tvst

Special Issue

Deep Learning for Assessing the Corneal Endothelium from Specular Microscopy Images up to 1 Year after Ultrathin-DSAEK Surgery

Juan P. Vigueras-Guillén[1,2], Jeroen van Rooij[3], Angela Engel[2], Hans G. Lemij[3], Lucas J. van Vliet[1], and Koenraad A. Vermeer[2]

[1] Department of Imaging Physics, Delft University of Technology, Delft, the Netherlands
[2] Rotterdam Ophthalmic Institute, Rotterdam Eye Hospital, Rotterdam, the Netherlands
[3] Rotterdam Eye Hospital, Rotterdam, the Netherlands

Correspondence: Juan P. Vigueras-Guillén, Rotterdam Ophthalmic Institute (ROI). Schiedamse Vest 160d, 3011 BH, Rotterdam, the Netherlands. e-mail: j.p.vigueras guillen@tudelft.nl

Received: May 6, 2020
Accepted: July 6, 2020
Published: August 21, 2020

Keywords: corneal transplantation; corneal endothelial cells; in vivo imaging; artificial intelligence

Citation: Vigueras-Guillén JP, van Rooij J, Engel A, Lemij HG, van Vliet LJ, Vermeer KA. Deep learning for assessing the corneal endothelium from specular microscopy images up to 1 year after ultrathin-DSAEK surgery. Trans Vis Sci Tech. 2020;9(2):49, https://doi.org/10.1167/tvst.9.2.49

Purpose: To present a fully automatic method to estimate the corneal endothelium parameters from specular microscopy images and to use it to study a one-year follow-up after ultrathin Descemet stripping automated endothelial keratoplasty.

Methods: We analyzed 383 post ultrathin Descemet stripping automated endothelial keratoplasty images from 41 eyes acquired with a Topcon SP-1P specular microscope at 1, 3, 6, and 12 months after surgery. The estimated parameters were endothelial cell density (ECD), coefficient of variation (CV), and hexagonality (HEX). Manual segmentation was performed in all images.

Results: Our method provided an estimate for ECD, CV, and HEX in 98.4% of the images, whereas Topcon's software had a success rate of 71.5% for ECD/CV and 30.5% for HEX. For the images with estimates, the percentage error in our method was 2.5% for ECD, 5.7% for CV, and 5.7% for HEX, whereas Topcon's software provided an error of 7.5% for ECD, 17.5% for CV, and 18.3% for HEX. Our method was significantly better than Topcon's ($P < 0.0001$) and was not statistically significantly different from the manual assessments ($P > 0.05$). At month 12, the subjects presented an average ECD $= 1377 \pm 483$ [cells/mm^2], CV $= 26.1 \pm 5.7$ [%], and HEX $= 58.1 \pm 7.1$ [%].

Conclusions: The proposed method obtains reliable and accurate estimations even in challenging specular images of pathologic corneas.

Translational Relevance: CV and HEX, not currently used in the clinic owing to a lack of reliability in automatic methods, are useful biomarkers to analyze the postoperative healing process. Our accurate estimations allow now for their clinical use.

Introduction

Descemet stripping automated endothelial keratoplasty (DSAEK) is a cornea transplant procedure that involves the removal of the cornea's Descemet membrane and endothelium, followed by the transplantation of donor tissue containing the endothelium, Descemet membrane, and a part of corneal stroma. This procedure was introduced in 2006[1,2] and has replaced conventional penetrating keratoplasty for most cases of endothelial dysfunction.[3,4] The most common indication for DSAEK is corneal edema in cases of Fuchs endothelial dystrophy (FED) or pseudophakic bullous keratopathy.[5] FED is characterized by degenerative changes in the endothelial cells, which trigger an abnormal, irregular growth of the Descemet membrane and a subsequent rapid loss of endothelial cells. Graft survival after DSAEK is reported to be good, especially in eyes with FED,[6] although the outcome with respect to visual function is variable.[7] Newer techniques have aimed to reduce the thickness of the donor stroma (ultrathin [UT]-DSAEK), reporting better visual outcomes

than conventional DSAEK in terms of speed of visual recovery and visual acuity.[8] A different procedure, Descemet membrane endothelial keratoplasty, aims to transplant donor tissue with corneal endothelium and Descemet membrane but without additional stromal tissue.

Regardless of the type of surgical procedure, long-term survival of corneal grafts is mainly dependent on the endothelial cell density (ECD),[9] which is defined as the number of cells per square millimeter. ECD decreases naturally with age, from an average density of 2800 to 3000 cells/mm^2 in healthy young people to a density of 2200 to 2600 cells/mm^2 for healthy elderly people.[9–11] This loss can be exacerbated by surgical trauma and certain diseases or stress factors, and it is assumed that 400 to 700 cells/mm^2 is the density at which corneal decompensation occurs and, thus, it defines the limit for graft survival.[9,12] In addition to the loss of endothelial cells, the remaining healthy cells lose their regular hexagonal form and become irregular in shape and size. The two parameters used to describe this irregularity are the coefficient of variation in cell size (CV or polymegethism, expressed as the ratio of the standard deviation [SD] of the cell size to its mean size, in percentage), and the hexagonality (HEX or pleomorphism, defined as the percentage of six-sided cells).

To measure these parameters, it is necessary to image the endothelium. A clinically used method is specular microscopy, a noncontact, noninvasive technique that sends light toward the cornea at a small angle of incidence and captures the light reflected from the interface between the endothelium and the aqueous humor.[10] The specular reflex requires a regular, smooth endothelial surface to provide high-quality images, which is the norm in healthy corneas. Furthermore, the curvature of the cornea limits the area that can be imaged, which translates into a rather small field of view. Postkeratoplasty specular images are prone to show low contrast, high image noise, and areas that are heavily blurred or out of focus because of the many optical distortions that may arise owing to a variable thickness of the graft tissue, an irregular surface of the graft endothelium, or a graft–recipient interface haze.[11] This factor can substantially decrease the number of visible cells in the images, especially during the early postoperative phase.

Current specular microscopy manufacturers usually provide built-in software that automatically segments the images and estimates the endothelial parameters. However, many recent studies have shown a lack of agreement between the estimates from different microscope's methods and sometimes an overestimation when compared with a manual segmentation, suggesting that the automatic results should be used with caution.[13–17] These studies mainly focused in ECD, in both healthy and diseased corneas. Accurate estimation of CV and HEX is even more challenging because it requires good quality segmentations, which is not reachable in most clinical cases. Several improved approaches have been proposed in the last years.[18–27] However, most of these methods show two main limitations that prevent them to be used in the clinic: (1) they require user intervention to manually discard the areas where cells are not distinguishable after segmentation and (2) they are mainly tested in rather high-quality images and/or healthy corneas.

In this study, we present a deep learning (DL) approach that automatically segments post–UT-DSAEK endothelial images, selects the trustworthy area where cells can reliably be detected, and computes the endothelial parameters with high accuracy, all in a few seconds, without the need for any user intervention. This method was compared with the estimates provided by the microscope's software and a manual segmentation (error analysis), and it was used to study the 1-year evolution of the endothelial parameters in post–UT-DSAEK corneas (clinical analysis).

Methods

Datasets

The main dataset contained 383 images of the central corneal endothelium from 41 eyes (41 patients) that underwent UT-DSAEK surgery in the Rotterdam Eye Hospital (Rotterdam, the Netherlands), acquired at 1, 3, 6, and 12 months after surgery. The included population for the study were patients over 18 years old with FED indicated for keratoplasty for visual rehabilitation and a visual acuity of less than 0.6 (Snellen chart). At the time of surgery, patients had an average age of 73 $\pm$ 7 years. The cornea grafts were provided by the Euro Tissue Bank (Beverwijk, the Netherlands). Written informed consent was obtained from all participants, and the study was conducted in accordance with the principles of the Declaration of Helsinki (October 2013). Trial registration NL4805 (www.trialregister.nl, registered on 15-12-2014). Endothelium images were obtained with the noncontact specular microscope Topcon SP-1P (Topcon Co., Tokyo, Japan), which included the software IMAGEnet i-base (version 1.32) to estimate the endothelial parameters. The images covered an area of approximately 0.25 mm $\times$ 0.55 mm and were saved as 8-bit grayscale images of 240 $\times$ 528 pixels. Our protocol for image acquisition stated to take one image of the cornea and to repeat the acquisition up to a maximum of five images if the quality was unsatisfactory. Following this

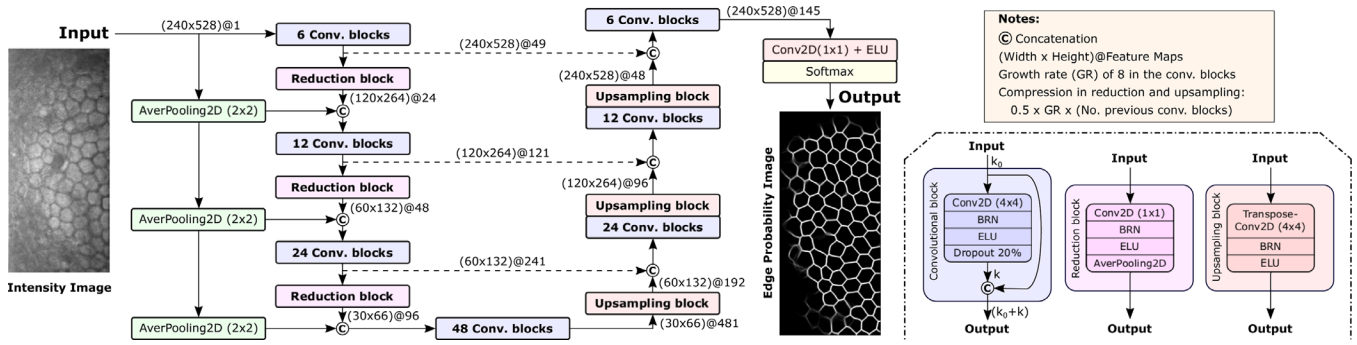

Figure 1. A schematic overview of the DL network. The different blocks in the network (convolutional, reduction, and upsampling) are depicted in the right-bottom. A dense block is defined by the concatenation of several convolutional blocks: 6 in the first resolution block, 12 in the second, etc. Each convolutional layer within the convolutional blocks has a growth rate GR = 8 (feature maps created by the layer), whereas the convolutional layers within the reduction and upsampling blocks create 0.5 × GR × (number of previous convolutional blocks) feature maps.

protocol, images were reacquired in 76% of the cases, with on average 2.3 images per session. For the clinical analysis, we used the parameters obtained from the image with the highest number of cells, and two patients were excluded because they missed one visit.

A secondary dataset was used to assist the training phase of the DL models. This dataset came from a clinical study concerning the implantation of a Baerveldt glaucoma device in the Rotterdam Eye Hospital (trial registration NL4823), referred as the Baerveldt dataset hereafter. This dataset contained 400 specular images of the central and temporal superior endothelium, obtained from 100 patients who were imaged before surgical implantation and 3, 6, 12, and 24 months after surgery, with the aim of observing whether the implant's tube was affecting the endothelium (the original Baerveldt dataset contained almost 8000 images from 200 patients, but only 400 images were manually annotated). The images were acquired with the same Topcon SP-1P microscope. Patients in this clinical study had no corneal disease, showed an average cell density of 2200 cells/mm^2, and the image quality was significantly better than the UT-DSAEK dataset. Thus, these images were useful for building robust DL models, because they provided examples not present in the UT-DSAEK dataset.

All images were manually segmented to create the gold standard, using the open-source image manipulation program GIMP (version 2.10). The DL networks were programmed in Python 3.7, and we used Tensorflow (version 2.10) to train and test the models. The parameter estimation and statistical analyses were done in Matlab 2018a (MathWorks, Natick, MA). Confidence intervals were computed with the bootstrap method 'bias corrected and accelerated percentile' in Matlab.

Development of the Network

DL is a class of machine learning algorithms that uses a sequence of mathematical operations (encoded in layers) to progressively transform the input data into more abstract representations to ultimately perform a task, such as classification, segmentation, or regression.[28] In our case, cell segmentation was done by a convolutional neural network (CNN), which are sliding-window filters that execute operations on images. This method basically entails transforming a specular endothelial image into another image where the pixels corresponding with the cell edges are given a high value (Fig. 1). In contrast with classic machine learning techniques, DL does not require to set specific rules to extract the relevant features from the images. Instead, DL will do both tasks—feature extraction and segmentation—on its own, in what is called the "black box paradigm." To this end, DL requires plenty of labeled data to learn the transformations.

An earlier version of the DL network was presented in a previous work,[27] in which we used a U-net[29] and 50 images from the Baerveldt dataset to produce the edge images (CNN-Edge), obtaining a pixel accuracy of 97.33%. Subsequently, we presented a different network[30] based on Dense U-Nets[31] to perform the segmentation of the region of interest (ROI) given the edge images (CNN-ROI); that is, the CNN-ROI identified the areas in an edge images where the cells were correctly detected. For that network, we extended the dataset up to 140 challenging images from the Baerveldt dataset, and we manually created the ROI gold standard images by selecting the trustworthy areas based on the edge images. Hence, CNN-ROI would mimic the evaluation that a human would do when observing the edge images. Our Dense U-Net[30] was a

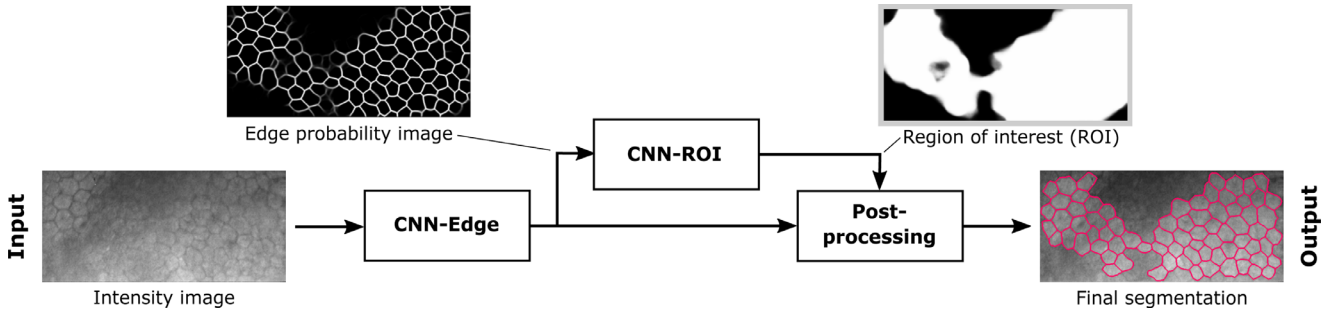

Figure 2. Overview diagram of the fully automatic method. The intensity image is the input of the first deep neural network (CNN-Edge), which provides an edge (probability) image as output. This is introduced as input to the second deep neural network (CNN-ROI), which determines the areas in the edge image where the edges were correctly detected. Finally, both images are input for the postprocessing method, which applies the ROI onto the edge image, fixes the blurry edges with the watershed algorithm, and removes the partial cells in contact to the image border, providing the final binary segmented image for computation of the clinical parameters (the final segmentation is superimposed to the intensity image for display purposes).

direct update from our previous U-Net,[27] achieving a pixel accuracy of 98.94% on the ROI problem.

In this work, we improved our Dense U-Net in the following way (Fig. 1): (1) we decreased the number of feature maps per convolutional block and added more blocks, therefore allowing for a higher reuse of features; (2) batch renormalization[32] layers were added in between the convolutional layers and the activation layers, which helped to stabilize the training; (3) we used exponential linear unit activations[33] and average pooling layers to improve performance; and (4) dropout layers[34] were placed before the concatenation to avoid additive dropout to earlier maps.

This network was used in both DL methods, the CNN-Edge and CNN-ROI (Fig. 2). That is, the CNN-Edge was trained with intensity images as input and the manually segmented images as target, providing the edge probability images as output, whereas the CNN-ROI was trained with edge probability images as input and the manually annotated ROI as target. Both networks were trained independently. Subsequently, a postprocessing method[27,35] combined both images (edge and ROI) to produce the final binary segmentation image. Briefly, the edge image was smoothed based on the average cell size (automatically obtained by Fourier analysis) and subsequently the watershed algorithm[36] was applied to create the binary segmentation. Finally, cells at the image border or with less than 75% of their area within the ROI were discarded. Once the final binary segmentation was created, the endothelial parameters were computed. To allow for a fair comparison with Topcon's software, we used the same Topcon's restrictions: ECD and CV were computed when at least six cells were segmented, whereas HEX was estimated from the inner cells (defined as segmented cells surrounded by other segmented cells) when there were at least six inner cells.

Although having three different stages to first perform the segmentation and later infer the clinical parameters might seem cumbersome, it provided a robust aproach. Earlier experiments on the design of a single DL network that would directly estimate the endothelial parameters from the specular images indicated a substantial decrease in accuracy.[37]

To evaluate the DL algorithms, a five-fold cross-validation was performed on the UT-DSAEK dataset: the images were randomly divided into five subsets (with all images from one eye in the same subset), using four subsets (plus the whole Baerveldt dataset) for training and the remaining subset for testing, and repeating the same process for the other subsets. The DL network hyperparameters were categorical cross-entropy as loss function, nadam optimizer,[38] 150 epochs, and flipping left-right and up-down as data augmentation. A Jupyter notebook with the code of the DL models, the weights, and a few example images can be found in GitHub: https://github.com/jpviguerasguillen/corneal_endothelium_TVST.

Statistical Analysis

To assess the statistical significance of our method against Topcon's software, we used the paired Wilcoxon test to compare the percentage error after assigning a 100% error if no parameter estimate was produced. To assess the clinical statistical significance of our method against the gold standard, we used the one-way analysis of variance. For this end, we used the Shapiro-Wilk normality test to establish normality and Levene's test for homogeneity. The 95% limits of agreement from the Bland-Altman analysis were also used to compare our estimates against the gold standard. Changes in the estimates between visits were evaluated with paired *t*-test. A Pearson correlation

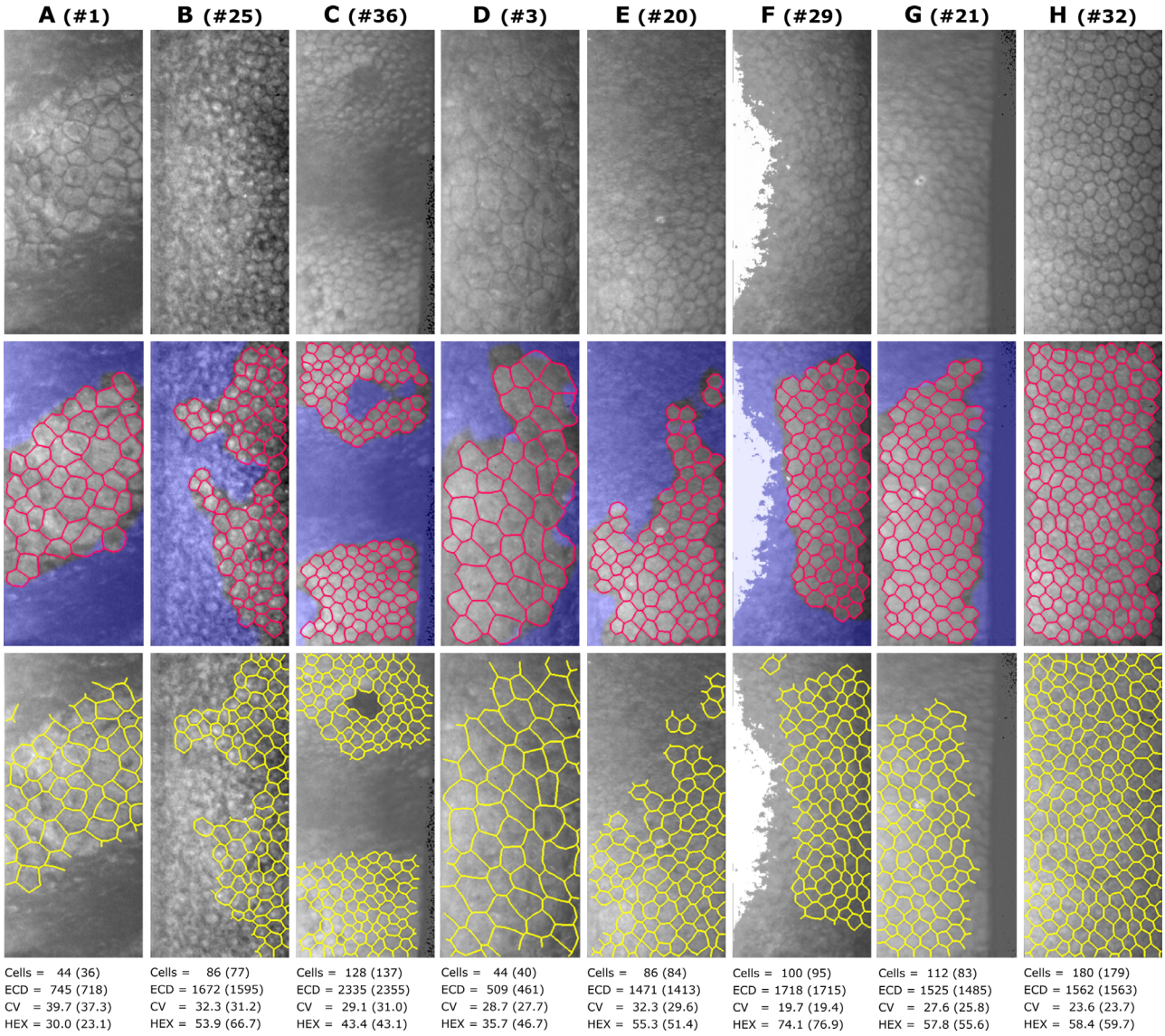

Figure 3. (Top) Eight endothelial microscopy images from different patients (patient number # is in relation to Fig. 10), acquired at month 1 (A, B), month 3 (C, D), month 6 (E, F), and month 12 (G, H). (Middle) Our fully automatic DL segmentation (*red*) superimposed on the specular image along with the areas that CNN-ROI identifies as not trustworthy (*blue*). (Bottom) The manual gold standard annotations (*yellow*). Our estimates are indicated underneath (gold standard values in parenthesis).

coefficient was used to evaluate the correlation among the changes over the 1-year follow-up. All tests used a statistical significance established at $\alpha = 0.05$.

Results

Quantitative Analysis

Our fully automatic method detected, on average, 115 cells per image whereas Topcon only detected 30 cells (paired Wilcoxon test, $P < 0.0001$). As a result, our method and Topcon's software were able to estimate the ECD/CV in 98.4% and 71.5% of the images, respectively. According to the gold standard, insufficient cells were captured in 25 images (6.5%) to be able to estimate HEX. For the remaining images, our method and Topcon's software provided HEX estimates in 99.7% and 30.5% of the images, respectively. This large discrepancy originated from the necessity of detecting at least 25 to 50 cells in an image to have six or more inner cells, which Topcon's software failed to reach in most cases. These differences are illustrated in Figures 3 and 4, showing how our method was able to detect as many cells as the gold standard in very challenging images, whereas Topcon's performance was unsatisfac-

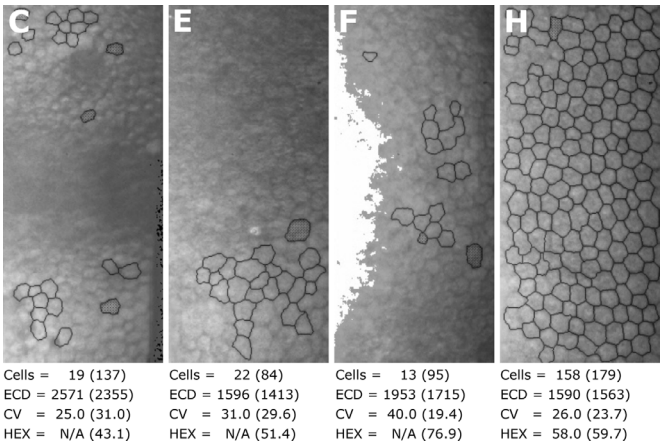

Figure 4. The only images from Figure 3 where Topcon's software could perform cell segmentation. Topcon's estimates are indicated underneath (gold standard values in parenthesis). N/A, not acquired.

tory unless the image had good quality (Fig. 4H), which was the exception in this dataset.

An error analysis on the endothelial parameters was performed (Fig. 5). Topcon's estimates displayed an ECD overestimation for the cases with high ECD and a large spread of the error in the three parameters. In contrast, our estimates showed an overall smaller error and a similar error spread for different ECD/CV/HEX reference values. Therefore, our method exhibited a robust performance regardless of the characteristics of the cells in the image.

A key factor in the method accuracy was the number of detected cells in the image, as a segmentation mistake (or discrepancy in the cells segmented) would distort the estimation significantly for images with a low number of cells. This finding can be observed in Figures 3A, 3C, where no major mistakes were detected but several cells in the border of the ROI were not included in the gold standard but appeared in our segmentation (or vice versa), and as a result there was a significant difference between our and gold standard's estimates, particularly in HEX. To model the behavior of our method in this respect, the error was plotted as a function of the number of cells (Fig. 6). Because the images in the UT-DSAEK dataset contained a rather low number of cells, we included the estimates that would be obtained from the images of the Baerveldt dataset so that we had an overview of the method's performance for images containing a large number of cells (200–400 cells). For this, we trained new models, now applying a similar five-fold cross-validation to the Baerveldt dataset and using the UT-DSAEK dataset as training assistance. Subsequently, we fitted two exponentials to the mean and SD of the error using the least-squares method (Fig. 6). The error showed a normal distribution along the y-axis for all three parameters and, thus, we could assume that the area within two SDs covered approximately the 95% of the error. This analysis showed that (1) for all three parameters, our error spread decreased as more cells were detected, (2) HEX required more cells to reduce the error spread, and (3) there was a small overestimation in CV for the images with less than 50 cells.

Error Analysis Over Time

As stated in the Methods, images were reacquired in the clinic when the quality was unsatisfactory. Specifically, we selected the image with more cells in each session, discarding the remaining images. For our method, a satisfactory image was obtained with the first acquisition in 50% of the visits, after the second in 27%, after the third in 20%, after the fourth in 2%, and after the fifth in 1%, which suggested that taking up to three images ensures acquiring a good clinical measure with high certainty. The average number of cells detected by our method was 124 ± 74, 132 ± 61, 134 ± 55, and 134 ± 55 cells for months 1, 3, 6, and 12, respectively, whereas Topcon's software detected 32 ± 42, 37 ± 36, 47 ± 39, and 54 ± 41 cells, respectively (paired Wilcoxon test, $P < 0.0001$). The mean absolute error (Table) and mean absolute percentage error (Fig. 7) were computed for both, our and Topcon's estimates. The gold standard did not display enough cells in 6.2% of images to calculate the HEX, and our method was able to estimate HEX in all other cases.

Our method provided a significantly smaller error than Topcon (Fig. 7). The statistical analysis yielded a statistically significant difference in favor of our approach for all parameters at all time points (paired Wilcoxon test, $P < 0.0001$). The percentage error was highest in the first month after surgery for all three parameters and decreased with time after surgery (Fig. 7). This behavior was mainly due to the improvement in image quality over the months, which resulted in more images with estimates, more visible cells, less segmentation mistakes, and hence better estimation accuracy (Fig. 8).

Clinical Analysis

Our method depicted the same evolution over time in the parameters as the gold standard except for CV, where a slight overestimation occurred (Fig. 9). The distributions of the estimated parameters from our method and the gold standard at the different time points resembled a normal distribution, and they passed the Shapiro-Wilk normality test after excluding one outlier for CV at month 1 and for HEX at

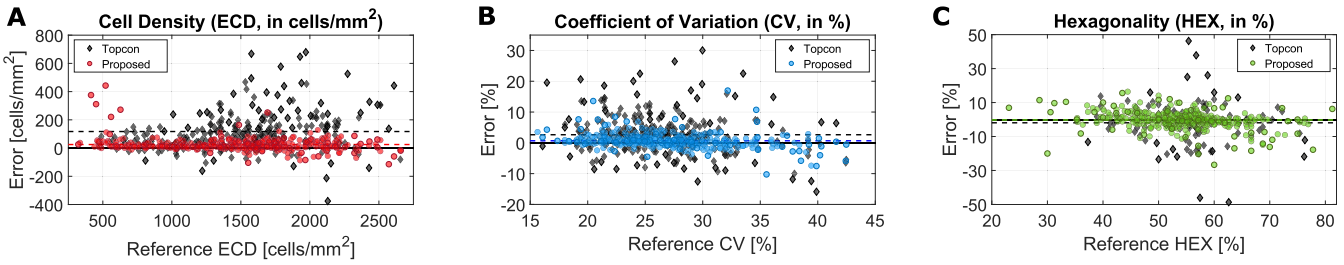

Figure 5. Error of the estimates of ECD (A), CV (B), and HEX (C) in the UT-DSAEK dataset for our DL approach (*colored circles*) and Topcon (*black diamonds*). The *x*-axis indicates the value for the gold standard images and the *y*-axis indicates the error computed as the difference between the estimates and the gold standard. The average error for each approach is depicted with a dashed line.

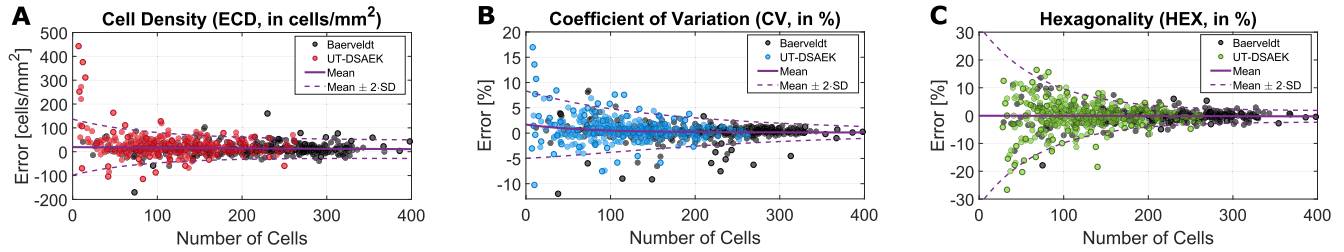

Figure 6. Error of the estimates of ECD (A), CV (B), and HEX (C) with our DL approach in the UT-DSAEK (*colored circles*, 383 images) and Baerveldt datasets (*black circles*, 400 images), displayed as a function of the number of detected cells. The *y*-axis indicates the error computed as the difference between the estimates and the gold standard. The mean (*solid line*) and two SDs (*dashed lines*) of the error function were modeled with exponentials.

Table. The Mean Absolute Error and the SD of the Endothelial Parameters for Topcon and Our Method, in the Selected Images. The Mean Absolute Error is Computed as $\frac{1}{n}\sum_{i=1}^{n}|E_i - A_i|$, Where A_i is the Actual Value, E_i is the Estimated Value, and n is the Number of Images. Only the Cases with Estimates were Used to Obtain The Error (% Indicated in Second Column)

	Percent Images Included (ECD/CV, HEX)	ECD [cells/mm^2]	CV [%]	HEX [%]
Our method	98.1%, 93.8%	30.0 ± 38.5	1.5 ± 2.0	3.0 ± 3.2
Topcon	82.1%, 42.6%	118.7 ± 166.2	4.6 ± 4.8	10.3 ± 11.1

month 3. All distributions also passed the Levene's test for homogeneity. A one-way analysis of variance indicated no statistically significant difference between the manual and our automatic assessments for any case ($P > 0.05$). A Bland-Altman analysis showed that more than 95% of the ECD estimates were within the 95% limit of agreement for all months except month 1 (94.8%). For CV, month 1 (92.1%), and month 3 (92.5%) did not reach the 95% limit of agreement. HEX was below the limit for all months (91.2%–94.9%).

The average estimated ECD at the cornea bank was 2705 ± 174 cells/mm^2, which decreased abruptly after the surgical intervention and it stabilized after the third month (Fig. 9), although the changes were statistically significant between all visits (paired *t*-test, $P < 0.03$). The average CV presented a similar evolution to ECD, with a substantial decrease between the first and third months ($P < 0.01$) and a subsequent stabilization ($P > 0.1$). In contrast, the average HEX displayed a continuous increase, with a statistically significant change between months 1 and 3 ($P = 0.01$).

We observed that the group of patients with the largest loss in ECD (Fig. 10, patients #1–#15, except #2) had a consistent large decrease in CV (–11.1 ± 9.9%) and increase in HEX (+13.2 ± 12.4%) between months 1 and 12 (values from gold standard estimates), which suggested a possible stabilization in the cellular health. The exception, patient #2, was the only case developing FED after surgery, with a bad progression in CV (+6.2%) and no estimates for HEX. Overall, we

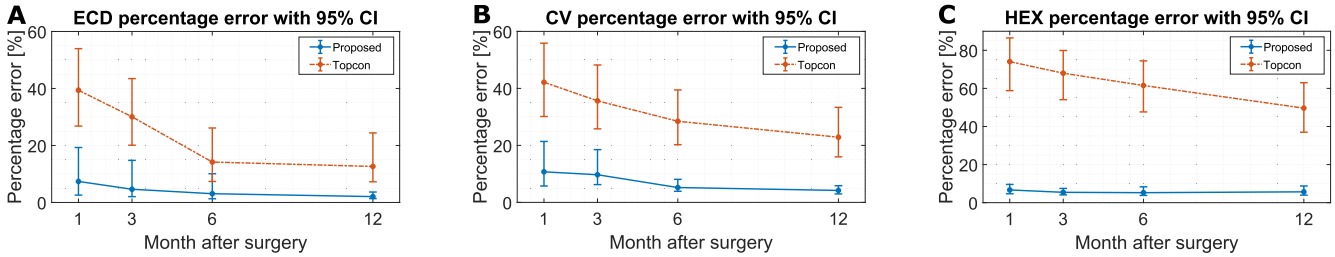

Figure 7. The mean absolute percentage error (computed as $\frac{100\%}{n}\sum_{i=1}^{n}|E_i - A_i|/A_i$, where A_i is the actual value, E_i is the estimated value, and n is the number of images) in the parameters ECD, CV, and HEX with the 95% confidence intervals, at the different time points, for Topcon (*red*) and our method (*blue*), in the selected images. Images without a parameter estimation were assigned a 100% percentage error (for HEX, the cases where the gold standard did not provide an estimate were discarded). If the images without estimates were not considered, the overall mean absolute percentage error would be 2.5% for ECD, 5.7% for CV, and 5.7% for HEX in our method, and 7.5% for ECD, 17.5% for CV, and 18.3% for HEX in Topcon's software.

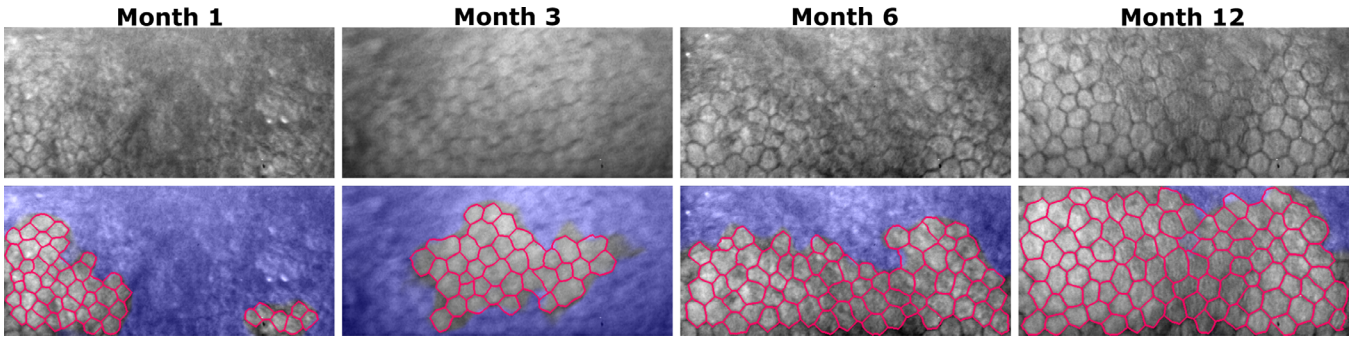

Figure 8. A representative example (patient #7) of the progression of the endothelium over the months, with a clear improvement in image quality and detected cells. Top row displays the intensity image and bottom row indicates our DL segmentation in *pink* (nontrustworthy areas in *blue*). Further information on the ECD for patient #7 in Figure 10.

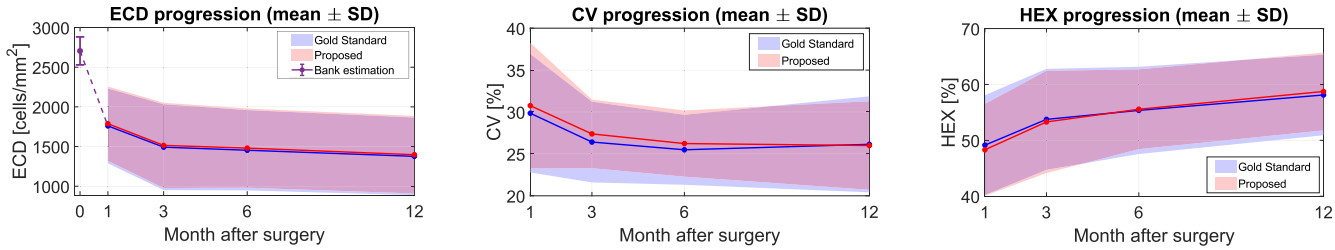

Figure 9. The evolution of the clinical parameters—ECD (A), CV (B), and HEX (C)—in the UT-DSAEK dataset over time, displayed as the average value (*solid line*) and one SD (*colored area*). Gold standard is depicted in *blue* and our proposed method in *red*.

found a statistically significant correlation between the changes in ECD and CV from month 1 to month 12 among all patients (Pearson's R = 0.356; $P = 0.031$) and between CV and HEX (R = –0.538; $P = 0.001$), but no correlation between ECD and HEX. This was also observed with our estimates.

Finally, we also observed that the patients that displayed a smaller ROI at month 1 usually had a larger ECD loss between months 1 and 12 (Pearson's R =–0.374; $P = 0.019$). This finding agrees with the idea that a smaller area with visible cells is an indication of an unhealthy tissue, which will in turn have an increased progressive loss of ECD.

Discussion

We have presented a robust, fully automatic method for the estimation of the corneal endothelium density and morphometric parameters from specular microscopy images in very challenging cases (i.e., post–

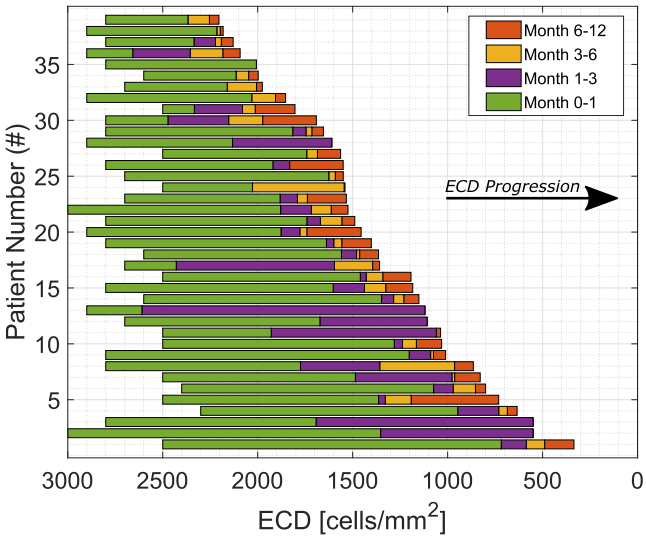

Figure 10. The ECD progression during the one-year follow-up for each patient (to read from left to right). The progression (decrease in ECD) for each time interval is depicted with a bar (values from the gold standard estimates). Patients are sorted based on the ECD at month 12. If there is an increase in ECD during a time interval, the corresponding bar is not depicted and instead the bar of the previous period is shortened.

UT-DSAEK images). The parameters estimates were in very high agreement with the gold standard. In comparison with Topcon's software, the improvement was considerable: our method detected cells in almost all images (Topcon failed in more than one-fourth of the images); the mean absolute error was more than three times smaller in the three parameters (Table), and the number of detected cells was almost three times higher, hence decreasing the estimation variability. This variability, particularly significant for HEX, is a well-known problem,[39] being widely accepted that 75 cells are required to estimate ECD with high reliability.[40]

As mentioned, just a few segmentation mistakes can significantly affect the parameters estimates. Therefore, the use of three stages (CNN-Edge, CNN-ROI, and postprocessing) is necessary to not only infer the corneal parameters from the trustworthy areas, but to also fix potential mistakes in the edge images. For example, the CNN-ROI in Figure 2 detected (in the left–central area) three cells with weak and/or duplicate edges (small dark blob), but the CNN-ROI combined with the postprocessing was able to fix this problem. A single DL network to generate both images simultaneously (edges and ROI) has been used in the literature,[41] but this approach provided unreliable ROI images for our image data. Indeed, selecting the trustworthy area based on the intensity images generated blurred ROI images that did not sufficiently match the edge images.[30] Alternatively, a single DL model to directly infer the parameters[37] may seem attractive because of its simplicity, but (1) it would require considerably more annotated data for training, which is very expensive to create, and (2) it would be impossible for the user to know whether the provided estimates are reliable (black box paradigm). In contrast, a segmentation can be visually inspected, allowing the user to assess its reliability and even allowing for manual corrections.

Previous work on automated endothelial cell parameter estimation found in the literature can be classified into two main groups: pre-DL and DL methods. Pre-DL methods[18–22] aimed mainly to segment the whole image with classic machine learning techniques, which would be clinically usable if cells were visible in the whole image or if the user would select the ROI manually. The DL approaches,[23–27] presented in the last 2 years, mainly focused on proving that a DL network is capable of segmenting endothelial images, addressing the segmentation accuracy on relatively good quality images, but rarely the clinical parameters. The closest work to ours was presented by Daniel et al.,[24] who used the U-net to segment 385 "real-world" specular images, including different ophthalmologic diseases and image qualities. Their ground truth was made by manually dotting the center of each cell and manually selecting the ROI, which only allows for ECD evaluation, and hence it was difficult to establish a comparison. Furthermore, their postprocessing was based on simple thresholding (prone to mistakes in edges not completely delineated), and their approach lacked a method to automatically select the ROI (blurred areas surrounded by detected cells would be considered cells).

One strength of our method is its ability to deal with a variety of imaging artifacts (Fig. 3). Soon after surgery, the irregular, unsmoothed surface of the endothelial graft makes it difficult to have an image in which all parts are in focus (Figs. 3A–C). Furthermore, studies in postoperative corneal thickness have indicated a presence of edema in the recipient corneal stroma soon after surgery and a continuous decrease over time,[42] which agrees with the improvement in image quality over time. In general, any lesion or abnormal tissue structure in the upper layers of the cornea can distort the optic path, creating all kind of noisy patterns. This includes saturation problems (Figs. 3F, 3G), in which the microscope internal software fails to transform the specular reflex into a suitable image because of those light aberrations. In contrast, ocular microsaccades, the movement of the patient, or simply respiration can produce heavy blurriness (Fig. 8, month 3), which is unrelated to the state of the endothelium. In all these cases, Topcon failed to detect any

cell (or detected barely a few), whereas our method performed almost as good as the gold standard. For the most complex cases, such as in images with FED, our approach needs further improvements, which probably could be tackled with more labeled images containing guttae to teach the DL models the structural and visual changes introduced by FED.

Clinically, we obtained an average ECD loss of 46.3% at month 6 and 49.1% at month 12, which was similar (or slightly higher) than other cases found in the literature. For instance, the ECD loss reported after 12 months from UT-DSAEK surgery was 49% by Feng et al.,[43] 38.9% by Graffi et al.,[44] and 35.6% by Busin et al.[8] For DSAEK, van Cleynenbreugel et al.[45] reported a 40.2% loss after 6 months, Guerra et al.[46] a 34.9% loss after 12 months, and Javadi et al.[47] a 42.8% at 2 years of follow-up. We observed that our cohort showed a large variability on ECD loss at the different time intervals, without a clear pattern (Fig. 10). Indeed, some patients showed a deceleration on ECD loss over time with an almost nonexistent progression within the last 6 months, whereas others depicted a small acceleration on ECD loss at that time. Regarding CV and HEX, barely any postkeratoplasty study includes them in their analysis because of the lack of reliability (up to now) in automatic methods. In our case, we estimated from the manual assessments a CV of 26.1 ± 5.7% and a HEX of 58.1 ± 7.1% at month 12. Existing literature in various ethnic groups indicates that the average CV is 26 ± 4%[48,49] and HEX is within 58% to 74%[50] in healthy population, and it is widely accepted that a CV of less than 30% and a HEX of greater than 60% is usually a sign of a healthy, stable endothelium. Hence, our post-transplant cohort showed an overall good outcome in terms of CV and HEX. Furthermore, we have shown that (1) our automatic method provides a high accuracy in CV and HEX (Table), (2) the existence of an evolution pattern after UT-DSAEK surgery (decrease of CV, increase of HEX; Fig. 9), particularly significant for the cases with the largest loss in ECD, and (3) a correlation between some of the parameters' evolution (stronger between CV and HEX), which might be an indication of good healing and cell loss stabilization. Indeed, the transplants that suffered a larger ECD loss in the first month had a better improvement in CV and HEX later. These two parameters can tell us something about the distress of the cells, but this is merely an extrapolation from biological science lacking proper confirmation from clinical trials. In contrast, ECD as a clinical parameter to evaluate the functionality of the cornea in terms of total thickness and clarity has been studied extensively, both in the natural state and after corneal transplantation. As we have developed a better method to analyze CV and HEX, it would be interesting to study these two parameters, for example in future follow-up studies of corneal grafting.

In summary, the results reported here demonstrated the ability of this DL method to estimate the endothelial parameters from images with different qualities and noise patterns. It also indicated a potential usefulness of employing all three endothelial parameters to study the evolution of the tissue after keratoplasty. Hence, our DL method presents itself as a valuable tool to be used in studies of corneal transplantation programs with more patients and larger follow-ups to analyze the relevance of CV and HEX and their potential roles as predictors for graft survival.

Acknowledgments

The authors thank Esma Islamaj, Caroline Jordaan, and Annemiek Krijnen for their contribution in acquiring the images.

Supported by the Dutch Organization for Health Research and Healthcare Innovation (ZonMw, The Hague, the Netherlands) under Grants 842005004 and 842005007, and by the Combined Ophthalmic Research Rotterdam (CoRR, Rotterdam, the Netherlands) under grant no. 2.1.0.

Disclosure: **J.P. Vigueras-Guillén,** None; **J. van Rooij,** None; **A. Engel,** None; **H.G. Lemij,** None; **L.J. van Vliet,** None; **K.A. Vermeer,** None

References

1. Price MO, Price FW, Jr. Descemet's stripping with endothelial keratoplasty: comparative outcomes with microkeratome-dissected and manually dissected donor tissue. *Ophthalmology*. 2006; 113:1936–1942.
2. Gorovoy MS. Descemet-stripping automated endothelial keratoplasty. *Cornea*. 2006;25:886–889.
3. Park CY, Lee JK, Gore PK, Lim CY, Chuck RS. Keratoplasty in the United States: a 10-year review from 2005 through 2014. *Ophthalmology*. 2015;122:2432–2442.
4. Anshu A, Price MO, Tan DT, Price FW, Jr. Endothelial keratoplasty: a revolution in evolution. *Surv Ophthalmol*. 2012;57:236–252.
5. Busin M. DSAEK for the treatment of endothelial disease: results in the initial 100 cases. *Klin Monbl Augenheilkd*. 2009;226:757–760.

6. Ang M, Soh Y, Htoon HM, Mehta JS, Tan D. Five-year graft survival comparing Descemet stripping automated endothelial keratoplasty and penetrating keratoplasty. *Ophthalmology*. 2016;123:1646–1652.
7. Turnbull AM, Tsatsos M, Hossain PN, Anderson DF. Determinants of visual quality after endothelial keratoplasty. *Surv Ophthalmol*. 2016;61:257–271.
8. Busin M, Madi S, Santorum P, Scorcia V, Beltz J. Ultrathin Descemet's stripping automated endothelial keratoplasty with the microkeratome double-pass technique (two-year outcomes). *Ophthalmology*. 2013;120:1186–1194.
9. Armitage WJ, Dick AD, Bourne WM. Predicting endothelial cell loss and long-term corneal graft survival. *Invest Ophthalmol Vis Sci*. 2003;44:3326–3331.
10. McCarey BE, Edelhauser HF, Lynn MJ. Review of corneal endothelial specular microscopy for FDA clinical trials of refractive procedures, surgical devices and new intraocular drugs and solutions. *Cornea*. 2008;27:1–16.
11. Hindman HB, Huxlin KR, Pantanelli SM, et al. Post-DSAEK optical changes: a comprehensive prospective analysis on the role of ocular wavefront aberrations, haze, and corneal thickness. *Cornea*. 2013;32:1567–1577.
12. Foster CS, Azar DT, Dohlman CH. *Smolin and Thoft's the cornea: scientific foundations & clinical practice*. 4th edition. Philadelphia, PA: Lippincott Williams & Wilkins; 2004:46–48.
13. Huang J, Maram J, Tepelus TC, Sadda SR, Chopra V, Lee OL. Comparison of noncontact specular and confocal microscopy for evaluation of corneal endothelium. *Eye Contact Lens*. 2018;44:S144–S150.
14. Price MO, Fairchild KM, Price FW. Comparison of manual and automated endothelial cell density analysis in normal eyes and DSEK eyes. *Cornea*. 2013;32:567–873.
15. Luft N, Hirnschall N, Schuschitz S, Draschl P, Findl O. Comparison of 4 specular microscopes in healthy eyes and eyes with cornea guttata or corneal grafts. *Cornea*. 2015;34:381–386.
16. Gasser L, Reinhard T, Böhringer D. Comparison of corneal endothelial cell measurements by two non-contact specular microscopes. *BMC Ophthalmol*. 2015;15:87.
17. Kitzmann AS, Winter EJ, Nau CB, McLaren JW, Bourne WJ. Comparison of corneal endothelial cell images using a noncontact specular microscope and the confoscan 3 confocal microscope. *Invest Ophthalmol Vis Sci*. 2004;45:155.
18. Scarpa F, Ruggeri A. Development of a reliable automated algorithm for the morphometric analysis of human corneal endothelium. *Cornea*. 2016;35:1222–1228.
19. Al-Fahdawi S, Quawaji R, Al-Waisy AS, et al. A fully automated cell segmentation and morphometric parameter system for quantifying corneal endothelial cell morphology. *Comput Methods Programs Biomed*. 2018;160:11–23.
20. Piórkowski A, Gronkowska-Serafin J. Towards precise segmentation of corneal endothelial cells. International Conference on Bioinformatics and Biomedical Engineering (IWBBIO), Granada, Spain, 2015. Lecture Notes in Computer Science. 2015;9043:240–249.
21. Selig B, Vermeer KA, Rieger B, Hillenaar T, Luengo-Hendriks CL. Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy. *BMC Medical Imaging*. 2015;15:13.
22. Vigueras-Guillén JP, Andrinopoulou ER, Engel A, et al. Corneal endothelial cell segmentation by classifier-driven merging of oversegmented images. *IEEE Trans Med Imag*. 2018;37:2278–2289.
23. Fabijańska A. Automatic segmentation of corneal endothelial cells from microscopy images. *Biomed Signal Process Control*. 2019;47:145–148.
24. Daniel MC, Atzrodt L, Bucher F, et al. Automated segmentation of the corneal endothelium in a large set of 'real-world' specular microscopy images using the U-net architecture. *Nature Sci Rep*. 2019;9:4752.
25. Kolluru C, Benetz BA, Joseph N, Menegay HJ, Lass JH, Wilson D. Machine learning for segmenting cells in corneal endothelium images. Proceedings of SPIE 2019. *San Diego, CA*. 2019;109500:109504G.
26. Nurzynska K. Deep learning as a tool for automatic segmentation of corneal endothelium images. *Symmetry*. 2018;10:60.
27. Vigueras-Guillén JP, Sari B, Goes SF, et al. Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation. *BMC Biom Engineer*. 2019;1:4.
28. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol*. 2016;12:878.
29. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 2015. Lecture Notes in Computer Science. 2015;9351:234–241.
30. Vigueras-Guillén JP, Lemij HG, van Rooij J, Vermeer KA, van Vliet LJ. Automatic detection of the

region of interest in corneal endothelium images using dense convolutional neural networks. Proceedings of SPIE, Medical Imaging 2019: Image Processing. San Diego, CA. 2019;10949:1094931.
31. Jégou S, Drozdzal M, Vázquez D, Romero A, Bengio Y. The one hundred layers tiramisu: fully convolutional DenseNets for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition Workshops 2017*. Honolulu, HI. 2017;1175–1183.
32. Ioffe S. Batch renormalization: towards reducing minibatch dependence in batch-normalized models. *31st Conference on Neural Information Processing Systems (NIPS)*. 2017, Long Beach, CA.
33. Clevert D-A, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). International Conference on Learning Representations (ICLR). San Juan, Puerto Rico, 2016.
34. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Machine Learn Res*. 2014;15:1929–1958.
35. Vigueras-Guillén JP, Engel A, Lemij HG, van Rooij J, Vermeer KA, van Vliet LJ. Improved accuracy and robustness of a corneal endothelial cell segmentation method based on merging superpixels. 15th International Conference on Image Analysis and Recognition (ICIAR), Póvoa de Varzim, Portugal. *Lecture Notes in Computer Science*. 2018;10882:631–638.
36. Beucher S, Meyer F. The morphological approach to segmentation: the watershed transformation. *Mathematical morphology in image processing*. New York, NY: Marcel Dekker; 1993;433–481.
37. Vigueras-Guillén JP, van Rooij J, Lemij HG, Vermeer KA, van Vliet LJ. Convolutional neural network-based regression for biomarker estimation in corneal endothelium microscopy images. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019;876–881.
38. Dozat T. Incorporating Nesterov momentum into Adam. International Conference on Learning Representations (ICLR) Workshop, San Juan, Puerto Rico, 2016:2013–2016.
39. Hirst LW, Ferris FL, Stark WJ, Fleishman JA. Clinical specular microscopy. *Invest Ophthalmol Vis Sci*. 1980;19:2–4.
40. Doughty MJ, Müller A, Zaman ML. Assessment of the reliability of human corneal endothelial cell-density estimates using a noncontact specular microscope. *Cornea*. 2000;19:148–158.
41. Heinzelmann S, Daniel MC, Maier PC, Reinhard T, Böhringer D. Automated cell counting using "deep learning" in donor corneas from organ culture achieves high precision and accuracy. *Klin Monatsbl Augenheilk*. 2019;236:1407–1412.
42. van Rooij J, Engel A, Remeijer L, van Vleijnenbreugel H, Wubbels R. Long-term functional and anatomical outcome after Descemet stripping automated endothelial keratoplasty: a prospective single-center study. *J Ophthalmol*. 2018;7320816:1–5.
43. Feng Y, Qu HQ, Ren J, Prahs P, Hong J. Corneal endothelial cell loss in femtosecond laser-assisted Descemet's stripping automated endothelial keratoplasty: a 12-month follow-up study. *Chinese Med J*. 2017;130:2927–2932.
44. Graffi S, Leon P, Nahum Y, et al. Outcomes of ultrathin Descemet stripping automated endothelial keratoplasty (UT-DSAEK) performed in eyes with failure of primary Descemet membrane endothelial keratoplasty (DMEK). *Br J Ophthalmol*. 2019;103:599–603.
45. van Cleynenbreugel H, Remeijer L, Hillenaar T. Descemet stripping automated endothelial keratoplasty: effect of intraoperative lenticule thickness on visual outcome and endothelial cell density. *Cornea*. 2011;30:1195–1200.
46. Guerra FP, Anshu A, Price MO, Price FW. Endothelial keratoplasty: fellow eyes comparison of Descemet stripping automated endothelial keratoplasty and Descemet membrane endothelial keratoplasty. *Cornea*. 2011;30:1382–1386.
47. Javadi MA, Feizi S, Jafari R, Hosseini SB, Safapour S. Factors influencing graft endothelial cell density after Descemet stripping automated endothelial keratoplasty. *J Ophthalm Vis Res*. 2018;13:10–16.
48. Doughty MJ. A prospective analysis of corneal endothelial polymegethism and cell density in young adult Asians. *Clin Exp Optom*. 2014;97:256–263.
49. Doughty MJ, Aakre BM. Further analysis of assessments of the coefficient of variation of corneal endothelial cell areas from specular microscopic images. *Clin Exp Optom*. 2008;91:438–446.
50. Doughty MJ, Fonn D. Pleomorphism and endothelial cell size in normal and polymegethous human corneal endothelium. *Int Contact Lens Clin*. 1993;20:116–123.