work has demonstrated the validity of manual annotations in terms of intra- and inter-observer agreements [32]. Dataset 3 was not annotated and only used for clinical testing using the deep learning segmentations.

Methods

In this section, the new method for automatically segmenting the corneal nerves in CCM images is presented. The preparation of our combined datasets for use in a training and testing approach is given and we define our automated segmentation method. This is then built on with ensemble learning and random sampling. Finally, clinically-relevant parameters were extracted and compared with the existing state-of-the-art ACCMetrics.

Dataset Preparation for Training-Testing Approach

The BioImLab and Beijing datasets have 1698 images in total and were used for the development of the model: 1494 ($\sim 90\%$) images from the Beijing dataset were used for training, while 84 images from Beijing and all the BioImLab dataset were used for testing. Each image in these datasets was used for either training or testing to avoid overfitting. Dataset 3 (ENA image dataset) was only used for clinical testing and validation but not to train the model. The images for training and testing were selected using a random permutation at the patient level determined using a Mersenne Twister method [33]. Note that splitting took place on the image (rather than patient) level in order to avoid potential bias.

All the images were standardised to have a pixel size of $1.04\mu\text{m}$ (384×384 pixels) by bilinear interpolation. To increase the dataset size, it is augmented by extracting patches of size 128×128 pixels with a 32-pixel overlap, creating 81 patches per image. The selection of patches used for training/testing was done on an image-level to avoid testing patches from images whose data have been used for training.

Image Segmentation Using Deep Learning

Corneal nerves were segmented adopting U-Net CNN architecture [34]. Unlike conventional CNNs, which aim to assign one classification (or more) to an image, this type of architecture aims to achieve full-image segmentation by determining a pixel-wise segmentation map. Fig. 2 illustrates the architecture of our proposed U-Net model. It can be visualised as a U-shape, whose left side is an encoding path and right side is a decoding path. At the end of the architecture, a sigmoid activation function is employed to create a segmentation map. A key feature of U-Net is direct connectivity

between the encoding and decoding layers, allowing extracted feature reuse and strengthening feature propagation. The Dice Similarity Coefficient (DSC) was used as a cost function, i.e. to measure error during training.

The models were developed with Python 3.5.2 (<https://www.python.org/>), Tensorflow 1.0.0 (<https://www.tensorflow.org/>, Google Inc, Mountain View, CA), and Keras 1.2.1 (<https://keras.io/>). The model was trained for 200 iterations using an NVIDIA K40 GPU (NVIDIA, Santa Clara, CA). Following the training step, the trained model weights were used to obtain the segmentation maps of each previously unseen testing patch.

Ensemble Classification

In order to improve the accuracy of the model by using multiple copies, an ensemble of 5 U-Net networks were trained on the same training data using a random-sample-with-replacement approach. The final prediction was computed by a majority vote over the predictions of the ensemble network. In the following sections, training with our preliminary model, a single U-Net model, is referred to as “LCNN” (Liverpool Convolutional Neural Network) while training with our refined ensemble deep learning approach is referred to as “LDLA” (Liverpool Deep Learning Algorithm).

Image Reconstruction and Parameter Extraction

The trained models were able to produce segmentations on a patch basis. The segmentation of a whole CCM image was obtained by combining the segmentations of all its patches using majority voting on the overlap regions. From the image-level segmentation result, further analysis was carried out to derive the clinically relevant parameters including the corneal nerve length, branch points, tail points, and fractal number [35].

Evaluation

Our models, LCNN and LDLA, together with the state-of-the-art ACCMetrics approach [24], denoted as “ACCM” were compared with the manual annotation. The performance of the algorithm was measured using the Bland-Altman approach. Agreement between the automatic segmentations and manual annotations were assessed using the intra-class coefficient (ICC). For the clinical evaluations, analysis of variance (ANOVA) with Tukey post-hoc analysis was performed for comparison between different patient groups. Area under curve (AUC) was calculated to compare the detection performance of different models. SPSS for Windows, version 22.0 (IBM-SPSS, Chicago, IL) was used for the statistical analysis with a p-value of <0.05 deemed statistically significant.

Results

Fig. 3 shows four example testing images along with their ‘ground-truth’ manual annotations (MA) and segmentation results obtained by LCNN, LDLA, and ACCM. LCNN and LDLA produced more faithful results to the manual annotations than ACCM, particularly in example 4 where ACCM failed to detect the nerves at the top middle and right bottom. Overall, the segmentation performance is consistent and there is no obvious failed case. For illustration, the supplementary Fig SM1 shows the results of all the first 30 images of Dataset 1, Fig SM2 shows the results of 12 randomly chosen images from Dataset 2, and Fig SM3 shows the results of 12 randomly chosen images from Dataset 3.

Analysis of dataset 1 and 2 shows the mean total CNF length from the manual ‘ground-truth’ annotation was highest ($2441.4 \pm 919.5 \mu\text{m}$) compared to the 3 automated approaches (LCNN:

2089.4 ± 804.6 μm, LDLA: 2260.3 ± 835.3 μm, ACCM: 2394.1 ± 768.1 μm). Total CNF length was greater in ACCM and closer to manual annotation than LCNN or LDLA. However, the ICC analysis (Table 1) demonstrated that both LCNN and LDLA produce more consistent results to the manual annotations when compared to ACCM. Furthermore, our two methods performed consistently better than ACCM in terms of correct segment length, number of branching points, and fractal numbers. Bland-Altman analysis (Fig. 4) further confirmed that the limits of agreement (LoA) of ACCM were greater than both LCNN and LDLA, implying greater variability despite ACCM's mean total corneal lengths. In other words, although the results of ACCM are closer to the manual annotation in this case, the variation due to over- and under-segmentation was much larger than both LCNN and LDLA; ACCM may therefore have produced heterogeneous results.

Based on the 95% confidence interval of the ICC estimate, values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively [36].

Table 2 shows the comparisons of the root mean squared error and standard deviations of the derived measures v_i over each image i , against the manual annotations, in terms of (i) number of branching points, (ii) number of terminal points, (iii) number of segments, (iv) total nerve fibre length, (v) mean nerve fibre length, (vi) standard deviation of nerve fibre length, and (vii) fractal number for each of the methods M using:

$$RMSE = \sqrt{\frac{1}{n} \sum_i V_{i,M}^2}, \quad SD = \sqrt{\frac{1}{n-1} \sum_i (V_{i,M} - \bar{V}_M)^2}, \quad \bar{V}_M = \frac{1}{n} \sum_i V_{i,M}, \quad V = v_{i,M} - v_{i,MA}$$

As shown in Table 2, LDLA has lower values for every measure, indicating closer agreement with the ground-truth annotation. For each measure, LCNN has the second-lowest RMSE and ACCM has the highest, indicating weaker agreement. LDLA has the lowest standard deviation for all except the number of terminal points, indicating more consistent agreement with the ground-truth over the set of images, while ACCM has the highest deviation for all measures. From this, it can be concluded that both LCNN and LDLA outperform ACCM, while LDLA clearly has the best performance.

Given the convincing performance of LDLA, which outperforms both LCNN and ACCM in each metric, it was applied to the third dataset and the results were used for clinical evaluation.

Clinical Testing and Validation Based on ENA Image Dataset

ANOVA analysis demonstrated that differences in the total CNF length between the six groups (of dataset 3) are in keeping with their neuropathy phenotype (Table 3 and Fig. 5). A Tukey post-hoc analysis was performed, which demonstrated that length in healthy-volunteer controls is higher than all the other groups ($p < 0.01$) while total CNF length in people with type 1 diabetes and neuropathy (Group 3) was lower than all other groups ($p < 0.001$). ACCM consistently yielded higher total CNF length than LDLA.

Furthermore, area under curve (AUC) analysis was undertaken to compare LDLA and ACCM (Fig. 5b and c). First, total corneal nerve length alone was used to classify control and neuropathy cases. There were 132 patients with neuropathy in total (from groups 2, 3, 5 and 6) and 90 controls (group 1). The resulting Receiver Operating characteristic (ROC) curve in Fig. 5 (b) shows that the AUC is 0.826 for LDLA and 0.801 for ACCM, respectively. To determine the sensitivity and specificity of the model, optimal cut points are determined by the commonly used Youden index [37], i.e. the sum of sensitivity and specificity minus one. In a perfect test, Youden's index is equal to 1. For LDLA, the optimal cut determined a specificity of 0.867 and sensitivity of 0.677 while ACCM achieved

specificity 0.800 and sensitivity 0.699. LDLA showed better prediction performance than ACCM when utilising CNF length. Similarly, Fig. 5(c) shows that LDLA has better prediction performance in classifying controls (n=90) and all patients with diabetes (n=301 from group 3, 4, 5 and 6) than ACCM when utilising CNF length: The AUC is 0.806 for LDLA and 0.780 for ACCM. The optimal cut points of LDLA are: specificity 0.7222 and sensitivity 0.784 while for ACCM: specificity 0.7222 and sensitivity 0.745.

Discussion

In this study, an artificial intelligence-based DLA has been developed for the analysis and quantification of corneal nerves in CCM images. To our knowledge, this is the first DLA for the analysis of corneal nerve morphology and pathology. This study validates our DLA and demonstrates its superior performance compared to ACCMetrics, the existing state-of-the-art system. In particular, there are more consistent results as evidenced by a superior intra-class correlation for a number of metrics including total CNF length. In addition to the total CNF length, this DLA is also capable of producing the number of branching and tail points, fractal numbers, tortuosity and segment length. As such, these quantitative parameters may provide additional utility to diagnose diabetic neuropathy and neuropathic severity.

A fractal is a visual product of non-linear system characterised by its complexity and by the quality of self-similarity or scale invariance. Fractal analysis of the corneal subbasal nerve plexus has been proposed by several authors [38, 39]. We believe that the additional utility of fractal dimensions provide an additional means of differentiating patients with early/ subclinical DPN. Corneal nerve fibre length is a robust measure of DPN and small fibre neuropathy. A large multicentre pooled concurrent diagnostic validity study revealed that CNFL was the optimal CCM variable [40]. CNFL has also shown to be a measure of early small fibre regeneration [41]. From published data, CNFL and CNFD are the most robust measures of DPN. Our data confirms the validity of CNFL. However, we feel other metrics are also of importance and require further scientific interrogation in a real world clinically-oriented study.

In this study, the quantification of images in healthy-volunteer controls and groups with and without diabetic neuropathy (group 1-6) demonstrates a reduction in total CNF length in patients with neuropathy compared to healthy controls. This study is in keeping with other data on the utility of CNF length as a valid biomarker of diabetic neuropathy [11, 12, 15, 18]. The sensitivity and specificity of our DLA for gold standard DPN diagnosis with CCM (using the Toronto Criteria) is far superior to currently used clinical methods such as the 10g monofilament and 128Hz tuning fork [42] (with rudimentary clinical assessments) thus providing a strong rationale for its use in clinical screening/practice.

This work extends our preliminary work on 584 CCM images where the initial DLA demonstrated good localisation performance for the detection of corneal nerves [43]. This study has further refined our preliminary model with the ensemble (LDLA) model, which is now validated in large image datasets to diagnose diabetic neuropathy using CCM.

The strength of deep learning is echoed by Oakley et al [44] who used corneal nerve segmentation in macaque images. Deep learning-based approaches make the segmentation task relatively easier for the end-user compared to conventional approaches with various filters and graph approaches [23]. In particular, compared to conventional machine learning methods such as support vector machines (SVM), deep learning reduces the need and additional complexity of feature selection and extraction, allowing the computer to learn features alongside the segmentation. The training of deep

learning approaches is computationally expensive; for example, it takes approximately 30 minutes per epoch to train a single U-Net model. The advantage is that, once the model is trained, the segmentation is very fast, taking milliseconds to segment CCM images.

In recent years, CNNs and DLAs have been added to the algorithms used to screen for diabetic retinopathy. DLAs promise to leverage the large amounts of images for physician interpreted screening and learn from raw pixels. The high variance and low bias of these models will allow DLAs to diagnose diabetic neuropathy through the utilisation of CCM images without the pre-processing requirements and more likely overfitting of earlier approaches such as [25]. This automated DLA for the detection of diabetic neuropathy offers a number of advantages including consistency of interpretation, high sensitivity and specificity, and near instantaneous reporting of results. In this study, good sensitivity and adequate specificities were achieved using our DLA.

This is the largest study to date for the development and validation of corneal nerve segmentation and supersedes the numbers in the study by Chen et al [24], who used 1088 images from 176 subjects with 200 images for training and 888 for testing. Our study utilised a robust dataset; however, further development of this DLA requires the prerequisite that a developmental set of images with large numbers (tens of thousands) of normal and abnormal cases is required. An area of further research is that of interrupted CNF segments which are challenging cases that often appear in the CNF segmentation results in previous methods [24]. This problem is mainly caused by non-uniform illumination and contrast variations of CNF in images. Since quantitative biomarkers like CNF length and density are important measures for computer-aided diagnosis, missing CNF segments may theoretically reduce the diagnostic reliability of any automated system. In our previous work, the automatic gap reconnection method proposed by Zhang et al. [45] was employed to bridge the interrupted nerve fibre structures. The gap-filling task is achieved by enforcing line propagation using the stochastic contour completion process with iterative group convolutions. Geometric connectivity of local CNF fibre structures can be easily recovered based on their contextual information [45]. However, this connection step was not included in this model as there was only a modest improvement in the quantification of CNF length despite extra computation time of about 1 minute per image. This is an area of future development of the DLA. It will also be important to investigate the potential for bias to be introduced by factors such as camera type. The major advantage of this DLA is the continual learning and refinement of the algorithm over standard automated techniques.

Given that four hundred and twenty million people worldwide have been diagnosed with diabetes mellitus [46] and the prevalence of diabetic neuropathy is ~50% [47] there is a need for valid quantitative population screening of diabetic neuropathy to prevent or limit sequelae such as foot ulcers and amputations. Skin biopsy with quantification of intra-epidermal nerve fibres has been considered the 'reference standard' test for the diagnosis of small fibre neuropathy [48]. It is an invasive test, however, which needs specialist diagnostic facilities and repeated tests at the same site are not feasible. CCM is a rapid non-invasive ophthalmic imaging modality, which quantifies early axonal damage in diabetic neuropathy with high sensitivity and specificity [12-16, 49]. CCM also predicts incident neuropathy [17] and accurately detects CNF regeneration [18, 50]. The utility of CCM in diagnosing and monitoring the progression of diabetic neuropathy has been extensively evaluated [11, 12, 15, 16, 18, 38].

Further studies are required to determine the feasibility of applying this algorithm in clinical settings and to determine outcomes compared with currently used diabetic neuropathy screening methods as they typically have low sensitivity for detection except in advanced neuropathy. There is also a need to compare the diagnostic ability of this DLA with tests of small fibre dysfunction (thermal

thresholds/sudomotor/autonomic) and IENFD in skin biopsy in diabetic and other peripheral neuropathies. The next key step is also to utilise the DLA alongside clinical neuropathy screening in a multi-centre primary care study.

Conclusion

Automated detection and screening offer a unique opportunity to detect early neuropathy and prevent the sequelae of advanced diabetic neuropathy. Our results demonstrate that this artificial intelligence-based DLA provides excellent localisation performance for the quantification of corneal nerve parameters and therefore has potential to be adopted for screening and assessment of diabetic neuropathy.

Funding

This research was partly funded by The National Natural Science Foundation of China (NSFC81570813). The authors would like to thank NVIDIA Inc for sponsoring the K40 GPU card used in this work. The study sponsor was not involved in the design of the study; the collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication.

Duality of Interest

The authors declare that there is no duality of interest associated with this manuscript.

Data Availability

The publically shared cornea nerve dataset (Dataset 1) is available at <http://bioimlab.dei.unipd.it/Corneal%20Nerve%20Tortuosity%20Data%20Set.htm> and <http://bioimlab.dei.unipd.it/Corneal%20Nerve%20Data%20Set.htm>. The remaining data is publicly available upon request with restrictions in accordance to ethical approvals.

Author Contributions

BW, YLZ, YTZ, JZ worked on the proposed model and conducted experimental testing. RL, BM, QH worked on the acquisition of Dataset 2. JL, MF, UA, IP, GP, RM worked on ACCMetrics and the collection and compilation of Dataset 3. DB, VR, SBK worked on the annotations of Dataset 3. All authors were involved in discussions regarding the work, the writing and revisions of the manuscript and approve of the submitted version.

YLZ and UA act as joint guarantors of this manuscript.

References

- [1] Tan H, Yang M, Wu Z, et al. (2004) Rapid screening method for *Schistosoma japonicum* infection using questionnaires in flood area of the People's Republic of China. *Acta Trop* 90(1): 1-9
- [2] Lauria G, Cornblath D, Johansson O, et al. (2005) EFNS guidelines on the use of skin biopsy in the diagnosis of peripheral neuropathy. *Eur J Neurol* 12(10): 747-758
- [3] Smith AG, Russell J, Feldman EL, et al. (2006) Lifestyle intervention for pre-diabetic neuropathy. *Diabetes Care* 29(6): 1294-1299
- [4] Quattrini C, Tavakoli M, Jeziorska M, et al. (2007) Surrogate markers of small fiber damage in human diabetic neuropathy. *Diabetes* 56(8): 2148-2154

- [5] Dehghani C, Pritchard N, Edwards K, et al. (2014) Morphometric stability of the corneal subbasal nerve plexus in healthy individuals: a 3-year longitudinal study using corneal confocal microscopy. *Invest Ophthalmol Vis Sci* 55(5): 3195-3199
- [6] Sivaskandarajah GA, Halpern EM, Lovblom LE, et al. (2013) Structure-function relationship between corneal nerves and conventional small-fiber tests in type 1 diabetes. *Diabetes Care* 36(9): 2748-2755
- [7] Tavakoli M, Quattrini C, Abbott C, et al. (2010) Corneal confocal microscopy: a novel noninvasive test to diagnose and stratify the severity of human diabetic neuropathy. *Diabetes Care* 33(8): 1792-1797
- [8] Kemp HI, Petropoulos IN, Rice AS, et al. (2017) Use of corneal confocal microscopy to evaluate small nerve fibers in patients with human immunodeficiency virus. *JAMA Ophthalmol* 135(7): 795-800
- [9] Ferdousi M, Azmi S, Petropoulos IN, et al. (2015) Corneal confocal microscopy detects small fibre neuropathy in patients with upper gastrointestinal cancer and nerve regeneration in chemotherapy induced peripheral neuropathy. *PLoS One* 10(10): e0139394
- [10] Tavakoli M, Marshall A, Thompson L, et al. (2009) Corneal confocal microscopy: a novel noninvasive means to diagnose neuropathy in patients with Fabry disease. *Muscle Nerve* 40(6): 976-984
- [11] Tavakoli M, Marshall A, Pitceathly R, et al. (2010) Corneal confocal microscopy: a novel means to detect nerve fibre damage in idiopathic small fibre neuropathy. *Exp Neurol* 223(1): 245-250
- [12] Petropoulos IN, Alam U, Fadavi H, et al. (2013) Corneal nerve loss detected with corneal confocal microscopy is symmetrical and related to the severity of diabetic polyneuropathy. *Diabetes Care* 36(11): 3646-3651
- [13] Petropoulos IN, Manzoor T, Morgan P, et al. (2013) Repeatability of in vivo corneal confocal microscopy to quantify corneal nerve morphology. *Cornea* 32(5): e83-e89
- [14] Petropoulos IN, Alam U, Fadavi H, et al. (2014) Rapid automated diagnosis of diabetic peripheral neuropathy with in vivo corneal confocal microscopy. *Invest Ophthalmol Vis Sci* 55(4): 2071-2078
- [15] Alam U, Jeziorska M, Petropoulos IN, et al. (2017) Diagnostic utility of corneal confocal microscopy and intra-epidermal nerve fibre density in diabetic neuropathy. *PLoS One* 12(7): e0180175
- [16] Chen X, Graham J, Dabbah MA, et al. (2015) Small nerve fiber quantification in the diagnosis of diabetic sensorimotor polyneuropathy: comparing corneal confocal microscopy with intraepidermal nerve fiber density. *Diabetes Care* 38(6): 1138-1144

- [17] Pritchard N, Edwards K, Russell AW, Perkins BA, Malik RA, Efron N (2015) Corneal confocal microscopy predicts 4-year incident peripheral neuropathy in type 1 diabetes. *Diabetes Care* 38(4): 671-675
- [18] Tavakoli M, Mitu-Pretorian M, Petropoulos IN, et al. (2013) Corneal confocal microscopy detects early nerve regeneration in diabetic neuropathy after simultaneous pancreas and kidney transplantation. *Diabetes* 62(1): 254-260
- [19] Ziegler D, Papanas N, Zhivov A, et al. (2014) Early detection of nerve fiber loss by corneal confocal microscopy and skin biopsy in recently diagnosed type 2 diabetes. *Diabetes* 63(7): 2454-2463
- [20] Brines M, Culver DA, Ferdousi M, et al. (2018) Corneal nerve fiber size adds utility to the diagnosis and assessment of therapeutic response in patients with small fiber neuropathy. *Sci Rep-UK* 8(1): 4734
- [21] Dabbah MA, Graham J, Petropoulos IN, Tavakoli M, Malik RA (2011) Automatic analysis of diabetic peripheral neuropathy using multi-scale quantitative morphology of nerve fibres in corneal confocal microscopy imaging. *Med Image Anal* 15(5): 738-747
- [22] Dabbah MA, Graham J, Petropoulos I, Tavakoli M, Malik RA (2010) Dual-model automatic detection of nerve-fibres in corneal confocal microscopy images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, pp 300-307
- [23] Kim J, Markoulli M (2018) Automatic analysis of corneal nerves imaged using in vivo confocal microscopy. *Clin Exp Optom* 101(2): 147-161
- [24] Chen X, Graham J, Dabbah MA, Petropoulos IN, Tavakoli M, Malik RA (2017) An automatic tool for quantification of nerve fibers in corneal confocal microscopy images. *IEEE T Bio-med Eng* 64(4): 786-794
- [25] LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553): 436
- [26] Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195(1): 215-243
- [27] Shen D, Wu G, Suk H-I (2017) Deep learning in medical image analysis. *Annu Rev Biomed Eng* 19: 221-248
- [28] Scarpa F, Zheng X, Ohashi Y, Ruggeri A (2011) Automatic evaluation of corneal nerve tortuosity in images from in vivo confocal microscopy. *Invest Ophth Vis Sci* 52(9): 6404-6408
- [29] Scarpa F, Grisan E, Ruggeri A (2008) Automatic recognition of corneal nerve structures in images from confocal microscopy. *Invest Ophth Vis Sci* 49(11): 4801-4807
- [30] Tesfaye S, Boulton AJ, Dyck PJ, et al. (2010) Diabetic neuropathies: update on definitions, diagnostic criteria, estimation of severity, and treatments. *Diabetes Care* 33(10): 2285-2293

- [31] Marfurt CF, Cox J, Deek S, Dvorscak L (2010) Anatomy of the human corneal innervation. *Experimental eye research* 90(4): 478-492
- [32] Borroni D, Beech M, Williams B, et al. (2018) Building a validated in vivo confocal microscopy (IVCM) dataset for the study of corneal nerves. *Invest Ophth Vis Sci* 59(9): 5719-5719
- [33] Matsumoto M, Nishimura T (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 8(1): 3-30
- [34] Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*
Springer, pp 234-241
- [35] Brunner M, Romano V, Steger B, et al. (2018) Imaging of corneal neovascularization: optical coherence tomography angiography and fluorescence angiography. *Invest Ophth Vis Sci* 59(3): 1263-1269
- [36] Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15(2): 155-163.
<https://doi.org/10.1016/j.jcm.2016.02.012>
- [37] Youden WJ (1950) Index for rating diagnostic tests. *Cancer* 3(1): 32-35
- [38] Chen X, Graham J, Petropoulos IN, et al. (2018) Corneal nerve fractal dimension: a novel corneal nerve metric for the diagnosis of diabetic sensorimotor polyneuropathy. *Invest Ophth Vis Sci* 59(2): 1113-1118
- [39] Bianciardi G, Latronico M, Traversi C (2015) Decreased geometric complexity of corneal nerve fibers distribution in Sjogren's syndrome patients. *Int J Ophthalmic Pathol Res* 4: 2
- [40] Perkins BA, Lovblom LE, Bril V, et al. (2018) Corneal confocal microscopy for identification of diabetic sensorimotor polyneuropathy: a pooled multinational consortium study. *Diabetologia* 61(8): 1856-1861
- [41] Azmi S, Jeziorska M, Ferdousi M, et al. (2019) Early nerve fibre regeneration in individuals with type 1 diabetes after simultaneous pancreas and kidney transplantation. *Diabetologia*: 1-10
- [42] Baraz S, Zarea K, Shahbazian HB, Latifi SM (2014) Comparison of the accuracy of monofilament testing at various points of feet in peripheral diabetic neuropathy screening. *Journal of Diabetes & Metabolic Disorders* 13(1): 19
- [43] Qi H, Borroni D, Liu R, et al. (2018) Automated detection of corneal nerves using deep learning. *Invest Ophth Vis Sci* 59(9): 5721-5721
- [44] Oakley JD, Russakoff DB, Weinberg R, et al. (2018) Automated Analysis of In Vivo Confocal Microscopy Corneal Images Using Deep Learning. *Investigative Ophthalmology & Visual Science* 59(9): 1799-1799

- [45] Zhang J, Bekkers E, Chen D, et al. (2018) Reconnection of Interrupted Curvilinear Structures via Cortically Inspired Completion for Ophthalmologic Images. *IEEE T Bio-med Eng* 65(5): 1151-1165
- [46] Cho N, Shaw J, Karuranga S, et al. (2018) IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pr* 138: 271-281
- [47] Iqbal Z, Azmi S, Yadav R, et al. (2018) Diabetic peripheral neuropathy: epidemiology, diagnosis, and pharmacotherapy. *Clin Ther* 40(6): 828-849
- [48] Lauria G, Hsieh ST, Johansson O, et al. (2010) European Federation of Neurological Societies/Peripheral Nerve Society Guideline on the use of skin biopsy in the diagnosis of small fiber neuropathy. Report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society. *Eur J Neurol* 17(7): 903-e949
- [49] Alam U, Riley DR, Jugdey RS, et al. (2017) Diabetic neuropathy and gait: a review. *Diabetes Ther* 8(6): 1253-1264
- [50] Culver DA, Dahan A, Bajorunas D, et al. (2017) Cibinetide improves corneal nerve fiber abundance in patients with sarcoidosis-associated small nerve fiber loss and neuropathic pain. *Invest Ophth Vis Sci* 58(6): BIO52-BIO60

Tables

Table 1. Absolute agreement measured by intraclass correlation coefficient (ICC)

Method	Total length	Mean length per segment	Number of branch points	Number of tail points	Number of nerve segments	Fractal
LCNN	0.867	0.596	0.809	0.647	0.844	0.887
LDLA	0.933	0.656	0.891	0.623	0.878	0.927
ACCM	0.825	0.352	0.570	0.257	0.504	0.758

Table 2: Root mean squared error (RMSE) and standard deviations of the error (SD) of each of the methods for different measures. Lower results indicate closer agreement with the manual annotation.

		LCNN	LDLA	ACCM
Number of Branching Points	RMSE	5.1326	4.1603	7.3667
	SD	4.3934	4.1652	6.9110
Number of Terminal Points	RMSE	8.7127	8.6766	13.0271
	SD	8.0985	8.3479	11.6407

Number of Segments	RMSE	8.9521	8.3752	15.0708
	SD	8.4248	8.3929	15.0296
Total Fibre Length	RMSE	463.4712	326.0016	501.6230
	SD	302.2312	271.7448	500.6295
Mean Fibre Length	RMSE	40.0348	37.6684	51.1406
	SD	37.9770	34.3762	50.6172
Standard Deviation of Fibre Length	RMSE	27.6372	24.0555	33.4168
	SD	26.8020	22.6133	32.4616
Fractal Number	RMSE	0.0403	0.0307	0.0518
	SD	0.0278	0.0273	0.0519

Table 3. Total corneal nerve fibre length (μm) for Dataset 3 utilising the LDLA (Deep Learning Algorithm) model.

Group	Participant numbers	Mean (μm)	Std. Deviation
1	90	2695.2	606.8
2	53	2245.2	648.6
3	37	1229.0	710.4
4	53	1917.1	732.2
5	108	2000.4	710.0
6	41	2131.7	803.7
Total	382	2125.9	800.6

Figures

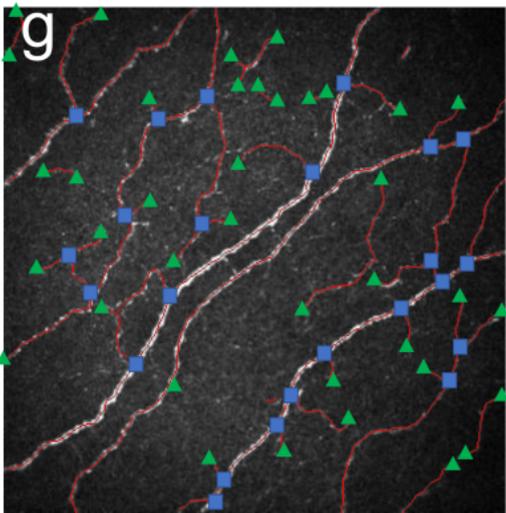
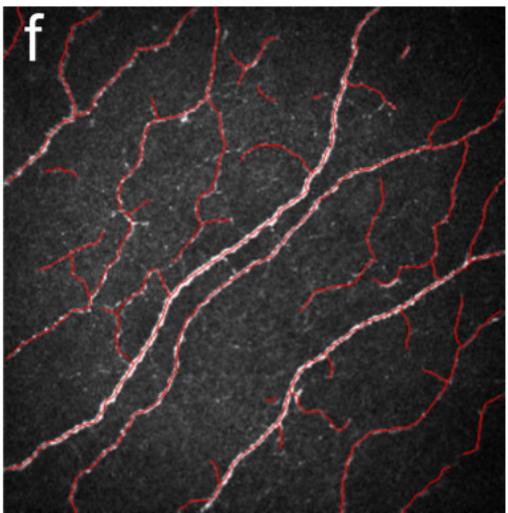
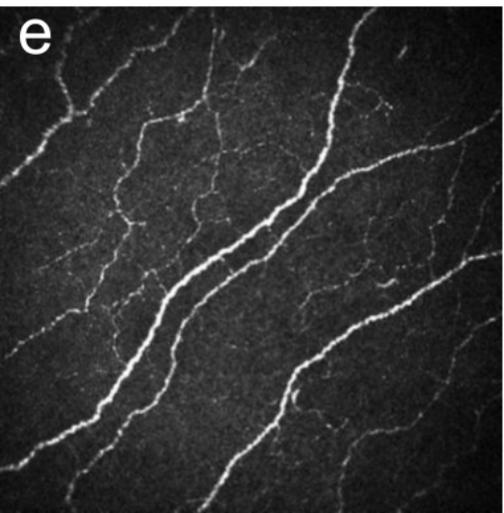
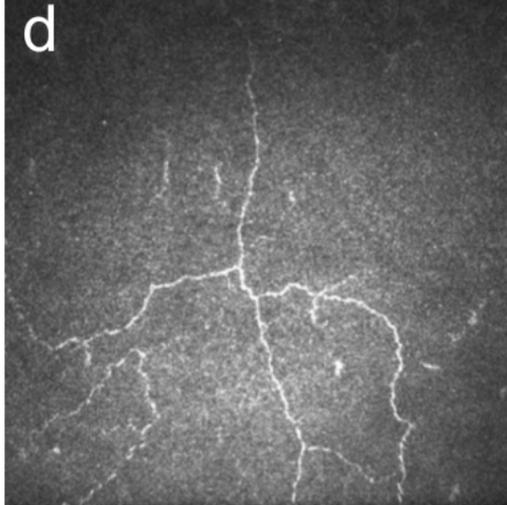
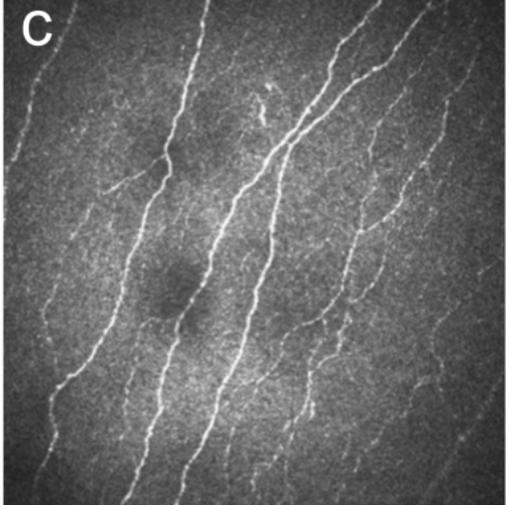
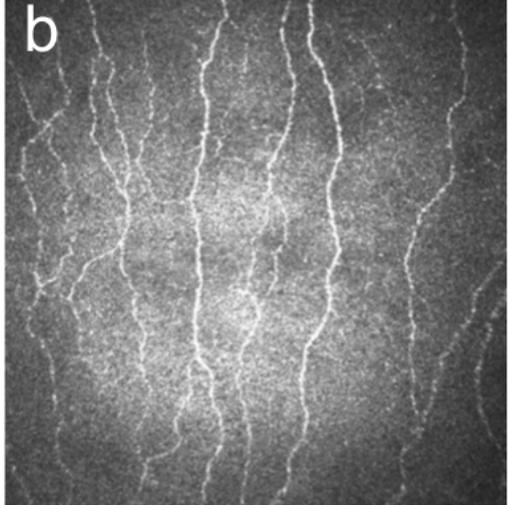
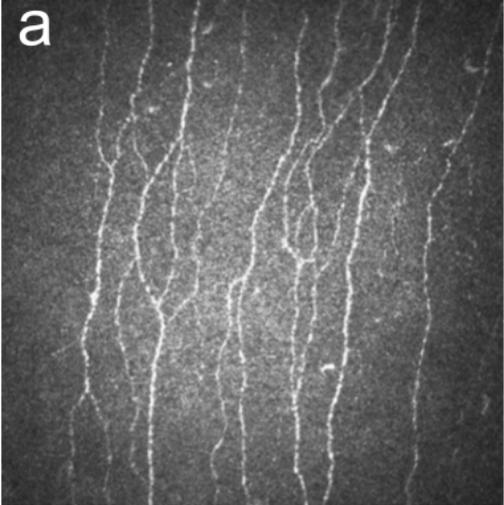
Figure 1. CCM image examples from control (a,b) and diabetic neuropathy (c,d) cases. An example image (e) with manual annotation (f). Branch and terminal points (manually added) are shown in (g), where green triangles denote tail points and blue squares denote branching points.

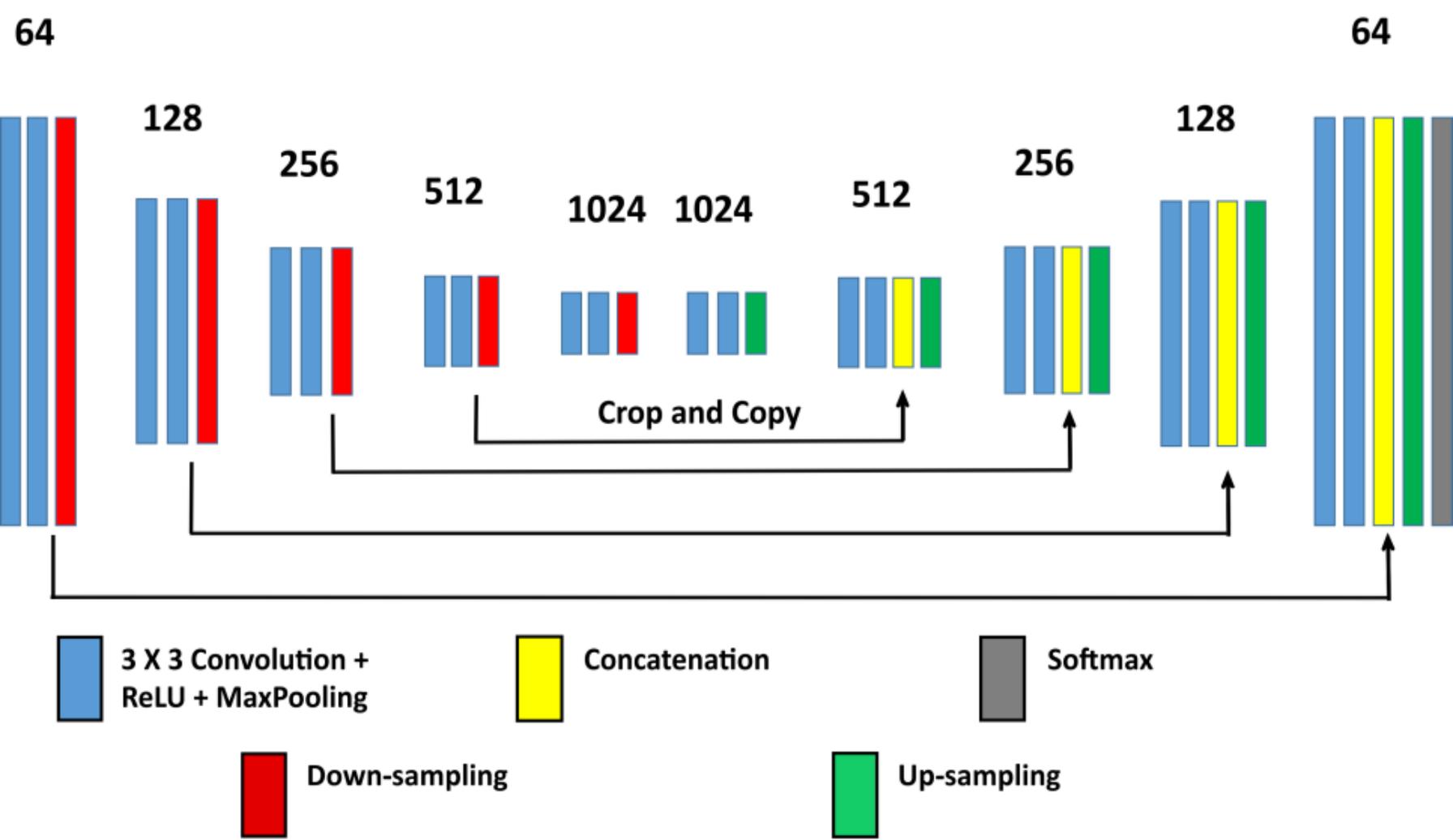
Figure 2. Diagram of the proposed U-Net architecture.

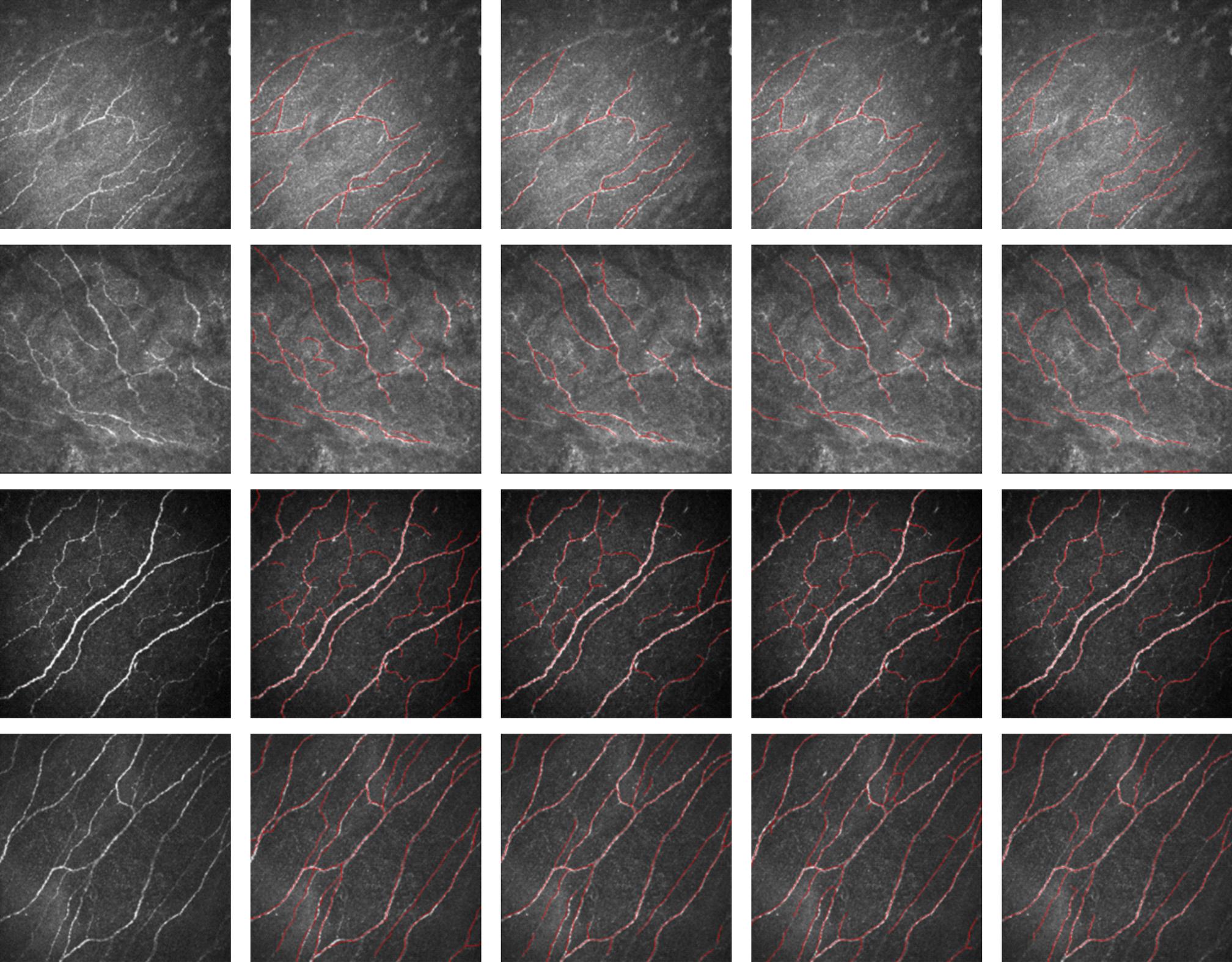
Figure 3. Four example of segmentation of corneal nerves. From the left to right columns are the original images, manual annotations, segmentation results of LCNN, LDLA and ACCM respectively where red lines denotes the centrelines of the segmented nerves.

Figure 4. Bland Altman plots on the total corneal fibre length (μm) for the LCNN, LDLA and ACCM methods respectively. The limits of agreement (LoA) are defined as the mean difference ± 1.96 standard deviation of differences. Error bars representing the 95% confidence interval for both the upper and lower limits of agreement.

Figure 5. Graphs showing the results on Dataset 3. (a) Box plot in combination of dot plot of the total corneal nerve fibre length of 6 groups with our LDLA model and ACCM. The line within each box represents the median, and the top and bottom of the box represent the 75th and 25th percentiles, respectively. The whiskers indicate the maximum and minimum values excluding outliers. (b) ROC curves of classification of controls and participants with diabetic neuropathy comparing LDLA and ACCM. (c) ROC curves of classification of controls and participants with diabetes comparing LDLA and ACCM.







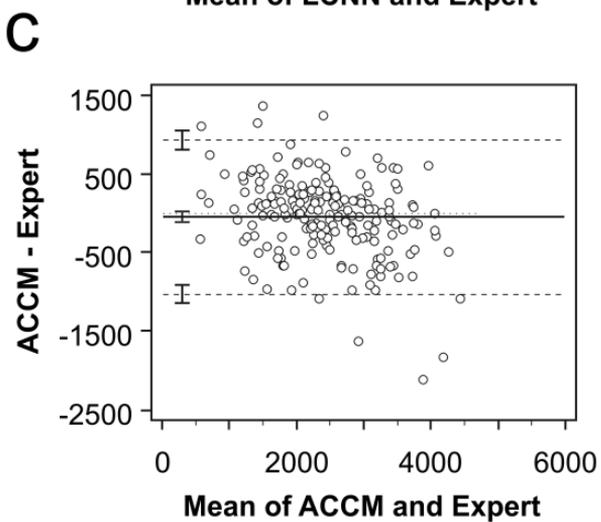
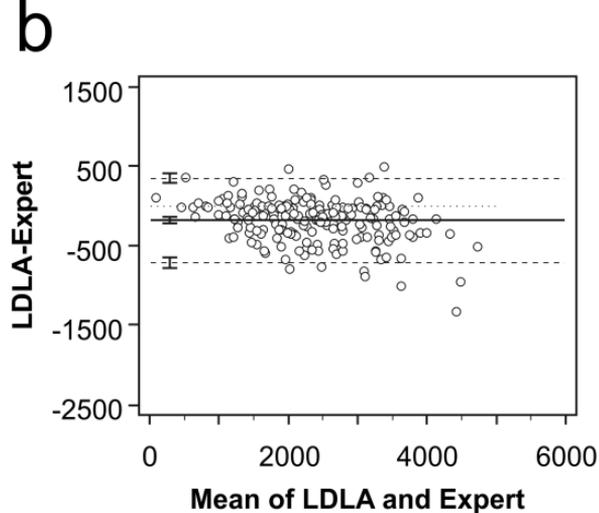
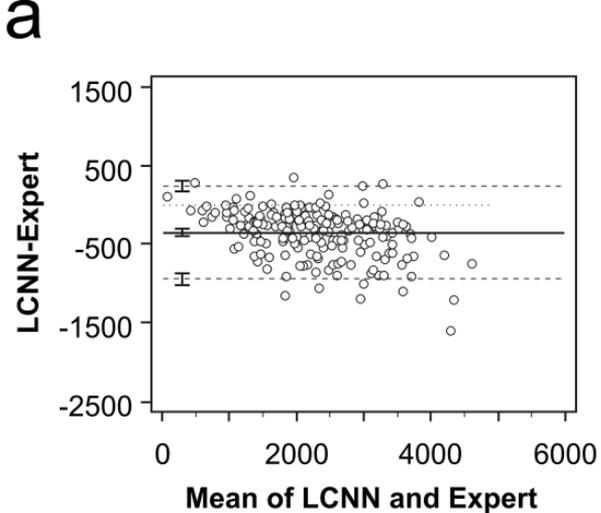
Original

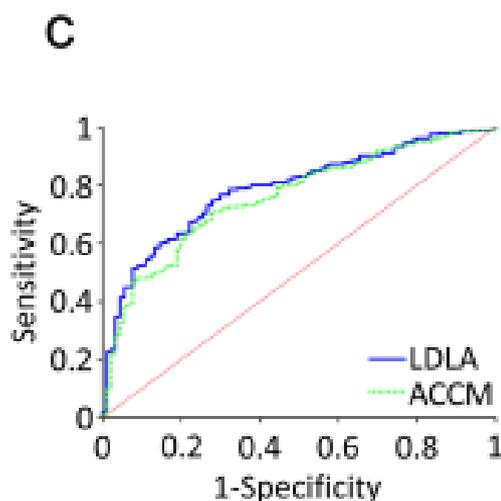
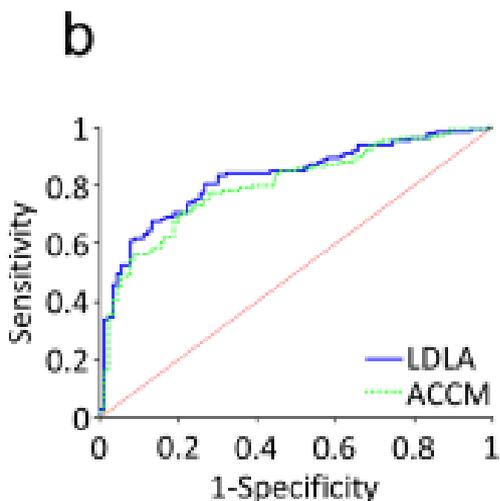
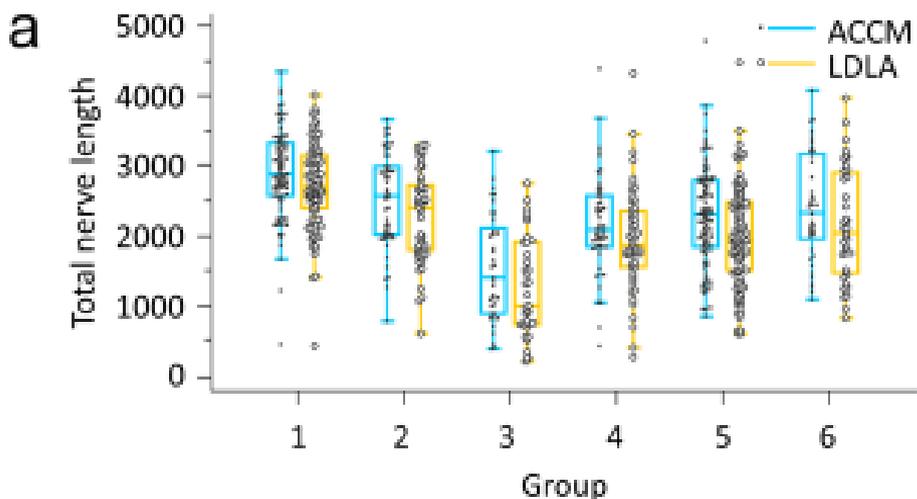
Manual

LCNN

LDLA

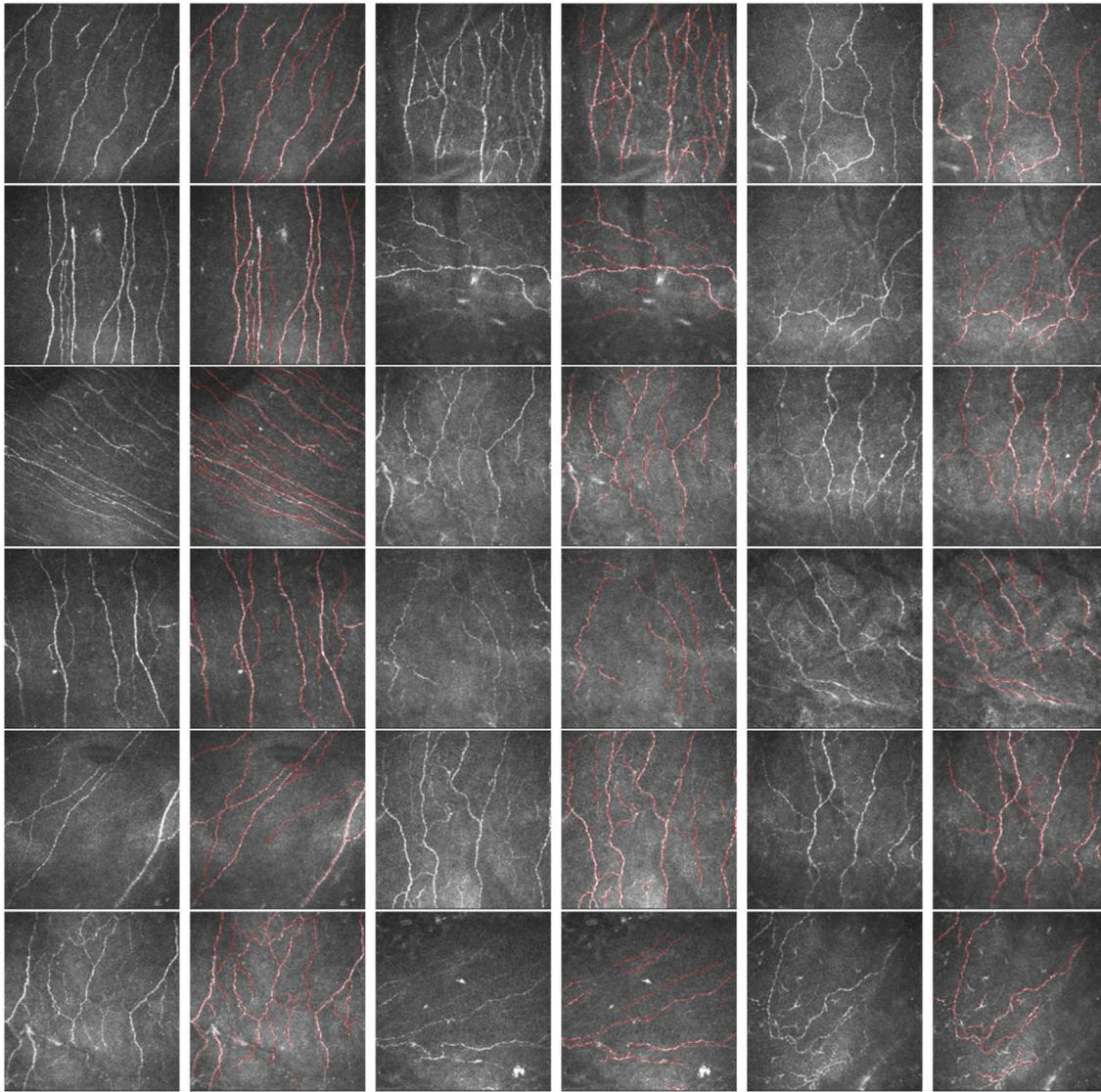
ACCM





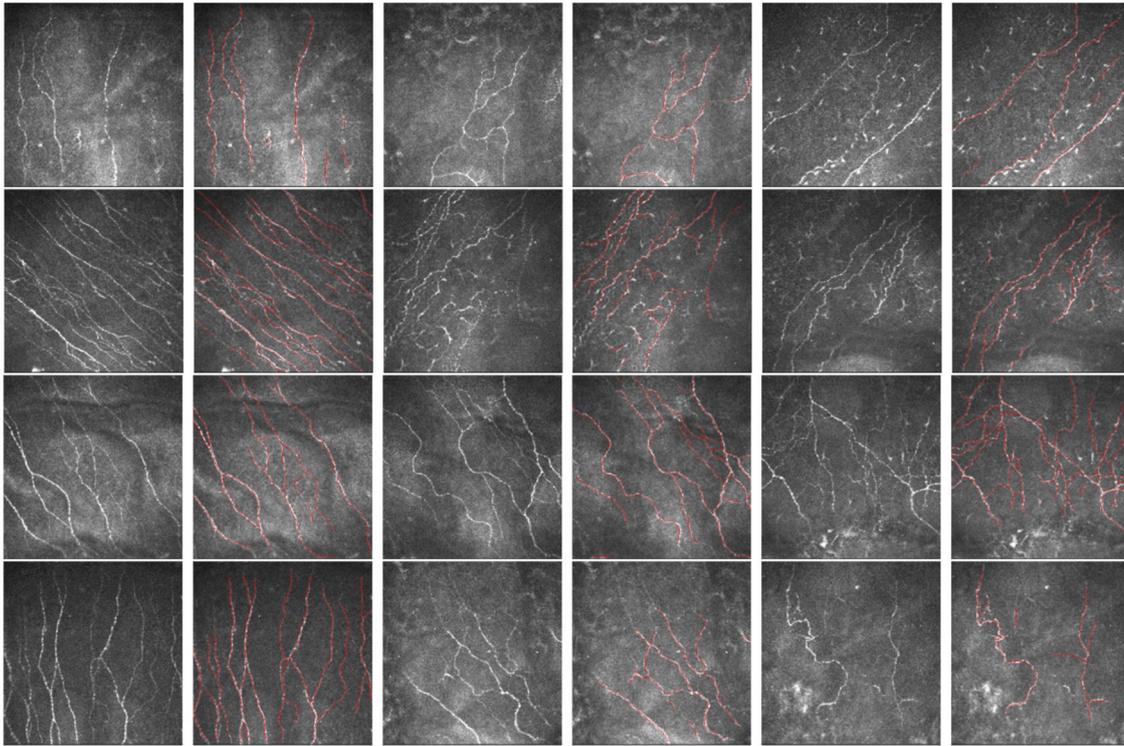
In this supplementary material, we show further examples of our LDLA on all 30 images forming the first part of Dataset 1 from BioImLab [34] in ESM Figs. 1a and 1b, on 10 randomly selected images from Dataset 2 from Beijing in ESM Fig. 2, and on 10 randomly selected images from Dataset 3 in ESM Fig. 3.

ESM Fig. 1a



ESM Fig. 1a: Original and segmentation results from our LDLA overlaid (red) on 18 of the 30 images forming the first part of Dataset 1 from BiolumLab [34].

ESM Fig. 1b



ESM Fig. 1b: Original and segmentation results from our LDLA overlaid (red) on 12 of the 30 images forming the first part of Dataset 1 from BiolmLab [34].

ESM Fig. 2

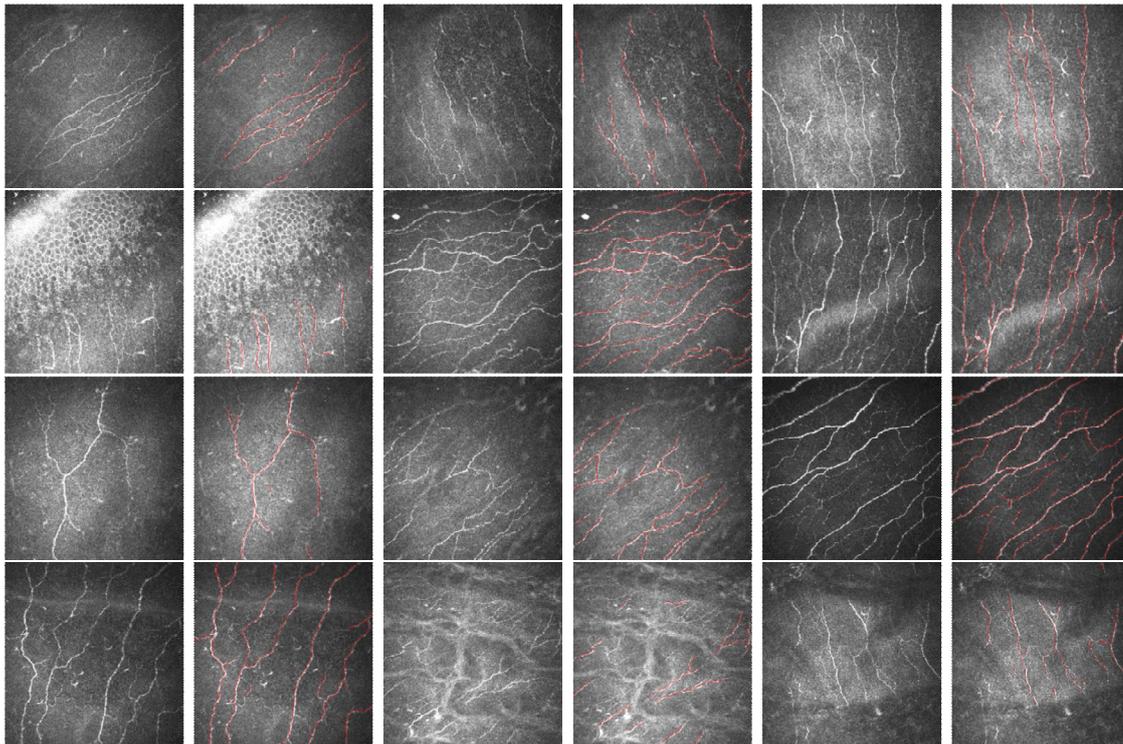


Figure SM2: Original and segmentation results from our LDLA overlaid (red) on 10 randomly selected images from Dataset 2 from Beijing.

ESM Fig. 3

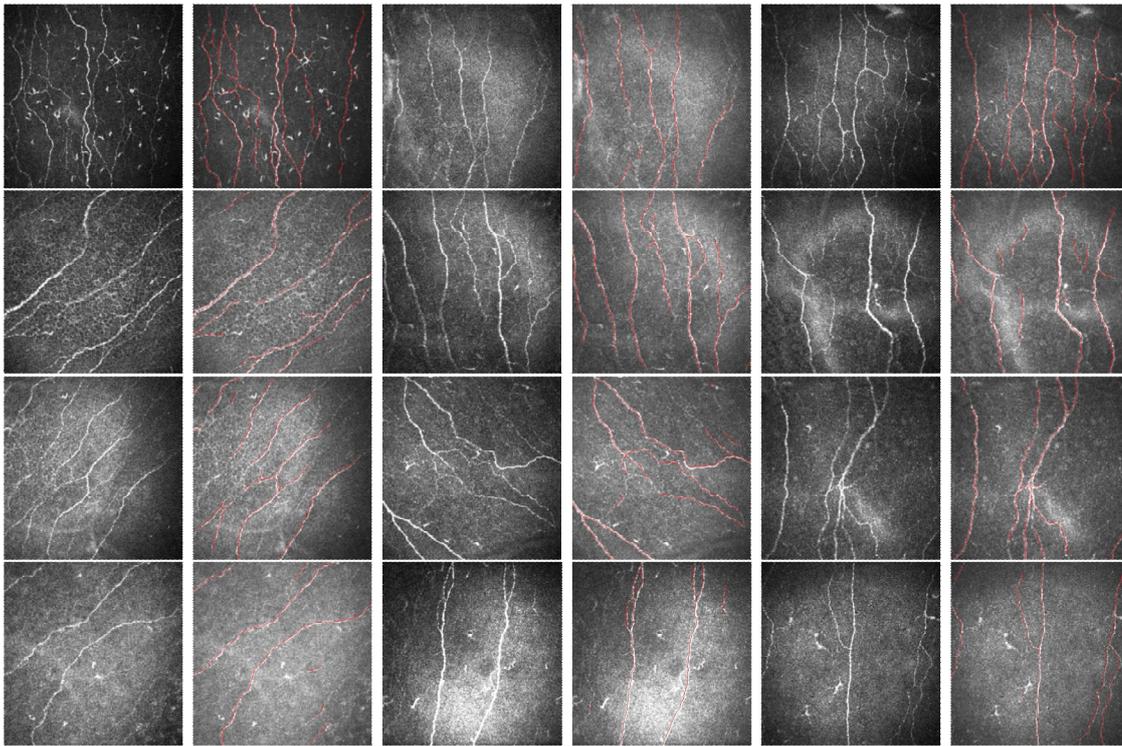


Figure SM3: Original and segmentation results from our LDLA overlaid (red) on 10 randomly selected images from Dataset 3.