Development of Novel Diagnostic Tools for Dry Eye Disease using Infrared Meibography and In Vivo Confocal Microscopy

Inaugural Dissertation

zur

Erlangung des Doktorgrades
philosophiae doctor (PhD) in Health Sciences
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von
Md Asif Khan Setu
aus Narail, Bangladesh

Digital Express 24
Köln, 2022

Betreuer:	Prof. Dr. Philipp Steven
Gutachter:	Prof. Dr. Joseph Kambeitz
	Prof. Dr. Achim Tresch
Datum der Mündlichen Prüfung:	22.11.2022

Table of contents

Summary

Dry eye disease (DED) is a multifactorial disease of the ocular surface where tear film instability, hyperosmolarity, neurosensory abnormalities, meibomian gland dysfunction, ocular surface inflammation and damage play a dedicated etiological role. Estimated 5 to 50% of the world population in different demographic locations, age and gender are currently affected by DED. The risk and occurrence of DED increases at a significant rate with age, which makes dry eye a major growing public health issue. DED not only impacts the patient's quality of vision and life, but also creates a socio-economic burden of millions of euros per year.

DED diagnosis and monitoring can be a challenging task in clinical practice due to the multifactorial nature and the poor correlation between signs and symptoms. Key clinical diagnostic tests and techniques for DED diagnosis include tearfilm break up time, tear secretion – Schirmer's test, ocular surface staining, measurement of osmolarity, conjunctival impression cytology. However, these clinical diagnostic techniques are subjective, selective, require contact, and are unpleasant for the patient's eye. Currently, new advances in different state-of-the-art imaging modalities provide non-invasive, non- or semi-contact, and objective parameters that enable objective evaluation of DED diagnosis. Among the different and constantly evolving imaging modalities, some techniques are developed to assess morphology and function of meibomian glands, and microanatomy and alteration of the different ocular surface tissues such as corneal nerves, immune cells, microneuromas, and conjunctival blood vessels. These clinical parameters cannot be measured by conventional clinical assessment alone. The combination of these imaging modalities with clinical feedback provides unparalleled quantification information of the dynamic properties and functional parameters of different ocular surface tissues. Moreover, image-based biomarkers provide objective, specific, and non / marginal contact diagnosis, which is faster and less unpleasant to the patient's eye than the clinical assessment techniques.

The aim of this PhD thesis was to introduced deep learning-based novel computational methods to segment and quantify meibomian glands (both upper and lower eyelids), corneal nerves, and dendritic cells. The developed methods used raw images, directly export from the clinical devices without any image pre-processing to generate segmentation masks. Afterward, it provides fully automatic morphometric quantification parameters for more reliable disease diagnosis. Noteworthily, the developed methods provide complete segmentation and quantification information for faster disease characterization. Thus, the developed methods are the first methods (especially for meibomian gland and dendritic cells) to provide complete morphometric analysis.

Taken together, we have developed deep learning based automatic system to segment and quantify different ocular surface tissues related to DED namely, meibomian gland, corneal nerves, and dendritic cells to provide reliable and faster disease characterization. The developed system overcomes the current limitations of subjective image analysis and enables precise, accurate, reliable, and reproducible ocular surface tissue analysis. These systems have the potential to make an impact clinically and in the research environment by specifying faster disease diagnosis, facilitating new drug development, and standardizing clinical trials. Moreover, it will allow both researcher and clinicians to analyze meibomian glands, corneal nerves, and dendritic cells more reliably while reducing the time needed to analyze patient images significantly. Finally, the methods developed in this research significantly increase the efficiency of evaluating clinical images, thereby supporting and potentially improving diagnosis and treatment of ocular surface disease.

Zusammenfassung

Das Syndrom des trockenen Auges (DED) ist eine multifaktorielle Erkrankung der Augenoberfläche, bei der Tränenfilminstabilität, Hyperosmolarität, neurosensorische Anomalien, Funktionsstörungen der Meibomschen Drüsen sowie Entzündungen und Schädigungen der Augenoberfläche eine wichtige Rolle spielen. Derzeit sind schätzungsweise 5 bis 50 % der Weltbevölkerung in verschiedenen demografischen Regionen, Altersgruppen und Geschlechtern von DED betroffen. Das Risiko und Auftreten von DED nimmt mit dem Alter deutlich zu, was das trockene Auge Syndrom zu einem wichtigen und wachsenden Problem der öffentlichen Gesundheit macht. DED beeinträchtigt nicht nur die Seh- und Lebensqualität der Patienten, sondern verursacht auch eine sozioökonomische Belastung in Höhe von Millionen Euro pro Jahr.

Die Diagnose und Überwachung von DED kann in der klinischen Praxis aufgrund des multifaktoriellen Charakters und der geringen Korrelation zwischen Anzeichen und Symptomen eine schwierige Aufgabe darstellen. Zu den wichtigsten klinischen Diagnosetests und -techniken für die DED-Diagnose gehören die Tränenfilmaufreißszeit, die Tränensekretion (Schirmer-Test), die Färbung der Augenoberfläche, die Messung der Osmolarität und die Bindehautabdruckzytologie. Diese klinischen Diagnoseverfahren sind jedoch subjektiv, selektiv, erfordern Kontakt und sind für das Auge des Patienten unangenehm. Gegenwärtig bieten neue Fortschritte bei verschiedenen modernen bildgebenden Verfahren nicht-invasive, berührungsfreie oder halb-berührungsfreie und objektive Parameter, die eine objektive Bewertung der DED-Diagnose ermöglichen. Unter den verschiedenen und sich ständig weiterentwickelnden bildgebenden Verfahren wurden einige Techniken entwickelt, um die Morphologie und Funktion der Meibom-Drüsen sowie die Mikroanatomie und Veränderungen der verschiedenen Gewebe der Augenoberfläche wie Hornhautnerven, Immunzellen, Mikroneurome und konjunktivale Blutgefäße zu beurteilen. Diese klinischen Parameter können

nicht allein durch die herkömmliche klinische Beurteilung gemessen werden. Die Kombination dieser bildgebenden Verfahren mit klinischem Feedback liefert unvergleichliche Informationen zur Quantifizierung der dynamischen Eigenschaften und funktionellen Parameter der verschiedenen Gewebe der Augenoberfläche. Darüber hinaus bieten bildbasierte Biomarker eine objektive, spezifische und kontaktfreie Diagnose, die schneller und weniger unangenehm für das Auge des Patienten ist als die klinischen Beurteilungsverfahren.

Ziel dieser Dissertation war die Einführung neuer, auf Deep Learning basierender Berechnungsmethoden zur Segmentierung und Quantifizierung von Meibom-Drüsen (Ober- und Unterlider), Hornhautnerven und dendritischen Zellen. Die entwickelten Methoden verwendeten Rohbilder, die direkt von den klinischen Geräten exportiert wurden, ohne jegliche Bildvorverarbeitung, um Segmentierungsmasken zu erzeugen. Anschließend liefert es vollautomatische morphometrische Quantifizierungsparameter für eine zuverlässigere Krankheitsdiagnose. Bemerkenswert ist, dass die entwickelten Methoden vollständige Segmentierungs- und Quantifizierungsinformationen für eine schnellere Krankheitscharakterisierung liefern. Somit sind die entwickelten Methoden die ersten (insbesondere für Meibom-Drüsen und dendritische Zellen), die eine vollständige morphometrische Analyse ermöglichen.

Zusammengenommen haben wir ein auf Deep Learning basierendes automatisches System zur Segmentierung und Quantifizierung verschiedener Augenoberflächengewebe im Zusammenhang mit DED entwickelt, nämlich Meibom-Drüse, Hornhautnerven und dendritische Zellen, um eine zuverlässige und schnellere Charakterisierung der Krankheit zu ermöglichen. Das entwickelte System überwindet die derzeitigen Grenzen der subjektiven Bildanalyse und ermöglicht eine präzise, genaue, zuverlässige und reproduzierbare Analyse des Augenoberflächengewebes. Diese Systeme haben das Potenzial, sich sowohl in der Klinik als auch in der Forschung auszuwirken, indem sie eine schnellere Krankheitsdiagnose

ermöglichen, die Entwicklung neuer Medikamente erleichtern und klinische Studien standardisieren. Darüber hinaus können Forscher und Kliniker Meibom-Drüsen, Hornhautnerven und dendritische Zellen zuverlässiger analysieren und gleichzeitig die für die Analyse von Patientenbildern benötigte Zeit erheblich reduzieren. Schließlich erhöhen die in dieser Forschungsarbeit entwickelten Methoden die Effizienz der Auswertung klinischer Bilder erheblich, wodurch die Diagnose und Behandlung von Augenoberflächenerkrankungen unterstützt und potenziell verbessert werden kann.

1 Introduction

1.1 Dry eye disease

Over the last three decades, the prevalence of Dry Eye Disease (DED) has significantly increased in the population and therefore understanding of DED has grown [1]. This immune-based inflammatory disease of the ocular surface and tears is characterized by its multifactorial nature where tear film instability, hyperosmolarity, neurosensory abnormalities, meibomian gland dysfunction (MGD), ocular surface inflammation, and damage play etiological role [2]. An estimated 5% to 50% of the world population in different geographic locations, age, and gender are currently affected by DED [3]. The risk and occurrence of DED increases at a significant rate with age, which makes dry eye a major growing public health issue as the population of older people is expected double from its current number by 2050 [4]. DED not only impacts the patient's quality of vision and life, but also creates a socio-economic burden [5,6] of millions of euros per year. Dry Eye is a disease of the lacrimal functional unit [7] resulting in inadequate tear film composition. The disturbance of tear production and tear evaporation causes ocular discomfort, dryness, pain, and alteration of tear composition. DED is mainly categorized into two categories, aqueous-deficiency dry eye (ADDE) and evaporative dry eye (EDE), however these two categories exist on a continuum rather than as different entities [8]. Epidemiological and clinical evidence suggested that evaporative category is preponderance of DED in nature which makes EDE the most common type of DED[3,8]. Most patients demonstrate features of enhanced tear film evaporation, clinical signs of MGD, and ocular surface inflammation [8,9].

1.1.1 Role of the meibomian gland in dry eye disease

Meibomian glands (MG) are the large sebaceous glands that are located below the tarsal conjunctiva of both upper and lower eyelid and within the tarsal plate (Fig. 1) [10]. They produce lipids (an oily substance) which are the key component of the lipid layer of the tear film. In the tear film, the lipid layer is superficial to the aqueous layer and prevents early evaporation

of the aqueous phase. The lipid layer also balances the tear film by lowering surface tension [11]. Therefore, lipids produced by MG are an indispensable part of ocular surface health. In general, a healthy upper eyelid contains 25-40 glands with an average of around 31 and the lower eyelid contains 20-30 glands with an average of around 26. The individual gland length is approximately 5.5 mm and 2 mm in the middle of the upper and lower eyelid, respectively. The volume of the total MGs are approximately twice as large in the upper eyelid (26 µL) than in the lower eyelid (13 µL) and the number of acini is also higher in the upper eyelid. However, MG of the lower eyelid are wider than the upper eyelid [12].

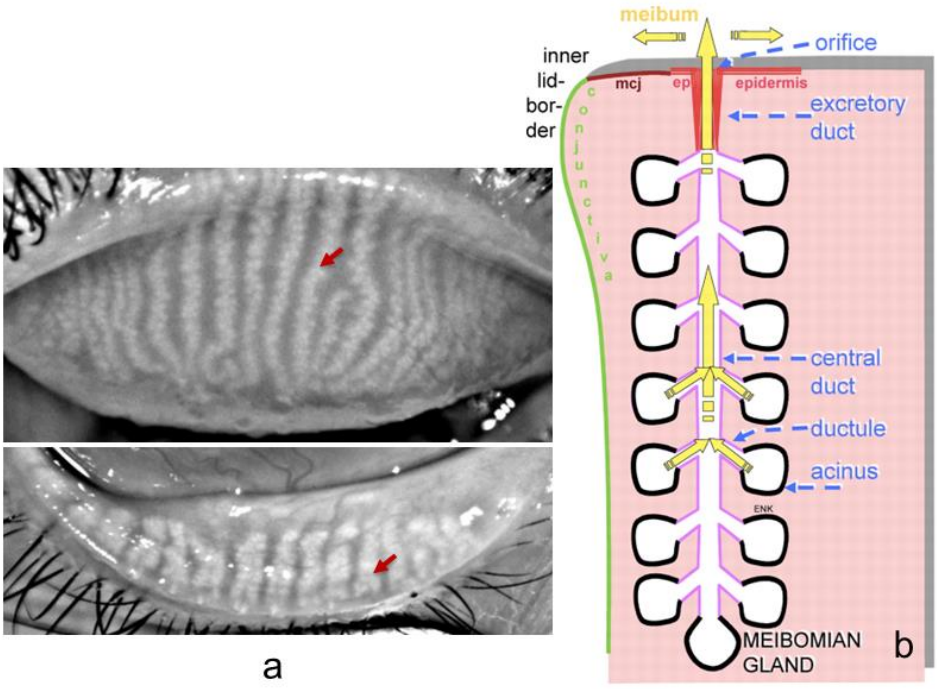

Figure 1. Meibomian glands with cellular physiology. (a) MG within the tarsal plate of the upper and lower eyelids. Red arrow indicates single MG. (b) Morphology of a single MG. The oily secretum (meibum) is synthesized within the secretory acini and transported (yellow arrows) in a distal direction toward the orifice. Figure 1b is reproduced from [13] with kind permission of Springer Science.

MGD is commonly described by functional abnormalities of MGs. It is a chronic, diffuse abnormality of MGs, which is caused by terminal duct obstruction and quantitative or qualitative changes in glandular secretion [9]. The term dysfunction is used as the systematic function of

MGs is interrupted and the term diffuse abnormality is used as the disorder associates with most of the MGs. MGD causes alteration of the tear film, eye irritation, inflammation, and ocular surface disease, including dry eye. Obstruction of the MG orifices and ducts and quantitative or qualitative changes in MG secretions are the most well-known clinical characteristics of MGD. These lead to an unstable lipid layer of the tear film and cause an increase in the evaporation of the underlying aqueous layer [14]. Therefore, any change in the structure or function of MGs could have an important clinical impact [1]. MGD is considered the major cause of EDE and lipid layer deficiency [9,12,15]. Therefore, thorough analysis of MGD in patients with DED provides insight of changes to the phenotype of disease.

1.1.2 Role of corneal nerve fibers and dendritic cells in dry eye disease

The human cornea is a transparent, avascular tissue [16] and has a higher distribution of sensory nerves than tooth pulp or finger tips. The sensory nerve fibers largely originate from the ophthalmic trigeminal nerve with free nerve endings terminating in the corneal epithelium [17]. In addition to these sensory nerve fibers, the cornea has autonomic nerves who play a role in important corneal function such as corneal epithelial proliferation, integrality, wound healing and homeostasis [17,18] . For instance, if any foreign body enter the ocular surface and stimulate mechano-nociceptors, thus cause production of overflow tears that wash away harmful materials. Hence sensory nerves are able to rapidly response to a wide range of environmental changes by altering volume of tear production [17]. Histologically, the cornea consists of five basic layers: (1) Epithelium, (2) Bowman's layer, (3) Stroma, (4) Descemet's membrane, and (5) Endothelium (Fig. 2a). The sensory and autonomic nerve fibers are located between Bowman's layer and basal epithelium [19]. Corneal nerve fibers (CNF) (subbasal nerve plexus) derive from the anterior stroma and run forward to the central area of the cornea in a radial fashion parallel to the corneal surface [20] and giving rise to branches that innervate the anterior and mid stroma [18,21]. Peripheral stromal nerves contain both myelinated and unmyelinated nerve

fascicles. These fascicles turn anteriorly towards the corneal surface in a 90-degree angle, penetrate Bowman's layer, after which they extend as nerve axons in the form of epithelial leashes parallel to the corneal surface. The subbasal nerve plexus forms between Bowman's and basal epithelial layer that supplies the overlying corneal epithelium (Fig. 2b)[21,22].

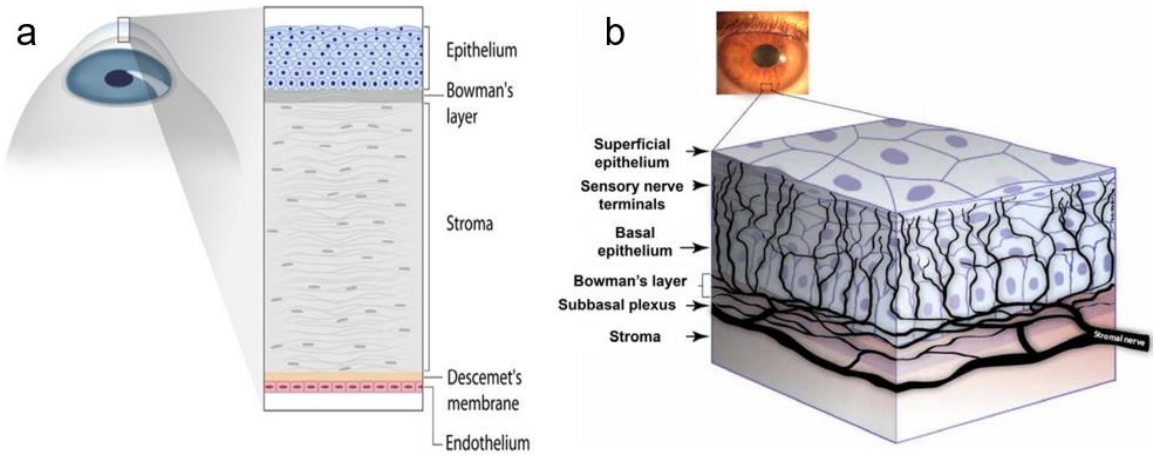

Figure 2. Corneal layers and nerve fibers. (a) Five basic layers of the cornea. Reproduced from [19] with kind permission of Elsevier. (b) Diagrammatic representation of human corneal nerves. Reproduced from [22] with kind persion of Elsevier.

Corneal dendritic cells (DCs) are antigen-presenting cells which are derived from bone marrow and play a vital role in the corneal immune defense system against the external environment [23,24]. DC are broadly distributed on the ocular surface and are mainly processing and presenting antigens to other immune cells, hereby play a key role in ocular surface immunity [24]. DCs appear as a white reflective structures (Fig. 3) when imaged in-vivo using confocal microscopy [23].

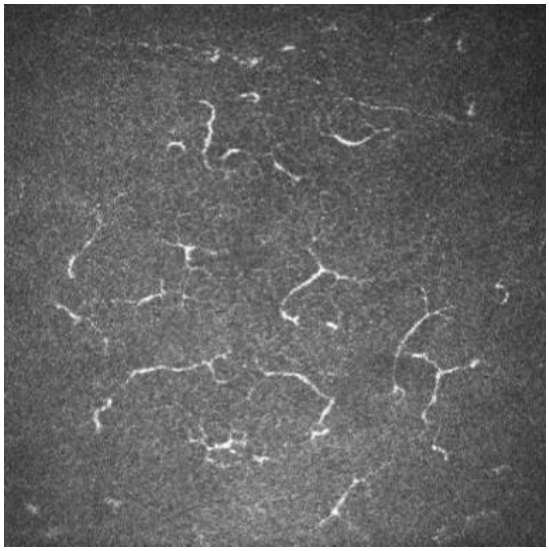

Figure 3. In vivo confocal microscopy image of cornea. White reflective structures are DCs.

DCs are categorized into immature and mature based on their location and dendrites (branch-like structure) [16]. In the healthy state, most of the immature cells without dendrites are located at the center of the cornea and mature cells with dendrites are located at the peripheral cornea. Zhivov et al. further classified DCs into three types: (1) immature cells without dendrites, (2) cells with small dendrites (transition stage cells which start dendrite process), and (3) mature cells with long dendrites [25]. Recent research showed that the changes in DCs play an important role in the pathogenesis of DED which induces T-cell activation and create an inflammatory cascade in DED [26].

Corneal nerve fibers play an integral role by mediating touch, pain, chemical and mechanical stimuli, and temperature signals to our brain. Moreover, they induce reflex tear production, blinking, and release tropic factors with the aim of preserving the structural and functional integrity of the ocular surface [27]. In addition to CNF, DCs are also an essential regulator of both innate and adaptive immune systems to maintain a healthy ocular surface [24]. Altered morphological structure of CNFs and DCs could destabilize the tear film and lead to ocular surface disease such as DED. These abnormalities of CNFs and DCs are generally found to correlate with DED severity leading to further ocular surface damage and also corneal

neuropathic pain [28] which has gained increasing attention in the near past and has developed into an own field of research.

Patients with corneal neuropathic pain may suffer from diverse symptoms such as light sensitivity, irritation, and a vague sensation of pressure. Routine activities like reading, driving, and work are heavily affected which impairment patient's overall quality of life [27]. In general pain is defined as "an unpleasant sensory and emotional experience associated with potential or actual tissue damage which might exist over a short or a prolonged period" [29]. This also applies to the eye. Ocular pain may be caused by a wide range of ocular disorders. However, the most common causes of ocular pain are ocular surface and corneal disorders [30–32], including DED [22,33,34], refractive surgery [35], Sjögren's syndrome [36], diabetes [37], small fiber neuropathy [32,38]. Pain from DED not only causes irritation but can also indicate the severity of the disease. Several studies [30] [39] have shown that patients who suffer from DED with ocular pain have more severe signs (such as decreased CNF number and density, increased DCs density) and symptoms (such as burning, hypersensitivity to wind, sensitivity to light and temperature) than patients without ocular pain. Patients suffer from small fiber neuropathy demonstrated reduced sensitivity, nerve fiber density, length, branching points, and increased tortuosity [40]. Furthermore, patients often report burning sensation and other somatosensory features [41]. It could also serve as a driver of centralized ocular pain [37]. Therefore, evaluation of these corneal nerves and DC in patients with ocular surface disease including DED, and neuropathic corneal pain could provide better insight into the changes of the clinical expression.

1.2 Dry eye imaging in clinical practice

DED diagnosis and monitoring can be a challenging task in clinical practice [42]. This is due to the multifactorial nature of the disease and the poor correlation between signs and symptoms of DED [43,44]. The TFOS DEWS II [45] Diagnostic Methodology Subcommittee

recommended key diagnostic tests and techniques for DED diagnosis. These include tearfilm break up time, tear secretion – Schirmer's test, ocular surface staining with fluorescein and lissamine (observing the cornea, conjunctiva, and eyelid margin), measurement of osmolarity, conjunctival impression cytology [45]. However, these clinical diagnostic techniques are subjective, selective, in part require contact to the eye, and are therefore unpleasant or bear risks of damaging the delicate ocular surface. Furthermore, sensitivity and specificity of these techniques heavily rely on the inclusion criteria, population examined, and severity of the DED group.

At present, there is a growing need in developing biomarkers for ocular surface disease diagnosis in particular DED, and advancements in imaging technologies allowing image based biomarkers to serve as a prospective solution and most promising approach to this need [46,47]. Different imaging modalities allow clinicians to image ocular surface tissues for quantitative and/or qualitative evaluation. Currently, new advances in different state-of-the-art imaging modalities provide non-invasive, non or semi contact, and objective parameters that enable objective evaluation of DED diagnosis. Among the different and constantly evolving imaging modalities, namely, meibomian gland imaging, in vivo confocal microscopy, and optical coherence tomography are developed to assess tear film stability and volume, morphology and function of meibomian gland (structural abnormalities or gland dropout), and microanatomy and alteration of the different ocular surface tissues such as corneal nerves, immune cells, microneuromas, and conjunctival blood vessels [46]. These clinical parameters cannot be measured by conventional clinical assessment alone. The combination of these imaging modalities with clinical feedback provides unparalleled quantification information of the dynamic properties and functional parameters of different ocular surface tissues. Moreover, image-based biomarkers provide objective, specific, and non / marginal contact diagnosis, which is faster and less unpleasant to patient's eye than the clinical assessment techniques [44].

1.2.1 Meibomian gland imaging

Infrared (IR) meibography is routinely used and widely accepted in clinical practice as a method for imaging and quantifying MGs during ocular surface disease examination. IR meibography is a non-contact MG imaging technique, where a camera and an IR light source are used to image MG by not touching patient eyelid during examination [48,49]. It was first introduced by Mathers et al. [50] in 1994 where the authors developed a video meibography system to observe MG real time on a computer. An IR charged-coupled camera was used to develop this system. Later, Arita et al. [48,49,51–53] used non-contact infrared meibography device which consisted of a regular slit-lamp microscope, with an IR charge-coupled device (CCD) video camera, and an IR transmitting filter. It allows clinicians to observe both upper and lower eyelids, when they are everted, without the contact of a probe as the lids are illuminated by an IR light source. The silicon-based CCD digital video camera's chip is sensitive to IR light around 850 nm. IR images enhance the visibility of MG morphology as the glands are IR hyperreflective. It currently not known what elements of MG (connective tissue, cell type, and secretion) are IR hyperreflective, therefore knowledge of such would be beneficial for clinicians to assess the MG using meibography. Overall, non-contact IR meibography is preferable for the patient's comfort during examination [49].

To date, several commercially available clinical IR meibography devices have been developed for the European market: (1) Oculus Keratograph 5M (OCULUS Optikgeräte GmbH, Wetzlar, Germany), (2) bon ANTARES – the High-End Dry Eye Topographer (bon Optic Vertriebsges GmbH, Lübeck, Germany), and (3) ANTARES (CSO, Costruzione Strumenti Oftalmici, Florence, Italy). These three different IR meibography devices technical and optical information are tabulated in table 1.

Table 1: Different IR meibography devices - technical and optical information [49,54,55].

IR meibography device	Camera type	IR light source	Accuracy	Number of rings	Number of data points
Keratograph 5M	Digital CCD	840 nm	± 0.1D	22	22000
Bon ANTERAS	Digital CCD	No information	± 0.01D	24	6144
CSO ANTERES	Digital CCD	875 nm	No information	24	6144

IR meibography provides a physiological visualization of the eyelid along with MG. For reliable disease diagnosis, clinicians need to segment MG accurately in IR images. Traditionally, the techniques used to segment MG were manual or semi-automated, which require experienced personnel. However, manual or semi-automatic techniques were laborious, cost-ineffective, and potentially subject to user bias. To address these issues, several automatic MG segmentation methods have been published by Koh et al. [56], Llorens-Quintana et al. [57], Arita et al. [58], Celik et al. [59], and Koprowski et al. [60]. However, all of these methods heavily depend on traditional image processing techniques which performance relies on IR meibography image quality. It is quite challenging for examiners and clinicians to repeatedly acquire sufficient good quality images [10]. Figure 4 represents the oculus keratography 5M device (Fig. 4a) and both upper and lower eyelids IR meibography images (Fig. 4b and Fig. 4c).

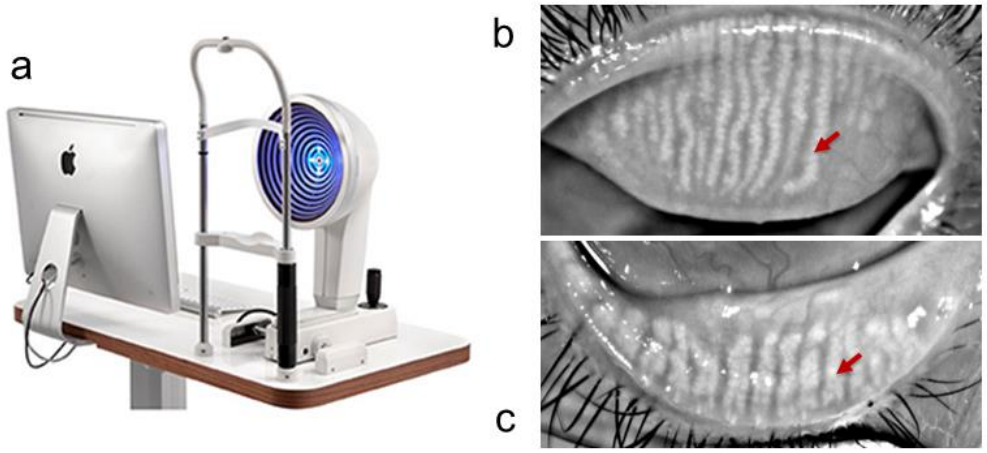

Figure 4. IR meibography device and meibomian gland imaging. (a) IR meibography device, Oculus Keratograph 5M, (b) Upper eyelid IR meibography image, and (c) Lower eyelid IR meibography image. Red arrow indicates individual MG in both upper and lower eyelids. Oculus Keratograph 5M device image reproduced from [61] with kind permission of Frontiers in Medicine (Ophthalmology).

1.2.2 In vivo confocal microscopy

In vivo confocal microscopy (IVCM) is clinically widely accepted and routinely used in the diagnosis of various ocular surface disorders including dry eye, neuropathic corneal pain (NCP) and diabetic neuropathy by providing detailed morphometric information of CNF and DC [62,63]. Confocal microscopy known as confocal laser scanning microscopy (CLSM) is a non-invasive optical imaging technique that provides high-resolution images with a better rejection of out-of-focus information than regular light microscopy. It provides multiple advantages over optical microscopy by controlling the depth of field, eliminating background information from the focal plane, and collecting a series of optical sections from thick tissue specimens. It uses spatial filtering techniques to reduce out-of-focus light in tissue with thickness exceeding the immediate plane of focus. Thereby, it is uniquely fitted to image and analysis tissue specimens in the living subjects. IVCM has been used over the last 25 years since it was developed in various areas of corneal research. Unlike other non-invasive anterior segment imaging techniques such as optical coherence tomography, IVCM provides high resolution en face images of various corneal structures and cells [64].

To date, three important confocal microscopy systems have been introduced for in vivo corneal imaging: (1) Heidelberg Retina Tomograph with Rostock Cornea Module (HRT-RCM, Heidelberg Engineering GmbH, Heidelberg, Germany) [64,65], (2) Tandem Scanning Confocal Microscope (TSCM) [64,66], and (3) Confoscan 4 (Nidek Technologies Srl, Padova, Italy) [64,67], and (3). The HRT-RCM is a CLSM that uses a 670 nm laser beam in a raster pattern over the field of view. It uses a very high numerical aperture of 63× objective lens which provides images of high contrast and better axial resolution (7.6 µm) than the other mentioned IVCM systems (9 µm resolution for TSCM and 24 µm resolution for confoscan) [64,66,68].

IVCM enables live cellular physiological visualization of corneal nerve fibers (CNF) and dendritic cells (DC). For accurate quantification of CNF morphology and detection of DC, nerves and DCs must be accurately segmented in the IVCM images. Traditionally, the methods used to segment CNFs and DCs have been manual or semi-automated techniques. These require experienced personnel, which is laborious, cost-ineffective, and potentially subject to user bias. To address these issues, several automatic CNF segmentation and quantification software tools have been developed [69–72] using traditional image processing or machine learning where manually selected features were used. For example, Dabbah et al. developed a dual-model automated CNF detection method which showed excellent correlation with manual grading ($r = 0.92$) and further extension of this method using dual-model property in a multi scale framework to generate feature vectors at every pixel and achieved high correlation with manual grading ($r = 0.95$) [69,73]. Figure 5 demonstrates an in-vivo confocal microscopy device (HRTIII-RCM) (Fig. 5a) and an example IVCM image of CNF and DCs (fig. 5b).

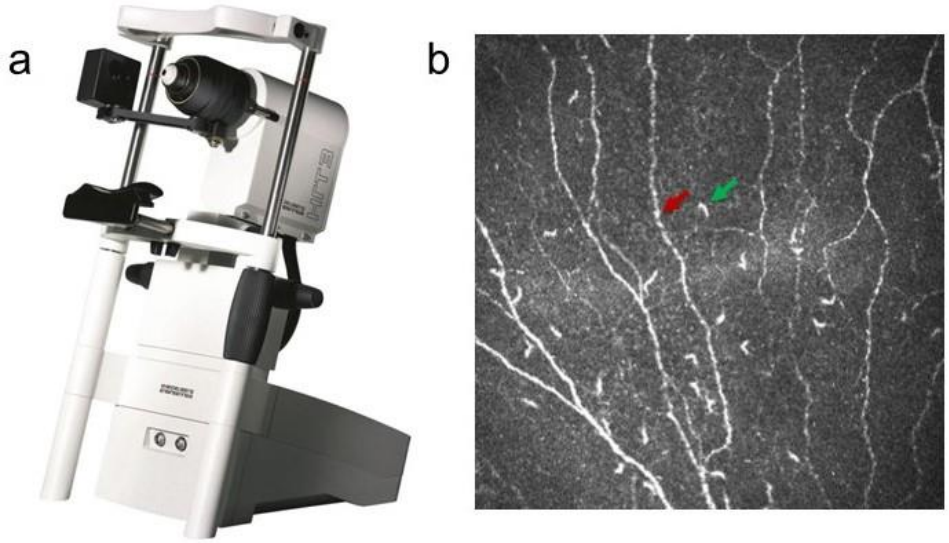

Figure 5. IVCM device along with corneal nerves and DCs. (a) Heidelberg Retina Tomograph - Rostock Cornea Module (HRT-RCM), in vivo confocal microscopy device image reproduced from [74] with kind permission of Springer Science (b) IVCM image. Red arrow indicates CNF and green arrow indicates DC.

1.3 Artificial intelligence in ophthalmology

The recent advancements in computer hardware, software, and data have enabled the use of artificial intelligence (AI) using deep learning (DL) to achieve high performance in several applications including healthcare. These advancements include decreasing the cost of a graphics processing unit (GPU), significant improvement of mathematical modeling, and availability of high volume of big data. Utilization of DL provides robust performance and improved capability in image segmentation and evaluation. In medicine, DL systems have demonstrated significant progress in image-based disease diagnosis in radiology, dermatology, pathology, and ophthalmology [75]. Ophthalmology is spearheading AI applications in medicine [76,77], and, currently, DL techniques have been increasing in popularity in the field of ophthalmology [75,76,78,79].

Deep learning proves the ability of deep neural networks to learn useful features and visual representations from a big dataset of manually labeled images, without explicitly defining hand crafted rules. Deep neural network is a cascade of processing layers which mimic the

human visual cortex signaling processes. Each neuron of the visual cortex response to a specific region or object of an image, similar to this, each layer of a deep neural networks extracts specific features of a particular region or object of an image [75]. In contrast to traditional image processing techniques, a DL system has an advantage of extracting and representing useful features from the raw images automatically and demonstrated significant performance on image segmentation and classification tasks [80]. Thus, it overcomes the need for manual cumbersome feature selection or use of several filters for image enhancement to noise reduction. DL is more preferable to traditional techniques or machine learning techniques as manually selected features for a specific disease may not be transferable to another disease [81,82]. Using CNN and a high volume of image data, a DL system can learn to segment specific objects from several ophthalmic images including retinal blood vessels [83], retinal layers [84], optic disc [85], and anterior chamber angle [86]. To date, several DL based MG and CNF segmentation models have been proposed. There are only few commercially established DL-based automatic software that were developed specially for retinal disease diagnosis such as diabetic retinopathy, glaucoma, and age-related-macular degeneration. However, these software are not available for all markets worldwide [87]. Thus, there is a need for the anterior segment and ocular surface disease diagnosis DL-based commercial software. Prabhu et al. [88] developed the first DL based MG segmentation where they achieved acceptable *p* values between manual and automatic segmentation. Wang et al. [89] developed another DL system to analyze gland atrophy, however, their model could not segment or evaluate individual MG which is clinically important for reliable ocular surface disease diagnosis. Maruoka et al. [90] proposed a DL method to detect obstructive MGD using IVCM images, however, IVCM is not routinely used to image and analyze MG in clinical practice. In addition, a small field of view of IVCM images refrains to image full length of MG or total eyelid area. The first DL based CNF segmentation model was developed by Williams et al. [91]. They demonstrated the efficacy of deep learning models to identify DPN with high interclass correlation with manual ground truth

annotation. Wei et al. [92] and Colonna et al. [93] developed a model that provides high sensitivity and specificity in the segmentation task of corneal nerve fibers (CNF). Oakley et al. [94] developed a DL model to analyze macaque CNF which achieved a high correlation between manual and automatic segmentation. Yıldız et al. [95] proposed CNF segmentation using generative adversarial network (GAN), which is different from traditional CNN, and they achieved similar correlation and Bland-Altman analysis results with the traditional DL model, U-Net. Though several DL based MG and CNF segmentation models have been developed by different research groups, the developed models provide only segmentation without further morphometric evaluation. Additionally, for IVCM image analysis, researchers have focused on CNF segmentation alone, and not DC segmentation. Therefore, there is a need for fully automatic segmentation and morphometric evaluation of MG, CNF, and DC.

In this PhD thesis, deep learning-based novel computational methods for the segmentation and quantification of meibomian glands (both upper and lower eyelids), corneal nerves, and dendritic cells are introduced. The developed methods use raw images, directly export from the clinical devices without any image pre-processing to generate segmentation masks. Afterward, it provides morphometric quantification parameters for more reliable and faster disease diagnosis. Noteworthily, the developed methods provide complete and automatic segmentation and quantification information for faster disease characterization. This is a significant improvement compared to current traditional image processing or AI-based segmentation methods. Most of the deep learning-based methods provide segmentation without further morphometric analysis. Thus, the here developed methods are the first methods (especially for meibomian gland and dendritic cells) to provide complete morphometric analysis information.

Finally, the developed deep learning-based method's performance was analyzed and compared with the manual annotation and currently available state-of-the-art software (for corneal nerves analysis).

2 Results

2.1 Deep learning-based automatic meibomian gland segmentation and morphology assessment in infrared meibography

Md Asif Khan Setu, Jens Horstmann, Stefan Schmidt, Michael E. Stern, Philipp Steven

Objectives: Meibomian glands (MG) are large sebaceous glands located below the tarsal conjunctiva and the abnormalities of these glands cause Meibomian gland dysfunction (MGD) which is responsible for evaporative dry eye disease (DED). Accurate MG segmentation is a key prerequisite for automated imaging-based MGD related DED diagnosis. However, Automatic MG segmentation in infrared meibography is a challenging task due to image artifacts.

Methods: A deep learning-based MG segmentation has been proposed which directly learns MG features from the training image dataset without any image pre-processing. The model is trained and evaluated using 728 anonymized clinical meibography images. Additionally, automatic MG morphometric parameters, gland number, length, width, and tortuosity assessment were proposed.

Results: The average precision, recall, and F1 score were achieved 83%, 81%, and 84% respectively on the testing dataset with an AUC value of 0.96 based on ROC curve and dice coefficient of 84%. Single image segmentation and morphometric parameter evaluation took on average 1.33 s.

Conclusions: The trained U-Net model is able to learn specific and prominent features of MG from input training meibography images which may improve the MG segmentation performance.

There is a need for validation procedures by expert clinicians. To the best of our knowledge, this is the first time that a validated deep learning-based approach is applied in MG segmentation and evaluation for both upper and lower eyelids.

Translational Relevance: This automatic MG segmentation method could overcome current limitations such as time and accuracy. It provides accurate, faster, and reliable automatic MG segmentation than conventional MG segmentation methods.

Own Contribution to publication 1:

Md Asif Khan Setu designed and developed the deep learning algorithm and automatic morphometric parameters assessment software for meibomian glands, wrote python-based code to develop deep learning algorithm and automatic morphometric parameter assessment software, anonymized the clinical meibography image data, pre and post processed images, generated ground truth mask images, algorithm and software performance evaluated, data evaluated and interpreted, statistical analysis, and wrote the main manuscript text.

Co-authors contribution to publication 1:

Jens Horstmann and Stefan Schmidt provided technical knowledge and reviewed the manuscript text. Mike E. Stern participated in study design and reviewed and commented on the manuscript. Philipp Steven generated the main idea of this research, provided clinical background knowledge, collected clinical meibography image data, verified ground truth masks and wrote and reviewed the manuscript.

2.2 Segmentation and Evaluation of Corneal Nerves and Dendritic Cells from In Vivo Confocal Microscopy Images using Deep Learning

Md Asif Khan Setu, Stefan Schmidt, Gwen Musial, Michael E. Stern, Philipp Steven

Purpose: Segmentation and evaluation of in vivo confocal microscopy (IVCM) images requires manual intervention, which is time consuming, laborious, and non-reproducible. The aim of this research was to develop and validate deep learning–based methods that could automatically segment and evaluate corneal nerve fibers (CNFs) and dendritic cells (DCs) in IVCM images, thereby reducing processing time to analyze larger volumes of clinical images.

Methods: CNF and DC segmentation models were developed based on U-Net and Mask R-CNN architectures, respectively; 10-fold cross-validation was used to evaluate both models. The CNF model was trained and tested using 1097 and 122 images, and the DC model was trained and tested using 679 and 75 images, respectively, at each fold. The CNF morphology, number of nerves, number of branching points, nerve length, and tortuosity were analyzed; for DCs, number, size, and immature–mature cells were analyzed. Python-based software was written for model training, testing, and automatic morphometric parameters evaluation.

Results: The CNF model achieved on average 86.1% sensitivity and 90.1% specificity, and the DC model achieved on average 89.37% precision, 94.43% recall, and 91.83% *F*1 score. The interclass correlation coefficient (ICC) between manual annotation and automatic segmentation were 0.85, 0.87, 0.95, and 0.88 for CNF number, length, branching points, and tortuosity, respectively, and the ICC for DC number and size were 0.95 and 0.92, respectively.

Conclusions: Our proposed methods demonstrated reliable consistency between manual annotation and automatic segmentation of CNF and DC with rapid speed. The results showed that these approaches have the potential to be implemented into clinical practice in IVCM images.

Translational Relevance: The deep learning–based automatic segmentation and quantification algorithm significantly increases the efficiency of evaluating IVCM images, thereby supporting

and potentially improving the diagnosis and treatment of ocular surface disease associated with corneal nerves and dendritic cells.

Own contribution to publication 2:

Md Asif Khan Setu designed and developed both deep learning algorithms and automatic morphometric parameters assessment software for corneal nerves and dendritic cells, wrote python-based code to develop deep learning algorithm and automatic morphometric parameter assessment software, anonymized the clinical confocal microscopy image data, pre and post processed images, generated ground truth mask images, algorithms, and software performance evaluated, data evaluated and interpreted statistical analysis, and wrote the main manuscript text.

Co-authors contribution to publication 2:

Stefan Schmidt and Gwen Musial provided technical knowledge and reviewed the manuscript text. Mike E. Stern participated in study design and reviewed and commented on the manuscript. Philipp Steven generated the main idea of this research, provided clinical background knowledge, collected clinical confocal microscopy image data, verified ground truth masks images, and wrote and reviewed the manuscript.

3 Discussion

3.1 Main findings of the two presented research studies in summary

Fully automatic deep neural network-based novel computational methods for MG, CNF, and DC segmentation and quantification in infrared meibography and corneal IVCM images have been developed in this PhD research. The developed methods use a novel strategy where it first segments the desired objects from the raw images and then provides morphometric

quantification parameters for reliable disease diagnosis. To the best of the authors knowledge, this is the first deep neural network-based MG (both upper and lower eyelids) and DC segmentation and evaluation method which have been developed and validated. This PhD research study validates the here developed deep-learning methods and confirms that the segmentation performance is comparable with manual ground truth images while significantly reducing amount of time to analyze those images. In particular, the here developed methods provide high accuracy between manual ground truth images and automatic segmentation in the quantification metrics even in images with reduced acquisition quality. In addition to automatic segmentation, the newly developed methods provide fully automatic, morphometric, clinical variables which have utility in the diagnosis of DED, corneal neuropathic pain, and other ocular surface diseases.

3.2 Meibomian gland segmentation and quantification

Infrared meibography images play an important role to analyze MG of both upper and lower eyelids in the diagnosis of MGD. In clinical practice, ophthalmologists need to analyze the images multiple times to ensure accurate results for reliable disease diagnosis. During the manual analysis, the ophthalmologist may evaluate the MG differently at different patient appointments which could lead to within observer variability. Furthermore, eversion of the upper eyelid is never the same and differences in positioning of the eyelid have vast impact on two-dimensional topography of MGs. In addition, two different observers could quantify the same image differently leading to observer variability. Therefore, a fully automatic MG segmentation and evaluation software is needed for reliable, faster, and reproducible results. In contrast to traditional image processing techniques, where various filters were used [60-63], DL-based techniques are much faster and more reliable to segment and quantify desired objects from the clinical images. Unlike the traditional machine learning methods where manually selected hand

crafted features were used to segment or classify object [61], the DL method waives the need for complex manual feature selection by using automatic feature extraction and representation.

The developed deep neural network-based method in this PhD research provides fully automatic MGs segmentation and morphometric evaluation from infrared meibography images without any image pre-processing. DL model achieved high performance (97% accuracy, 84% dice similarity, 83% precision, and 81% recall) with 0% variability and 100% repeatability (test-retest reliability). The average segmentation and evaluation time per meibography image was approximately 1.33 seconds which means the model could evaluate 1000 clinical images only in 22 minutes without any user involvement. It reduces significant image processing time and computational cost to analyze clinical images and assist clinicians to take faster and better disease diagnosis decision. The single trained model is able to segment MGs from both upper and lower eyelids. There is no failed case to segment MG in the test dataset. From the visualization of the automated segmentation of MGs has high visual similarity with manual ground truth MGs. Furthermore, statistical analysis (Bland-Altman method) supports the similarity between automatic segmentation and manual annotation. Overall, the research results demonstrate a fully automated and reliable deep learning-based technique for MG segmentation and morphometric evaluation from IR meibography images.

MGs segmentation and quantification model has some limitations. Firstly, the developed method is device and location specific, where all the images (training, validation, and testing) were acquired using an Oculus Keratograph 5 M (Oculus GmbH, Wetzlar, Germany) from the Department of Ophthalmology, University Hospital Cologne, Germany. Secondly, the ground truth images were created manually where there is a chance of observer variability. However, all manually labeled images were verified and corrected by a senior experienced ophthalmologist. Thirdly, DL model did not perform well in all test images. In some cases, a single gland is separated into multiple one and in other cases two glands are connected and considered as a

single gland. In all of these unusual cases, images were not acquired correctly and had strong artifacts such as specular reflection and MGs morphological deformation. Moreover, the model should be validated and tested using patient longitudinal data, which will make the model more robust to diagnosis follow-up cases. Finally, the developed model does not provide any clinical correlation to differentiate disease versus health and meibography score to distinguish disease severity stage.

3.3 Corneal nerves and dendritic cells segmentation and quantification

IVCM images play an important role in clinical practice to diagnosis and analyze many corneal diseases. To ensure accurate results, ophthalmologists often need to analyze the patient images multiple times for disease diagnosis or scientific research. The ophthalmologist may evaluate CNF and DC differently at different patient appointments during their manual image analysis which could lead to within observer variability. Therefore, there is a need for reliable, faster, and reproducible fully automatic CNF and DC segmentation and evaluation methods. Unlike the traditional image processing techniques where various filters and graphs were used [69–72], deep learning-based methods are comparatively faster to segment objects from the clinical images. In contrast to traditional machine learning method where handcrafted features are needed to segment or classify the object [70], the deep neural network-based method replaces the manual complex feature selection with automatic feature extraction using hundreds of filters.

In this PhD research study, a novel method was developed to classify corneal nerves into main nerves and branch nerves for further quantitative evaluation after the binary segmentation. CNF deep neural network model achieved on average 84% sensitivity, 91% specificity, and AUC of 0.88 on the test dataset. Corneal nerve fiber length, fiber density, number of branching points, and tortuosity were calculated automatically using the newly developed software which

are useful clinical parameters to diagnose dry eye and neuropathic corneal pain severity [30,31,34].

The research results demonstrate that the developed deep learning models can reliably segment and evaluate the CNF and DC with rapid speed. The CNF and DC model took on average 4.5 seconds and 3 seconds per image, respectively, to segment and evaluate IVCM images. Both deep learning-based models significantly reduce clinical image analysis time of a large volume of patient data. Visibly, both manual annotation and automatic segmentation of CNF and DC appear similar with important clinical features. Overall, the developed methods provide an objective, fast, and reliable approach to segment and morphometric analysis of CNF and DC in IVCM images.

CNF and DC segmentation and quantification models have some limitations. Firstly, CNF and DC images were acquired from Peking University Third Hospital, China and Department of Ophthalmology, University Hospital Cologne, Germany, respectively using the Heidelberg Retina Tomograph (HRT3) with the Rostock Cornea Module (RCM) (Heidelberg Engineering GmbH, Heidelberg, Germany) based on small patient cohort. Therefore, a large and more diverse patient cohort with regards to age, gender and regional origin, and different types of IVCM devices could potentially increase the generalizability and robustness of the developed model. Moreover, the model should be validated and tested using patient longitudinal data, which will make the model more specific to diagnose follow-up cases. Finally, the developed model could not automatically differentiate disease versus healthy cornea instead of providing different morphometric quantification parameters.

3.4 Future studies

Further research and clinical validation (clinical trials) are needed to be carried out to identify the feasibility of introducing this system into ophthalmic practice and commercially

available diagnostic devices. Noteworthily, the DL-based model should provide clinical explainability and interpretability instead of some deep learning probability score to assist medical doctors make better and faster diagnosis decision. Furthermore, the DL model should meet regulatory guidelines in order to be trustworthy, namely accuracy, resiliency, reliability, safety, and accountability. This could significantly help to close the critical gap between deep learning model development and deployment into clinical practice. In order to gain trust, different stakeholders such as AI designers, medical doctors, reading centers, hospitals, patients, regulatory bodies, payers should work together closely from the development to integration phase of the deep learning model. Collectively, deep learning could be used to identify, process and define image-based biomarkers and help to better diagnose and differentiate ocular disease, as DL is able to seek for useful characteristics or pattern by themselves and is not limited to identify clinical features only.

3.5 Conclusion

In this PhD research study, deep neural network-based fully automatic segmentation and quantification methods were developed for MG, CNF, and DC. The developed techniques overcome current limitations of subjective image analysis and enable faster, precise, accurate, reliable, and reproducible ocular surface tissue characterization. These techniques have the potential to make an impact clinically and in the research environment by specifying faster disease diagnosis, facilitating new drug development, and standardizing clinical trials for a new drug of ocular surface disease related to DE. The developed automatic and objective image analysis methods could assist researchers and clinicians in the diagnosis of several ocular surface diseases including DED, neuropathic corneal pain, and reduce inter- or intra-observer variability and time to analyze a large volume of clinical images. Thus, the methods demonstrate that deep neural network-based systems provide automatic segmentation and morphometric

evaluation of MG, CNF, and DC and have the potential to be implemented into clinical practice in infrared meibography and IVCM images.

The author believes that the developed methods would contribute towards initializing a new phase of ocular surface tissue segmentation and analysis. It will allow both researchers and clinicians to analyze MG, CNF, and DC more reliably while reducing a significant amount of time to analyze the patient images. Finally, the methods developed in this research significantly increase the efficiency of evaluating clinical images, thereby supporting and potentially improving diagnosis and treatment of ocular surface disease related to dry eye.

References

[1] Rico-del-Viejo L, Benítez-del-Castillo JM, Gómez-Sanz FJ, García-Montero M, Llorens-Quintana C, Madrid-Costa D. The influence of meibomian gland loss on ocular surface clinical parameters. Contact Lens Anterior Eye 2019;42:562–8. https://doi.org/10.1016/j.clae.2019.04.004.

[2] Craig JP, Nelson JD, Azar DT, Belmonte C, Bron AJ, Chauhan SK, et al. TFOS DEWS II Report Executive Summary. Ocul Surf 2017;15:802–12. https://doi.org/10.1016/j.jtos.2017.08.003.

[3] Stapleton F, Alves M, Bunya VY, Jalbert I, Lekhanont K, Malet F, et al. TFOS DEWS II Epidemiology Report. Ocul Surf 2017;15:334–65. https://doi.org/10.1016/j.jtos.2017.05.003.

[4] Nations U. Department of Economic and Social Affairs, population division. Int Migr Rep 2015.

[5] McDonald M, Patel DA, Keith MS, Snedecor SJ. Economic and Humanistic Burden of Dry Eye Disease in Europe, North America, and Asia: A Systematic Literature Review. Ocul Surf 2016;14:144–67. https://doi.org/10.1016/j.jtos.2015.11.002.

[6] Uchino M, Schaumberg DA. Dry Eye Disease: Impact on Quality of Life and Vision. Curr Ophthalmol Rep 2013;1:51–7. https://doi.org/10.1007/s40135-013-0009-1.

[7] Stern ME, Gao J, Siemasko KF, Beuerman RW, Pflugfelder SC. The role of the lacrimal functional unit in the pathophysiology of dry eye. Exp Eye Res 2004;78:409–16. https://doi.org/10.1016/j.exer.2003.09.003.

[8] Craig JP, Nichols KK, Akpek EK, Caffery B, Dua HS, Joo CK, et al. TFOS DEWS II Definition and Classification Report. Ocul Surf 2017;15:276–83. https://doi.org/10.1016/j.jtos.2017.05.008.

[9] Nichols KK, Foulks GN, Bron AJ, Glasgow BJ, Dogru M, Tsubota K, et al. The international workshop on meibomian gland dysfunction: Executive summary. Investig Ophthalmol Vis Sci 2011;52:1922–9. https://doi.org/10.1167/iovs.10-6997a.

[10] Setu MAKMAK, Horstmann J, Schmidt S, Stern MEME, Steven P. Deep learning-based automatic meibomian gland segmentation and morphology assessment in infrared meibography. Sci Rep 2021;11:7649. https://doi.org/10.1038/s41598-021-87314-8.

[11] Bron AJ, Tiffany JM, Gouveia SM, Yokoi N, Voon LW. Functional aspects of the tear film lipid layer. Exp Eye Res 2004;78:347–60. https://doi.org/10.1016/J.EXER.2003.09.019.

[12] Knop E, Knop N, Millar T, Obata H, Sullivan DA. The international workshop on meibomian gland dysfunction: Report of the subcommittee on anatomy, physiology, and pathophysiology of the meibomian gland. Investig Ophthalmol Vis Sci 2011;52:1938–78. https://doi.org/10.1167/iovs.10-6997c.

[13] Knop N, Knop E. Meibom-Drüsen. Teil I: Anatomie, Embryologie und Histologie der Meibom-Drüsen. Ophthalmologe 2009;106:872–83. https://doi.org/10.1007/S00347-009-2006-1/FIGURES/8.

[14] Bron AJ, de Paiva CS, Chauhan SK, Bonini S, Gabison EE, Jain S, et al. TFOS DEWS II

pathophysiology report. Ocul Surf 2017;15:438–510. https://doi.org/10.1016/j.jtos.2017.05.011.

[15] Mathers WD. Ocular Evaporation in Meibomian Gland Dysfunction and Dry Eye. Ophthalmology 1993;100:347–51. https://doi.org/10.1016/S0161-6420(93)31643-X.

[16] Hamrah P, Huq SO, Liu Y, Zhang Q, Dana MR. Corneal immunity is mediated by heterogeneous population of antigen-presenting cells. J Leukoc Biol 2003;74:172–8. https://doi.org/10.1189/jlb.1102544.

[17] Labetoulle M, Baudouin C, Calonge M, Merayo-Lloves J, Boboridis KG, Akova YA, et al. Role of corneal nerves in ocular surface homeostasis and disease. Acta Ophthalmol 2019;97:137–45. https://doi.org/10.1111/AOS.13844.

[18] Oliveira-Soto L, Efron N. Morphology of corneal nerves using confocal microscopy. Cornea 2001;20:374–84. https://doi.org/10.1097/00003226-200105000-00008.

[19] Al-Fahdawi S, Qahwaji R, Al-Waisy AS, Ipson S, Malik RA, Brahma A, et al. A fully automatic nerve segmentation and morphometric parameter quantification system for early diagnosis of diabetic neuropathy in corneal images. Comput Methods Programs Biomed 2016;135:151–66. https://doi.org/10.1016/j.cmpb.2016.07.032.

[20] Müller LJ, Marfurt CF, Kruse F, Tervo TMT. Corneal nerves: Structure, contents and function. Exp Eye Res 2003;76:521–42. https://doi.org/10.1016/S0014-4835(03)00050-2.

[21] Dehghani C, Pritchard N, Edwards K, Russell AW, Malik RA, Efron N. Fully automated, semiautomated, and manual morphometric analysis of corneal subbasal nerve plexus in individuals with and without diabetes. Cornea 2014;33:696–702. https://doi.org/10.1097/ICO.0000000000000152.

[22] Cruzat A, Qazi Y, Hamrah P. In Vivo Confocal Microscopy of Corneal Nerves in Health and Disease. Ocul Surf 2017;15:15–47. https://doi.org/10.1016/j.jtos.2016.09.004.

[23] Khan A, Li Y, Ponirakis G, Akhtar N, Gad H, George P, et al. Corneal Immune Cells Are Increased in Patients With Multiple Sclerosis. Transl Vis Sci Technol 2021;10:19. https://doi.org/10.1167/tvst.10.4.19.

[24] Cavalcanti BM, Cruzat A, Sahin A, Pavan-Langston D, Samayoa E, Hamrah P. In vivo confocal microscopy detects bilateral changes of corneal immune cells and nerves in unilateral herpes zoster ophthalmicus. Ocul Surf 2018;16:101–11. https://doi.org/10.1016/j.jtos.2017.09.004.

[25] Zhivov A, Stave J, Vollmar B, Guthoff R. In vivo confocal microscopic evaluation of Langerhans cell density and distribution in the normal human corneal epithelium. Graefe's Arch Clin Exp Ophthalmol 2005;243:1056–61. https://doi.org/10.1007/s00417-004-1075-8.

[26] Kheirkhah A, Rahimi Darabad R, Cruzat A, Hajrasouliha AR, Witkin D, Wong N, et al. Corneal Epithelial Immune Dendritic Cell Alterations in Subtypes of Dry Eye Disease: A Pilot In Vivo Confocal Microscopic Study. Investig Opthalmology Vis Sci 2015;56:7179. https://doi.org/10.1167/iovs.15-17433.

[27] Goyal S, Hamrah P. Understanding neuropathic corneal pain - Gaps and current therapeutic approaches. Semin Ophthalmol 2016;31:59–70. https://doi.org/10.3109/08820538.2015.1114853.

[28] Stevenson W, Chauhan SK, Dana R. Dry Eye Disease: An Immune-Mediated Ocular Surface Disorder. Arch Ophthalmol 2012;130:90–100. https://doi.org/10.1001/ARCHOPHTHALMOL.2011.364.

[29] Belmonte C, Acosta MC, Merayo-Lloves J, Gallar J. What Causes Eye Pain? Curr Ophthalmol Rep 2015;3:111–21. https://doi.org/10.1007/s40135-015-0073-9.

[30] Moein HR, Akhlaq A, Dieckmann G, Abbouda A, Pondelis N, Salem Z, et al. Visualization of microneuromas by using in vivo confocal microscopy: An objective biomarker for the diagnosis of neuropathic corneal pain? Ocul Surf 2020;18:651–6. https://doi.org/10.1016/j.jtos.2020.07.004.

[31] Hamrah P, Cruzat A, Dastjerdi MH, Zheng L, Shahatit BM, Bayhan HA, et al. Corneal sensation and subbasal nerve alterations in patients with herpes simplex keratitis: An in vivo confocal microscopy study. Ophthalmology 2010;117:1930–6. https://doi.org/10.1016/j.ophtha.2010.07.010.

[32] Dieckmann G, Goyal S, Hamrah P. Neuropathic Corneal Pain: Approaches for Management. Ophthalmology 2017;124:S34–47. https://doi.org/10.1016/j.ophtha.2017.08.004.

[33] Pflipsen M, Massaquoi M, Wolf S. Evaluation of the Painful Eye. vol. 93. 2016.

[34] Benítez Del Castillo JM, Wasfy MAS, Fernandez C, Garcia-Sanchez J. An in vivo confocal masked study on corneal epithelium and subbasal nerves in patients with dry eye. Investig Ophthalmol Vis Sci 2004;45:3030–5. https://doi.org/10.1167/iovs.04-0251.

[35] Toda I, Asano-Kato N, Komai-Hori Y, Tsubota K. Dry eye after laser in situ keratomileusis. Am J Ophthalmol 2001;132:1–7. https://doi.org/10.1016/S0002-9394(01)00959-X.

[36] Gøransson LG, Herigstad A, Tjensvoll AB, Harboe E, Mellgren SI, Omdal R. Peripheral neuropathy in primary sjogren syndrome: a population-based study. Arch Neurol 2006;63:1612–5. https://doi.org/10.1001/ARCHNEUR.63.11.1612.

[37] Rosenthal P, Borsook D. Ocular neuropathic pain. Br J Ophthalmol 2016;100:128–34. https://doi.org/10.1136/bjophthalmol-2014-306280.

[38] Goyal S, Hamrah P. Understanding Neuropathic Corneal Pain-Gaps and Current Therapeutic Approaches. Semin Ophthalmol 2016;31:59. https://doi.org/10.3109/08820538.2015.1114853.

[39] Galor A, Zlotcavitch L, Walter SD, Felix ER, Feuer W, Martin ER, et al. Dry eye symptom severity and persistence are associated with symptoms of neuropathic pain. Br J Ophthalmol 2015;99:665–8. https://doi.org/10.1136/bjophthalmol-2014-306057.

[40] Tavakoli M, Marshall A, Pitceathly R, Fadavi H, Gow D, Roberts ME, et al. Corneal confocal microscopy: A novel means to detect nerve fibre damage in idiopathic small fibre neuropathy. Exp Neurol 2010;223:245. https://doi.org/10.1016/J.EXPNEUROL.2009.08.033.

[41] Birnbaum J. The nervous system in rheumatic disease. Rheumatol Sixth Ed 2015;1–2:298–305. https://doi.org/10.1016/B978-0-323-09138-1.00039-5.

[42] Bron AJ, Abelson MB, Ousler G, Pearce E, Tomlinson A, Yokoi N, et al. Methodologies to diagnose and monitor dry eye disease: report of the Diagnostic Methodology

Subcommittee of the International Dry Eye WorkShop (2007). Ocul Surf 2007;5:108–52. https://doi.org/10.1016/S1542-0124(12)70083-6.

[43] Sullivan BD, Crews LA, Messmer EM, Foulks GN, Nichols KK, Baenninger P, et al. Correlations between commonly used objective signs and symptoms for the diagnosis of dry eye disease: clinical implications. Acta Ophthalmol 2014;92:161–6. https://doi.org/10.1111/AOS.12012.

[44] Chan TCY, Wan KH, Shih KC, Jhanji V. Advances in dry eye imaging: the present and beyond. Br J Ophthalmol 2018;102:295–301. https://doi.org/10.1136/BJOPHTHALMOL-2017-310759.

[45] Wolffsohn JS, Arita R, Chalmers R, Djalilian A, Dogru M, Dumbleton K, et al. TFOS DEWS II Diagnostic Methodology report. vol. 15. 2017.

[46] Villani E, Bonsignore F, Cantalamessa E, Serafino M, Nucci P. Imaging Biomarkers for Dry Eye Disease. Eye Contact Lens 2020;46:S141–5. https://doi.org/10.1097/ICL.0000000000000650.

[47] Binotti WW, Bayraktutar B, Ozmen MC, Cox SM, Hamrah P. A Review of Imaging Biomarkers of the Ocular Surface. Eye Contact Lens 2020;46:S84. https://doi.org/10.1097/ICL.0000000000000684.

[48] Arita R, Itoh K, Inoue K, Amano S. Noncontact Infrared Meibography to Document Age-Related Changes of the Meibomian Glands in a Normal Population. Ophthalmology 2008;115:911–5. https://doi.org/10.1016/j.ophtha.2007.06.031.

[49] Pult H, Nichols JJ. A review of meibography. Optom Vis Sci 2012;89. https://doi.org/10.1097/OPX.0b013e3182512ac1.

[50] Mathers WD, Daley T, Verdick R. Video Imaging of the Meibomian Gland. Arch Ophthalmol 1994;112:448–9. https://doi.org/10.1001/ARCHOPHT.1994.01090160022008.

[51] Arita R, Itoh K, Maeda S, Maeda K, Furuta A, Fukuoka S, et al. Proposed diagnostic criteria for obstructive meibomian gland dysfunction. Ophthalmology 2009;116. https://doi.org/10.1016/J.OPHTHA.2009.04.037.

[52] Arita R, Itoh K, Maeda S, Maeda K, Tomidokoro A, Amano S. Efficacy of diagnostic criteria for the differential diagnosis between obstructive meibomian gland dysfunction and aqueous deficiency dry eye. Jpn J Ophthalmol 2010;54:387–91. https://doi.org/10.1007/S10384-010-0858-1.

[53] Arita R, Itoh K, Inoue K, Kuchiba A, Yamaguchi T, Amano S. Contact lens wear is associated with decrease of meibomian glands. Ophthalmology 2009;116:379–84. https://doi.org/10.1016/J.OPHTHA.2008.10.012.

[54] Srinivasan S, Menzies K, Sorbara L, Jones L. Infrared imaging of meibomian gland structure using a novel keratograph. Optom Vis Sci 2012;89:788–94.

[55] IMAGING MEIBOMIAN GLAND STRUCTURES USING THE OCULUS KERATOGRAPH n.d. https://www.aaopt.org/detail/knowledge-base-article/imaging-meibomian-gland-structures-using-oculus-keratograph (accessed January 2, 2022).

[56] Koh YW. Detection of meibomian glands and classification of meibography images. J Biomed Opt 2012;17:086008. https://doi.org/10.1117/1.jbo.17.8.086008.

[57] Llorens-Quintana C, Rico-Del-Viejo L, Syga P, Madrid-Costa D, Iskander DR. A novel automated approach for infrared-based assessment of meibomian gland morphology. Transl Vis Sci Technol 2019;8:17–17. https://doi.org/10.1167/tvst.8.4.17.

[58] Arita R, Suehiro J, Haraguchi T, Shirakawa R, Tokoro H, Amano S. Objective image analysis of the meibomian gland area. Br J Ophthalmol 2014;98:746–55. https://doi.org/10.1136/bjophthalmol-2012-303014.

[59] Celik T, Lee HK, Petznick A, Tong L. Bioimage informatics approach to automated meibomian gland analysis in infrared images of meibography. J Optom 2013;6:194–204. https://doi.org/10.1016/j.optom.2013.09.001.

[60] Koprowski R, Wilczyński S, Olczyk P, Nowińska A, Weglarz B, Wylegała E. A quantitative method for assessing the quality of meibomian glands. Comput Biol Med 2016;75:130–8. https://doi.org/10.1016/j.compbiomed.2016.06.001.

[61] Wu Y, Wang C, Wang X, Mou Y, Yuan K, Huang X, et al. Advances in Dry Eye Disease Examination Techniques. Front Med 2022;8:3048. https://doi.org/10.3389/FMED.2021.826530/BIBTEX.

[62] Villani E, Baudouin C, Efron N, Hamrah P, Kojima T, Patel S V., et al. In vivo confocal microscopy of the ocular surface: From bench to bedside. Curr Eye Res 2014;39:213–31. https://doi.org/10.3109/02713683.2013.842592.

[63] Niederer RL, McGhee CNJ. Clinical in vivo confocal microscopy of the human cornea in health and disease. Prog Retin Eye Res 2010;29:30–58. https://doi.org/10.1016/j.preteyeres.2009.11.001.

[64] Petroll WM, Robertson DM. In Vivo Confocal Microscopy of the Cornea: New Developments in Image Acquisition, Reconstruction, and Analysis Using the HRT-Rostock Corneal Module. Ocul Surf 2015;13:187–203. https://doi.org/10.1016/j.jtos.2015.05.002.

[65] Guthoff RF, Baudouin C, Stave J. Atlas of confocal laser scanning in-vivo microscopy in opthalmology: Principles and applications in diagnostic and therapeutic ophtalmology. Atlas Confocal Laser Scanning In-Vivo Microsc Ophthalmol Princ Appl Diagnostic Ther Ophthalmol 2006:1–200. https://doi.org/10.1007/3-540-32707-X.

[66] PETROLL WM, CAVANAGH HD, JESTER J V. Three-dimensional imaging of corneal cells using in vivo confocal microscopy. J Microsc 1993;170:213–9. https://doi.org/10.1111/J.1365-2818.1993.TB03344.X.

[67] Masters BR, Thaer AA. Real-time scanning slit confocal microscopy of the in vivo human cornea. Appl Opt 1994;33:695. https://doi.org/10.1364/AO.33.000695.

[68] Erie EA, McLaren JW, Kittleson KM, Patel S V., Erie JC, Bourne WM. Corneal subbasal nerve density: a comparison of two confocal microscopes. Eye Contact Lens 2008;34:322–5. https://doi.org/10.1097/ICL.0B013E31818B74F4.

[69] Dabbah MA, Graham J, Petropoulos IN, Tavakoli M, Malik RA. Automatic analysis of diabetic peripheral neuropathy using multi-scale quantitative morphology of nerve fibres in corneal confocal microscopy imaging. Med Image Anal 2011;15:738–47. https://doi.org/10.1016/j.media.2011.05.016.

[70] Chen X, Graham J, Dabbah MA, Petropoulos IN, Tavakoli M, Malik RA. An automatic tool for quantification of nerve fibers in corneal confocal microscopy images. IEEE Trans

Biomed Eng 2017;64:786–94. https://doi.org/10.1109/TBME.2016.2573642.

[71] Scarpa F, Grisan E, Ruggeri A. Automatic recognition of corneal nerve structures in images from confocal microscopy. Investig Ophthalmol Vis Sci 2008;49:4801–7. https://doi.org/10.1167/iovs.08-2061.

[72] Kim J, Markoulli M. Automatic analysis of corneal nerves imaged using in vivo confocal microscopy. Clin Exp Optom 2018;101:147–61. https://doi.org/10.1111/cxo.12640.

[73] Dabbah MA, Graham J, Petropoulos I, Tavakoli M, Malik RA. Dual-model automatic detection of nerve-fibres in corneal confocal microscopy images. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6361 LNCS, 2010, p. 300–7. https://doi.org/10.1007/978-3-642-15705-9_37.

[74] Bille PDJF. High Resolution Imaging in Microscopy and Ophthalmology. High Resolut Imaging Microsc Ophthalmol 2019. https://doi.org/10.1007/978-3-030-16638-0.

[75] Ting DSW, Peng L, Varadarajan A V., Keane PA, Burlina PM, Chiang MF, et al. Deep learning in ophthalmology: The technical and clinical considerations. vol. 72. 2019.

[76] Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. vol. 103. 2019.

[77] Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. Prog Retin Eye Res 2018;67:1–29. https://doi.org/10.1016/J.PRETEYERES.2018.07.004.

[78] Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. Ophthalmology 2017;124:962–9. https://doi.org/10.1016/j.ophtha.2017.02.008.

[79] Grewal PS, Oloumi F, Rubin U, Tennant MTS. Deep learning in ophthalmology: a review. Can J Ophthalmol 2018;53:309–13. https://doi.org/10.1016/j.jcjo.2018.04.019.

[80] Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys 2019;29:102–27. https://doi.org/10.1016/j.zemedi.2018.11.002.

[81] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44. https://doi.org/10.1038/nature14539.

[82] Shen D, Wu G, Suk H Il. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng 2017;19:221–48. https://doi.org/10.1146/annurev-bioeng-071516-044442.

[83] Jiang Z, Zhang H, Wang Y, Ko SB. Retinal blood vessel segmentation using fully convolutional network with transfer learning. Comput Med Imaging Graph 2018;68:1–15. https://doi.org/10.1016/j.compmedimag.2018.04.005.

[84] Mishra Z, Ganegoda A, Selicha J, Wang Z, Sadda SVR, Hu Z. Automated Retinal Layer Segmentation Using Graph-based Algorithm Incorporating Deep-learning-derived Information. Sci Rep 2020;10:1–8. https://doi.org/10.1038/s41598-020-66355-5.

[85] Sreng S, Maneerat N, Hamamoto K, Win KY. Deep Learning for Optic Disc Segmentation and Glaucoma Diagnosis on Retinal Images. Appl Sci 2020;10:4916. https://doi.org/10.3390/app10144916.

[86] Fu H, Baskaran M, Xu Y, Lin S, Wong DWK, Liu J, et al. A Deep Learning System for Automated Angle-Closure Detection in Anterior Segment Optical Coherence Tomography Images. Am J Ophthalmol 2019;203:37–45. https://doi.org/10.1016/j.ajo.2019.02.028.

[87] Hsieh YT, Chuang LM, Jiang Y Der, Chang TJ, Yang CM, Yang CH, et al. Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening. J Formos Med Assoc 2021;120:165–71. https://doi.org/10.1016/J.JFMA.2020.03.024.

[88] Prabhu SM, Chakiat A, S S, Vunnava KP, Shetty R. Deep learning segmentation and quantification of Meibomian glands. Biomed Signal Process Control 2020;57:101776. https://doi.org/10.1016/j.bspc.2019.101776.

[89] Wang J, Yeh TN, Chakraborty R, Yu SX, Lin MC. A deep learning approach for meibomian gland atrophy evaluation in meibography images. Transl Vis Sci Technol 2019;8:37–37. https://doi.org/10.1167/tvst.8.6.37.

[90] Maruoka S, Tabuchi H, Nagasato D, Masumoto H, Chikama T, Kawai A, et al. Deep Neural Network-Based Method for Detecting Obstructive Meibomian Gland Dysfunction With in Vivo Laser Confocal Microscopy. Cornea 2020;39:720–5. https://doi.org/10.1097/ICO.0000000000002279.

[91] Williams BM, Borroni D, Liu R, Zhao Y, Zhang J, Lim J, et al. An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study. Diabetologia 2020;63:419–30. https://doi.org/10.1007/s00125-019-05023-4.

[92] Wei S, Shi F, Wang Y, Chou Y, Li X. A Deep Learning Model for Automated Sub-Basal Corneal Nerve Segmentation and Evaluation Using In Vivo Confocal Microscopy. Transl Vis Sci Technol 2020;9:32. https://doi.org/10.1167/tvst.9.2.32.

[93] Colonna A, Scarpa F, Ruggeri A. Segmentation of Corneal Nerves Using a U-Net-Based Convolutional Neural Network. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2018;11039 LNCS:185–92. https://doi.org/10.1007/978-3-030-00949-6_22.

[94] Oakley JD, Russakoff DB, McCarron ME, Weinberg RL, Izzi JM, Misra SL, et al. Deep learning-based analysis of macaque corneal sub-basal nerve fibers in confocal microscopy images. Eye Vis 2020 71 2020;7:1–11. https://doi.org/10.1186/S40662-020-00192-5.

[95] Yıldız E, Arslan AT, Taş AY, Acer AF, Demir S, Şahin A, et al. Generative Adversarial Network Based Automatic Segmentation of Corneal Subbasal Nerves on In Vivo Confocal Microscopy Images. Transl Vis Sci Technol 2021;10:33–33. https://doi.org/10.1167/TVST.10.6.33.

Acknowledgements

First of all, I want to thank my PhD supervisor and mentor **Prof. Dr. med. Philipp Steven**. I am truly honored for his extensive and continuous support, guidance, and suggestions, both scientific and non-scientific, throughout my PhD journey. I am grateful for his trust, his kindness, and his great interest in my professional and personal growth. I am lucky to be a PhD student under his direct supervision and be a part of his talented and diverse research group.

I also want to thank my two IPHS tutors **Prof. Dr. Achim Tresch** and **Prof. Dr. Joseph Kambeitz** for their special interest in my research work and productive discussions during our tutor meetings. I am really happy to get the opportunity to learn from such great and humble scientist.

Furthermore, I want to thank my PhD secondment supervisor **Dr. Stefan Schmidt** (Heidelberg Engineering GmbH) for his significant support during and after my secondment. Also thank other team members from R&D and Product Management team for their warm welcome and support during my stay at Heidelberg Engineering.

Also, many thanks to all of our current and former ocular surface group lab members, namely: **Dr. Jens Hosrtmann, Dr. Uta Gehlsen, Dr. Gwen Musial, Dr. Martina Maaß, Bao Tran, Daniela Heß. Sabine Manz, Anna-Lena Zachert, Timur Bilgin, Volkan Tahmaz**.

Many thanks to all of our EU Horizon 2020 ITN (IT-DED^3) colleagues, PIs, and coordinators to arrange great and knowledgeable training courses, workshops, summer/winter schools which greatly enhance both academic and professional growth.

Special thanks to my parents who always supported me and most importantly, I want to thank my beloved wife, **Bodrunnesa Tutia**, for her great and constant support from the beginning to the end of this PhD journey.

Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen - einschließlich Tabellen, Karten und Abbildungen -, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, sind in jedem Einzelfall als Entlehnung kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertationsschrift noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss der Promotion nicht ohne Genehmigung der / des Vorsitzenden des IPHS-Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Philipp Steven betreut worden.

Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

Übersicht der Publikationen:

Md Asif Khan Setu, Jens Horstmann, Stefan Schmidt, Michael E. Stern, Philipp Steven, “Deep Learning-based Automatic Meibomian Gland Segmentation and Morphology Assessment in Infrared Meibography”, Scientific Reports 11, 7649 (2021) https://doi.org/10.1038/s41598-021-87314-8

Md Asif Khan Setu, Stefan Schmidt, Gwen Musial, Michael E. Stern, Philipp Steven, “Segmentation and Evaluation of Corneal Nerves and Dendritic Cells from In Vivo Confocal Microscopy Images using Deep Learning”, Translational Vision Science and Technology Vol. 11, 24 (2022) https://doi.org/10.1167/tvst.11.6.24

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem IPHS-Promotionsausschuss unverzüglich mitzuteilen.

24.11.2022

Datum Unterschrift

Curriculum Vitae

Personal data

Md Asif Khan Setu

Department of Ophthalmology

University Hospital Cologne

50931 Cologne, Germany

Email: setu.asif@gmail.com

Education

Sep. 2018 – Sep. 2022	PhD in Interdisciplinary Program Health Science (IPHS) University of Cologne, Cologne, Germany
Sep. 2016 – Aug. 2018	M.Sc. in Information Technology (major: Computer Science) University of Eastern Finland, Joensuu, Finland
Jun. 2007 – Mar. 2012	B.Sc. in Computer Science and Engineering Bangladesh University of Engineering and Technology

Scientific Career

Oct. 2021 – Aug. 2022	Postdoctoral researcher Data Science for Bioimages Center for Molecular Medicine Cologne (CMMC) University Hopsital Cologne
Sep. 2018 – Sep. 2021	EU Researcher (PhD Student) Ocular Surface Group Department of Ophthalmology University Hospital Cologne
Sep. 2019 – Jun. 2020	Industrial Secondment Research and Development Team Heidelberg Engineering GmbH

May. 2017 – Aug. 2018	Research Assistant Computational Spectral Imaging Group School of Computing University of Eastern Finland
Oct. 2017 – Feb. 2018	International Visiting Research Student Visual Perception & Cognition Group Department of Computer Science Toyohashi University of Technology, Japan

Professional Experience

Oct. 2012 – Aug. 2016	Software Engineer IQVIA Asia Pacific, Dhaka, Bangladesh

Scholarships and Awards

Mar. 2021	Association for Research in Vision and Ophthalmology (ARVO) International Travel Grant for the ARVO 2021 Annual Meeting, USA
Sep. 2018	Marie Sklodowska-Curie Actions (MSCA) PhD fellowship as ESR 11 from European Union H2020-MSCA-ITN program IT-DED3
Oct. 2017	International Visiting Research Students fellowship from Japan Student Services Organization (JASSO)
Sep. 2016	M.Sc. scholarship from University of Eastern Finland as one of the top 5 internationals selecting master's degree student
Jun. 2011	Achieved 1st position in "SPAANDANB Entrepreneurship Contest 2011", a national level entrepreneurship contest, Bangladesh

Publications

Peer reviewed journals:

Md Asif Khan Setu, Jens Horstmann, Stefan Schmidt, Michael E. Stern, Philipp Steven, "Deep Learning-based Automatic Meibomian Gland Segmentation and Morphology Assessment in Infrared Meibography", Scientific Reports 11, 7649 (2021)

Md Asif Khan Setu, Stefan Schmidt, Gwen Musial, Michael E. Stern, Philipp Steven, “Segmentation and Evaluation of Corneal Nerves and Dendritic Cells from In Vivo Confocal Microscopy Images using Deep Learning”, Translational Vision Science and Technology Vol. 11, 24 (2022)

Conference abstracts:

Md Asif Khan Setu, Jens Horstmann, Michael E. Stern, Philipp Steven, “Automated analysis of Meibography images: comparison between intensity, region growing and deep learning-based methods”, Abstractband DOG 2019. Ophthalmologe 116, 25–218 (2019)

Md Asif Khan Setu, Stefan Schmidt, Gwen Musial, Volkan Tahmaz, Michael E. Stern, Philipp Steven, “Deep Learning-based Segmentation and Quantification of Nerve Fibers and Dendritic cells in Confocal Microscopy of the Cornea”, Investigative Ophthalmology & Visual Science, Vol.62, 2142 (2021)

Md Asif Khan Setu, Stefan Schmidt, Gwen Musial, Michael E. Stern, Philipp Steven, “Automatic corneal dendritic cell density and size evaluation in confocal micros copy images using deep learning”, European Dry Eye Society (EuDES) Congress 2021, Paris, France

Gwen Musial, **Md Asif Khan Setu**, Uta Gehlsen, Timur Bilgin, Anna Lina Zachert, Philipp Steven, “Automatic segmentation of corneal vasculature from cornea whole mount images”, European Dry Eye Society (EuDES) Congress 2021, Paris, France

scientific reports

Check for updates

OPEN

Deep learning-based automatic meibomian gland segmentation and morphology assessment in infrared meibography

Md Asif Khan Setu[1,2], Jens Horstmann[1], Stefan Schmidt[4], Michael E. Stern[1,2,3] & Philipp Steven[1,2]✉

Meibomian glands (MG) are large sebaceous glands located below the tarsal conjunctiva and the abnormalities of these glands cause Meibomian gland dysfunction (MGD) which is responsible for evaporative dry eye disease (DED). Accurate MG segmentation is a key prerequisite for automated imaging based MGD related DED diagnosis. However, Automatic MG segmentation in infrared meibography is a challenging task due to image artifacts. A deep learning-based MG segmentation has been proposed which directly learns MG features from the training image dataset without any image pre-processing. The model is trained and evaluated using 728 anonymized clinical meibography images. Additionally, automatic MG morphometric parameters, gland number, length, width, and tortuosity assessment were proposed. The average precision, recall, and F1 score were achieved 83%, 81%, and 84% respectively on the testing dataset with AUC value of 0.96 based on ROC curve and dice coefficient of 84%. Single image segmentation and morphometric parameter evaluation took on average 1.33 s. To the best of our knowledge, this is the first time that a validated deep learning-based approach is applied in MG segmentation and evaluation for both upper and lower eyelids.

Dry eye disease (DED) is one of the pervasive diseases of the ocular surface due to its multifactorial nature where tear film instability, hyperosmolarity, neurosensory abnormalities, ocular surface inflammation, ocular surface damage, and meibomian gland dysfunction (MGD) play etiological roles[1–4]. DED affects visual acuity and causes ocular discomfort and other symptoms, leading to changes in the quality of life[2,5].

MGD is a chronic disease and caused by diffuse abnormalities of meibomian glands, terminal duct obstruction, and changes in the glandular secretion. Obstructive MGD is a common cause of evaporative dry eye and lipid layer deficiency[5–7]. Infrared (IR) Meibography is a well-established non-contact optical imaging technique, which uses IR illumination to depict MG morphology by examining the everted eyelid[7]. Clinically, it is widely accepted, and also recommended to image and quantify MG during dry-eye examination. Meibography is non-invasive, provides large image areas, detailed morphometric information of MG, and is easy to operate for clinicians and technicians.

Several grading schemes and methods to analyze MG have been published by Pult et al.[8], Srinivasan et al.[9], Arita et al.[10], and Engel et al.[11] However, these MG segmentation methods are subjective or semi-automatic and require user interaction, which is laborious, time-consuming and non-repeatable. In previous research[8–11], meibography images have been segmented and quantified using ImageJ (National Institute of Health; http://imagej.nih.gov/ij) software where clinicians need to involve manually to identify the MG. In addition, different clinicians identify glands differently, which causes inter-observer variability[12].

A reliable automatic MG segmentation technique may overcome the difficulties of manual image segmentation, as infrared meibography images often contain various artifacts such as low contrast, non-uniform illumination, defocus gland area, or specular reflections which make image segmentation more challenging[13]. Until now, Koh et al.[12], Llorens-Quintana et al.[13], Arita et al.[14], Celik et al.[15], and Koprowski et al.[16] have proposed automatic MG segmentation methods. However, all of these methods rely on intensity-thresholding based image

[1]Department of Ophthalmology, Faculty of Medicine, University Hospital Cologne, University of Cologne, 50937 Cologne, Germany. [2]Division of Dry Eye and Ocular GvHD, University Hospital Cologne, 50937 Cologne, Germany. [3]ImmunEyez LLC, Irvine, CA, USA. [4]Heidelberg Engineering GmbH, 69115 Heidelberg, Germany. ✉email: philipp.steven@uk-koeln.de

segmentation which performance heavily depends on image quality. At the same time, it is quite challenging for examiners to repeatedly acquire sufficient quality images.

In contrast to intensity-thresholding based image segmentation methods, a deep learning-based method has the advantages of learning useful features and representations from the raw images automatically[17] and could thus overcome the above-mentioned restrictions. In a previous study[18], we have compared intensity-thresholding, region growing, and deep learning method to automatic analysis of IR meibography images. Among these three implemented approaches, we have demonstrated that deep learning could produce high quality and reliable results for the challenging task of automated IR meibography image segmentation and quantification. In the past years, deep learning has been paid increasing attention in the field of ophthalmology[19–22]. Using a neural network and a high volume of image data, a deep learning algorithm can learn to detect specific objects from the images. Nowadays, deep learning algorithms perform well due to increased computational power and image data. U-net, a state-of-the-art biomedical image segmentation method was first introduced by Ronneberger et al.[23] The algorithm provides high performance while requiring less training image data and gaining more accuracy than the conventional neural network.

In this research, an automatic MG segmentation method is proposed based on U-net. The detailed deep learning model training procedure and application to new images is illustrated in supplementary Fig. S1.

Unlike previous image segmentation techniques whose performance heavily depends on the image quality, the proposed deep learning-based MG segmentation method directly learns the MG features from the training images, which does not require any pre-processing such as artifact removal or image enhancement. It can automatically segment MG on new images. We demonstrate the good performance of the proposed technique by assessing various evaluation metrics.

Materials and methods

Meibography data collection. In total 728 anonymized clinical infrared meibography images of both upper (398) and lower (330) eyelids of adults male and female humans (age > 18 years) were randomly collected using the database of an Oculus Keratograph 5 M (Oculus GmbH, Wetzlar, Germany) device at the Eye Hospital, Department of Ophthalmology, University Hospital Cologne, Germany. The image collection adhered to the tenets of the Declaration of Helsinki and was approved by the Ethics and Institutional review board (IRB) from the University of Cologne—Germany (approval number #16-045). As this research study was conducted retrospectively using anonymized non-biometric data, the need for participant's informed consent was waived by the Ethics committee and IRB.

Data annotations. All collected clinical meibography images were manually annotated using Fiji (fiji.sc), which is an open-source image processing software, with the Segmentation Editor (https://imagej.net/Segmentation_Editor) plugin to generate ground truth masks. In the segmentation editor plugin window, the Exterior was set to black, and the Interior was set to white to generate a binary ground truth mask. One of the authors (M.A.K.S.) created a polygon boundary around the glands and saved the annotation mask images as JPG file format. Ground truth masks were generated for both upper and lower eyelids. Examples of training images and their corresponding masks of upper and lower eyelids are presented in Fig. 1. Finally, all annotated ground truth mask images were verified and corrected by an experienced senior ophthalmologist (P.S.) from the University Hospital Cologne before they were used in deep learning model training and testing.

Data preparation and allocation. From the overall collected 728 anonymized clinical meibography images, 100 were prepared as a test dataset to measure the true performance of our trained model. The remaining 628 images were randomly divided into training and validation datasets. Approximately, 80% (502 images) and 20% (126 images) of image data were used to train and validate respectively the deep learning model. Due to variations in the size of data set images, all images were resized to 256 × 256 pixels (width × height). Training images were used for deep learning model training and validation images were used to fine-tune the hyperparameters such as epochs, learning rate, momentum, validation steps.

Deep learning model design and training. We have applied transfer learning with a pre-trained backbone to increase the learning efficiency of the deep learning model. In this research work, Inception-ResNet-v2[24], a pre-trained backbone based on a publicly available ImageNet[25] dataset was used for transfer learning. In the beginning, we pre-trained our model with the Montgomery Chest X-ray images[26] for five epochs, which helps transfer learning performance as such medical image data is closer to our target application than the original ImageNet dataset. There are 138 publicly available X-ray images acquired by the Department of Health and Human Services of Montgomery County, MD, USA. The approach is to pre-train the baseline U-Net model to increase learning efficiency. After that, we re-trained the U-net model again using the meibography images and their respective masks. To optimize the deep learning model, both cross-entropy (CE) and Dice loss were used. The CE loss, Dice loss, and total loss were computed by the following equations,

$$\mathrm{CE} = \sum_{x \in \Omega} w(x) g_l(x) \log\left(p_l(x)\right) \tag{1}$$

$$Dice = 1 - \frac{2\Sigma_{x \in \Omega} p_l(\mathrm{x}) g_l(\mathrm{x})}{\Sigma_{x \in \Omega} g_l^2(\mathrm{x}) + \Sigma_{x \in \Omega} p_l^2(\mathrm{x})} \tag{2}$$

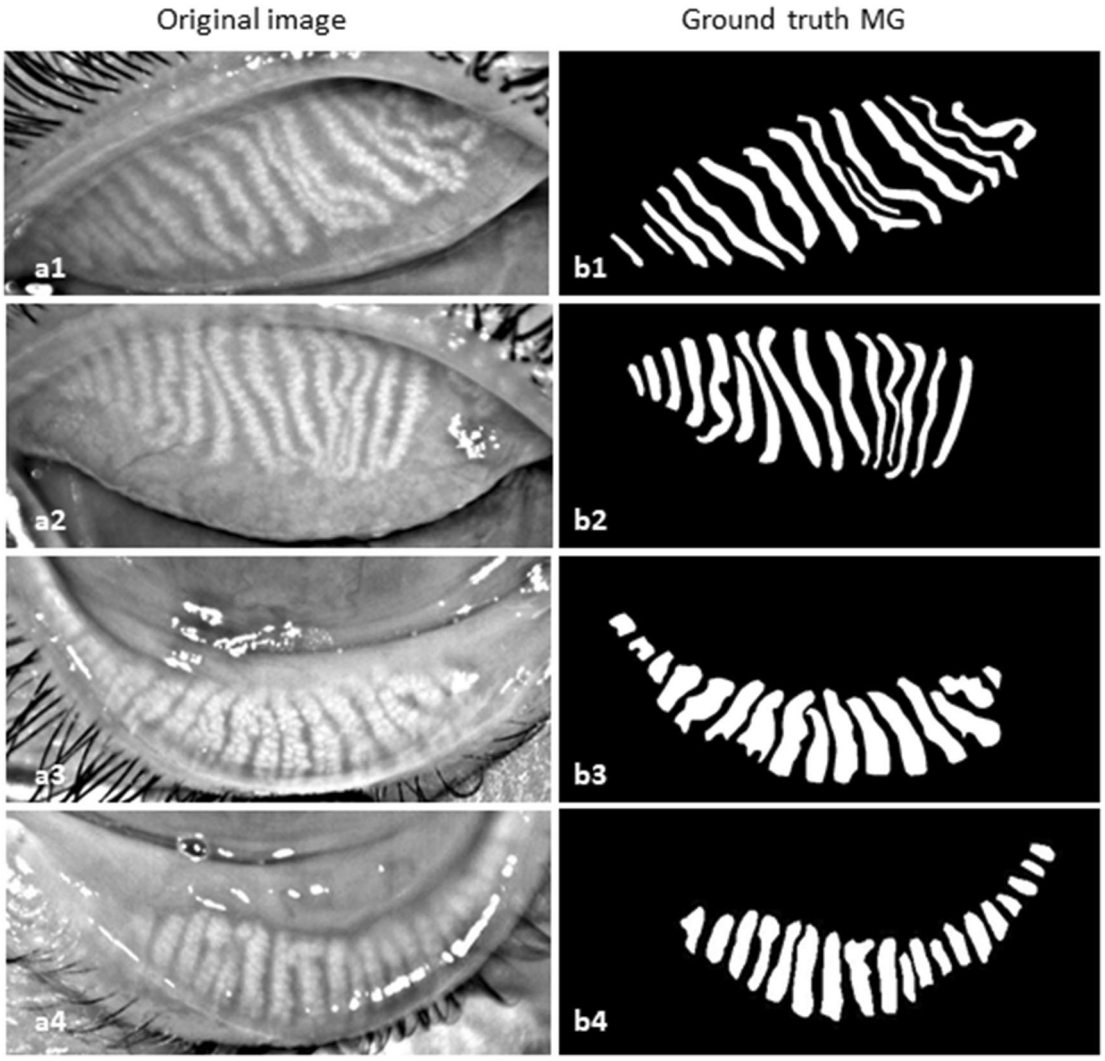

Figure 1. Example of training images and their corresponding ground truth images. (**a1–a2**) anonymized upper eyelid images, (**a3–a4**) anonymized lower eyelid images, (**b1–b2**) upper eyelid ground truth images, and (**b3–b4**) lower eyelid ground truth images.

$$Total\ loss = CE + Dice \tag{3}$$

where w(x) is the weight assigned to the pixel $x \in \Omega$, $g_l(x)$ is the ground truth pixel for layer l, $p_l(x)$ is the segmented pixel for layer l and total loss is the summation of the cross-entropy and dice loss.

Adam[27], a gradient-based stochastic optimizer that is one of the most efficient optimization algorithms for deep learning model optimization was also used to optimize the deep learning model.

Cross-validation study. To evaluate a deep learning model k-Fold cross-validation is generally used. In this research, k = 5 was applied. The total 628 images were randomly split into 5 sub-groups and each time four groups were selected as a training dataset and one group was selected as a validation dataset. In total 502 and 126 images were used as training and validation datasets respectively on every fold for training the model.

Evaluation metrics. For the MG segmentation task in this research, the common evaluation metrics Precision (P), Recall (R), and F1 score have been used. The MG pixels (e.g. white pixels in the binary segmented masks) are considered as positive instances. Based on the combination of the ground truth masks and segmented masks, these pixels are categorized into four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Then we can interpret Precision, and Recall in the following equations:

$$P = \frac{TP}{TP + FP} \tag{4}$$

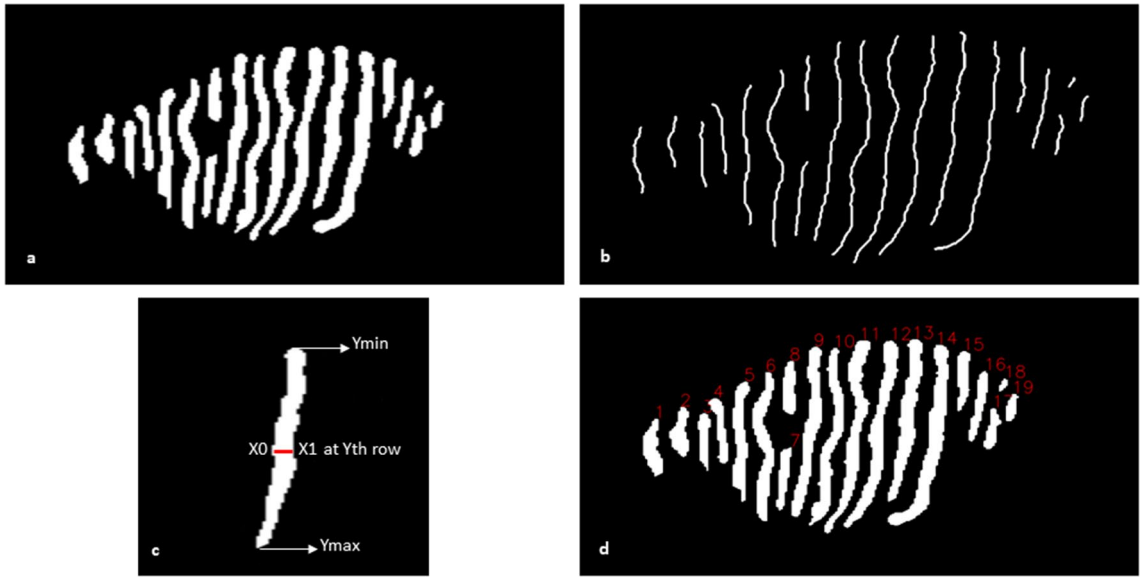

Figure 2. Morphometric parameters assessment of MG. (**a**) Binary segmented image, (**b**) skeletonized MGs of one-pixel width, (**c**) Individual gland width measurement. The Red line indicates the leftmost white pixel (X_0) and rightmost white pixel (X_1) at the Y_{th} row, and (**d**) Number of each MG is presented on top of each gland.

$$R = \frac{TP}{TP + FN} \tag{5}$$

F1 score is interpreted as a weighted average of precision and recall and based on the following equation:

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{6}$$

Morphometric parameters assessment. To better analyze the MG's morphology, gland number, length, width, and tortuosity were measured. These morphometric parameters were computed directly from the output of a deep learning-based binary segmentation image. To calculate the morphometric parameters, each segmented gland was analyzed individually. However, to generalize these morphometric parameters, the mean gland length, width, and tortuosity were calculated for an individual eyelid.

To compute the individual gland width, the difference of the leftmost white pixel (X_0) and the rightmost white pixel (X_1) at each row (Y) within the range of the length of the gland (Y_{min}, Y_{max}) were measured and then averaged (Fig. 2c). The number of MGs were computed based on the total number of separated white objects in the binary segmented image and put the number on top of each MG (Fig. 2d).

To compute the length and tortuosity of the gland, first, the binary-segmented image (Fig. 2a) was skeletonized (Fig. 2b) into one-pixel width using a custom-made python-based software. Then the path length (distance over the gland's skeletonized path) and the chord length (Euclidean distance between the first and last pixel of an individual skeletonized gland) were calculated. Finally, tortuosity τ for each gland g was measured as:

$$\tau(g) = \frac{Path\ length}{Chord\ length} \tag{7}$$

The deep learning model training, validation, testing, data preparation, and morphometric parameters assessment were conducted on a laptop running on Windows 10 Professional, 64-bit Intel Core i7-9750H CPU @ 2.6 GHz with 12 MB of cache memory, SSD 512 GB M.2 Samsung 970 Pro PCIe 3.0 × 4 NVMe, RAM 32 GB DDR4 @ 2666 MHz and NVIDIA GeForce RTX 2070 Max-Q with 8 GB GDDR6 of memory. Data preparation, deep learning model design, and training, morphometric parameters assessment software was written in Python (version: 3.6.6) using Keras (version: 2.2.5) with TensorFlow (TensorFlow version: 1.4, CUDA version: 10.0, cuDNN version: 7.6.3) in the backend.

Test–retest reliability. To determine the variability and repeatability (test–retest reliability)[28,29] of the developed algorithm, we additionally have compared two manual annotations by the same observer with two automatic segmentations, using the same test dataset of 100 images (50 upper and 50 lower eyelids). The same observer (M.A.K.S.) manually re-annotated individual MG of the test dataset. Dice similarity coefficient (DSC)[30], Cohen's kappa coefficient[31], and inter-class correlation coefficient (ICC)[32] were used to measure the variability and repeatability between manual annotations and between automatic segmentations of the test dataset.

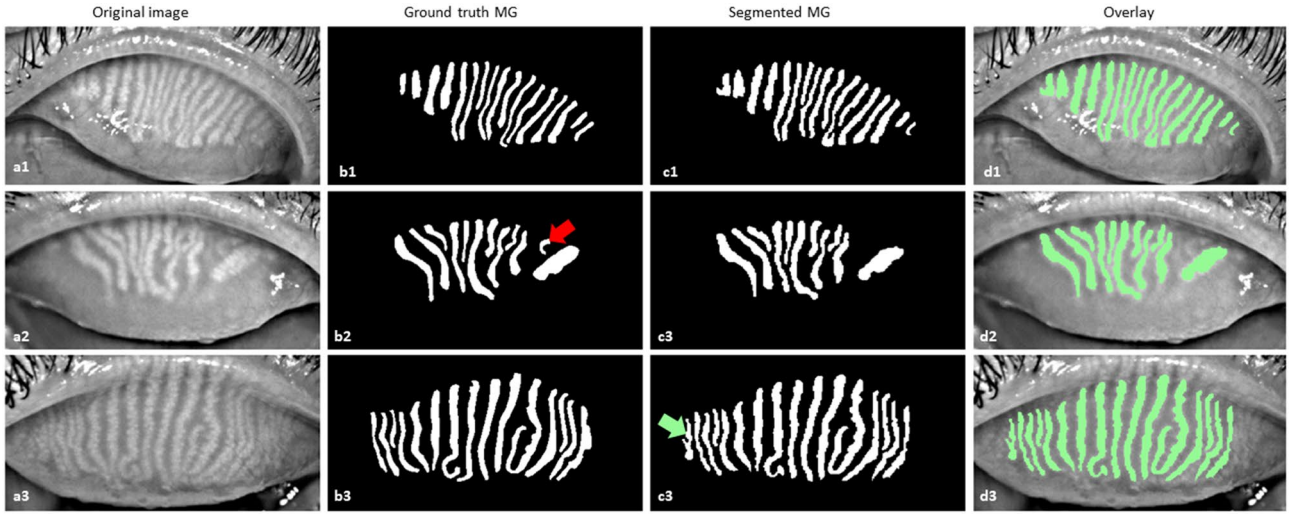

Figure 3. Example of MG segmentation from upper eyelid on test images. (**a1**–**a3**) Original images, (**b1**–**b3**) ground truth MG, (**c1**–**c3**) segmented MG, and (**d1**–**d3**) overlay of the original image and segmented MG. The red (**b2**) and green (**c3**) arrows indicate the missing gland and false gland segmentation respectively.

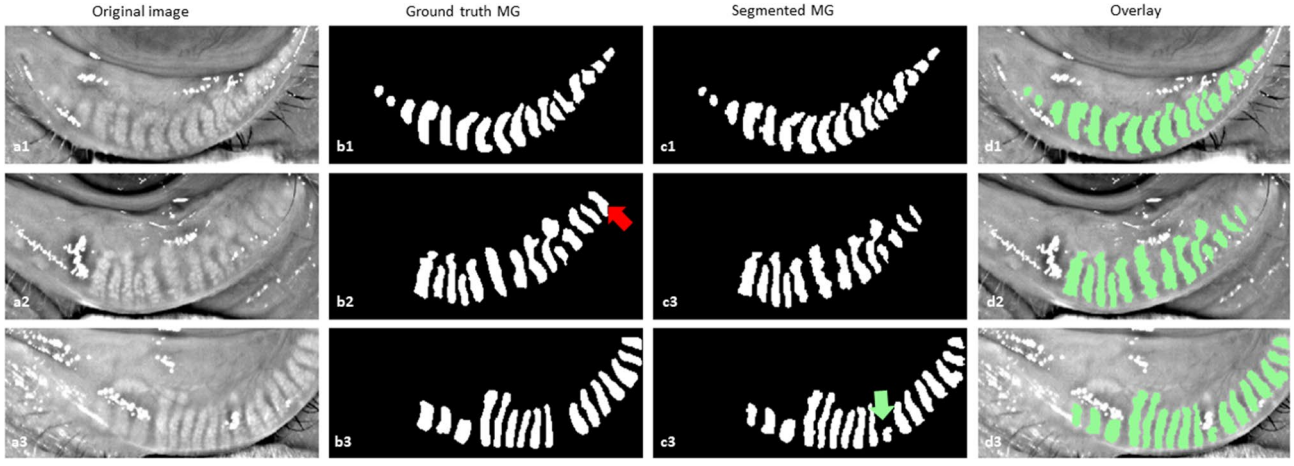

Figure 4. Example of MG segmentation from lower eyelid on test images. (**a1**–**a3**) original images, (**b1**–**b3**) ground truth MG, (**c1**–**c3**) Segmented MG, and (**d1**–**d3**) overlay of the original image and segmented MG. The red (**b2**) and green (**c3**) arrows indicate the missing gland and false gland segmentation respectively.

Statistics. The deep learning-based MG segmentation method was compared with the manually segmented MGs. The performance of the deep learning-based segmentation was measured using the Bland–Altman method[33]. Open-source program Python (version 3.6.6) based SciPy (version 1.5.2) library was used to perform the statistical analysis.

Results

Meibomian glands segmentation. MG segmentation model training for 30 epochs with a batch size of 5 took approximately 8 h and 24 min using the above described hard- and software. To segment and evaluate all 100 testing images, took 1 min 58 s on average 1.33 s per image. The average precision, recall, and F1 score were achieved 83%, 81%, and 84% respectively on the testing dataset with an AUC value of 0.96 based on the ROC curve (Supplementary Fig. S2) and dice coefficient of 84%. Examples of upper and lower eyelid testing images segmentation are presented in Figs. 3 and 4.

In general, MGs were reliably segmented in all testing images of both upper and lower eyelids using the trained deep learning model. However, the glands were not accurately segmented in all-test images. In some cases, two glands were connected which need to appear separately and the single gland was divided into two, which need to be a continuous gland. Examples of segmented MGs where glands are connected and separated are illustrated in Fig. 5. Among the 100 upper (50) and lower (50) eyelid test images, 39 upper and 37 lower eyelid images were segmented correctly. One MG was missing in 6 upper and 7 lower eyelid images while two MGs missing in only one lower eyelid image. Furthermore, one MG was falsely segmented in 4 upper and 5 lower eyelid images while two MGs falsely segmented only one upper eyelid image.

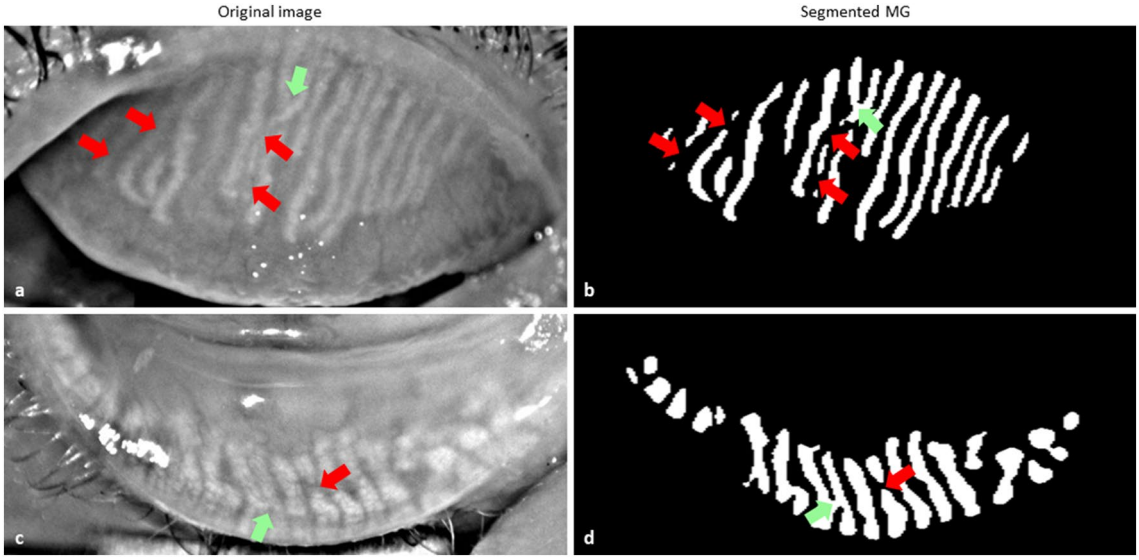

Figure 5. Example of MG segmentation where MG did not segment accurately. (**a**,**c**) Original images. (**b**,**d**) Segmented MGs. Green arrows indicate the MGs where glands need to appear separate and red arrows indicate where MGs need to become continuous.

Upper eyelid parameters	Ground truth MG	Segmented MG	*p* value
Number of glands	15.02 ± 2.68	15.00 ± 2.61	0.79
Mean gland length	269.87 ± 188.78	249.20 ± 181.65	0.58
Mean gland width	18.14 ± 2.95	19.20 ± 2.81	0.07
Mean tortuosity	1.74 ± 0.95	1.68 ± 0.91	0.18

Table 1. Morphometric parameters analysis of upper eyelid. *p* value is between ground truth versus segmented MGs.

Lower eyelid parameters	Ground truth MG	Segmented MG	*p* value
Number of glands	15.41 ± 3.18	15.32 ± 3.19	0.32
Mean gland length	105.60 ± 90.21	99.96 ± 87.36	0.09
Mean gland width	22.33 ± 5.40	22.66 ± 5.21	0.04
Mean tortuosity	1.78 ± 0.94	1.83 ± 0.94	0.61

Table 2. Morphometric parameters analysis of lower eyelid. *p* value is between ground truth versus segmented MGs.

Morphometric parameters assessment. The morphometric parameters MGs number, length, width, and tortuosity were computed from the binary-segmented images of test dataset. These are important clinical parameters to analysis the MG status. Our developed Python-based software provides automatic morphometric parameters assessment from the binary segmented image. All morphometric parameters of the binary segmented images of test dataset for both upper and lower eyelids were compared with the manually annotated images and the *p* value (using paired t-test) between ground truth and automatic segmentation is also presented. The results are tabulated in Tables 1 and 2.

To determine the consistency of the automatic MGs segmentation and manual annotation, Bland–Altman analysis was performed for all morphometric parameters. The statistical plots for upper eyelid are presented in Fig. 6. The total 50 upper eyelid testing images were used for this analysis. The middle solid line indicates the mean of manual and segmented MG which are 0.02, 20.67, − 1.06 and 0.07 for gland number, length, width and tortuosity respectively. All mean values are close to zero, thus the ground truth and automatic segmentation are not significantly different ($p > 0.005$). The two dotted lines indicate the limit of agreement (+ 1.96 SD) and the values were in between 1.13 and − 1.09 for gland number, 88.45 and − 47.11 for length, 1.27 and − 3.39 for width and 0.72 and − 0.59 for tortuosity. The gray bar indicates the confidence interval of 95%. Among the all 50 test images only 1, 2, 0, and 3 values were outside the limit of agreement thus the 98%, 96%, 100%, and 94% of the values are within the limit of agreement.

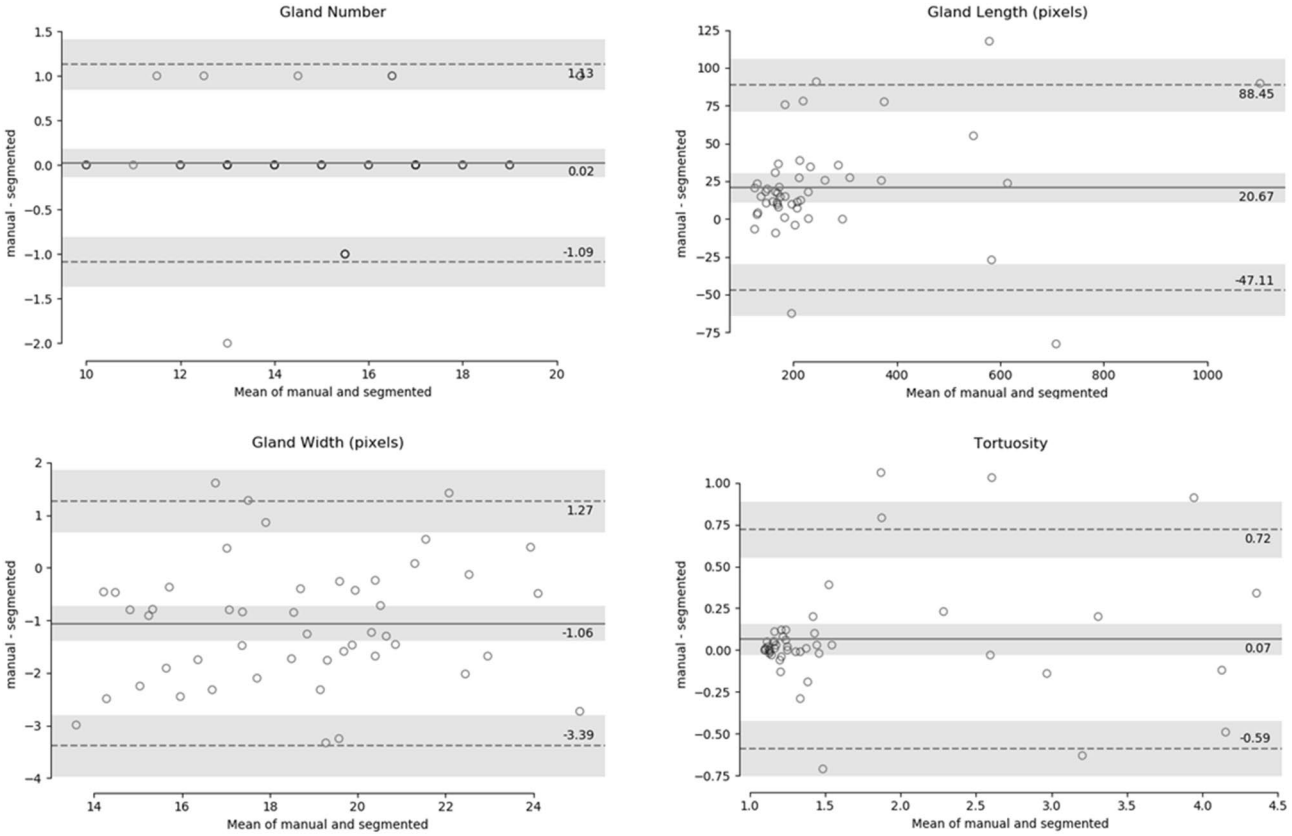

Figure 6. Bland–Altman plots of MGs number, length, width, and tortuosity between ground truth and deep learning segmentation for upper eyelid. The middle solid line indicates the mean value, the two dotted lines indicate the limit of agreement (± 1.96 SD), and the gray bar indicates the confidence interval of 95%.

Figure 7 represents the lower eyelid analysis. The total 50 lower eyelid testing images were used for this analysis. The middle solid line indicates the mean of manual and segmented MG which are 0.08, 5.64, − 0.34 and − 0.04 for gland number, length, width and tortuosity respectively. All mean values are close to zero, thus the ground truth and automatic segmentation are not significantly different ($p > 0.005$). The two dotted lines indicate the limit of agreement (± 1.96 SD) and the values were in between 1.20 and − 1.04 for gland number, 40.04 and − 28.75 for length, 1.72 and − 2.39 for width, and 0.93 and − 1.00 for tortuosity. The gray bar indicates the confidence interval of 95%. Among the all 50 test images only 1, 0, 1, and 2 values were outside the limit of agreement thus the 98%, 100%, 98%, and 96% of the values are within the limit of agreement.

Test–retest reliability. The developed algorithm achieved 0% variability and 100% repeatability, while running two tests, in contrast, there was variability between two manual annotations with the same observer and using the same test dataset. The variability and repeatability between manual annotations (intra-observer) and between automatic segmentations (intra-method) are presented in the Table 3.

Performance evaluation of inferior quality images. The trained deep learning model was also able to segment MGs from different inferior quality of meibography images, which were not considered during the deep learning model training, validation, and testing such as out of focus image, eyelashes in the glands area, partially everted upper eyelid, and artifact of everting finger in the top left of the image. Examples of MGs segmentation on inferior quality images are illustrated in Fig. 8.

Discussion

In several previous publications, MGs were detected using the image processing software ImageJ[8–11]. User interaction is needed with this software to define the MGs manually for each image to quantify the meibography images. Users may draw the gland regions differently and therefore intra- or inter-observer variability occurs. The first automatic MG detection algorithm was published by Koh et al.[12] and the algorithm was able to detect glands from both healthy and unhealthy glands. The most important part of this research work depended on how the meibography images were acquired. Small artifacts may cause complications during the MG detection and thus the final result was affected. Later, another automated algorithm for quantitative analysis of MG was developed by Arita et al.[14] based on image enhancement methods where various image processing techniques were applied. In their research work, user interaction was needed for a manual correction after the MG detection

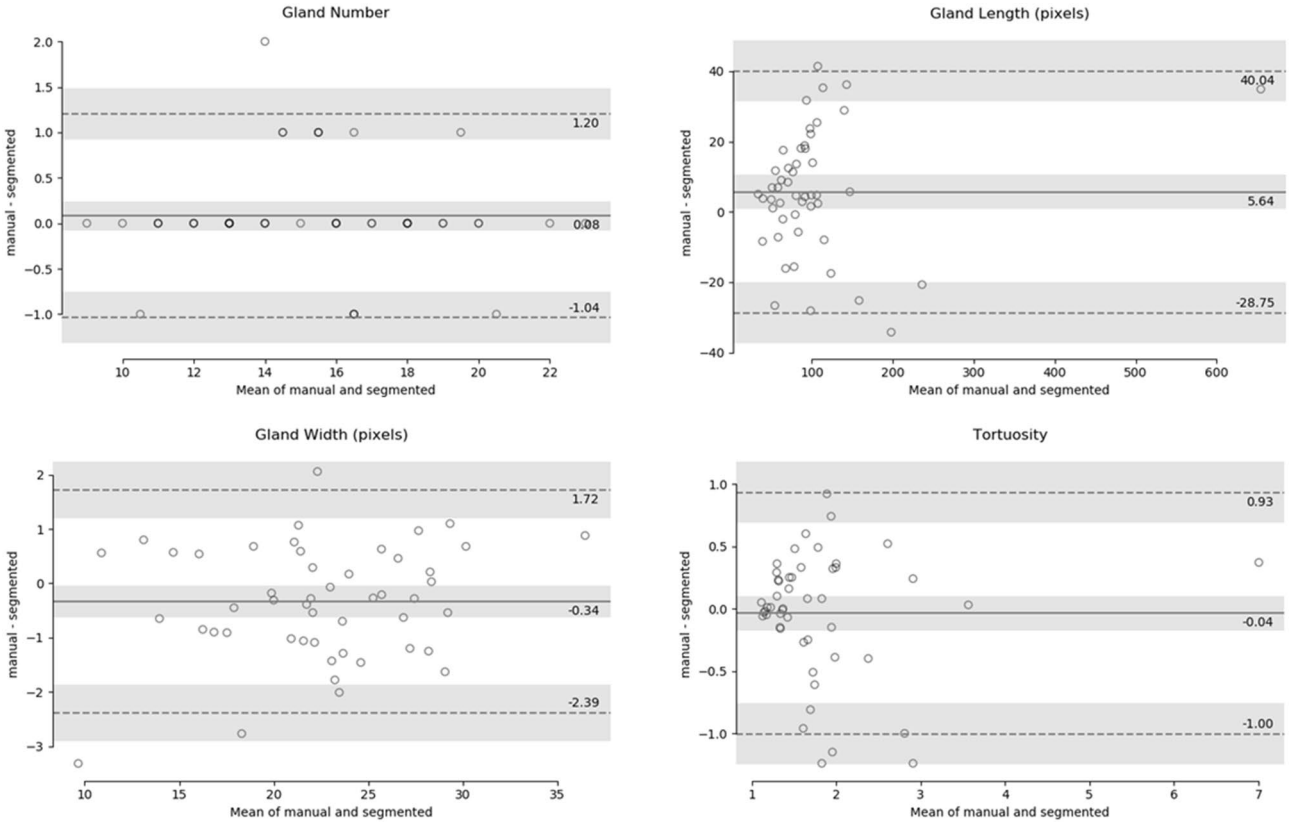

Figure 7. Bland–Altman plots of MGs number, length, width, and tortuosity between ground truth and deep learning segmentation for lower eyelid. The middle solid line indicates the mean value, the two dotted lines indicate the limit of agreement (± 1.96 SD), and the gray bar indicates the confidence interval of 95%.

Metrics	Upper eyelid		Lower eyelid	
	Intra-observer	Intra-method	Intra-observer	Intra-method
DSC	0.96 ± 0.01	1.0 ± 0.0	0.97 ± 0.01	1.0 ± 0.0
Kappa	0.87 ± 0.04	1.0 ± 0.0	0.83 ± 0.08	1.0 ± 0.0
ICC	0.90 ± 0.06	1.0 ± 0.0	0.86 ± 0.11	1.0 ± 0.0

Table 3. Variability and repeatability analysis of intra-observer and intra-method on test dataset images. Intra-observer = between manual annotations of the same observer, Intra-method = between automatic segmentations running two tests.

when the images had too many specular reflections and extensive MG loss. Thus, the algorithm was not fully automatic. In recent works, Llorens-Quintana et al.[13] proposed a fully automatic MG detection based on image enhancement methods using various image processing techniques. In their research work, they also proposed a gland irregularity measurement which was very helpful for clinicians for a follow-up checkup. However, user input was necessary, and the algorithm did not work for several images as they were not taken properly and there were some artifacts such as defocus areas, out of frame image and everted upper eyelid touching the lower eyelid. Additionally, this algorithm was designed only for the upper eyelid MG detection.

A machine learning-based automatic MG detection was developed by Celik et al.[15], which was able to detect MG using the Gabor filter. A supervised machine learning algorithm, support vector machine (SVM), was used to classify the pixels that belong to either glands or inter-glands areas. However, while our model is trained with a gland class and a general background class, their model solely learns from the gland and inter-gland areas, which could potentially lead to false detection. Furthermore, this algorithm was tested for the upper eyelid. Another artificial intelligence (AI) based MG detection was proposed by Koprowski et al.[16], where the SVM classifier was used to classify healthy, at-risk, and affected subjects with very high sensitivity (99.3%) and specificity (97.5%). The main limitations of this algorithm were the use of handcrafted features and preprocessing, which provided only the centerline information of each MG and there were no options to separate the glands. Also, the quantification of MGD was not possible using global analysis, and the AI (SVM classifier) sometimes produced false results due to the lack of classification features.

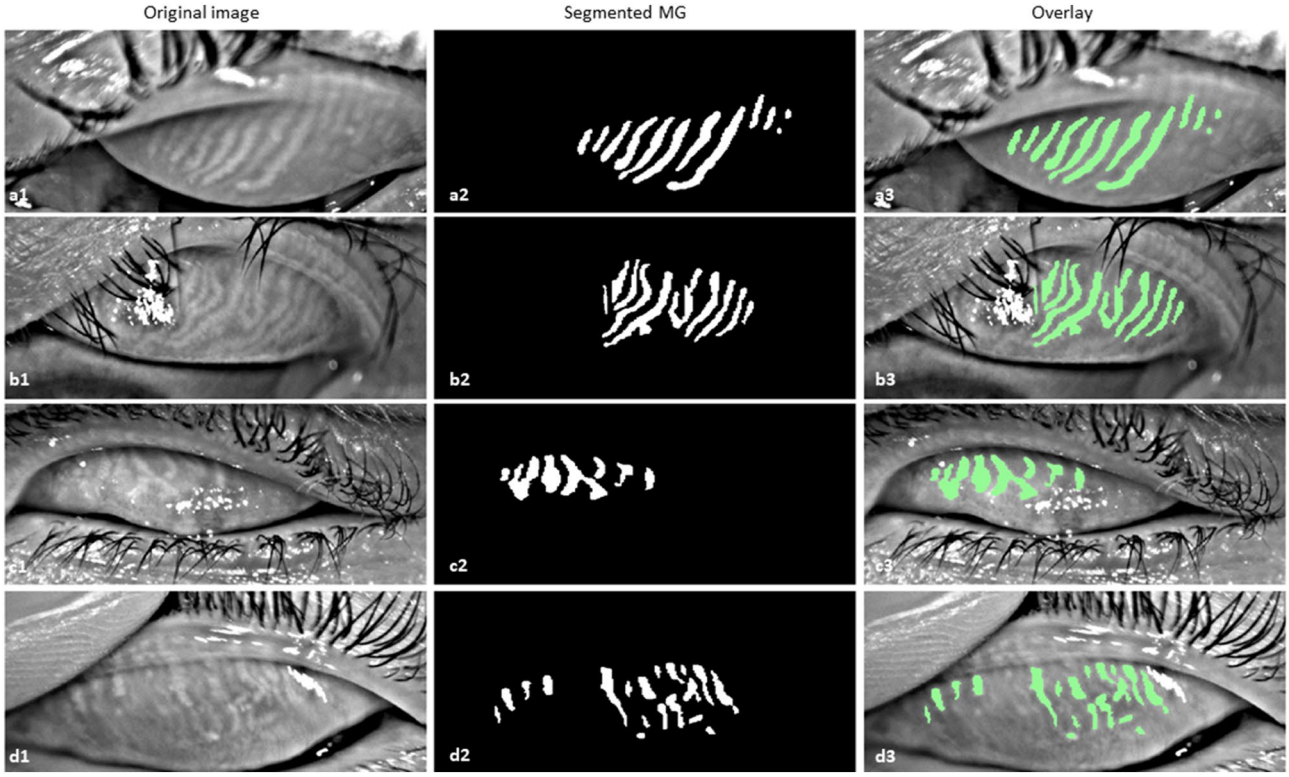

Figure 8. Example of MG segmentation on inferior quality images. (**a1**–**a3**) Out of focus image, segmented MGs, and overlay, (**b1**–**b3**) Eyelashes in the glands area image, segmented MGs, and overlay, (**c1**–**c3**) Partially everted upper eyelid image, segmented MGs, and overlay, and (**d1**–**d3**) Artifact of everting finger in the top left of the image, segmented MGs, and overlay.

Unlike the traditional AI, deep learning-based methods do not require handcrafted features or image pre-processing and it extracts the useful features from the raw images automatically[17] using the hidden layers. First, deep learning-based MG segmentation and morphometric quantification based on U-Net architecture was proposed by Prabhu et al.[34] They achieved acceptable p values ($p > 0.005$) between manual annotation and automatic segmentation method. However, comparison to our model performance is not possible as no deep learning performance metrics were mentioned. The authors calculated gland width by measuring the pixels exclusively at the mid-line of the automatic segmented image. In our research study, we instead calculated individual gland width by measuring the difference of the leftmost pixel and the rightmost pixel at each row within the range of the length of the gland and then averaged. Furthermore, in contrast to the previously published model, which was trained for 300 epochs without transfer learning, our model was trained using two stage transfer learning, enabling 30 epochs, which is more efficient and computationally inexpensive[35]. In addition, their proposed method tested only upper eyelid meibographies. Another deep learning-based MG segmentation based on pyramid scene parsing (PSP) network has been proposed by Wang et al.[36], where they achieved 95.4% accuracy and 66.7% intersection over union (IoU) to analyze gland atrophy regions. In contrast, our proposed U-Net based method achieved 84% dice coefficient to segment individual MGs. Maruoka et al.[37] proposed a deep learning-based method to detect obstructive MGD using in vivo confocal microscopy (IVCM). They achieved 94.2% sensitivity and 82.1% specificity to detect normal and obstructive MG acini. Though IVCM provides high resolution images but it is not routinely used to image meibomian gland, and their field of view is small to analyses the full length of MG or total area of the eyelid.

Our research study proposed a deep learning-based method to automatically segment and evaluate various morphometric features of MGs from IR meibography images without any image pre-processing. This deep learning model achieved high performance with 0% variability and 100% repeatability (test–retest reliability). The average computational time for segmentation and evaluation per IR meibography image was approximately 1.33 s (experiments were performed on above mentioned laptop). To evaluate 1000 disease images our method could take only 22 min without any clinicians' involvement within a single click which reduces significant processing time and computational cost. Furthermore, statistical analysis presents that more than 95% of all morphometric parameters value for both upper and lower eyelids are within the limit of agreement. Our proposed method achieved very high performance with on average 84% dice coefficient. The single trained model is able to segment MG of both upper and lower eyelids. From the visualization of the automated segmentation of MGs have high visual similarity with manual ground truth MGs. Overall, our research results demonstrate a fully automated and reliable deep learning-based technique for MG segmentation and morphometric evaluation from IR meibography images.

This research work has some limitations. Firstly, the proposed deep learning-based technique is device-specific, where training, validation, and testing images were acquired using a Keratography 5 M device. However, for the training of the neural network, MG features were automatically selected from training images by the model itself. Thus, the algorithm may segment MG from images of other IR meibography devices that need to be tested and validated. Secondly, all ground truth masks were generated manually where there is a possibility of inter-observer variability. However, all ground truth masks were validated and corrected by an experienced clinician. Finally, MGs are not accurately segmented in all test images. In some cases, two glands are connected, which need to appear separately and in other cases, single glands are separated into two or more glands, which need to be a continuous gland (Fig. 5). In all of these cases, meibography images were not acquired correctly and the acquisition problems were encountered due to the strong specular reflections, unfocused image, and MGs morphology. From Fig. 5a and c, it is clearly visible in the original images, where the glands are separated (red arrows), and the glands are connected (green arrows), that the images were not acquired correctly and the MGs are connected or overlapping. To overcome these limitations, in the future, we plan to increase the number of training dataset to more than 1000 patients and increase the number of inferior quality images thus the trained model could learn the MG features more precisely. This would also increase the learning efficiency of the deep learning model and able to segment MG more accurately from the inferior quality of images.

Conclusions

A deep learning-based automated IR meibography image segmentation method has been proposed in this research work. It reliably segments the MG as well as provides automated morphometric evaluation. The proposed technique overcomes the limitations of subjective MG segmentation and assessments which enables faster, non-invasive, precise, accurate and reproducible MG characterization thus specify better and faster disease diagnosis, new drug development, and clinical trials for a new drug of MGD related DED. These automated analyses provide valuable objective information regarding MG, which reduces, inter- or intra-observer variability and time associated with manual perception to analyze the large volume of clinical images.

Data availability

The image data utilized in this study are not publicly available due to the patients' privacy.

Received: 7 October 2020; Accepted: 21 March 2021
Published online: 07 April 2021

References

1. Craig, J. P. *et al.* TFOS DEWS II report executive summary. *Ocular Surf.* **15**, 802–812 (2017).
2. Stapleton, F. *et al.* TFOS DEWS II epidemiology report. *Ocular Surf.* **15**, 334–365 (2017).
3. Craig, J. P. *et al.* TFOS DEWS II definition and classification report. *Ocular Surf.* **15**, 276–283 (2017).
4. Stern, M. E., Gao, J., Siemasko, K. F., Beuerman, R. W. & Pflugfelder, S. C. The role of the lacrimal functional unit in the pathophysiology of dry eye. *Exp. Eye Res.* **78**, 409–416 (2004).
5. Nichols, K. K. *et al.* The international workshop on meibomian gland dysfunction: executive summary. *Investig. Ophthalmol. Vis. Sci.* **52**, 1922–1929 (2011).
6. Knop, E., Knop, N., Millar, T., Obata, H. & Sullivan, D. A. The international workshop on meibomian gland dysfunction: report of the subcommittee on anatomy, physiology, and pathophysiology of the meibomian gland. *Investig. Ophthalmol. Vis. Sci.* **52**, 1938–1978 (2011).
7. Mathers, W. D. Ocular evaporation in meibomian gland dysfunction and dry eye. *Ophthalmology* **100**, 347–351 (1993).
8. Pult, H. & Nichols, J. J. A review of meibography. *Optom. Vis. Sci.* **89**, E760–E769 (2012).
9. Srinivasan, S., Menzies, K., Sorbara, L. & Jones, L. Infrared imaging of meibomian gland structure using a novel keratograph. *Optom. Vis. Sci.* **89**, 788–794 (2012).
10. Arita, R., Itoh, K., Inoue, K. & Amano, S. Noncontact infrared meibography to document age-related changes of the meibomian glands in a normal population. *Ophthalmology* **115**, 911–915 (2008).
11. Engel, L. A. *et al.* Meibography and meibomian gland measurements in ocular graft-versus-host disease. *Bone Marrow Transplant.* **50**, 961–967 (2015).
12. Koh, Y. W. Detection of meibomian glands and classification of meibography images. *J. Biomed. Opt.* **17**, 086008 (2012).
13. Llorens-Quintana, C., Rico-Del-Viejo, L., Syga, P., Madrid-Costa, D. & Iskander, D. R. A novel automated approach for infrared-based assessment of meibomian gland morphology. *Transl. Vis. Sci. Technol.* **8**, 17–17 (2019).
14. Arita, R. *et al.* Objective image analysis of the meibomian gland area. *Br. J. Ophthalmol.* **98**, 746–755 (2014).
15. Celik, T., Lee, H. K., Petznick, A. & Tong, L. Bioimage informatics approach to automated meibomian gland analysis in infrared images of meibography. *J. Optom.* **6**, 194–204 (2013).
16. Koprowski, R. *et al.* A quantitative method for assessing the quality of meibomian glands. *Comput. Biol. Med.* **75**, 130–138 (2016).
17. Lundervold, A. S. & Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* **29**, 102–127 (2019).
18. Abstractband DOG 2019. *Ophthalmologe* **116**, 25–218 (2019).
19. Ting, D. S. W. *et al.* Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* **103**, 167–175 (2019).
20. Gargeya, R. & Leng, T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* **124**, 962–969 (2017).
21. Grewal, P. S., Oloumi, F., Rubin, U. & Tennant, M. T. S. Deep learning in ophthalmology: a review. *Can. J. Ophthalmol.* **53**, 309–313 (2018).
22. Ting, D. S. W. *et al.* Deep learning in ophthalmology: the technical and clinical considerations. *Prog. Retin. Eye Res.* **72**, 100759 (2019).
23. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 9351, 234–241 (Springer Verlag, 2015).
24. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In *31st AAAI Conference on Artificial Intelligence, AAAI 2017* 4278–4284 (AAAI Press, 2017).
25. Russakovsky, O. *et al.* ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).

26. Jaeger, S. *et al.* Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant. Imaging Med. Surg.* **4**, 475–477 (2014).
27. Kingma, D. P. & Ba, J. L. Adam: a method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings* (International Conference on Learning Representations, ICLR, 2015).
28. Haarburger, C. *et al.* Radiomics feature reproducibility under inter-rater variability in segmentations of CT images. *Sci. Rep.* **10**, 12688 (2020).
29. Gerig, G., Jomier, M. & Chakos, M. Valmet: a new validation tool for assessing and improving 3D object segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 2208, 516–523 (Springer, 2001).
30. Zou, K. H. *et al.* Statistical validation of image segmentation quality based on a spatial overlap index. *Acad. Radiol.* **11**, 178–189 (2004).
31. McHugh, M. L. Interrater reliability: the kappa statistic. *Biochem. Medica* **22**, 276–282 (2012).
32. Koo, T. K. & Li, M. Y. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
33. Martin Bland, J. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327**, 307–310 (1986).
34. Prabhu, S. M., Chakiat, A., Shashank, S., Vunnava, K. P. & Shetty, R. Deep learning segmentation and quantification of Meibomian glands. *Biomed. Signal Process. Control* **57**, 101776 (2020).
35. Raghu, M., Zhang, C., Kleinberg, J. & Bengio, S. Transfusion: understanding transfer learning for medical imaging. *arXiv* (2019).
36. Wang, J., Yeh, T. N., Chakraborty, R., Yu, S. X. & Lin, M. C. A deep learning approach for meibomian gland atrophy evaluation in meibography images. *Transl. Vis. Sci. Technol.* **8**, 37–37 (2019).
37. Maruoka, S. *et al.* Deep neural network-based method for detecting obstructive meibomian gland dysfunction with in vivo laser confocal microscopy. *Cornea* **39**, 720–725 (2020).

Acknowledgements

This research work is funded by the European Union's Horizon 2020 Innovative Training Network (ITN) under the Marie Skłodowska-Curie Actions [Integrated Training in Dry Eye Disease Drug Development (IT-DED^3), Grant Number 765608].

Author contributions

M.A.K.S. developed the deep learning algorithm and morphology assessment software, processed, and anonymized the meibography data, generated ground truth mask images, and wrote the main manuscript text. J.H. and S.S. provided technical knowledge and reviewed the manuscript. M.E.S. participated in study design and reviewed and commented on the manuscript. P.S. generated the main idea of this research, provided clinical background knowledge, collected clinical meibography data, and wrote and reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

MAKS, JH has no conflict of interest. SS is an employee of Heidelberg Engineering GmbH. MES is the chief scientific officer of ImmunEyez LLC and a member of the scientific advisory board of Novaliq GmbH. PS has received financial support from Novaliq GmbH, Roche, Bausch&Lomb, Ursapharm. The Division of Dry-eye and Ocular GvHD received donations from Novaliq, Ursapharm, and Juergen, and Monika Ziehm.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-87314-8.

Correspondence and requests for materials should be addressed to P.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

tvst

Article

Segmentation and Evaluation of Corneal Nerves and Dendritic Cells From In Vivo Confocal Microscopy Images Using Deep Learning

Md Asif Khan Setu[1,2], Stefan Schmidt[4], Gwen Musial[1,2], Michael E. Stern[1,2,5], and Philipp Steven[1–3]

[1] Department of Ophthalmology, Medical Faculty and University Hospital, University of Cologne, Cologne, Germany
[2] Division of Dry Eye and Ocular GvHD, University Hospital Cologne, University of Cologne, Cologne, Germany
[3] Cluster of Excellence: Cellular Stress Response in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany
[4] Heidelberg Engineering GmbH, Heidelberg, Germany
[5] ImmunEyez LLC, Irvine, CA, USA

Correspondence: Philipp Steven, Uniklinik Koeln, Kerpenerstrasse 62, 50937 Koeln, Germany. e-mail: philipp.steven@uk-koeln.de

Received: September 10, 2021
Accepted: May 25, 2022
Published: June 28, 2022

Keywords: deep learning; confocal microscopy; segmentation; corneal nerves; dendritic cells

Citation: Setu MAK, Schmidt S, Musial G, Stern ME, Steven P. Segmentation and evaluation of corneal nerves and dendritic cells from in vivo confocal microscopy images using deep learning. Transl Vis Sci Technol. 2022;11(6):24, https://doi.org/10.1167/tvst.11.6.24

Purpose: Segmentation and evaluation of in vivo confocal microscopy (IVCM) images requires manual intervention, which is time consuming, laborious, and non-reproducible. The aim of this research was to develop and validate deep learning–based methods that could automatically segment and evaluate corneal nerve fibers (CNFs) and dendritic cells (DCs) in IVCM images, thereby reducing processing time to analyze larger volumes of clinical images.

Methods: CNF and DC segmentation models were developed based on U-Net and Mask R-CNN architectures, respectively; 10-fold cross-validation was used to evaluate both models. The CNF model was trained and tested using 1097 and 122 images, and the DC model was trained and tested using 679 and 75 images, respectively, at each fold. The CNF morphology, number of nerves, number of branching points, nerve length, and tortuosity were analyzed; for DCs, number, size, and immature–mature cells were analyzed. Python-based software was written for model training, testing, and automatic morphometric parameters evaluation.

Results: The CNF model achieved on average 86.1% sensitivity and 90.1% specificity, and the DC model achieved on average 89.37% precision, 94.43% recall, and 91.83% F_1 score. The interclass correlation coefficient (ICC) between manual annotation and automatic segmentation were 0.85, 0.87, 0.95, and 0.88 for CNF number, length, branching points, and tortuosity, respectively, and the ICC for DC number and size were 0.95 and 0.92, respectively.

Conclusions: Our proposed methods demonstrated reliable consistency between manual annotation and automatic segmentation of CNF and DC with rapid speed. The results showed that these approaches have the potential to be implemented into clinical practice in IVCM images.

Translational Relevance: The deep learning–based automatic segmentation and quantification algorithm significantly increases the efficiency of evaluating IVCM images, thereby supporting and potentially improving the diagnosis and treatment of ocular surface disease associated with corneal nerves and dendritic cells.

Introduction

Dry eye disease (DED) is a multifactorial, immune-based inflammatory disease of the ocular surface and tears that includes ocular discomfort, dryness, pain, and alteration of tear composition, resulting in disturbance of tear production and tear evaporation. DED currently impacts 5% to 35% of the world's population, with variation in prevalence due to geographic location,

age, and gender.[1–3] Currently, DED is considered a major health problem due to its significant impact on the patient's quality of vision and life, which leads to socioeconomic burdens.[4,5]

Pain is defined as an unpleasant sensory and emotional experience that might exist or occur over a short or a prolonged period.[6] A wide range of ocular disorders can cause ocular pain; however, the most common denominator of ocular pain is ocular surface and corneal disorders,[7–9] including dry eye.[10–12] In DED patients, this ocular pain not only is creating irritation but also indicates the severity of DED. Moein et al.[7] and Galor et al.[13] have shown that, compared with patients without ocular pain, patients who suffer from DED with ocular pain have more severe signs (such as decreased corneal nerve fibers [CNFs] number and density and increased DCs density) and symptoms (such as burning, hypersensitivity to wind, and sensitivity to light and temperature). Diabetic peripheral neuropathy (DPN) affects roughly 50% of the patients that suffer from diabetes and is the main driving factor for painful diabetic neuropathy. A major challenge is identifying early neuropathy, which predominantly affects small nerve fibers first, rather than advanced neuropathy, which later affects the large nerve fibers. Furthermore, DPN showed a progressive reduction in corneal nerve fiber density, length, and branching points in patients with increasing severity of diabetic neuropathy.[14]

In vivo confocal microscopy (IVCM) is a non-invasive optical imaging modality that enables histological visualization of the CNFs and dendritic cells (DCs). IVCM is clinically widely accepted and commonly used in the diagnosis of various ocular surface disorders because it provides high resolution and detailed morphometric information regarding CNFs and DCs.[15,16] Previous work has utilized IVCM to demonstrate that the density of CNF is significantly reduced and the number of DCs is significantly higher in patients with dry eye.[7,12,17–19] Furthermore, CNFs demonstrate early and progressive pathology in a range of peripheral and central neurodegenerative conditions.[20–23] Previous studies have demonstrated analytical validation by showing that IVCM reliably quantifies early axonal damage in diabetic peripheral[24,25] neuropathy with high sensitivity and specificity.[26,27]

For accurate quantification of CNF morphology and detection of DC, the nerves and DCs must be accurately segmented in the IVCM images. Traditionally, the methods used to segment CNFs and DCs have been manual or semi-automated techniques that require experienced personnel and are laborious, cost ineffective, and potentially subject to user bias. To address these issues, several automatic CNF segmentation and quantification software tools have been developed.[28–31] For example, Dabbah et al.[28,32] developed a dual-model automated CNF detection method that showed excellent correlation with manual grading ($r = 0.92$); further extension of this method used a dual-model property in a multi-scale framework to generate feature vectors at every pixel. The authors achieved high correlation with manual grading ($r = 0.95$).

In contrast to traditional image processing or machine learning–based image segmentation techniques, a deep learning–based method offers the advantage of learning useful features and representations from raw images automatically. This method is preferable, as manually extracted features selected for a specific corneal disease may not be generally transferable to other corneal diseases.[33,34] The implementation of deep learning in the field of ophthalmology has increased in recent years.[35–37] Using a convolutional neural network (CNN) and a large volume of image data, a neural network can learn to segment and detect specific objects from ophthalmic images such as retinal blood vessels,[38] retinal layers,[39] optic disc,[40] anterior chamber angle,[41] and meibomian glands.[42] Several CNN-based CNF segmentation models have been proposed. Williams et al.[43] demonstrated the efficacy of deep learning models to identify DPN with high interclass correlation with manual ground-truth annotation. Wei et al.[44] and Colonna et al.[45] developed a U-Net–based CNN model to segment CNF that provides high sensitivity and specificity in the segmentation task. Oakley et al.[46] developed a U-Net–based model to analyze macaque corneal sub-basal nerve fibers which achieved high correlation between readers and CNN segmentation. Yıldız et al.[47] proposed CNF segmentation using a generative adversarial network (GAN), and they achieved correlation and Bland–Altman analysis results similar to those for U-Net; however, the GAN showed higher accuracy compared to U-Net in receiver operating characteristic (ROC) curves. Most of the previous studies, however, have provided only nerve segmentation without further quantification, with the exception of Williams et al.,[43] and all CNN-based models focus on CNFs, not DCs. Therefore, there is a need for fully automatic segmentation and quantification of both CNFs and DCs.

The aim of this research study was to develop and validate deep learning methods for automatic CNF and DC segmentation and the morphometric evaluation of IVCM images.

Materials and Methods

Datasets

Corneal Nerve Fiber Dataset

In total, 1578 images were acquired from both normal ($n = 90$) and pathological ($n = 105$, including 52 patients with diabetes) subjects using the Heidelberg Retina Tomograph 3 (HRT3) with Rostock Cornea Module (RCM) (Heidelberg Engineering GmbH, Heidelberg, Germany) at the Peking University Third Hospital, Beijing, China. This dataset is available for further research purposes on request.[48,49] After removing the 359 duplicate images, the remaining 1219 images were used for our model training and testing. All images cover an area of 400 × 400 µm (384 × 384 pixels). On average, eight images were obtained per subject. The images were from both central and inferior whorl regions.

Dendritic Cells Dataset

A total of 754 clinical IVCM images of both males and females from patients with dry eye and neuropathic corneal pain ($n = 54$) were collected using the HRT3-RCM at the Eye Hospital, University Hospital Cologne, Germany. These images have the same dimensions as the cornea nerve dataset (400 × 400 µm; 384 × 384 pixels). On average, 14 images were obtained per subject. The image collection adhered to the tenets of the Declaration of Helsinki and was approved by the ethics and institutional review board (IRB) at the University of Cologne, Germany (#16-405). As this study was conducted retrospectively using completely anonymized non-biometric image data, the ethics committee and IRB waived the need to obtain informed consent from the participants.

Image Annotation

Corneal Nerve Fiber Annotation

The CNF dataset from Peking University Third Hospital, Beijing, China, was manually annotated using ImageJ (National Institutes of Health, Bethesda, MD) with the NeuronJ (Biomedical Imaging Group, Lausanne, Switzerland) plug-in. All ground-truth images were verified and corrected by experienced ophthalmologists, thus providing validated ground-truth masks images along with raw IVCM images for further research.[48,49]

Dendritic Cells Annotation

All collected DC images were manually annotated using VGG Image Annotator (VIA),[50] an open-source, web-based, image annotation software. The user draws a polygon region around the DC body and hyper-reflective (dendrite) area. The VIA software tool saves all polygons in a JSON file format, which forms the ground-truth mask for the DC image dataset. One author (MAKS) created all of the initial ground-truth masks. The initial ground-truth masks were then verified and corrected by an experienced senior ophthalmologist (PS) from the University Hospital Cologne, Germany, before they were used in deep learning model training, validation, and testing.

Data Allocation

Corneal Nerve Fiber Data Allocation

The CNF image dataset, comprised of a total of 1219 images, was divided into training and testing datasets. Among 1219 images, ∼1097 were used for training, and ∼122 were used to test the model at each fold during the cross-validation.

Dendritic Cells Data Allocation

Images of DCs were also divided into training, and testing datasets. Among 754 images, ∼679 were used for model training and ∼75 images were used to test the model at each fold during cross-validation. To increase the number of training images and reduce the risk of model overfitting during the model training phase, data augmentation techniques were applied. The applied augmentations were horizontal flip, vertical flip, rotation (90°), gamma contrast (±30%), and random crop (25%) with subsequent resizing to original image size (Supplementary Fig. S1). Data augmentation increased the total number of training images to around 4753 (679 × 7).

Deep Learning Model Design and Training

Corneal Nerve Fiber Segmentation Model

The CNF segmentation model was based on the U-Net architecture.[51] A U-Net CNN architecture establishes a pixel-wise segmentation map to attain full-image resolution segmentation which makes it an ideal choice for medical image segmentation.[52–54] The detailed U-Net network architecture is shown in Figure 1. The ability of U-Net to provide pixel-level segmentation is a feature of the two sides of the "U" shape which form an encoding and a decoding path, respectively. The general pattern of encoding path includes a repeating group of convolution, dropout, convolution, batch normalization, and max pooling layer; the pattern of decoding path includes a repeating group of transpose convolution, concatenation, convolution, dropout, convolution, and batch normalization.

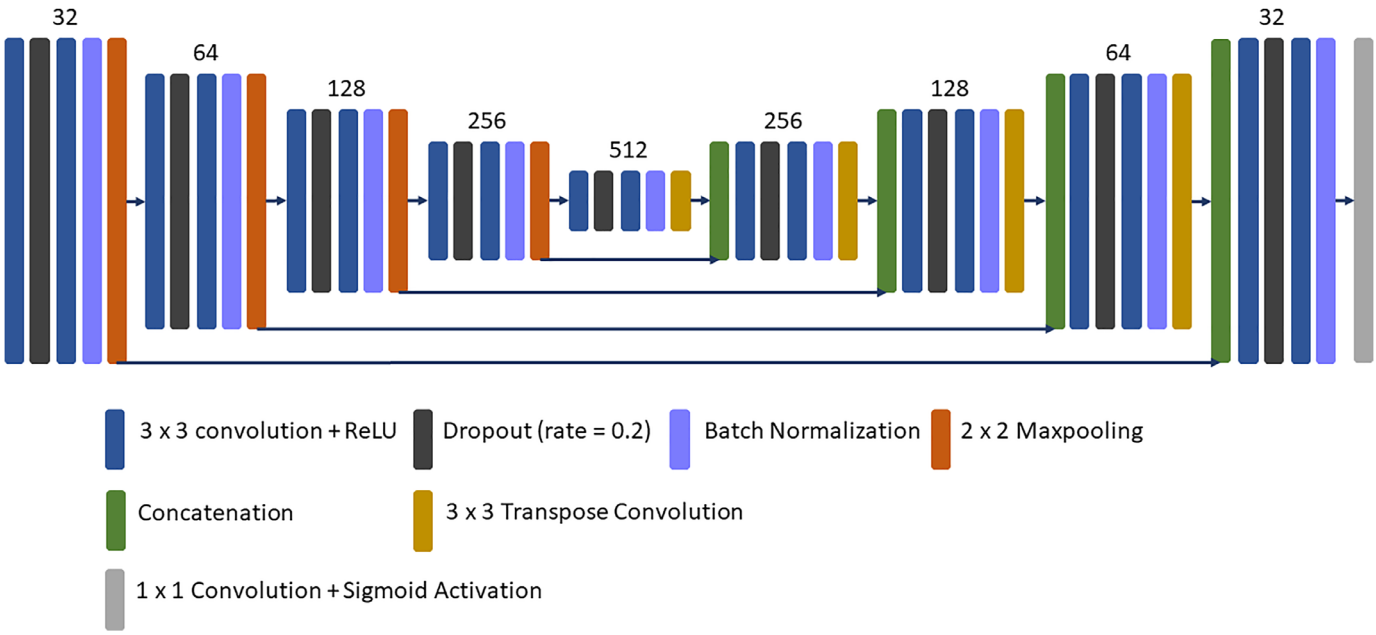

Figure 1. Detailed U-Net architecture. Each *dark blue rectangular block* represents a multi-channel feature map passing through 3 × 3 convolution followed by rectified linear unit (ReLU) operations. *Dark gray* and *light blue blocks* denote dropout with a rate of 0.2 and batch normalization. *Orange* and *dark yellow blocks* denote 2 × 2 max pooling and 3 × 3 transpose convolution, respectively. *Green blocks* denote the concatenation of feature maps. The *light gray block* denotes a 1 × 1 convolution operation followed by sigmoid activation. The number of convolution filters is indicated at the top of each column.

The direct connections between the encoding path and the decoding path allow reuse of the extracted features and strengthen feature propagation. At the end of the network architecture, the fully connected layer used a sigmoid activation function to produce the probabilistic segmentation map. This probabilistic segmentation map was converted into a binary image using a cut-off threshold value of 0.1.

A topology-preserving loss function, clDice,[55] was used during the CNF segmentation model training. Briefly, clDice preserves connectivity while segmenting tubular-like structures. One of the most efficient gradient-based stochastic optimization algorithms, Adam,[56] was used to optimize the CNF segmentation model during training. It optimizes individual learning rates for individual parameters used in the model training. The CNF segmentation model was trained for 50 epochs with an initial learning rate of 0.0001, a momentum of 0.9, and a batch size of 32 per fold during cross-validation.

Dendritic Cells Detection Model

The DC detection model was adopted from the Mask R-CNN architecture.[57] Mask R-CNN is divided into two stages: a Region Proposal Network (RPN), which proposes bounding boxes and objects, followed by a binary mask classifier to generate segmentation masks for each detected object inside the bounding box. The detailed Mask R-CNN network architecture is presented in Figure 2. First, the CNN (light yellow box in Fig. 2), based on the ResNet101[58] backbone that was pretrained with the MS COCO dataset,[59] generates a feature map (Fig. 2b) from the input image (Fig. 2a). Then, the RPN network (denoted by the purple box in Fig. 2) generates multiple regions of interest (ROIs) (dotted bounding box in Fig. 2c) using predefined bounding boxes referred to as anchors. Then, the ROI align network (green box in Fig. 2) takes both the proposed bounding boxes from the RPN network and the feature maps from the ResNet101 CNN as inputs and uses this information to find the best-fitting bounding box (Fig. 2d) for each proposed dendritic cell.

These aligned bounding box maps are then fed into fully connected layers (gray box in Fig. 2) to predict object class and bounding boxes (Fig. 2f) using softmax and regression models, respectively. Finally, the aligned bounding box maps are also fed into another CNN (light blue box in Fig. 2) consisting of four convolutional layers, transpose convolution, and sigmoid activation. This CNN is named the mask classifier, and it generates binary masks (Fig. 2e) for every detected DC. The complete Mask R-CNN uses a multi-task loss

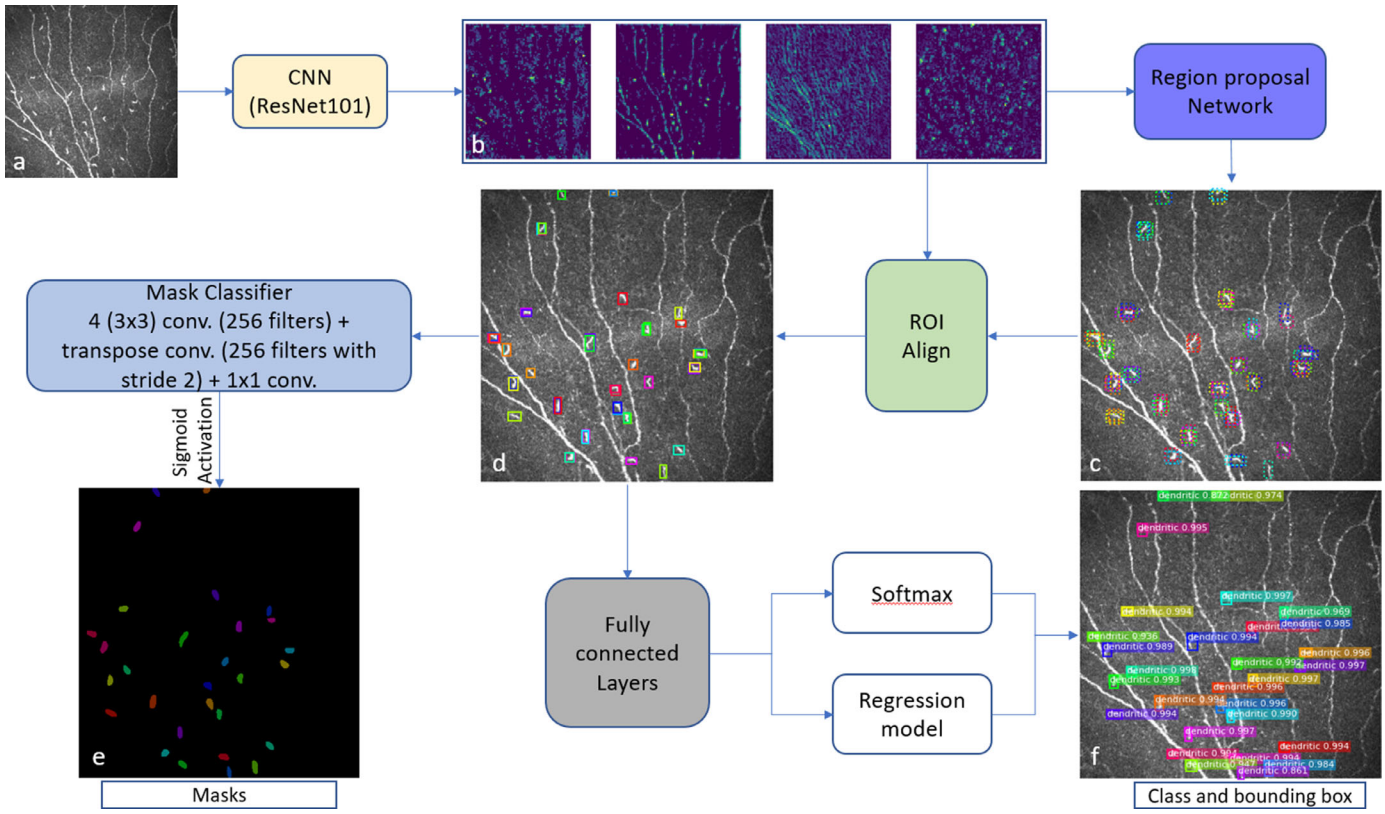

Figure 2. Mask R-CNN architecture. The pretrained ResNet101 (*light yellow box*) generates feature maps (b) from the input image (a). From the feature maps, the RPN (*purple box*; 3 × 3 convolution with 512 filters, padding the same) generates multiple ROIs (*dotted bounding box*) with the help of predefined bounding boxes referred to as anchors (c). The *green box* denotes the ROI Align network, which takes both the proposed bounding boxes from the RPN network and the feature maps as inputs and uses this information to find the best-fitting bounding box (d) for each proposed DC. These aligned box maps are fed into fully connected layers (7 × 7 convolution with 1024 filters + 1 × 1 convolution with 1024 filters), denoted by the *gray box*, and then generates a class and bounding box for each object using softmax and a regression model, respectively (f). Finally, the aligned box maps are fed into the Mask classifier (4 3 × 3 convolution with 256 filters + transpose convolution with 256 filters and stride = 2 + 1 × 1 convolution + sigmoid activation), denoted by the *light blue box*, to generate binary masks for each object (e).

function that combines object class, bounding box, and segmentation mask. The loss function is

$$L_{loss} = L_{class} + L_{box} + L_{mask} \quad (1)$$

where L_{loss} is the total loss, and L_{class}, L_{box}, and L_{mask} are the loss of object class, bounding box, and segmented masks, respectively. The model training occurred in two steps with the Adam[56] optimizer. First, the model was trained for 25 epochs with a learning rate of 0.0001 and a momentum of 0.9 and without data augmentations. In the second stage, we trained our model for another 25 epochs with a reduced initial learning rate of 0.00001, with momentum of 0.9, and with data augmentations per fold during cross-validation. We removed all DC detections with less than 90% confidence.

Both deep learning models training and testing were conducted on a laptop computer running Windows 10 Professional on a 64-bit Intel Core i7-9750H processor at 2.6 GHz with 12 MB of cache memory; Samsung 970 PRO NVMe Series SSD 512 GB M.2; Samsung 970 PRO PCIe 3.0 × 4 NVMe, RAM 32 GB DDR4 at 2666 MHz; and NVIDIA GeForce RTX 2070 Max-Q with 8 GB of GDDR6 memory. Data preparation, deep learning model design, training, evaluation, and testing were written in Python 3.6.6 using Keras 2.3.1 with TensorFlow 1.14.0 (CUDA 10.0, cuDNN 7.6.2) as the backend.

Cross-Validation Study

To evaluate our deep learning models, *k*-fold cross-validation was used. In this study, $k = 10$ was applied. The entire dataset of images (1219 for CNFs and 754 for DCs) was randomly split into 10 subgroups, and each time nine groups were selected as a training

dataset and one group was selected as a test dataset. In total ~1097 and ~122 images were randomly selected as training and testing datasets, and ~679 and ~75 images were randomly selected as training and testing datasets, respectively, for the CNF and DC segmentation models on every fold for training. To improve the accuracy of the CNF and DC models, an ensemble network of 10 trained models obtained using 10-fold cross-validation was used. The final segmentation was computed by a majority vote over the segmentation of the ensemble network.

Evaluation Metrics

For the CNF segmentation and DC detection tasks, sensitivity (*Sen*) and specificity (*Spe*) were used for CNFs, and precision (*P*), recall (*R*), and F_1 score were used for DC detection. The CNF or DC pixels (e.g., white pixels in the binary predicted masks) are considered as positive instances. Based on the combination of the ground-truth masks and predicted masks, these pixels are categorized into four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Sensitivity, specificity, precision, and recall are defined by the following equations, respectively:

$$Sen = \frac{TP}{TP + FN} \tag{2}$$

$$Spe = \frac{TN}{TN + FP} \tag{3}$$

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

The F_1 score is a weighted average of precision and recall and is defined by the following equation:

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{6}$$

Morphometric Parameter Assessment

To better analyze the CNF morphology, the number of nerves (number/frame), number of branching points (number/frame), length (mm), and tortuosity were measured; for the DC morphology, total cell number, cell size (μm^2), and mature–immature cell number were determined. All of these morphometric parameters were directly computed from the binary segmented image.

Nerve Length

To compute the nerve length (mm), the binary segmented image was first skeletonized. The branching points were then found to break up the total nerve into branches segments. Finally, the total CNF length was calculated by summing up the distance between two consecutive pixels in the nerve branch segments using the following equation:

$$CNF_{length} = \sum_{i=1}^{N-1} \sqrt{(x_{i-1} - x_i)^2 + (y_{i-1} - y_i)^2} \tag{7}$$

where N is the total number of pixels of a nerve segment, and (x_i, y_i) is the coordinate of the corresponding pixel; the inter-pixel distance is 1.0416 µm.

Nerve Tortuosity

For a curvilinear structure such as corneal nerves, tortuosity is a useful metric to calculate the curvature changes of the nerves. In this study, the average tortuosity of all detected CNFs in the image was calculated. First, the nerve length was calculated for each nerve segments. Then, the tortuosity[60] τ was calculated for each nerve segment (n) dividing the path length by the straight distance (Euclidean distance) between the start and end points of that nerve segment as follows:

$$\tau(n) = \frac{Path\ length}{Straight\ length} \tag{8}$$

Finally, the average tortuosity was calculated for the whole image by calculating the arithmetic average of the tortuosity values derived from each nerve segment. To calculate the average tortuosity, both main nerves (red) and branch nerves (orange) (Fig. 3c) were considered.

Branching Points

To calculate the branching points, we first automatically defined the start and end points for each nerve segment. All of these points are stored and evaluated. Then, if the same point is both a start and an end point for one nerve segment and a start and en end point for a different nerve segment, then this point is classified as a branch point. For example, in cluster A (Fig. 3a), there are three nerve segments (A1–A3, A2–A3, and A3–A4) and four points. A1–A2 is the starting point, and A4 is the end point. However, A3 is both an end point for

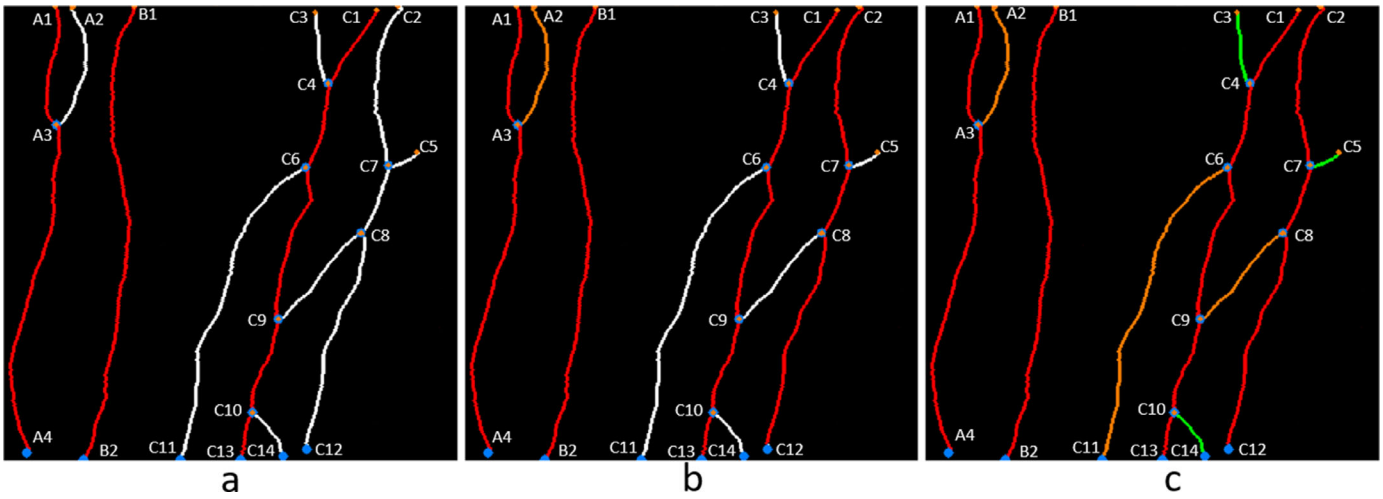

Figure 3. Calculation of number of corneal nerves. First, the corneal nerves were divided into clusters (A, B, and C) and then each start and end point was provided with a cluster point number (A1, A2, …, B1, B2, …, C1, C2, …). After that, an abstract graph was created for each cluster using the associated nodes and edges. Then, all possible paths (if branch nerves exist in a cluster such as clusters A and C) within the abstract graph were created, and the tortuosity was calculated for each path. The main nerve was selected with the lowest tortuosity for all clusters (a). All nodes creating this main nerve were removed and the process was run again to find another main nerve if one exists (cluster C); otherwise, the remaining nerve was considered the branch nerve (cluster A) (b). If there was no additional main nerve, then the remaining nodes and edges were considered branch nerves (*orange* and *green* in cluster C) (c). Finally, the total number of corneal nerves was calculated by summing the *red* and *orange nerves* while discarding the branch nerves (*green*) with length less than 20% (80 µm) of the image.

the A1–A3 and the A2–A3 nerve segments and a start point for the A3–A4 nerve segments. So, A3 is considered a branch point.

Number of Corneal Nerves

To calculate the total number of nerves presented in a frame, we first created an abstract graph using the nodes (corresponding to the start or end points of nerve segments) and edges (corresponding to the nerve segments between the start and end points) for each separated CNF cluster. Then, we extracted all of the possible paths in the graph between the isolated nodes (the start/end points that are not branching points), thus ensuring that all of the paths were "long"—more specifically, that they extended through the entire CNF cluster. Next, the tortuosities for each path were calculated. The main nerve was selected as the path with the lowest tortuosity (depicted as the red nerves in Fig. 3a). We then removed the main nerve and the branch nerves connected to it from the abstract graph and iteratively re-ran the process to find other main nerves on that CNF cluster, if existing (cluster C in Fig. 3b), until there are only two nodes left. Otherwise, the remaining edges were considered branch nerves (cluster A in Fig. 3b). Finally, we calculated the total number of nerves in a frame by summing the number of main nerves (red) and branch nerves (orange) that were longer than a predetermined threshold length. Here, we discarded nerves whose length was less than 20% (80 µm) of the image height (green). However, these discarded nerves were still considered for total nerve length, branching points, and tortuosity calculations.

Comparison with ACCMetrics

We compared corneal nerve fiber length (CNFL) using the widely used automated CNF analysis software, ACCMetrics (Early Neuropathy Assessment Group, University of Manchester, Manchester, UK).[29]

DC Size

The DC size, representing the total area surrounding the cell body and hyperreflective (dendrites) area,[61,62] was measured by calculating the segmented polygon area around the DC and reported as square micrometers.

Immature and Mature DCs

Hamrah et al.[63] differentiated immature DCs from mature DCs by their location and the absence or presence of dendrites; immature cells were located at the center of the cornea and had no dendrites, and

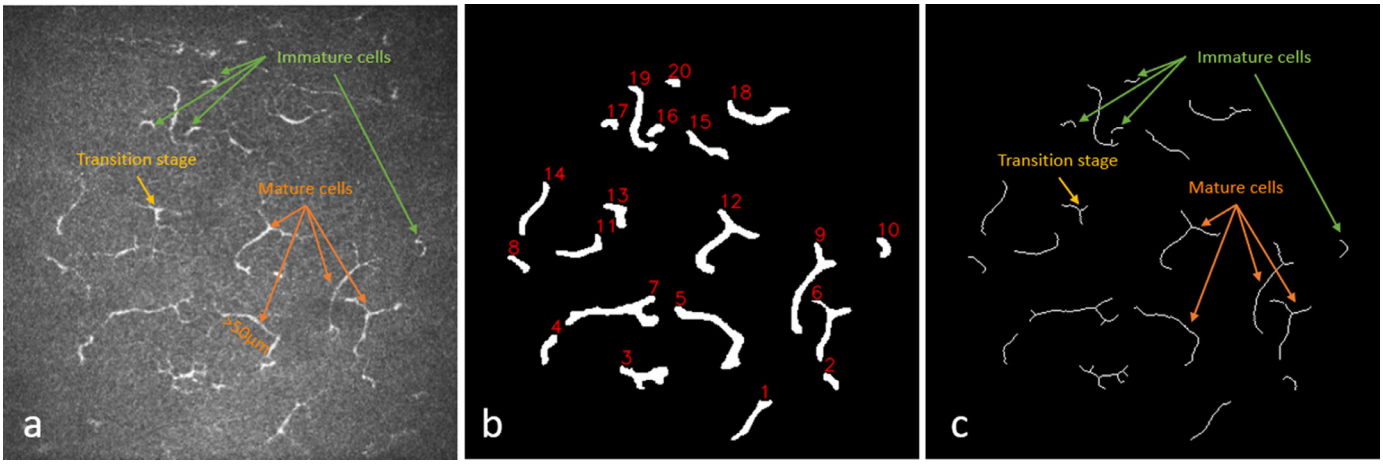

Figure 4. Immature and mature cell number calculation. (a) Original image, (b) binary segmented image with individual cell identification number, and (c) skeletonize image. The *green arrows* indicate immature cells (<50 µm) without dendrites, *yellow arrows* indicate transition-stage cells (<50 µm) with dendrites, and *orange arrows* indicate mature cells (>50 µm) with or without dendrites.

mature cells were located in the peripheral cornea and had dendrites. Building on this classification system, Zhivov et al.[64] proposed three types of DCs: (1) individual cell without dendrites (immature cell), (2) cell with small dendrites in the transition stage (starting dendrite processes), and (3) cell with long dendrites (mature cell). To differentiate among the three types of DCs, a total cell length threshold and the presence of dendrites were used as parameters. Mature cells were defined as greater than 50 µm in length with or without the presence of dendrites; immature cells were defined as cells without the presence of dendrites and less than 50 µm in length.[65,66] Finally, dendritic cells in the transition stage were defined as less than 50 µm in length with the presence of dendrites (Fig. 4a). To calculate the cell length and number of dendrites, first the binary segmented image (Fig. 4b) was skeletonized into one pixel width (Fig. 4c). Then, total cell length including dendrites and the presence of dendrites (branch numbers) were calculated for each individual cell presented in the skeletonized image using the Python-based library FilFinder 1.7.

Statistics

Our proposed CNF and DC models were compared with manual segmentation and detection. The performance of both deep learning models was measured using the Bland–Altman method.[67] Interclass correlation coefficient (ICC)[68] was used to measure agreement between manual annotation and deep learning segmentation. Python 3.6.6–based SciPy 1.5.2 and NumPy 1.18.1 libraries were used for the statistical analysis.

Results

CNF Segmentation

Training the CNF segmentation model for 50 epochs with a batch size of 32 per epoch on the full training dataset took approximately 10 hours and 28 minutes to complete, and segmenting the 122 testing images took around 9 minutes per fold on the described hardware and software. Figure 5 presents an example of testing images, along with their respective manual annotations, and automated segmentation obtained by our developed model. In general, the CNF segmentation model reliably segmented all testing images in each fold. The model achieved on average 86.1% ± 0.008% sensitivity and 90.1% ± 0.005% specificity with an average area under the ROC curve (AUC) value of 0.88 ± 0.01 (Supplementary Fig. S2) during 10-fold cross-validation.

Morphometric Parameter Assessment

The morphometric parameters of CNF number, CNF length (mm), number of branching points, and tortuosity were computed from the binary segmented image. These are important clinical parameters to analyze the CNF health status. Our software provides all of these morphometric parameters automatically from the binary segmented image. Figure 6 shows an example image where the total CNFs number (sum of red and orange) is 11, total nerve length is 3.59 mm (density, 22.43 mm/mm^2), number of branching points

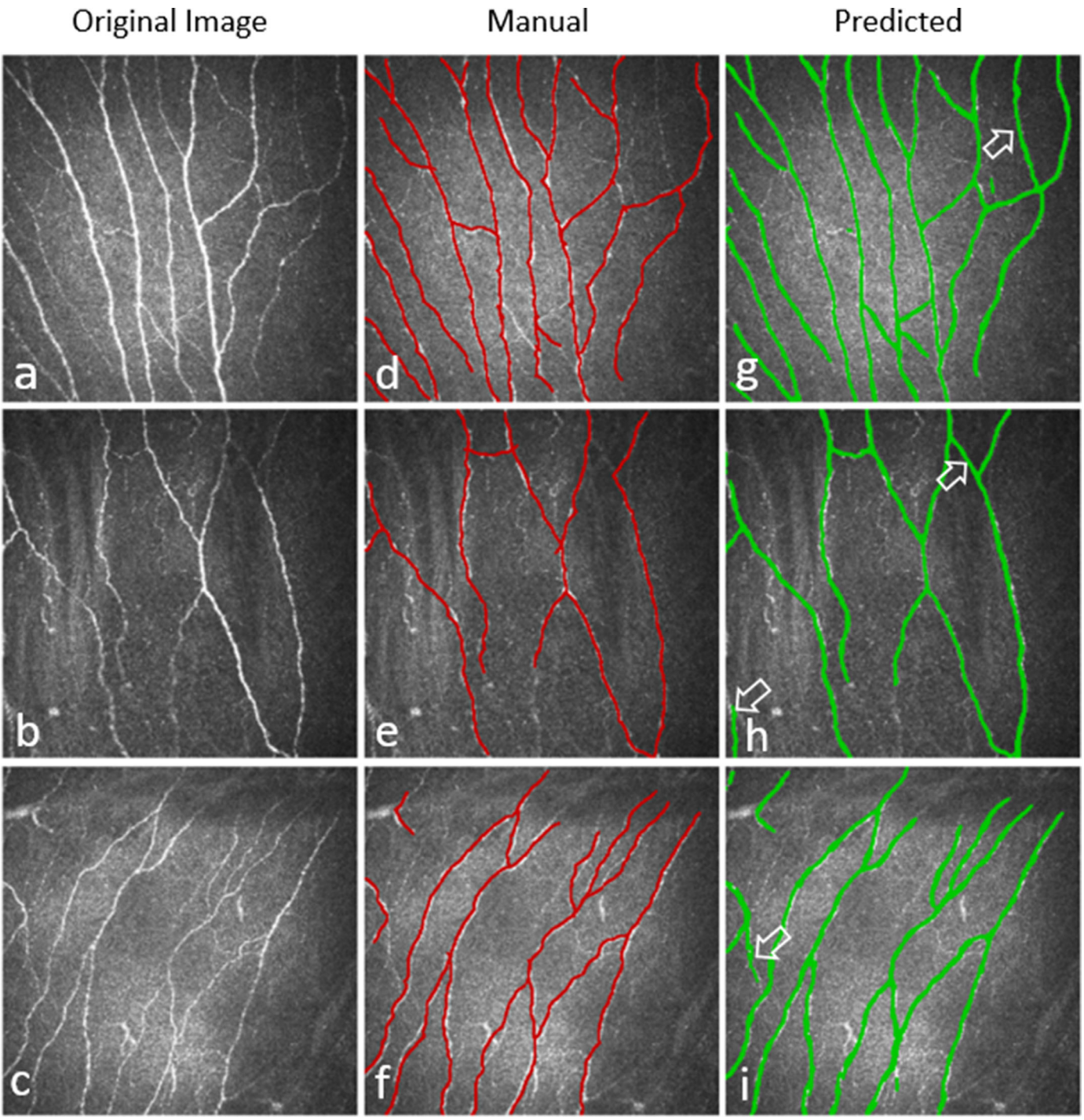

Figure 5. Three examples of CNF segmentation. (a–c) Original image, (d–f) manually segmented CNFs, and (g–i) predicted CNFs by the deep learning model. *White marked arrows* in the predicted images indicate the thin CNFs that were predicted by the deep learning model but were not annotated in the ground-truth images.

is 9, and average tortuosity of the nerves present in the frame is 1.18.

Morphological parameters from the automatically segmented images from the test dataset of 10-fold cross-validation were compared with the parameters from the manually annotated images with the *P* values (paired *t*-test) presented in the Table. The ICC of CNF number, length, branching points, and tortuosity were compared between automatic segmentation and manual annotation, and the values were 0.85, 0.87,

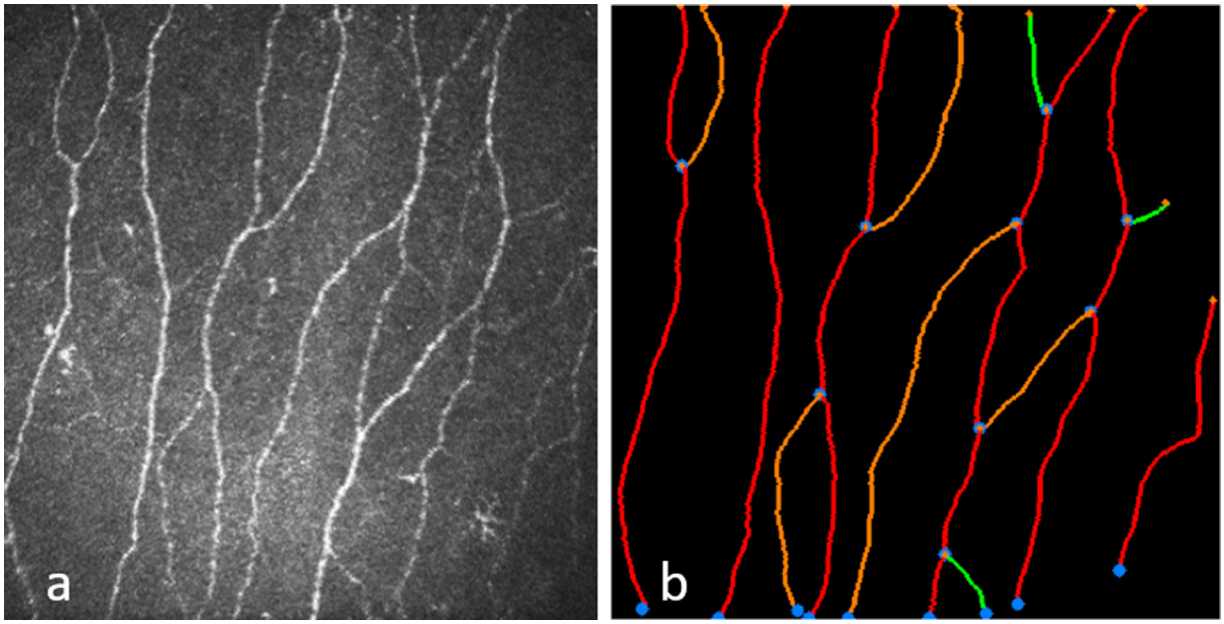

Figure 6. Example of an automatic CNF quantification from the segmented binary image of the deep learning model. (a) Original image, and (b) binary segmented image with automatic quantification.

Table. Morphometric Parameter Analysis of 10-Fold CNF Testing Dataset

	Mean ± SD		
Parameter	Manual Annotation	Automatic Segmentation	*P*
CNF number	10.14 ± 3.37	9.97 ± 3.26	0.06
CNF length (mm)	4.29 ± 3.81	4.30 ± 3.85	0.97
Branching points, *n*	6.45 ± 5.99	6.36 ± 5.85	0.31
Tortuosity	2.05 ± 1.34	2.00 ± 1.32	0.40

P values are between manual annotation and automatic segmentation.

0.95, and 0.88, respectively, for 1219 (~122 × 10) testing images from 10-fold cross-validation.

To determine the consistency between automatic and manual segmentation, Bland–Altman analysis was performed for CNF number, length (mm), branching points, and tortuosity. The results are presented in Figure 7. A total of 1219 (~122 × 10) test images from 10-fold cross-validation were used for this analysis.

Comparison With ACCMetrics

The total mean CNFL of 1219 (~122 × 10) test images from 10-fold cross-validation were 26.81 mm/mm^2, 26.87 mm/mm^2, and 13.94 mm/mm^2 for segmentation of manual ground truth, our proposed method, and ACCMetrics, respectively. The remaining quantification parameters from ACCMetrics software are tabulated in the Supplementary Table.

To better understand the difference between the proposed method and ACCMetrics, we performed further in-depth analyses by grouping the image dataset of 1219 images according to image quality: (1) high-/average-quality images (~691), defined by uniform illumination and high contrast; (2) low-quality images (~428), defined by non-uniform illumination, contrast variations, and artifacts; and (3) images of the endothelial layer, epithelial layer, and stromal keratocytes (~100). For the high-/average-quality images, there was better concordance between ACCMetrics (20.03 mm/mm^2) and the deep learning–based method (29.84 mm/mm^2). Also, there was concordance between both methods in images of the endothelial layer, epithelial layer, and stromal keratocytes (where there were no nerves present). However, there was clear discordance between ACCMetrics (7.36 mm/mm^2) and the deep learning–based method (22.98 mm/mm^2) for low-quality images containing artifacts such as nerves

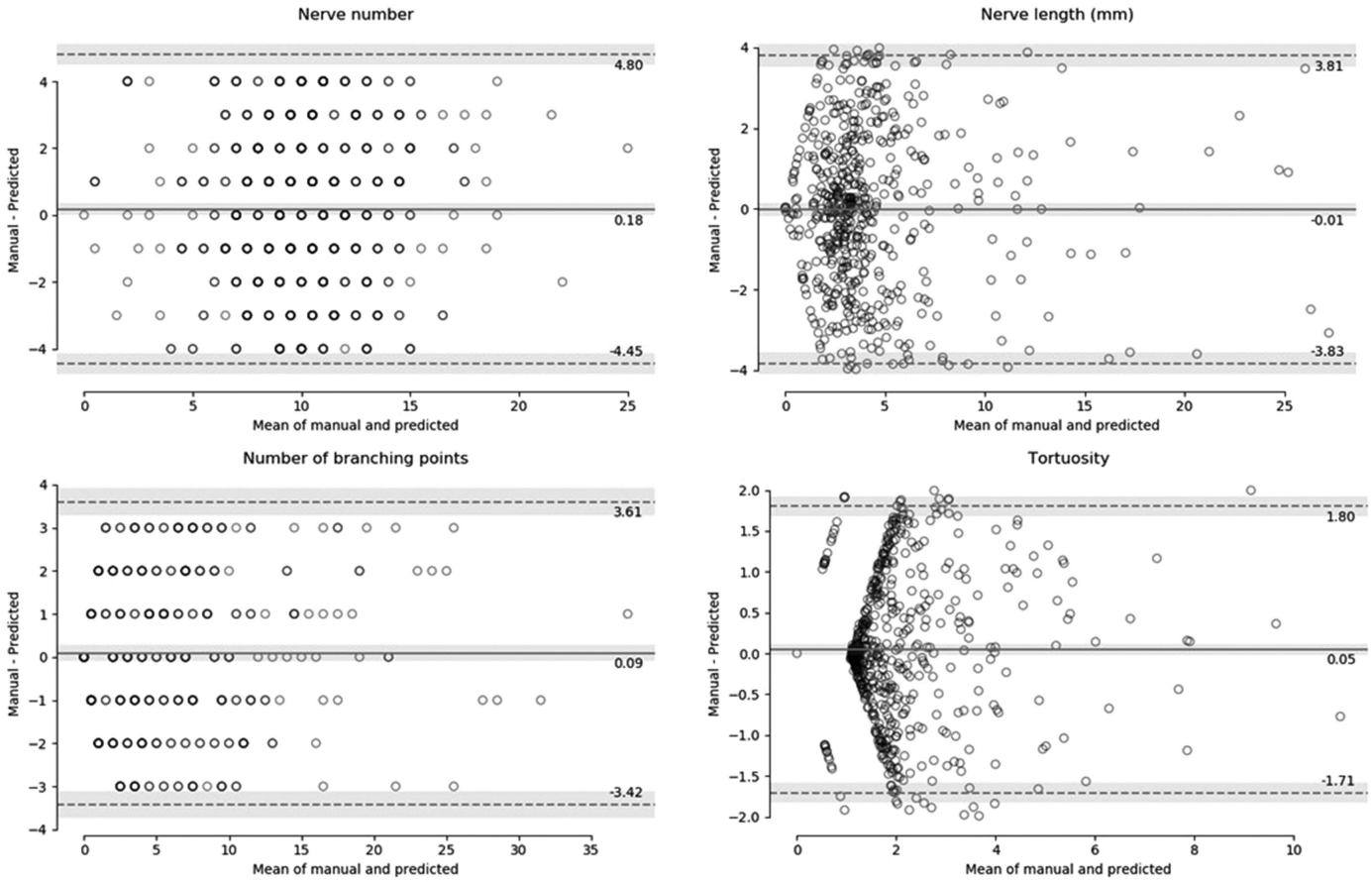

Figure 7. Bland–Altman plots present the consistency of CNF number, CNF length (mm), number of branching points, and tortuosity between manual annotation and deep learning segmentation methods. The *middle solid line* indicates the mean value of the two methods, and the two *dotted lines* indicate the limits of agreement (±1.96 SD). The *gray bands* indicate a confidence interval of 95%.

slightly out of focus, strong specular reflection, faint nerves, nerves that are not continuous (interrupted), speckle noise, and low contrast. Comparisons between the ACCMetrics and deep learning–based segmentation methods for low-quality images are presented in Supplementary Figure S3.

Dendritic Cell Detection

Training the DC detection model for 50 epochs, in two stages, on the full training dataset with data augmentation took approximately 3 hours per fold, and DC detection from 75 test images took around 3 minutes 45 seconds on the described hardware and software. A comparison between manual and automated DC detection of the test images is presented in Figure 8. The model achieved on average 89.37% ± 0.12% precision, 94.43% ± 0.07% recall, and 91.83% ± 0.09% F_1 score. The means of the total number of DCs and DC size were 9.74 ± 7.74 and 11.41 ± 7.75 ($P = 0.38$) and 1269.40 ± 939.65 μm^2 and 1296.84 ± 832.37 μm^2 ($P = 0.11$), respectively, for 754 (~75 × 10) testing images from 10-fold cross-validation of manual annotation and automatic segmentation.

Morphometric Parameter Assessment

The morphometric parameters of DC number, size (μm^2), and number of immature, transition-stage, and mature cells were computed directly from the binary segmented image. These are important clinical parameters for analyzing corneal health status. Our software provides all of these morphometric parameters automatically from binary segmented images. Figure 9 shows an example image of automatic DC segmentation with morphometric evaluation where the total number of DCs was 20, with seven immature cells (cell identification numbers 2, 4, 8, 10, 16, 17, and 20), one transition-stage cell (cell identification number 13), and 12 mature cells (cell identification numbers 1,

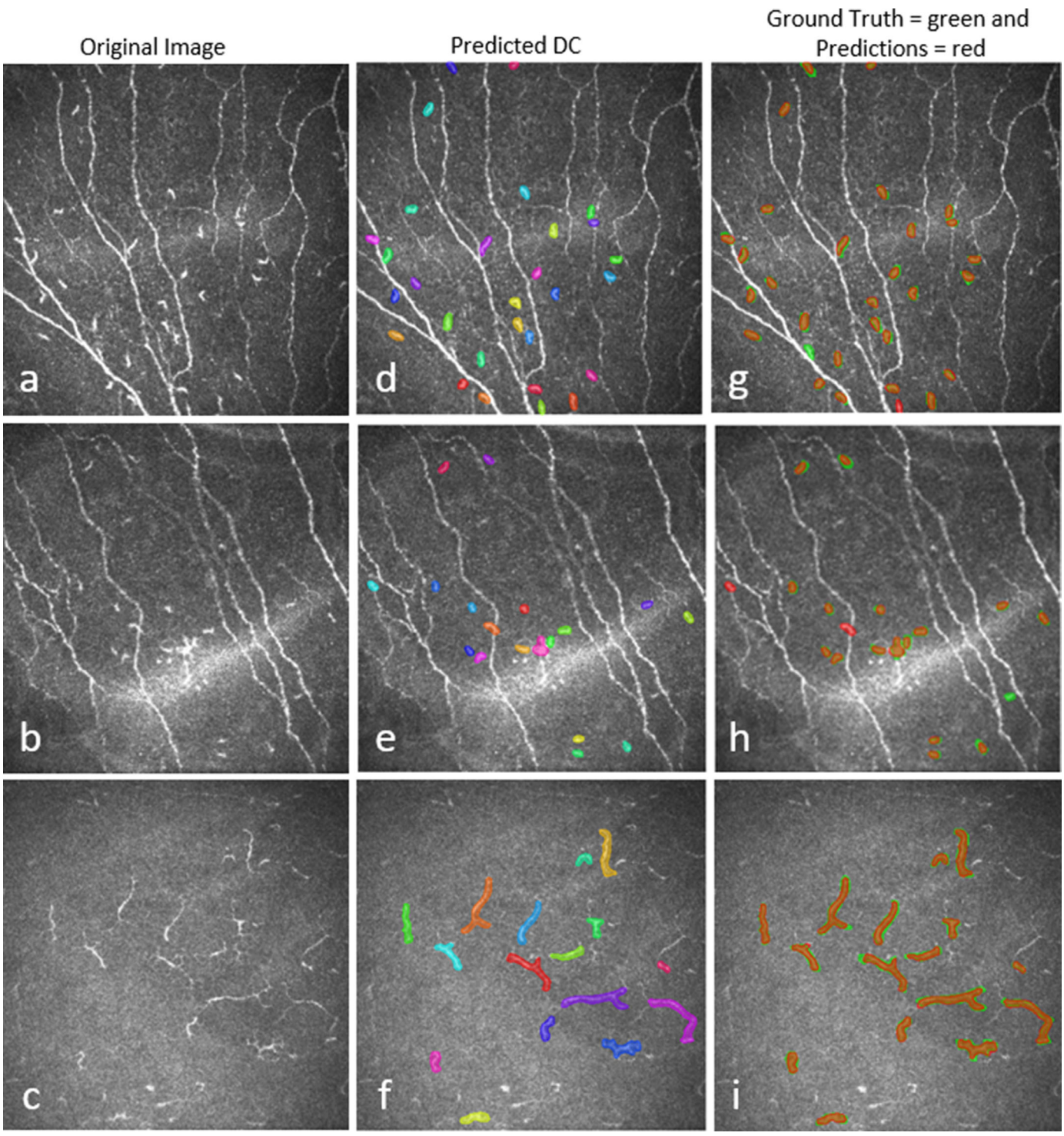

Figure 8. Three examples of DC prediction by the deep learning model. (a–c) Original image, d–f) predicted DC regions, and (g–i) overlay of manual annotation and predicted DCs.

3, 5, 6, 7, 9, 11, 12, 14, 15, 18, and 19); DC density was 125.0 cells/mm^2, and total cell size was 7555.34 μm^2.

The Bland–Altman analysis was also performed on 754 (~75 × 10) images from 10-fold cross-validation to determine the consistency of manual annotation and automatic segmentation of total DC number and size (μm^2), and the results are shown in Figures 10a and 10b. Furthermore, the total number of segmented cells without branches (dendrites) is shown in a scatter-box plot (Fig. 10c), and the immature, transition-stage,

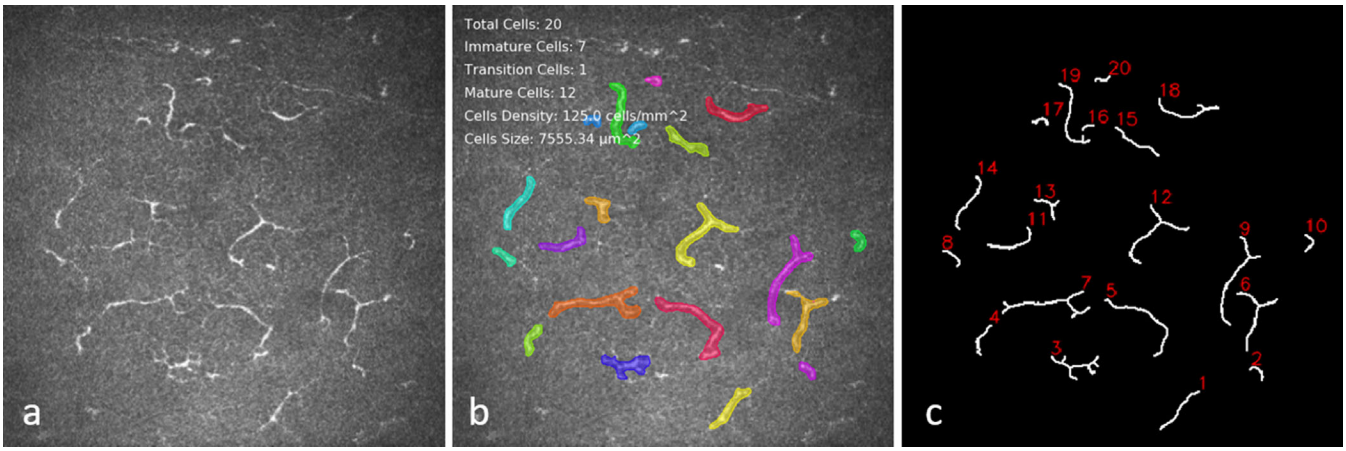

Figure 9. DC segmentation and quantification. (a) Original image, (b) segmentation overlay with quantification parameters, and (c) skeletonized binary segmented image with cell identification numbers. The immature cell identification numbers are 2, 4, 8, 10, 16, 17, and 20; the transition-stage cell identification number is 13; and the mature cell identification numbers are 1, 3, 5, 6, 7, 9, 11, 12, 14, 15, 18, 19.

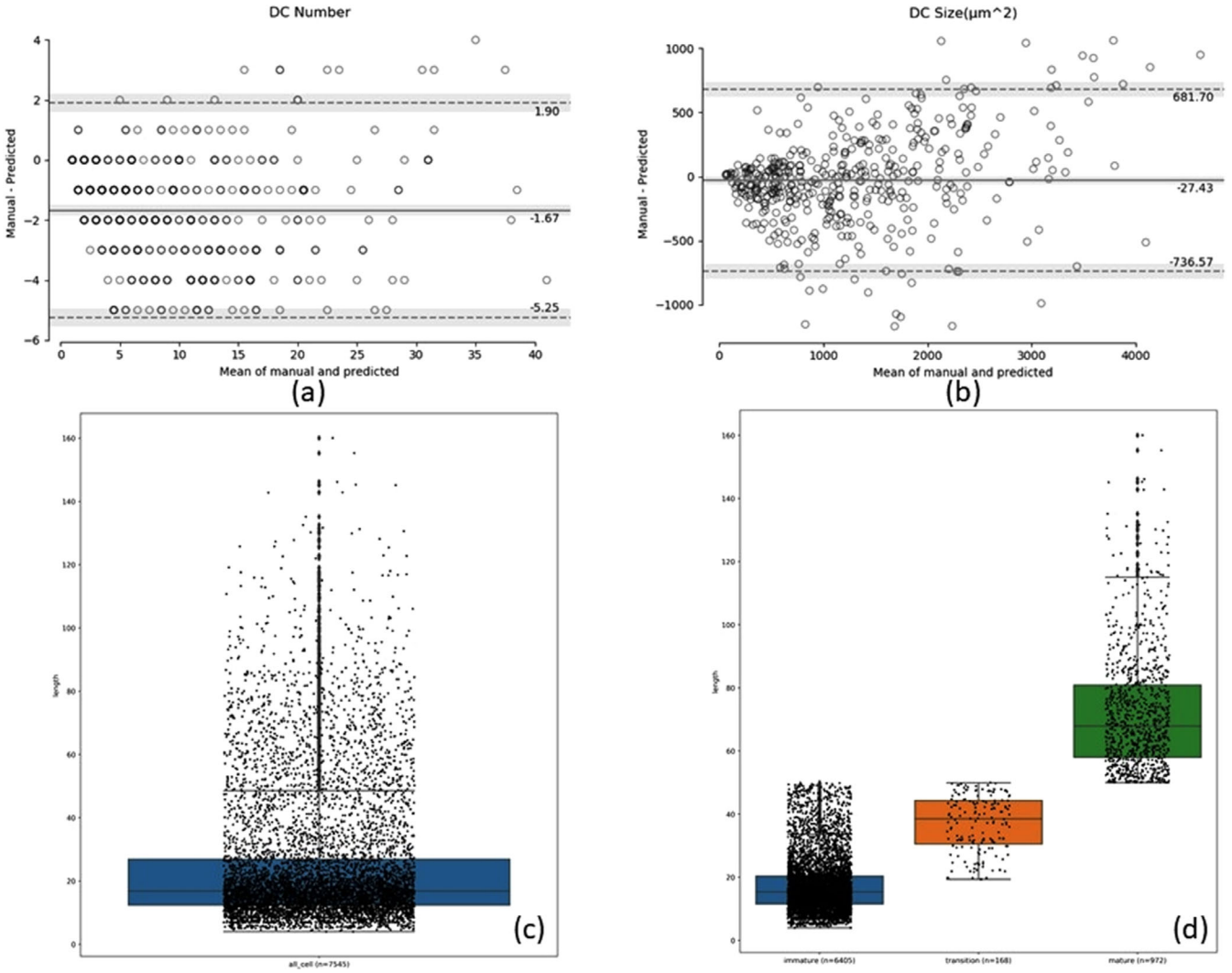

Figure 10. Bland–Altman plots indicate the consistency of the total DC number (a) and size (μm^2) (b) between manual annotation and automatic segmentation. The *middle solid line* indicates the mean value of the two methods, and the two *dotted lines* indicate the limits of agreement (±1.96 SD). The *gray bands* indicate a confidence interval of 95%. (c) Scatterbox plot of all segmented cells from the test images, and (d) scatterbox plot of the number of immature, transition-stage, and mature cells identified in the test images.

and mature dendritic cells are shown in a scatter-box plot (Fig. 10d) where the length of most of the immature cells (n = 6405) was between 10 and 20 µm (dense data point cloud). The lengths of the mature cells (n = 972) and transition-stage cells (n = 168) were evenly distributed. All 754 (~75 × 10) test images from 10-fold cross-validation were used to analyze the results for immature, transition-stage, and mature cells.

Discussion

In this study, fully automatic CNF and DC segmentation and quantification methods have been developed for corneal IVCM images. To the best of our knowledge, this is the first deep neural network–based DC segmentation and evaluation method to be proposed. This research study validates our deep learning methods and demonstrates segmentation performance that is comparable with manual annotation while reducing the amount of time required for analysis. In particular, the methods we developed show high interclass correlation between manual annotation and automatic segmentation in quantification metrics. In addition to automatic segmentation, our developed methods provide fully automatic, quantitative, clinical variables that can have utility in the diagnosis of dry eye disease, neuropathic corneal pain, and other corneal diseases.

IVCM images play an important role in the diagnosis of many corneal diseases in clinical practice. Clinicians must analyze the images multiple times to ensure accurate results for disease diagnosis or scientific research. During their manual analyses, clinicians may quantify the CNFs or DCs differently at different patient appointments which could lead to within-observer variability. In addition, two different clinicians could quantify the same image of CNFs or DCs differently, leading to between-observer variability. Therefore, a fully automatic quantitative evaluation of IVCM images is needed to obtain stable, constant, fast, and reproducible results. Unlike traditional feature engineering, deep neural network–based automatic segmentation is comparatively easier and faster than the conventional image processing approaches where various types of filters and graphs have been used.[28–31] In contrast to conventional machine learning methods such as support vector machine,[29] the deep learning–based method eliminates the need for manual feature selection and extraction and allows the machine to learn complex features using hundreds of filters.

We developed a new method of automatically characterizing corneal nerves into main trunk nerves and branching nerves for further morphometric analysis after segmentation using a deep neural network. Our trained model achieved on average 86.1% sensitivity and 90.1% specificity and AUC of 0.88 on the test dataset during cross-validation. We automatically calculated average nerve tortuosity, total nerve density, and total number of branch points, which have been shown to be useful clinical parameters to measure the severity of DED and ocular pain.[7,8,12] Our work builds upon the first deep learning–based CNF segmentation and evaluation method based on U-Net architecture, which was proposed by Williams et al.[43] This approach achieved ICC of 0.933 for total CNF length, 0.891 for branching points, 0.878 for number of nerve segments, and 0.927 for fractals between manual annotation and automatic segmentation (Liverpool Deep Learning Algorithm (LDLA) method) but had a different method for corneal nerve characterization. Williams et al.[43] calculated the total number of nerve segments by calculating the nerve segments between two branching points, two end points, or an end and a branching point, and they used fractal dimensions to describe the nerve curvature. Supplementary Figure S4 illustrates the difference between the method of Williams et al.[43] and our new proposal for detecting the trunk and branch nerves, highlighting the difference in total nerve count. Our work also differs in the loss function used. Williams et al.[43] used the Dice similarity coefficient as a loss function, whereas we used clDice,[55] a state-of-the-art topology preserving loss function that preserves connectivity among the segmented nerves. Another recent deep learning–based CNF segmentation based on U-Net was proposed by Wei et al.[44] This method achieved 96% sensitivity and 75% specificity for segmenting CNFs from IVCM images. In contrast, our U-Net–based model has achieved 86.1% sensitivity and 90.1% specificity for segmenting CNFs and provides validated automatic morphometric evaluation parameters such as the number of nerves, nerve density, nerve length, branching points, and tortuosity. Furthermore, our proposed method demonstrates better performance compared to ACCMetrics for low-quality images. In particular, average CNFL per segment obtained by our proposed method are closer to those obtained by manual annotation than the values obtained by ACCMetrics. By comparing both methods with regard to the image quality of the dataset used, we found that there is a higher requirement for image quality when using ACCMetrics to calculate CNFL, as out-of-focus, faint, or thinner nerves on the images cannot be detected properly. This is in agreement with previous studies using ACCMetrics

in which only images with high optical quality with regard to brightness, contrast, and sharpness were selected for the analysis.[69,70] In this context, Williams et al.[43] proposed further research on interrupted CNF segments. Our proposed deep learning model was trained using low-quality images; thus, it seems able to segment more nerves in the category of low-quality images.

We have proposed a new method of automatically differentiating among immature, transition-stage, and mature DCs for different pathological (dry eye, diabetic, and neuropathic corneal pain) patient data, along with cell density and cell size after the segmentation, using a deep neural network. This method could be a potential image-based biomarker to differentiate the severity of patients among the different pathological patient groups. However, in our future work, we will correlate these findings with clinical information for various patient groups. To segment DCs, we used the instance segmentation algorithm Mask R-CNN,[57] which is different from the semantic segmentation algorithm U-Net.[51] Mask R-CNN first predicts the bounding boxes that contain a DC, then segments the DC inside the bounding box. Therefore, the Mask R-CNN has the potential to detect objects more accurately than U-Net[71,72]; however, it struggles to predict good segmentation masks inside the bounding box.[71] A previous study[71] in microscopy image segmentation found that the two-step CNN process of the Mask R-CNN enables more precise localization compared with the single-step CNN used in typical U-Net architectures. For DCs, determining accurate localization and number is important; therefore, we compromised on accurate segmentation of dendritic cell border in favor of focusing on precise localization and determining accurate numbers.

In this research study, our results demonstrate that the methods we have developed can reliably and automatically segment and quantify CNFs and DCs with rapid speed. The average individual image segmentation and quantification time for CNFs was approximately 4.5 seconds, whereas for DCs it was approximately 3 seconds (based on the mentioned software and hardware). Our developed deep neural network–based methods significantly reduce image analysis time when applied to a large volume of clinical images. Manual annotation and automated segmentation of the same images appear similar, with important features segmented in both methods. Overall, the newly developed CNN models significantly reduce computational processing time while providing an objective approach to segmenting and evaluating CNFs and DCs.

Further studies are needed to identify the feasibility of implementing these methods in clinical practice and diagnostic devices. The CNF images that were acquired from Peking University Third Hospital in China and the DC images obtained from University Hospital Cologne in Germany were based on a small group of patients. Therefore, a larger patient cohort data in other racial populations could potentially enhance the strength and generalizability of the developed methods. Finally, the developed software should be assessed using different types of IVCM devices.

Conclusions

Deep neural network–based fully automatic CNF and DC segmentation and quantification methods have been proposed in this work. Automatic and objective analysis of IVCM images can assist clinicians in the diagnosis of several corneal diseases, thereby reducing user variability and time required to analyze a large volume of clinical images. Our results demonstrate that the deep learning–based approaches provide automatic quantification of CNFs and DCs and have the potential to be implemented in clinical practice for IVCM imaging.

Acknowledgments

The authors thank Yitian Zhao (Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences) for providing the CNF images along with verified ground-truth mask images to support our research.

Supported by the European Union's Horizon 2020 research and innovation program (under Marie Skłodowska-Curie grant agreement number 765608 for Integrated Training in Dry Eye Disease Drug Development [IT-DED3]) and Deutsche Forschungsgemeinschaft Research Unit FOR2240 (DFG STE1928/7-1, DFG GE3108/1-1). The Division of Dry-Eye and Ocular GvHD received donations from Novaliq, Ursapharm, and Juergen and Monika Ziehm.

Author contributions: MAKS designed and developed the deep learning algorithm and morphology assessment software, processed, and anonymized the DC data, generated ground truth mask images, and wrote the main manuscript text. SS and GM provided technical knowledge and reviewed the manuscript. MES participated in study design and reviewed and commented on the manuscript. PS generated the main

idea of this research, provided clinical background knowledge, collected and verified the clinical IVCM data, and wrote and reviewed the manuscript.

Disclosure: **M.A.K. Setu**, None; **S. Schmidt**, Heidelberg Engineering (E); **G. Musial**, None; **M.E. Stern**, ImmunEyez (E), Novaliq (C); **P. Steven**, Novaliq (R), Roche (R), Bausch & Lomb (R), Ursapharm (R)

References

1. Craig JP, Nelson JD, Azar DT, et al. TFOS DEWS II Report Executive Summary. *Ocul Surf*. 2017;15(4):802–812.
2. Stapleton F, Alves M, Bunya VY, et al. TFOS DEWS II Epidemiology Report. *Ocul Surf*. 2017;15(3):334–365.
3. Craig JP, Nichols KK, Akpek EK, et al. TFOS DEWS II Definition and Classification Report. *Ocul Surf*. 2017;15(3):276–283.
4. McDonald M, Patel DA, Keith MS, Snedecor SJ. Economic and humanistic burden of dry eye disease in Europe, North America, and Asia: a systematic literature review. *Ocul Surf*. 2016;14(2):144–167.
5. Uchino M, Schaumberg DA. Dry eye disease: impact on quality of life and vision. *Curr Ophthalmol Rep*. 2013;1(2):51–57.
6. Belmonte C, Acosta MC, Merayo-Lloves J, Gallar J. What causes eye pain? *Curr Ophthalmol Rep*. 2015;3(2):111–121.
7. Moein HR, Akhlaq A, Dieckmann G, et al. Visualization of microneuromas by using in vivo confocal microscopy: an objective biomarker for the diagnosis of neuropathic corneal pain? *Ocul Surf*. 2020;18(4):651–656.
8. Hamrah P, Cruzat A, Dastjerdi MH, et al. Corneal sensation and subbasal nerve alterations in patients with herpes simplex keratitis: an in vivo confocal microscopy study. *Ophthalmology*. 2010;117(10):1930–1936.
9. Dieckmann G, Goyal S, Hamrah P. Neuropathic corneal pain: approaches for management. *Ophthalmology*. 2017;124(11):S34–S47.
10. Pflipsen M, Massaquoi M, Wolf S. Evaluation of the painful eye. *Am Fam Physician*. 2016;93:991–998.
11. Cruzat A, Qazi Y, Hamrah P. In vivo confocal microscopy of corneal nerves in health and disease. *Ocul Surf*. 2017;15(1):15–47.
12. Benítez Del Castillo JM, Wasfy MAS, Fernandez C, Garcia-Sanchez J. An in vivo confocal masked study on corneal epithelium and subbasal nerves in patients with dry eye. *Invest Ophthalmol Vis Sci*. 2004;45(9):3030–3035.
13. Galor A, Zlotcavitch L, Walter SD, et al. Dry eye symptom severity and persistence are associated with symptoms of neuropathic pain. *Br J Ophthalmol*. 2015;99(5):665–668.
14. Petropoulos IN, Ponirakis G, Khan A, et al. Corneal confocal microscopy: ready for prime time. *Clin Exp Optom*. 2020;103(3):265–277.
15. Villani E, Baudouin C, Efron N, et al. In vivo confocal microscopy of the ocular surface: from bench to bedside. *Curr Eye Res*. 2014;39(3):213–231.
16. Niederer RL, McGhee CNJ. Clinical in vivo confocal microscopy of the human cornea in health and disease. *Prog Retin Eye Res*. 2010;29(1):30–58.
17. Benítez-Del-Castillo JM, Acosta MC, Wassfi MA, et al. Relation between corneal innervation with confocal microscopy and corneal sensitivity with noncontact esthesiometry in patients with dry eye. *Invest Ophthalmol Vis Sci*. 2007;48(1):173–181.
18. Erdélyi B, Kraak R, Zhivov A, Guthoff R, Németh J. In vivo confocal laser scanning microscopy of the cornea in dry eye. *Graefes Arch Clin Exp Ophthalmol*. 2007;245(1):39–44.
19. Khamar P, Nair AP, Shetty R, et al. Dysregulated tear fluid nociception-associated factors, corneal dendritic cell density, and vitamin D levels in evaporative dry eye. *Invest Ophthalmol Vis Sci*. 2019;60(7):2532–2542.
20. Sivaskandarajah GA, Halpern EM, Lovblom LE, et al. Structure-function relationship between corneal nerves and conventional small-fiber tests in type 1 diabetes. *Diabetes Care*. 2013;36(9):2748.
21. Tavakoli M, Quattrini C, Abbott C, et al. Corneal confocal microscopy: a novel noninvasive test to diagnose and stratify the severity of human diabetic neuropathy. *Diabetes Care*. 2010;33(8):1792.
22. Kemp HI, Petropoulos IN, Rice ASC, et al. Use of corneal confocal microscopy to evaluate small nerve fibers in patients with human immunodeficiency virus. *JAMA Ophthalmol*. 2017;135(7):795.
23. Ferdousi M, Azmi S, Petropoulos IN, et al. Corneal confocal microscopy detects small fibre neuropathy in patients with upper gastrointestinal cancer and nerve regeneration in chemotherapy induced peripheral neuropathy. *PLoS One*. 2015;10(10):e0139394.
24. Petropoulos IN, Alam U, Fadavi H, et al. Corneal nerve loss detected with corneal confocal microscopy is symmetrical and related to the severity of diabetic polyneuropathy. *Diabetes Care*. 2013;36(11):3646.

25. Petropoulos IN, Manzoor T, Morgan P, et al. Repeatability of in vivo corneal confocal microscopy to quantify corneal nerve morphology. *Cornea*. 2013;32(5):83–89.
26. Petropoulos IN, Alam U, Fadavi H, et al. Rapid automated diagnosis of diabetic peripheral neuropathy with in vivo corneal confocal microscopy. *Invest Ophthalmol Vis Sci*. 2014;55(4):2071–2078.
27. Alam U, Jeziorska M, Petropoulos IN, et al. Diagnostic utility of corneal confocal microscopy and intra-epidermal nerve fibre density in diabetic neuropathy. *PLoS One*. 2017;12(7):e0180175.
28. Dabbah MA, Graham J, Petropoulos IN, Tavakoli M, Malik RA. Automatic analysis of diabetic peripheral neuropathy using multi-scale quantitative morphology of nerve fibres in corneal confocal microscopy imaging. *Med Image Anal*. 2011;15(5):738–747.
29. Chen X, Graham J, Dabbah MA, Petropoulos IN, Tavakoli M, Malik RA. An automatic tool for quantification of nerve fibers in corneal confocal microscopy images. *IEEE Trans Biomed Eng*. 2017;64(4):786–794.
30. Scarpa F, Grisan E, Ruggeri A. Automatic recognition of corneal nerve structures in images from confocal microscopy. *Invest Ophthalmol Vis Sci*. 2008;49(11):4801–4807.
31. Kim J, Markoulli M. Automatic analysis of corneal nerves imaged using in vivo confocal microscopy. *Clin Exp Optom*. 2018;101(2):147–161.
32. Dabbah MA, Graham J, Petropoulos I, Tavakoli M, Malik RA. Dual-model automatic detection of nerve-fibres in corneal confocal microscopy images. *Med Image Comput Comput Assist Interv*. 2010;13(pt 1):300–307.
33. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444.
34. Shen D, Wu G, Suk H, II. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. 2017;19:221–248.
35. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res*. 2019;72:100759.
36. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167–175.
37. Grewal PS, Oloumi F, Rubin U, Tennant MTS. Deep learning in ophthalmology: a review. *Can J Ophthalmol*. 2018;53(4):309–313.
38. Jiang Z, Zhang H, Wang Y, Ko SB. Retinal blood vessel segmentation using fully convolutional network with transfer learning. *Comput Med Imaging Graph*. 2018;68:1–15.
39. Mishra Z, Ganegoda A, Selicha J, Wang Z, Sadda SVR, Hu Z. Automated retinal layer segmentation using graph-based algorithm incorporating deep-learning-derived information. *Sci Rep*. 2020;10(1):9541.
40. Sreng S, Maneerat N, Hamamoto K, Win KY. Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images. *Appl Sci*. 2020;10(14):4916.
41. Fu H, Baskaran M, Xu Y, et al. A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images. *Am J Ophthalmol*. 2019;203:37–45.
42. Setu MAK, Horstmann J, Schmidt S, Stern ME, Steven P. Deep learning-based automatic meibomian gland segmentation and morphology assessment in infrared meibography. *Sci Rep*. 2021;11(1):7649.
43. Williams BM, Borroni D, Liu R, et al. An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study. *Diabetologia*. 2020;63(2):419–430.
44. Wei S, Shi F, Wang Y, Chou Y, Li X. A deep learning model for automated sub-basal corneal nerve segmentation and evaluation using in vivo confocal microscopy. *Transl Vis Sci Technol*. 2020;9(2):32.
45. Colonna A, Scarpa F, Ruggeri A. Segmentation of corneal nerves using a U-Net-based convolutional neural network. In: Stoyanov D, Taylor Z, Ciompi F, et al., eds. *Computational Pathology and Ophthalmic Medical Image Analysis*. Cham: Springer; 2018:185–192.
46. Oakley JD, Russakoff DB, McCarron ME, et al. Deep learning-based analysis of macaque corneal sub-basal nerve fibers in confocal microscopy images. *Eye Vis (Lond)*. 2020;7(1):27.
47. Yıldız E, Arslan AT, Taş AY, et al. Generative adversarial network based automatic segmentation of corneal subbasal nerves on in vivo confocal microscopy images. *Transl Vis Sci Technol*. 2021;10(6):33.
48. Mou L, Zhao Y, Chen L, et al. CS-Net: channel and spatial attention network for curvilinear structure segmentation. In: Shen D, Liu T, Peters TM, et al., eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2019, LNCS Vol. 11764*. Cham: Springer; 2019:721–730.
49. Zhao YY, Zhang J, Pereira E, et al. Automated tortuosity analysis of nerve fibers in corneal

confocal microscopy. *IEEE Trans Med Imaging*. 2020;39(9):2725–2737.
50. Dutta A, Zisserman A. The VIA annotation software for images, audio and video. In: *MM '19: Proceedings of the 27th ACM International Conference on Multimedia*. New York: Association for Computing Machinery; 2019:2276–2279.
51. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, LNCS Vol. 9351*. Cham: Springer; 2015:234–241.
52. Zhu W, Huang Y, Zeng L, et al. AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med Phys*. 2019;46(2):576–589.
53. Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors (Basel)*. 2019;19(17):3722.
54. Norman B, Pedoia V, Majumdar S. Use of 2D U-net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. *Radiology*. 2018;288(1):177–185.
55. Shit S, Paetzold JC, Sekuboyina A, et al. clDice – a topology-preserving loss function for tubular structure segmentation. *arXiv*. 2021, https://doi.org/10.48550/arXiv.2003.07311.
56. Kingma DP, Ba J.Adam: a method for stochastic optimization. *arXiv*. 2017, https://doi.org/10.48550/arXiv.1412.6980.
57. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2017:2980–2988.
58. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2016:770–778.
59. Lin TY, Maire M, Belongie S, et al. Microsoft COCO: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, eds. *European Conference on Computer Vision—ECCV 2014, LNCS Vol. 8693*. Cham: Springer; 2014:740–755.
60. Smedbyörjan, Högman N, Nilsson S, Erikson U, Olssori AG, Walldius G. Two-dimensional tortuosity of the superficial femoral artery in early atherosclerosis. *J Vasc Res*. 1993;30(4):181–191.
61. Cavalcanti BM, Cruzat A, Sahin A, Pavan-Langston D, Samayoa E, Hamrah P. In vivo confocal microscopy detects bilateral changes of corneal immune cells and nerves in unilateral herpes zoster ophthalmicus. *Ocul Surf*. 2018;16(1):101–111.
62. 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. *Invest Ophthalmol Vis Sci*. 2015;56(12):7179.
63. Hamrah P, Huq SO, Liu Y, Zhang Q, Dana MR. Corneal immunity is mediated by heterogeneous population of antigen-presenting cells. *J Leukoc Biol*. 2003;74(2):172–178.
64. Zhivov A, Stave J, Vollmar B, Guthoff R. In vivo confocal microscopic evaluation of Langerhans cell density and distribution in the normal human corneal epithelium. *Graefes Arch Clin Exp Ophthalmol*. 2005;243(10):1056–1061.
65. Ferdousi M, Romanchuk K, Mah JK, et al. Early corneal nerve fibre damage and increased Langerhans cell density in children with type 1 diabetes mellitus. *Sci Rep*. 2019;9(1):8758.
66. Khan A, Li Y, Ponirakis G, et al. Corneal immune cells are increased in patients with multiple sclerosis. *Transl Vis Sci Technol*. 2021;10(4):19.
67. Martin Bland J, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307–310.
68. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163.
69. Chin JY, Yang LWY, Ji AJS, et al. Validation of the use of automated and manual quantitative analysis of corneal nerve plexus following refractive surgery. *Diagnostics*. 2020;10(7):493.
70. Wu PY, Wu JH, Hsieh YT, et al. Comparing the results of manual and automated quantitative corneal neuroanalysing modules for beginners. *Sci Rep*. 2021;11(1):18208.
71. Vuola AO, Akram SU, Kannala J. Mask-RCNN and U-Net ensembled for nuclei segmentation. In: *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging*. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2019:208–212.
72. Zhao T, Yang Y, Niu H, Chen Y, Wang D. Comparing U-Net convolutional network with Mask R-CNN in the performances of pomegranate tree canopy segmentation. *Proc SPIE*. 2018;10780:J-1–J-9.