Establishing a deep learning algorithm for the classification and segmentation of in vivo confocal microscopy images

Abstract

In vivo confocal microscopy (IVCM) has been gaining popularity as a supplementary diagnostic tool in a variety of ocular surface diseases by providing images of the cornea on a cellular level in a non-invasive manner.[1] The efficiency of IVCM imaging has been shown in numerous clinical studies, especially elucidating the changes in the sub-basal corneal nerves and detection of immune dendritiform cells (DCs).[2-4] Neuropathic corneal pain (NCP) is one of the most underdiagnosed ocular surface diseases due to lack of clinical signs explaining patients' symptoms, effecting corneal nerves. Our recent preliminary data showed that NCP patients without clinical signs of dry eye disease (DED) may present with corneal nerve alterations, referred as micro-neuromas, on in vivo confocal microscopy (IVCM) images suggesting can be used as an adjunct to making the differential diagnosis possible. We hypothesize that a form of artificial intelligence known as deep neural networks can be utilized in the automated analyses of corneal subbasal nerve alterations and identifying micro-neuromas as seen in NCP. Ocular surface inflammation controlled by DCs play a critical role in the pathogenesis and are significantly correlated with both symptoms and signs of dry eye disease (DED).[5] Moreover, visualization of DCs via IVCM has the potential to be utilized as a clinical tool to assess the effects of inflammation on corneal structures and function in inflammatory diseases such as dry eye disease (DED). However the detection and analyses of DCs in IVCM images is highly subjective although performed by trained professionals and very time consuming even when analyzed by semi-automated softwares currently available. Therefore, we hypothesize that a form of artificial intelligence known as deep neural networks can be utilized in the automated analyses of corneal subbasal nerve alterations and identifying micro-neuromas as seen in NCP as well as detection and quantification of cornea DCs would provide a rapid, standardized and objective evaluation of ocular surface inflammation enabling improved diagnostic and treatment accuracy. Our aim with this study was

to develop a deep learning algorithm able to classify IVCM images according to their respective layers and further segment and analyze DCs in the entire sequence and identify microneuromas. We also show the utilization of this algorithm in clinical practices by developing a graphical user interface (GUI) implementing the segmentation and analyses of DCs, used as possible supporting diagnostic tool.

Introduction

IVCM enables the non-invasive examination of corneal layers on a cellular level, comparable to those as ex-vivo histochemical tests.[1] Since its introduction to ophthalmology, IVCM has been utilized to image a variety of corneal pathologies. It has specifically become a reliable supporting diagnostic tool, in the quantitative analyses of corneal inflammation thanks to its ability in imaging corneal DCs.[4] DCs are corneal antigen presenting cells (APCs) able to activate T-cells, and have been shown to manage inflammation in corneal pathologies such as microbial keratitis, contact lens wear and dry eye syndrome.[6-9] Dry eye disease is one of the most frequently encountered ophthalmic disorders in clinical practices and inflammation of the cornea has been associated with signs and symptoms of the disease.[5] DCs specifically play an essential role in the inflammatory cascade contributing to the pathogenesis of DED. Changes in DC parameters have previously been evaluated in a number of studies conducted on DED patients via IVCM.[4,7,10-12] In light of these studies, imaging DCs may be utilized as a non-invasive and responsive surrogate biomarker in corneal inflammation in DED.

Various semi-automated softwares have been developed and utilized in the analyses of DCs in IVCM images. However, from the selection of representative images to DC tracing and segmentation, the evaluation process is observer dependent therefore highly subjective and also requires a meticulous attention making the analyses very time consuming and non-reproducible. Therefore, quantitative DC analyses is currently not a practical parameter in clinical setting but rather only suitable for studies and trials.

Machine learning has been described as "the automated detection of meaningful patterns in data" and these programs are endowed with the ability of learning and improving from experience without being explicitly programmed.[13,14] Deep learning, also referred as deep artificial neural networks, is a subset of machine learning, able to yield more rapid and precise results.[14] Hereby, we introduce a deep learning algorithm in the classification of IVCM images into specific corneal layers and further able to segment and quantify DCs in the entire sequence without the need of representative images.

Material and Methods

This is a retrospective imaging study conducted on the images obtained from the patients seen at the Cornea Service of the New England Eye Center, Tufts Medical Center, Boston, MA that was approved by the Institutional Board Review of Tufts Medical Center/Tufts University Health Sciences. The protocol conformed to the Declaration of Helsinki and adhered to the Health Insurance Portability and Accountability Act (HIPAA).

IVCM imaging and data set

IVCM images of healthy and DED patients selected from our database were utilized for this study. Images were obtained via Heidelberg Retinal Tomograph (HRT-II) with the Rostock Cornea Module (Heidelberg Engineering GmbH, Heidelberg, Germany). All images were recorded in a 400x400µm (384x384 pixel) size in gray scale. 2140 images obtained from healthy epithelium, subbasal nerve plexus, stroma, and endothelium were randomly selected by a trained ophthalmologist for the training and the validation sets of the classification algorithm. Moreover, 258 images demonstrating DCs from DED patients were analyzed by 2 trained professionals using Image J software (http://imagej.nih.gov.ezproxy.library.tufts.edu/ij/; provided in the public domain of the National Institutes of Health, Bethesda, MD, USA) for density and area. Density was identified as; number of DCs per image multiplied by a constant and area was calculated as total size of DC surface per image.

Network specifications

The algorithm was designed on a U-net based convolutional neural network. CNNs are types of artificial neural networks primarily used in the field of pattern recognition within images, which the input is loaded into the algorithm, that is then distributed to the hidden layers. These hidden layers make decisions based on the results of previous layers adjusting the results of the output according to the data it retrieves through the layers.[15,16]

The architecture of the CNN used in this study consists of 19 convolutional layers, 1 Fully Connected (FC) layer, 4 max-pooling and 4 up-sampling layers with 492,118 parameters (of which 240 are non-trainable). An additional classification branch was attached to the bottom part of the vanilla U-net segmentation structure. To reduce the number of parameters (/10) the final FC layer was replaced by a Global Average Pooling (GAP) layer.

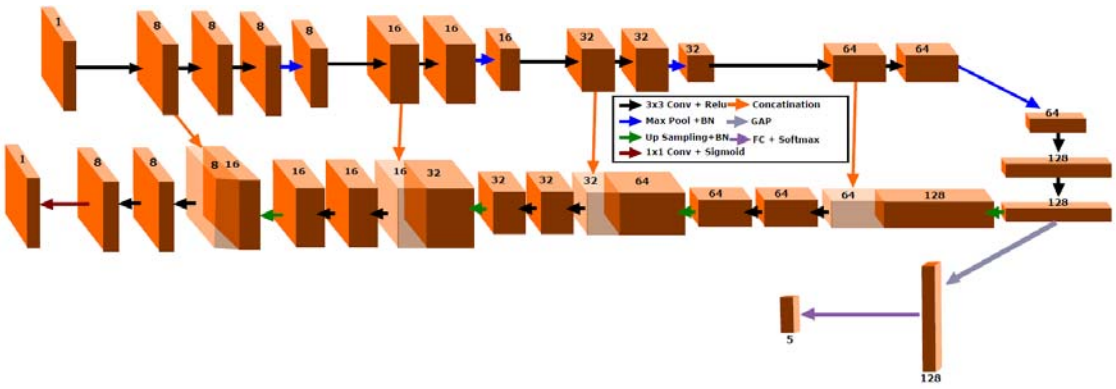

Figure 1: An illustration of the U-net CNN architecture.

DC segmentation

A total of 2335 images with the corresponding annotations of layers and DCs masked were used (training set: n=2073 images, validation: n=262). These images emerged from 5 corneal layers including the Endothelium (n=378), Epithelium (n=639), Subbasal nerve plexus (n=249), Stroma (n=671) and a non-corneal (gel) layer (n=136) (Figure 2, Table 1). The data distribution was balanced using different loss

function weights for each (table 1). Neural networks are known to achieve more precise results with larger sets of data. Even when the quality of the data is low, algorithms can actually perform better, with increased numbers of data from the original data set. Data augmentation is a strategy that enables significant increase in the diversity of data available for training models, without actually collecting new data. [17] To enrich the data, the images were augmented using 10 different approaches, selected randomly per image (Figure 3).

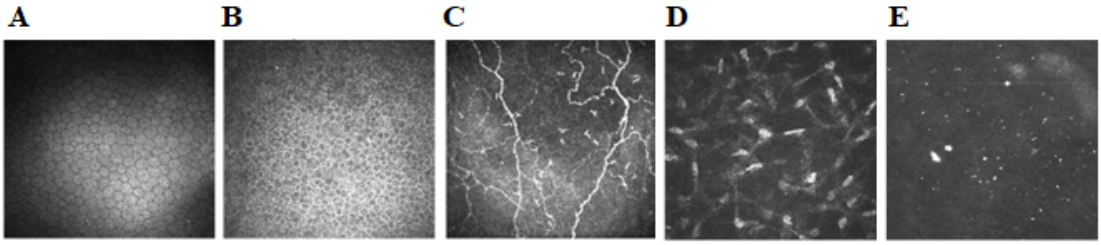

Figure 2: The different layers (classes) in the dataset. (A) Endothelium. **(B)** Epithelium. **(C)** Nerves. **(D)** Stroma. **(E)** Gel.

Layer	Number of slices	Weight
Endothelium	378	0.21
Epithelium	639	0.125
Sub-basal nerve plexus	249	0.25
Stroma	671	0.125
Gel	136	0.29

Table 1: Dataset distribution for the classification training and validation sets

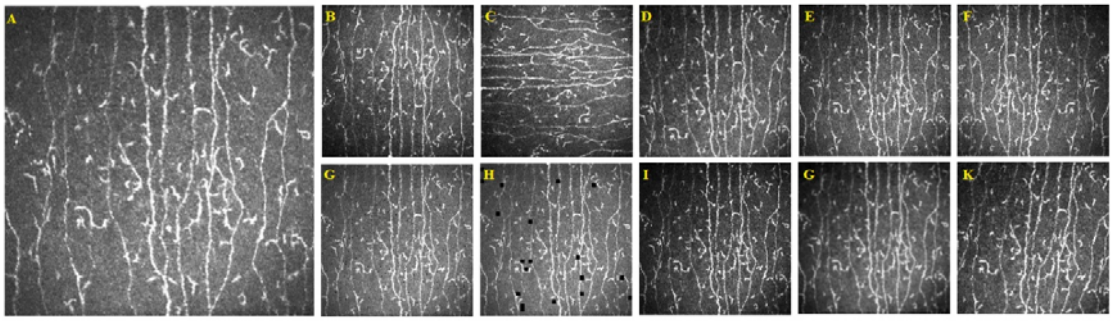

Figure 3: Data augmentation. (A) Raw image. **(B)** Flip up-down. **(C)** Rotate 90 degrees. **(D)** Scale. **(E)** Multiply. **(F)** Flip left-right. **(G)** Dropout. **(H)** Course dropout. **(I)** Contrast gamma. **(J)** Gaussian blur. **(K)** Affine transform.

Training and validation

Following 100 epochs of training, the network's performance was validated using a 5-fold cross validation procedure. The validation set consisted of 262 slices from 92 patients and their corresponding annotations (table 2). Each slice was independently annotated by two readers, when their intersection was used as a Ground Truth (GT).

Number of cases	Number of slices	Total amount of DCs in GTs	Total number of pixels marked as DCs in GTs
92	262	4457	281,978

Table 2: validation data information.

Statistical measures

The performance of the network was evaluated using various statistical measures including specificity, sensitivity, accuracy, precision, dice (F1), according to the following formulations.

$$Specificity = \frac{TN}{TN + FP} \quad Sensitivity = \frac{TP}{TP + FN} \quad Accuracy = \frac{TP + TN}{Total\ population}$$

$$Precision = \frac{TP}{TP+FP} \qquad Dice = 2 * \frac{\sum GT * Prediction}{\sum GT + \sum Prediction}$$

Morphological DC analysis

The area (total size of DCs surface per image) and density (number of DCs per image multiplied by a constant) were calculated for 33 cases and compared to the average score of the two readers. The error of each measure was calculated by subtraction of the CNN score from the average reader score divided by the readers' score. Accuracy was calculated by 1- error.

Results

Classification of corneal layers

For the classification of images, the CNN was trained on a total of 1540 images (epithelium: n=441, subbasal nerve plexus: n=510, stroma: n=411, endothelium: n=178). Validation sets consisted of 610 images (epithelium: n=99, subbasal nerve plexus: n=281, stroma: n=130, endothelium: n=100). The sensitivity, specificity and Area Under the Curve (AUC) measures were evaluated for the algorithm classification per each layer, yielding scores higher than 0.95 for all cases (table 3).

Corneal Layer	Epithelium	Subbasal Nerve plexus	Stroma	Endothelium
Sensitivity	0.97	0.95	1.0	0.98

Specificity	0.98	0.99	0.99	0.99
AUC	0.95	0.98	0.99	0.98

Table 3: Sensitivity, specificity and AUC of CNN corneal layer classification

Segmentation of DCs

The pixel wise segmentation of the DCs was calculated based on 262 sub-basal nerve plexus images (Figure 4). The CNN segmentation results yielded a false positive rate of 0.257 and a false negative rate of 0.243 comparing to the ground truth, resulting in an accuracy of 0.9957, specificity of 0.9986, sensitivity of 0.8499, dice score of 0.8782 and precision score of 0.9189 (table 4). A Receiver Operating Characteristic (ROC) curve was calculated yielding an AUC of 0.88.

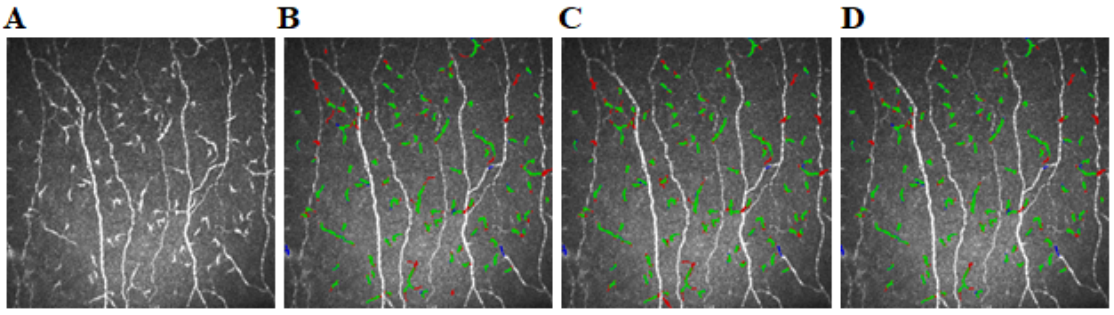

Figure 4: DCs segmentation. (A) A slice and its corresponding DC segmentation marked upon (**B)** reader 1 and (**C)** reader 2 annotations and (**D)** the intersection between them (green- true positive, blue- false positive, red- false negative).

Measure	CNN score accuracy				
Accuracy	**Specificity**	**Sensitivity**	**Dice**	**Precision**	**AUC**
Density		0.83		0.9189	0.88
Area		0.76			

Table 4: performance scores

Morphological DC analysis

The density and area of the DCs were calculated for the readers and CNN scores. A comparison between the scores indicated an accuracy of 0.83 for the density and 0.76 for the area calculations (table 5).

Table 5: DCs Density and Area scores

Graphical User Interface (GUI)

In the next stage, we implemented the above algorithms in a custom-written GUI (Figure 5). The GUI was designed to enable an easy access to the raw images and their corresponding analysis. This approach allows an implementation of a quick (10-30 seconds per case ranging from 200 to 800 images) automatic evaluation, alongside a manual examination by a human expert. This provides a standardized, accurate

and rapid analysis of DCs, improving the processes of diagnosis and treatment. The rapid segmentation and analysis of the DCs by the algorithm may allow an expansion of the calculations based on multiple images, providing further information and yielding even more accurate and less biased results.

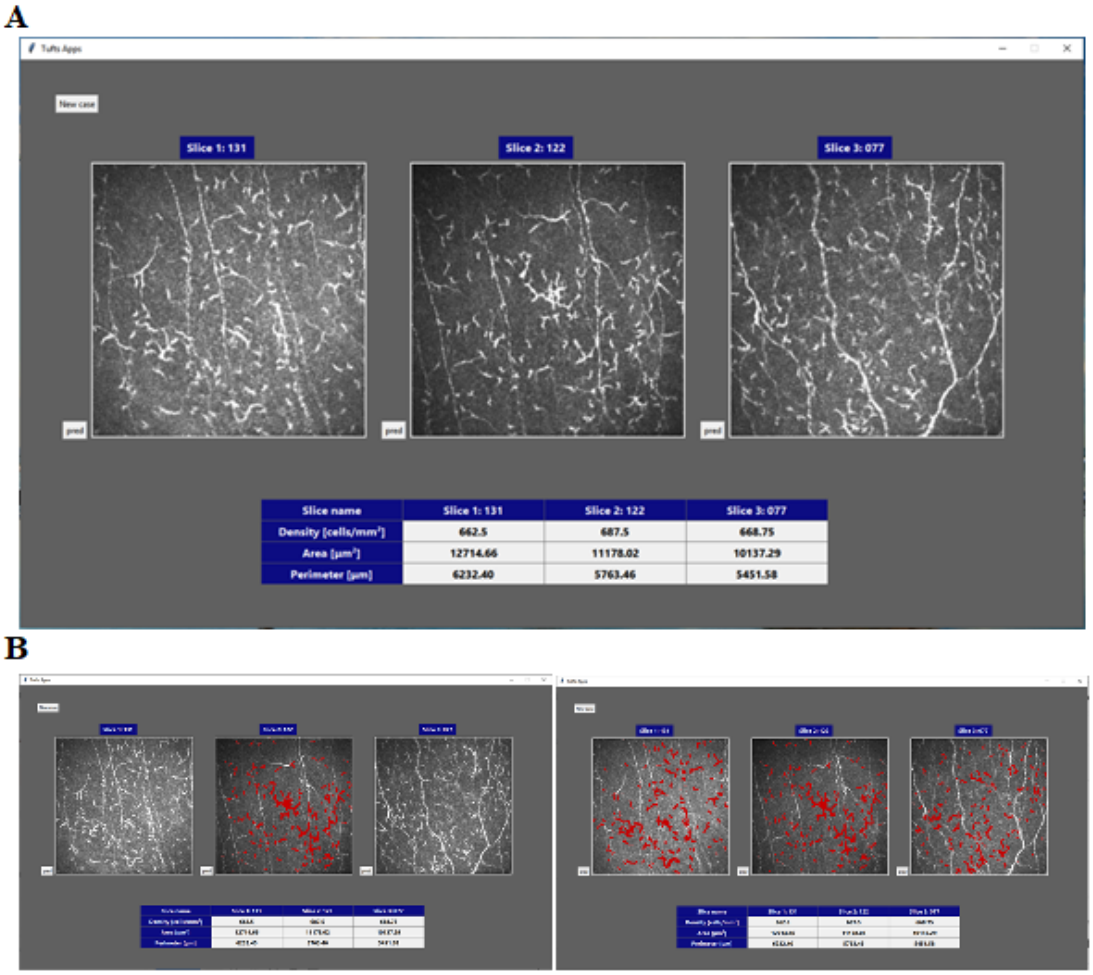

Figure 5: A print screen of the DCs analysis GUI. (A) The screen showing the raw images (top panel) and the automatic analysis per image (bottom panel). **(B)** Examples of an easy segmentation display overlaid on the raw images (red).

Micro-neuroma identification and detection

IVCM sequences consisting of 44,630 images from a mix of DED and NCP patients were utilized for this retrospective study. Images containing micro-neuromas were separated by a trained ophthalmologist and while 2,287 images were found to contain a micro-neuroma, 42,343 did not. A deep neural network with over 24 million parameters (ResNet-50, pre-trained on Image net) was trained to predict to presence or absence of micro-neuromas in each image. The model was trained using 8 Titan X graphics processing units over a 24-hour period. The network was trained on the data from 80% of the images chosen randomly from the entire set and validated on the remaining 20% of the images. This procedure was repeated 5 times as a standard cross-validation strategy to assess the performance of the algorithm.

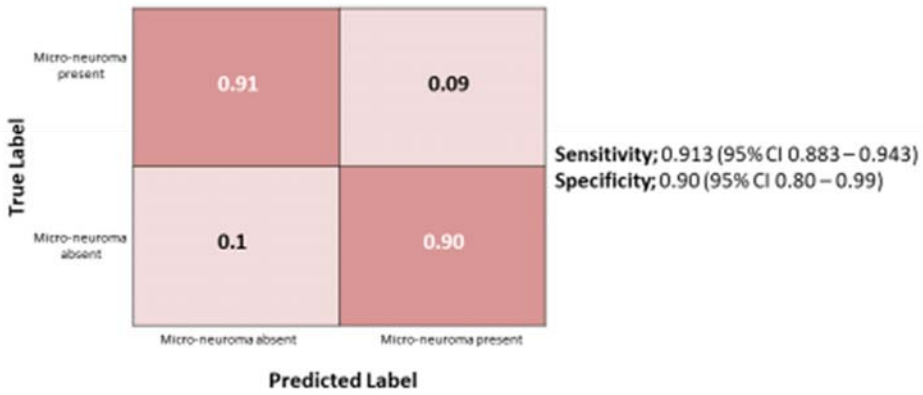

Figure 6: Sensitivity and specificity of micro-neuroma detection with the proposed CNN

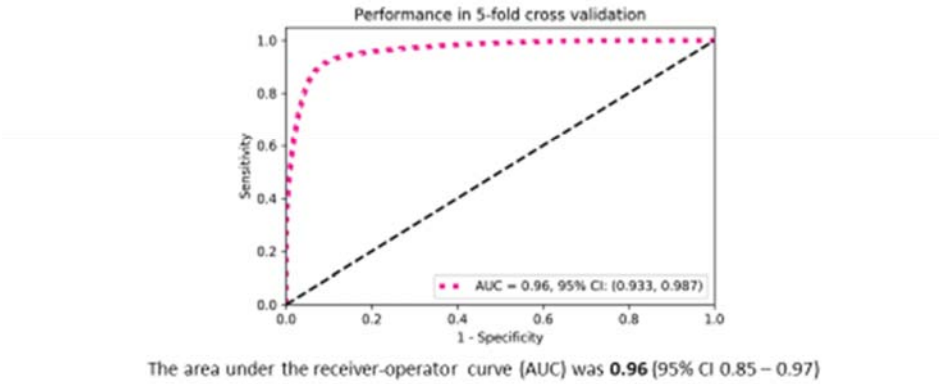

Figure 7: The accuracy of the proposed CNN algorithm with an AUC of 0.96

Discussion

IVCM is a practical, non-invasive, in-vivo imaging tool providing information about the ocular surface at a cellular level. Thanks to its capability of acquiring high resolution images of DCs, IVCM allows rapid diagnosis, follow-up, and management of inflammation and therefore may be considered a promising supplementary diagnostic tool for inflammatory ocular surface diseases such as DED.[18]

DED is one of the most prevalent diseases encountered in ophthalmic practices. Ocular surface inflammation has been shown to play a crucial role in the pathogenesis as well contributing to the signs and symptoms of the disease.[5] Management of inflammation has been proposed for the treatment of the disease and the visualization of inflammatory DCs in the cornea via IVCM has made it possible to make a more objective diagnosis and follow-up of these DED patients.[4,5,7] However, several challenges still remain in the analyses of these images as current softwares depend on manual or semi-automated approaches to

provide quantitative results on DCs. Other major drawbacks of these softwares are that the results are highly subjective and non-reproducible even performed by trained ophthalmologists and also extremely time consuming making it impractical for clinical use.

NCP is a recently acknowledged ocular pathology that still ill-defined and requires intensive research for identifying the pathophysiology. Neuropathic pain can simply be defined as clinical manifestation of chronic ectopic activity of damaged corneal nerves in response to both *innocuous* and *noxious* stimuli. NCP on the other hand has been interpreted as neuropathic pain without significant clinical slit lamp findings and failure of symptom resolution with conventional treatments. NCP diagnosis is typically based on a combination of clinical history, symptoms, thorough ophthalmologic examination and/or evidence of nerve injury or disease as visualized by IVCM. Our most recent preliminary data showed that micro-neuromas, benign growth or swelling of a nerve ending, indicating ineffective and unregulated nerve regeneration, often resulting in pain, were present in all NCP patients compared to DED and healthy subjects allowing differential diagnosis.

The term 'artificial intelligence' can be used for any device that can learn to take actions in order to solve a certain problem and can be broadly used to refer to any sort of machine learning program.[14,19] Deep learning is a form of machine learning that uses multilayered artificial neural networks inspired by biological nervous systems. Similar to traditional ANNs, CNNS are also comprised of neural layers that can self-optimize through learning and are able to automatically extract features from an image.[15] In recent years, CNNs have been getting more recognition in medical image analyses and have started to be implemented into ophthalmology as well. While the majority of advances are focusing on retinal image analyses (google), deep learning for corneal imaging modalities is also on the rise.[20-22]

In light of these recent advances, we hypothesized that a deep neural network would provide rapid and objective evaluation of DCs as well as detection of micro-neuromas in IVCM images therefore increase

diagnostic accuracy, ease follow-up and management of inflammatory ocular surface diseases. In this study we propose a U-net based CNN for the classification and segmentation of IVCM images. Our proposed method managed to achieve very high accuracy, specificity and sensitivity in the classification of IVCM images to their respective anatomical layers, easing further analyses regarding each different layer (>95% for all results). Moreover, the algorithm achieved an accuracy of 0.9957, specificity of 0.9986, and sensitivity of 0.8499 for the segmentation of DCs in IVCM images and eliminated the need of choosing 3 representative images by quantifying the DCs in the entire sequence in less than 30 seconds (200-800 images per sequence). The evaluation of DC morphological parameters also yielded high accuracy and interclass correlations with semi-automated analyses making it possible to evaluate the changes of DC morphology. The AI system also had a very high AUC of 0.96 for detecting micro-neuromas, and the inclusion of additional patients will likely increase this performance further therefore we suggest that the deep neural network shows great promise in identifying micro-neuromas associated with NCP suggesting that artificial intelligence can rapidly evaluate IVCM images, while maintaining a high degree of accuracy.

In conclusion, with our proposed algorithm IVCM image analyses would become more rapid by cutting down the time to assess images to under 30 seconds and accurate by providing objective results independent of an observer. Therefore, quantitative DC analyses would be more practical not only for studies and trials but also clinical use and easing management and follow-up of patients with inflammatory ocular surface diseases such as DED.

References

1. Cruzat A, Qazi Y, Hamrah P. In Vivo Confocal Microscopy of Corneal Nerves in Health and Disease. Ocul Surf 2017;15:15-47.
2. Moein HR, Kheirkhah A, Muller RT, Cruzat AC, Pavan-Langston D, Hamrah P. Corneal nerve regeneration after herpes simplex keratitis: A longitudinal in vivo confocal microscopy study. Ocul Surf 2018;16:218-25.
3. Müller RT, Abedi F, Cruzat A, et al. Degeneration and Regeneration of Subbasal Corneal Nerves after Infectious Keratitis: A Longitudinal In Vivo Confocal Microscopy Study. Ophthalmology 2015;122:2200-9.
4. Kheirkhah A, Rahimi Darabad R, Cruzat A, et al. Corneal Epithelial Immune Dendritic Cell Alterations in Subtypes of Dry Eye Disease: A Pilot In Vivo Confocal Microscopic Study. Investigative ophthalmology & visual science 2015;56:7179-85.
5. Baudouin C, Irkec M, Messmer EM, et al. Clinical impact of inflammation in dry eye disease: proceedings of the ODISSEY group meeting. Acta Ophthalmol 2018;96:111-9.
6. Cruzat A, Pavan-Langston D, Hamrah P. In vivo confocal microscopy of corneal nerves: analysis and clinical correlation. Seminars in ophthalmology 2010;25:171-7.
7. Lin H, Li W, Dong N, et al. Changes in corneal epithelial layer inflammatory cells in aqueous tear-deficient dry eye. Investigative ophthalmology & visual science 2010;51:122-8.
8. Cruzat A, Witkin D, Baniasadi N, et al. Inflammation and the nervous system: the connection in the cornea in patients with infectious keratitis. Investigative ophthalmology & visual science 2011;52:5136-43.
9. Cruzat A, Schrems WA, Schrems-Hoesl LM, et al. Contralateral Clinically Unaffected Eyes of Patients With Unilateral Infectious Keratitis Demonstrate a Sympathetic Immune Response. Investigative ophthalmology & visual science 2015;56:6612-20.
10. Marsovszky L, Resch MD, Nemeth J, et al. In vivo confocal microscopic evaluation of corneal Langerhans cell density, and distribution and evaluation of dry eye in rheumatoid arthritis. Innate immunity 2013;19:348-54.
11. Marsovszky L, Nemeth J, Resch MD, et al. Corneal Langerhans cell and dry eye examinations in ankylosing spondylitis. Innate immunity 2014;20:471-7.
12. He J, Ogawa Y, Mukai S, et al. In Vivo Confocal Microscopy Evaluation of Ocular Surface with Graft-Versus-Host Disease-Related Dry Eye Disease. Scientific reports 2017;7:10720.
13. Shalev-Shwartz S, Ben-David S. Understanding machine learning. From theory to algorithms. Understanding Machine Learning: From Theory to Algorithms 2013.
14. Wehle H-D. Machine Learning, Deep Learning, and AI: What's the Difference?2017.
15. O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. ArXiv e-prints 2015.
16. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.
17. Perez L, Wang J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. 2017.

18. Villani E, Baudouin C, Efron N, et al. In vivo confocal microscopy of the ocular surface: from bench to bedside. Current eye research 2014;39:213-31.
19. S.J R, Norvig P. Artificial Intelligence - A Modern Approach. Prentice Hall Inc, New Jersey 2003.
20. Lavric A, Valentin P. KeratoDetect: Keratoconus Detection Algorithm Using Convolutional Neural Networks. Computational intelligence and neuroscience 2019;2019:8162567.
21. Lopes BT, Ramos IC, Salomao MQ, et al. Enhanced Tomographic Assessment to Detect Corneal Ectasia Based on Artificial Intelligence. American journal of ophthalmology 2018;195:223-32.
22. Yousefi S, Yousefi E, Takahashi H, et al. Keratoconus severity identification using unsupervised machine learning. PloS one 2018;13:e0205998.