



OPEN

## ThicknessTool: automated ImageJ retinal layer thickness and profile in digital images

Daniel E. Maidana<sup>1,2,3</sup>, Shoji Notomi<sup>1</sup>, Takashi Ueta<sup>1</sup>, Tianna Zhou<sup>1,3</sup>, Danica Joseph<sup>1</sup>, Cassandra Kosmidou<sup>1</sup>, Josep Maria Caminal-Mitjana<sup>3</sup>, Joan W. Miller<sup>1</sup> & Demetrios G. Vavvas<sup>1</sup>✉

To develop an automated retina layer thickness measurement tool for the ImageJ platform, to quantitate nuclear layers following the retina contour. We developed the ThicknessTool (TT), an automated thickness measurement plugin for the ImageJ platform. To calibrate TT, we created a calibration dataset of mock binary skeletonized mask images with increasing thickness masks and different rotations. Following, we created a training dataset and performed an agreement analysis of thickness measurements between TT and two masked manual observers. Finally, we tested the performance of TT measurements in a validation dataset of retinal detachment images. In the calibration dataset, there were no differences in layer thickness between measured and known thickness masks, with an overall coefficient of variation of 0.00%. Training dataset measurements of immunofluorescence retina nuclear layers disclosed no significant differences between TT and any observer's average outer nuclear layer (ONL) ( $p = 0.998$ ), inner nuclear layer (INL) ( $p = 0.807$ ), and ONL/INL ratio ( $p = 0.944$ ) measurements. Agreement analysis showed that bias between TT vs. observers' mean was lower than between any observers' mean against each other in the ONL ( $0.77 \pm 0.34 \mu\text{m}$  vs  $3.25 \pm 0.33 \mu\text{m}$ ) and INL ( $1.59 \pm 0.28 \mu\text{m}$  vs  $2.82 \pm 0.36 \mu\text{m}$ ). Validation dataset showed that TT can detect significant and true ONL thinning ( $p = 0.006$ ), more sensitive than manual measurement capabilities ( $p = 0.069$ ). ThicknessTool can measure retina nuclear layers thickness in a fast, accurate, and precise manner with multi-platform capabilities. In addition, the TT can be customized to user preferences and is freely available to download.

Progressive photoreceptor cell death is a significant culprit in retinal degenerative diseases. Inasmuch, quantitation of this cell loss has been addressed by a myriad of approaches, predominantly in experimental models of retinal diseases, such as retinal detachment<sup>1,2</sup>. Among these methods, photoreceptor cell death assays<sup>3,4</sup>, outer nuclear layer cell counting<sup>5</sup>, and outer nuclear layer thickness<sup>6,7</sup>, have been used in animal models to quantitate photoreceptor degeneration. Given the technical and time constraints related to cell death assays and manual counting; to date, outer nuclear layer (ONL) thickness has been widely used as a practical proxy method to estimate the depth of photoreceptor cell death<sup>6-8</sup>.

The utility of retina ONL thickness quantitation to infer photoreceptor degeneration relies on manual calliper measurements. However, the advantage of this approach can be compromised as callipers are manually drawn and measured by an observer. In addition, given the curvature of the retina, manually positioning callipers perpendicular to the layer contour along the long axis can be difficult. Furthermore, given the high magnification of microscope images, a small area of interest is often analysed to expedite the analysis, which can lead to bias. As new imaging modalities also become available, a versatile method that can adapt to multiple imaging modalities is ideal. Therefore, a tool that can adapt to layer architecture and contour to measure thickness in a broad segment, in multiple layers, in either single images or large tiles, and in multiple platforms, is very compelling. Despite pioneering work done in this area<sup>9</sup>, to the best of our knowledge, there is currently no freely available tool for the ImageJ platform to automatically quantitate multiple layer thicknesses in large images and in different imaging modalities.

<sup>1</sup>From the Retina Service, Angiogenesis Lab, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA. <sup>2</sup>From the Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, Chicago, IL, USA. <sup>3</sup>From the Retina Service, Bellvitge Hospital, University of Barcelona, Barcelona, Spain. ✉email: Demetrios\_Vavvas@MEEI.HARVARD.EDU

The purpose of this work was to develop an automated retinal layer thickness measurement tool for the ImageJ platform, which can quantitate nuclear layers following the retina contour, with callipers as close to 1-pixel to each other. For this purpose, we developed the ThicknessTool (TT) and validated its accuracy by objective calibration and agreement analysis with two masked observers. We found that this measurement tool can provide accurate and precise thickness measurements. In addition, TT can process images from multiple imaging modalities.

## Materials and methods

All animals used in experiments and breeding adhered to the statement of the Association for Research in Vision and Ophthalmology (ARVO). Animal protocols were reviewed and approved by the Animal Care Committee of the Massachusetts Eye and Ear Infirmary. To validate the ThicknessTool in experimental model, a retinal detachment was induced in eight-week-old C57BL/6J mice, as previously described<sup>10</sup>. C57BL/6J mice were purchased from The Jackson Laboratories and maintained in a standard 12-h light/dark cycle.

**Digital image dataset.** *Calibration dataset.* We created an image dataset of fifty 8-bit mock images of  $1344 \times 1024$  pixels containing masks of increasing thickness, from 10 to 500-pixel thickness with 10-pixel increments. Following, we created a second dataset of images with a known mask area of  $1344 \times 200$  pixels with  $10^\circ$  rotation intervals, plus  $45^\circ$  and  $135^\circ$  rotations for a total of 21 images. Finally, to emulate layer thinning and thickening, we created a single mock image containing a mask with abrupt thickness changes, with segments of 200, 150, and 250 pixels. All images were stored in TIFF format, and the measured thickness was compared to the known standard.

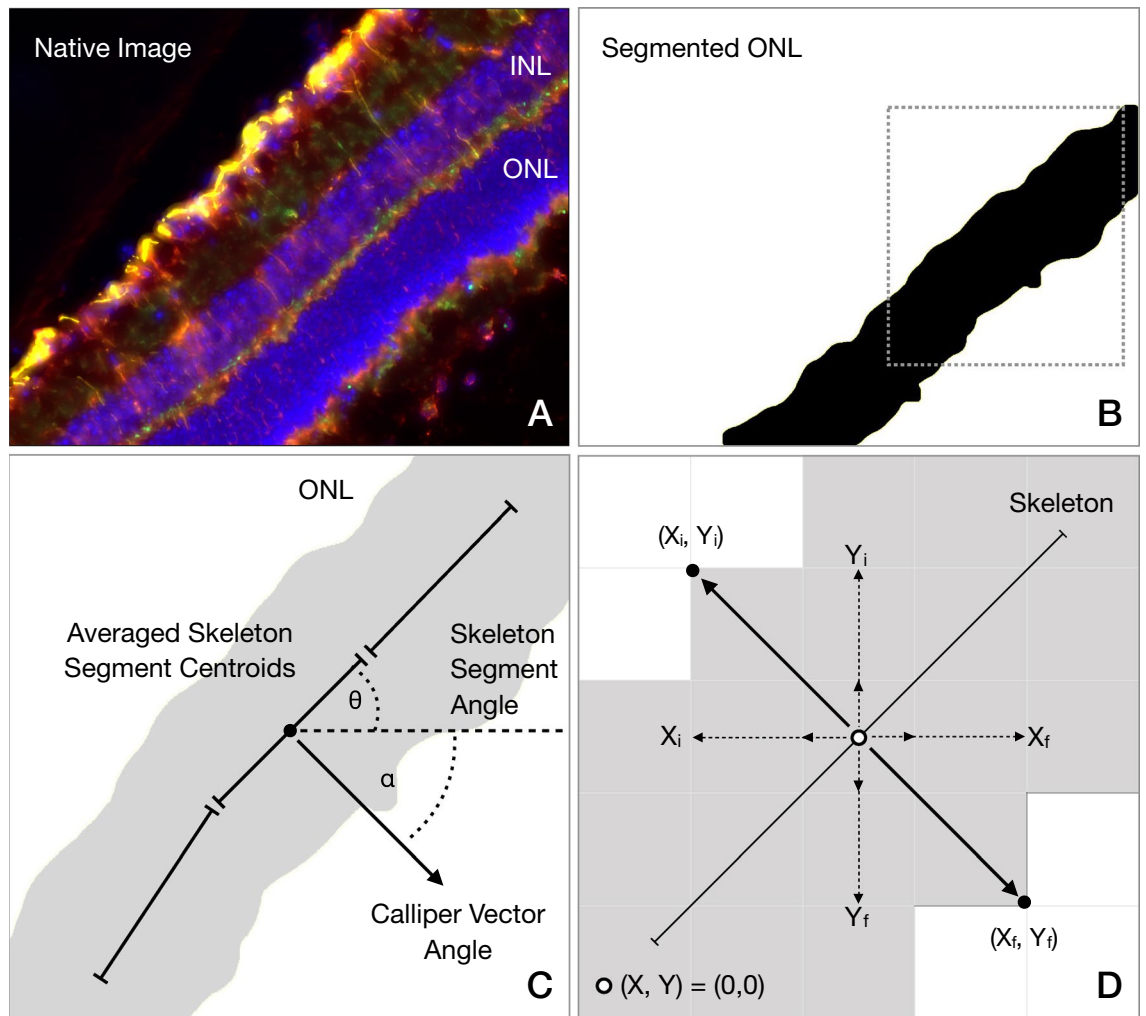
*Calibration criteria.* We defined the overall mean thickness limit of agreement to 1 pixel, and the following criteria for the ThicknessTool script calibration: (1) overall mean thickness equal to  $200 \pm 1$  pixel; (2) index of dispersion equal to 0, as the given area is a rectangle and all callipers should be parallel to each other with equal thickness; (3) single calliper measurement at  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ , and  $180^\circ$  rotations equal to 200 pixels; and (4) single calliper measurement equal to  $200 \text{ pixels} \pm 2 \text{ pixels}$  at the remaining rotations. This difference corresponds to the aliasing effect on area edges at rotations other than the specified on (3). The occurrence of uneven jagged edges requires a broader margin of error (2 pixels) than images with straight ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ) or even ( $45^\circ$  and  $135^\circ$ ) edges. For comparisons, the measured mean was tested against a hypothesized mean of 200 pixels. An equivalence test with a two one-sided tests (TOST) approach was used to assess minimum and maximum single calliper measurements, and the threshold difference considered equivalent to no difference was defined as 1 for images with even edges and 2 for images with uneven edges.

*Training dataset.* Immunofluorescence digital images of 4',6-diamidino-2-phenylindole (DAPI)-stained retinal cross-sections were obtained from the Angiogenesis Laboratory fluorescence microscopy database. We selected images from a murine retinal detachment model in order to evaluate TT performance with different ONL thicknesses and various retinal distortions. A total of 64 randomly selected images were included in this dataset and stored in TIFF format for further processing. Training dataset images were measured by an inexperienced observer who never performed the given task before, and by an experienced observer who had performed these measurements before. Images of murine retina displaying the outer (ONL) and inner nuclear layer (INL) were masked and given to observers. Each image was measured twice in a masked manner. Observers measured each layer with six individual callipers, placed across the retinal layer at their discretion. Single callipers were used to calculate mean, minimum, and maximum layer thickness.

*Validation dataset.* We randomly selected 16 images from 8 eyes of an experimental murine retinal detachment model. The layer thickness was manually assessed by an observer, who measured ONL and INL in the detached and attached retina, with six individual callipers for each layer.

**Thickness script.** We developed a script for the ImageJ platform (version 1.52p, <https://imagej.nih.gov/ij/>; provided in the public domain by the National Institutes of Health, Bethesda, MD, USA). To facilitate the use of the script, we designed a graphical user interface dialog to tailor performance parameters to users' preferences, including image scale, number of callipers, skeleton width, among others (Supplementary Fig. 1).

Briefly, retinal layers were segmented using a previously validated image processing algorithm (Fig. 1A–D)<sup>4</sup>. From segmented layers, a skeletonized mask was created with AnalyzeSkeleton function, as previously described by Arganda-Carreras and coauthors<sup>11</sup>. Following, thickness callipers were constructed as follows (Fig. 1D): (1) The segmented mask skeleton was fragmented in segments by user-defined  $n$  number of callipers; (2) Centroids from these segmented skeletons were isolated as single points and defined by the coordinates  $(X_0, Y_0)$ ; (3) The skeleton segment angle was calculated by an user-defined grid which evaluated the angle of the segmented mask skeleton in a given area; (4) From each centroid  $(X_0, Y_0)$ , a positive final vector and a negative initial vector were created perpendicularly to the segmented skeleton. The initial and final vectors were composed by a corresponding horizontal component vector  $(X_{step})$ , and a vertical component vector  $(Y_{step})$ ; (5) The  $X_{step}$  and  $Y_{step}$  components were created for the initial and final resolved vector at a 1-pixel distance from the centroid in horizontal and vertical directions, respectively. The preliminary initial  $(X_i, Y_i)$  and final  $(X_f, Y_f)$  coordinates were created from the centroid as described in Eq. (1); (6) Pixel values were evaluated in the retinal layer mask at *initial*  $(X_i, Y_i)$  and *final*  $(X_f, Y_f)$  coordinates. If the pixel value returned zero, this was interpreted as the retinal layer mask, and the  $X_{step}$  and  $Y_{step}$  components were increased by 1-pixel, to analyse the contiguous pixel; (7) This process was iterated until the pixel value was  $> 0$ , and this point was assumed as the *initial*  $(X_i, Y_i)$  and *final*



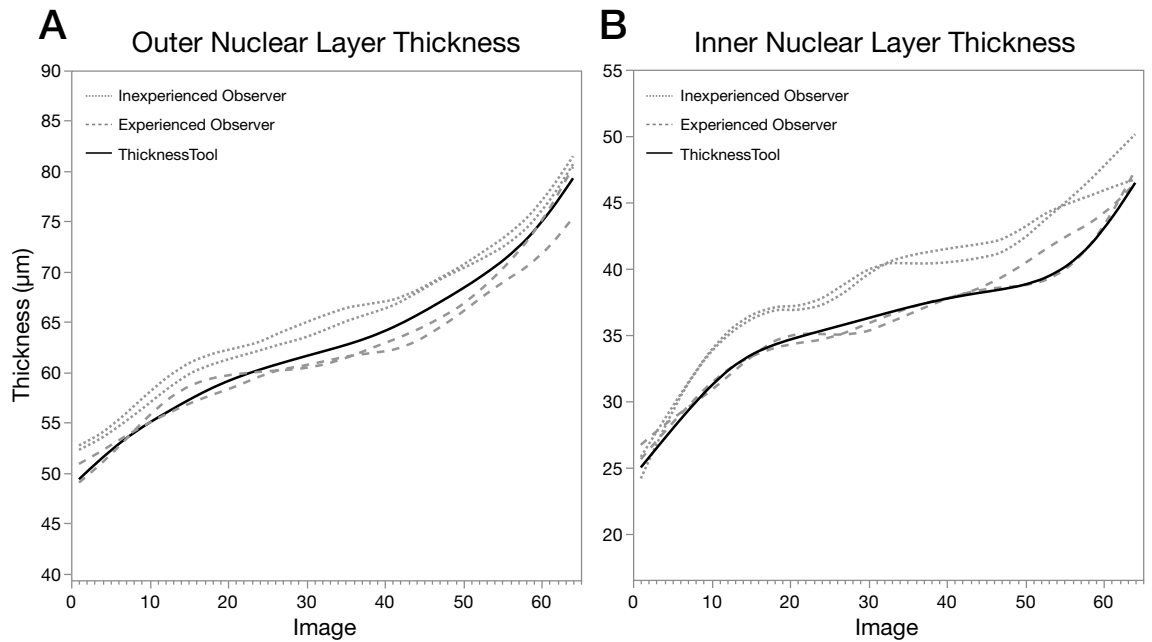
**Figure 1.** Outline of automated segmentation, thickness calliper constructs, and thickness measurements by ThicknessTool. **(A)** Representative native image acquired from mouse retina section and immunostaining. **(B)** Segmented outer nuclear layer (ONL) and mask. **(C)** Cropped area of the outer nuclear layer (ONL) mask. The angle of each skeleton segment ( $\theta$ ) was calculated respective to the zero-plane. The calliper angle ( $\alpha$ ), and resulting vector, were calculated as the supplementary angle of ( $\theta$ ). **(D)** Digital image representation displaying a pixel-grid and averaged skeleton segment calculated for the ONL mask, resulting vectors, and calliper endpoints.

$(X_f, Y_f)$  vector endpoint; (8) Thickness was calculated as the rounded Euclidean distance between *initial*  $(X_i, Y_i)$  and *final*  $(X_f, Y_f)$ , as seen in Eq. (2).

$$\begin{aligned}
 X_i &= \text{round}\left(\sin\left(\alpha \times \frac{\pi}{180}\right) \times X_{\text{step}} + X_0\right) \\
 Y_i &= \text{round}\left(Y_0 - \cos\left(\alpha \times \frac{\pi}{180}\right) \times Y_{\text{step}}\right) \\
 X_f &= \text{round}\left(X_0 - \sin\left(\alpha \times \frac{\pi}{180}\right) \times (X_{\text{step}} - \sqrt{1 + Y_{\text{step}}^2})\right) \\
 Y_f &= \text{round}\left(\cos\left(\alpha \times \frac{\pi}{180}\right) \times (Y_{\text{step}} - \sqrt{1 + Y_{\text{step}}^2}) + Y_0\right)
 \end{aligned}
 \tag{1}$$

$$\text{ThicknessCaliper} = \sqrt{(X_f - X_i)^2 + (Y_f - Y_i)^2}
 \tag{2}$$

where *i* = initial; *f* = final;  $\alpha$  = skeleton supplementary angle;  $X_0$  = x-axis centroid coordinate;  $Y_0$  = y-axis centroid coordinate;  $X_i$  = *initial* horizontal vector component endpoint x-axis coordinate;  $Y_i$  = *initial* vertical vector component endpoint y-axis coordinate;  $X_f$  = *final* horizontal vector component endpoint x-axis coordinate;  $Y_f$  = *final* vertical vector component endpoint y-axis coordinate;  $X_{\text{step}}$  = 1-pixel increment in the x-axis;  $Y_{\text{step}}$  = 1-pixel increment in the y-axis.



**Figure 2.** Parallelism plot of mean retinal thickness measurements of observers and ThicknessTool. Results were fitted to a cubic spline with a default lambda of 0.05. **(A)** Outer nuclear layer. **(B)** Inner nuclear layer. Dotted lines represent the first and second measurements by an inexperienced observer. Dashed lines represent the first and second measurements by an experienced observer. Solid lines represent measurements by ThicknessTool. Images are sorted by ThicknessTool values in ascending order ( $n = 64$ ).

**Statistical analyses.** Statistical analysis was performed with SAS software (2019 SAS, NC). Normality was assessed with Shapiro–Wilk test. Statistical significance for differences between groups was determined with t-test for matched pairs or Wilcoxon signed-rank test for matched pairs, t-test for independent samples, and one-way ANOVA with Tukey post hoc correction for multiple comparisons. For correlation analysis, values were expressed as Pearson correlation coefficient ( $r$ ) and statistical significance. Parallelism plots were constructed fitting a cubic spline with a default lambda of 0.05. Agreement analysis was performed by Bland–Altman plot method<sup>12,13</sup>, and as previously described<sup>4</sup>. Coefficient of variation (CoV) was calculated as the difference between measurements over their mean  $[(A - B)/(A + B)/2]$  and CoV% as  $\text{CoV} \times 100$ . The index of dispersion was calculated as variance-to-mean ratio. Results are expressed as mean  $\pm$  standard deviation (SD). A  $p$  value of  $< 0.050$  was considered statistically significant.

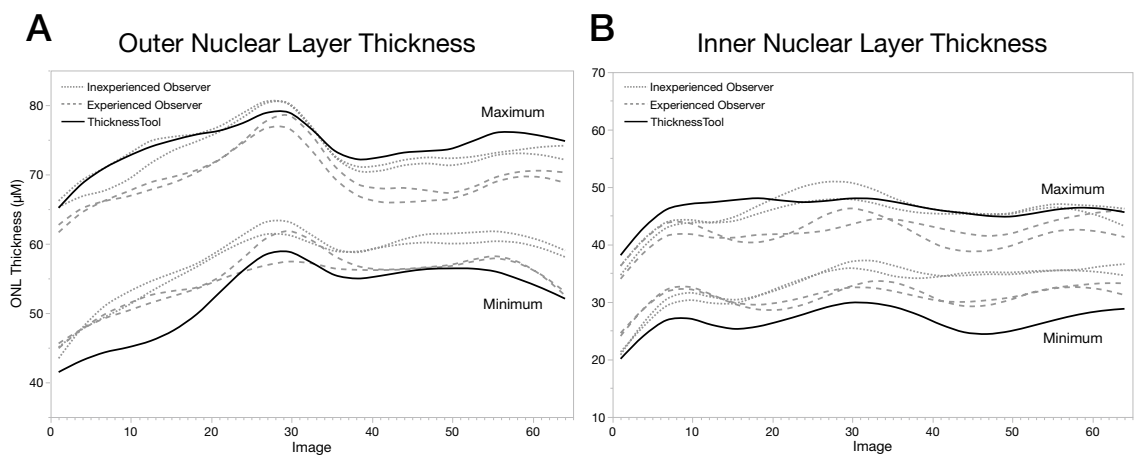
## Results

**ThicknessTool calibration.** To calibrate the ThicknessTool algorithm, we created three mock image datasets with masks of known area thickness and rotation. First, the TT was able to accurately measure areas of increasing thickness, from 10 to 500 pixels, with a bias of  $0.00 \pm 0.00$  (Supplementary Fig. 2) and thus, as expected, the correlation coefficient for the known vs. measured thickness was 1.00 ( $p < 0.001$ ). Following, we evaluated the performance of the algorithm at different layer rotations with 200-pixel masks at different angles. We predefined the known or theoretical mean, minimum, and maximum thickness to 200 pixels, a coefficient of variation (CoV) of 0.00, and a margin of error of 1-pixel. As seen in Supplementary Table 1, the mean thickness measured by ThicknessTool was  $199.88 \pm 0.25$  pixels, with a CoV of 0.00 for the overall sample. There were no statistically significant differences between the measured overall mean thickness and hypothesized mean ( $p = 0.06$ ). Mean thickness for images with even and jagged edges fulfilled the calibration criteria as previously defined, with CoV of 0.00 (Supplementary Fig. 2A–D). Equivalence Test of mean, minimum and maximum calliper thickness rejected both null hypotheses (both  $p < 0.001$ ), indicating that the observed difference does not exceed the defined threshold. Finally, we tested TT capability to detect true thinning and thickening in a mock image with decreasing and increasing thickness, respectively. TT accurately measured varying thickness ( $r = 1.00$ ,  $p < 0.001$ ), with a bias of  $0.00 \pm 0.00$  and no significant difference between known vs. measured markers ( $p = 1.000$ ) (Supplementary Fig. 2F). These results indicate that the ThicknessTool can automatically measure the different thicknesses and at different rotations in an accurate and precise manner.

**Training dataset mean layer thickness.** To evaluate the performance of observers and TT across different images and thickness, we first constructed a parallelism plot of mean ONL and INL thickness. As seen in Fig. 2, the inexperienced observer’s mean measurements were consistently higher than the experienced and TT, which in return were similar and parallel to each other in both ONL and INL layers, suggesting a proportionate performance. Correlation coefficient ( $r$ ) analysis between TT and observers’ measurements were  $> 0.88$  ( $p < 0.001$ ) for the ONL and  $> 0.84$  ( $p < 0.001$ ) for the INL (Supplementary Table 2 and 3).

	Outer nuclear layer			Inner nuclear layer			ONL/INL ratio	
	Mean $\pm$ SD	95% CI	<i>P</i> value*	Mean $\pm$ SD	95% CI	<i>P</i> value*	Mean $\pm$ SD	<i>P</i> value*
Inexperienced 1st	65.69 $\pm$ 7.94	63.70–67.67	0.394	39.23 $\pm$ 6.30	37.66–40.80	<b>0.044</b>	1.70 $\pm$ 0.26	0.910
Inexperienced 2nd	64.83 $\pm$ 7.78	62.88–66.77	0.814	39.03 $\pm$ 6.33	37.45–40.61	0.078	1.69 $\pm$ 0.26	0.768
Inexperienced Average	65.26 $\pm$ 7.78	63.31–67.20	0.613	39.13 $\pm$ 6.22	37.57–40.68	0.059	1.70 $\pm$ 0.26	0.832
Experienced 1st	62.44 $\pm$ 7.72	60.51–64.37	1.000	36.50 $\pm$ 5.70	35.07–37.92	1.000	1.74 $\pm$ 0.26	0.999
Experienced 2nd	61.56 $\pm$ 6.83	59.84–63.26	0.978	36.12 $\pm$ 5.29	34.79–37.44	1.000	1.73 $\pm$ 0.21	0.994
Experienced Average	62.00 $\pm$ 7.14	60.21–63.78	0.998	36.31 $\pm$ 5.17	35.01–37.60	1.000	1.73 $\pm$ 0.21	0.996
Observers average	63.47 $\pm$ 7.60	61.79–65.46	0.998	37.64 $\pm$ 5.68	36.34–39.10	0.807	1.71 $\pm$ 0.22	0.944
ThicknessTool	62.85 $\pm$ 7.33	61.01–64.68	–	36.12 $\pm$ 4.72	34.94–37.30	–	1.76 $\pm$ 0.22	–

**Table 1.** Training dataset mean thickness measured by observers and ThicknessTool. Values shown are means  $\pm$  standard deviations. ONL outer nuclear layer, INL inner nuclear layer, SD standard deviation, CI confidence interval, 1st first measurement, 2nd second measurement. \*Comparisons for all pairs vs ThicknessTool using Tukey–Kramer HSD test.



**Figure 3.** Parallelism plot of minimum and maximum retinal thickness measurements of observers and ThicknessTool. Results were fitted to a cubic spline with a default lambda of 0.05. (A) Outer nuclear layer. (B) Inner nuclear layer. Dotted lines represent the first and second measurements by an inexperienced observer. Dashed lines represent the first and second measurements by an experienced observer. Solid lines represent measurements by ThicknessTool ( $n = 64$ ).

The mean ONL thickness was  $65.26 \pm 7.78 \mu\text{m}$ ,  $62.00 \pm 7.14 \mu\text{m}$ , and  $62.85 \pm 7.33 \mu\text{m}$  for the inexperienced, experienced, and TT, respectively (Table 1). There were no significant differences between TT and any observer or average ONL measurements. The mean INL thickness was  $39.13 \pm 6.22 \mu\text{m}$ ,  $36.31 \pm 5.17 \mu\text{m}$ ,  $36.12 \pm 4.72 \mu\text{m}$  for the inexperienced, experienced and TT, respectively. In this group, there was only a significant difference between the inexperienced first measurement and TT ( $39.23 \pm 6.30 \mu\text{m}$  vs.  $36.12 \pm 4.72 \mu\text{m}$ ,  $p = 0.044$ ), and no significant differences between TT and any observer average INL measurements. The average for the inexperienced vs. experienced was not significantly different, in either the ONL ( $p = 0.219$ ) and INL ( $p = 0.097$ ), with overlapping confidence intervals. Moreover, there were no significant differences between TT and any observer or average ONL/INL ratio measurements, with overlapping 95% confidence intervals. Altogether, these results suggest that no significant difference was evidenced regarding the observers' prior experience. However, TT ONL and INL measurements stand more closely to experienced observer's results. Moreover, TT measurements are not significantly different from any observers' ONL and INL average.

**Training dataset minimum and maximum layer thickness.** Qualitative assessment of observers' thickness callipers showed no overlapping in repeated measurements, suggesting that measurements are not likely reproducible, as seen in Supplementary Fig. 3. Moreover, calliper over and undershoot was observed, which can lead to over and under measurement, respectively. In addition, the thickness vectors were not always perpendicular to the layer angle, hence prone to spurious results. Regardless of the lack of statistical differences in mean ONL and INL thickness between observers' average and TT measurements, we analysed the minimum and maximum calliper values for each image.

First, we constructed a parallelism plot of the minimum and maximum ONL and INL thickness calliper per image and observer. As seen in Fig. 3, the experienced observer had lower maximum and higher minimum ONL and INL values, without curve overlapping. Similarly, the inexperienced observer had higher minimum ONL

	Outer nuclear layer				Inner nuclear Layer			
	Minimum $\pm$ SD	P value*	Maximum $\pm$ SD	P value*	Minimum $\pm$ SD	P value*	Maximum $\pm$ SD	P value*
Inexperienced 1st	58.31 $\pm$ 8.76	<b>0.003</b>	73.94 $\pm$ 10.09	1.000	33.12 $\pm$ 7.30	< <b>0.001</b>	45.44 $\pm$ 6.90	0.997
Inexperienced 2nd	57.35 $\pm$ 8.14	<b>0.031</b>	72.78 $\pm$ 10.06	0.987	33.39 $\pm$ 6.97	< <b>0.001</b>	45.86 $\pm$ 6.91	1.000
Inexperienced overall	56.28 $\pm$ 8.54	0.218	75.00 $\pm$ 10.33	0.999	31.73 $\pm$ 7.12	< <b>0.001</b>	46.96 $\pm$ 6.97	0.999
Experienced 1st	55.51 $\pm$ 8.19	0.534	70.03 $\pm$ 9.67	0.198	31.01 $\pm$ 5.88	<b>0.002</b>	42.45 $\pm$ 7.09	<b>0.048</b>
Experienced 2nd	54.60 $\pm$ 7.67	0.899	69.09 $\pm$ 8.53	<b>0.049</b>	30.95 $\pm$ 4.92	<b>0.002</b>	41.91 $\pm$ 7.35	<b>0.012</b>
Experienced overall	53.32 $\pm$ 7.83	0.999	71.34 $\pm$ 9.43	0.667	29.41 $\pm$ 4.89	0.207	44.53 $\pm$ 7.83	0.858
Observers overall	52.44 $\pm$ 8.07	1.000	75.74 $\pm$ 10.29	0.990	28.36 $\pm$ 5.52	0.812	47.92 $\pm$ 7.78	0.894
ThicknessTool	52.73 $\pm$ 8.20	–	74.30 $\pm$ 8.88	–	26.76 $\pm$ 5.23	–	46.28 $\pm$ 5.72	–

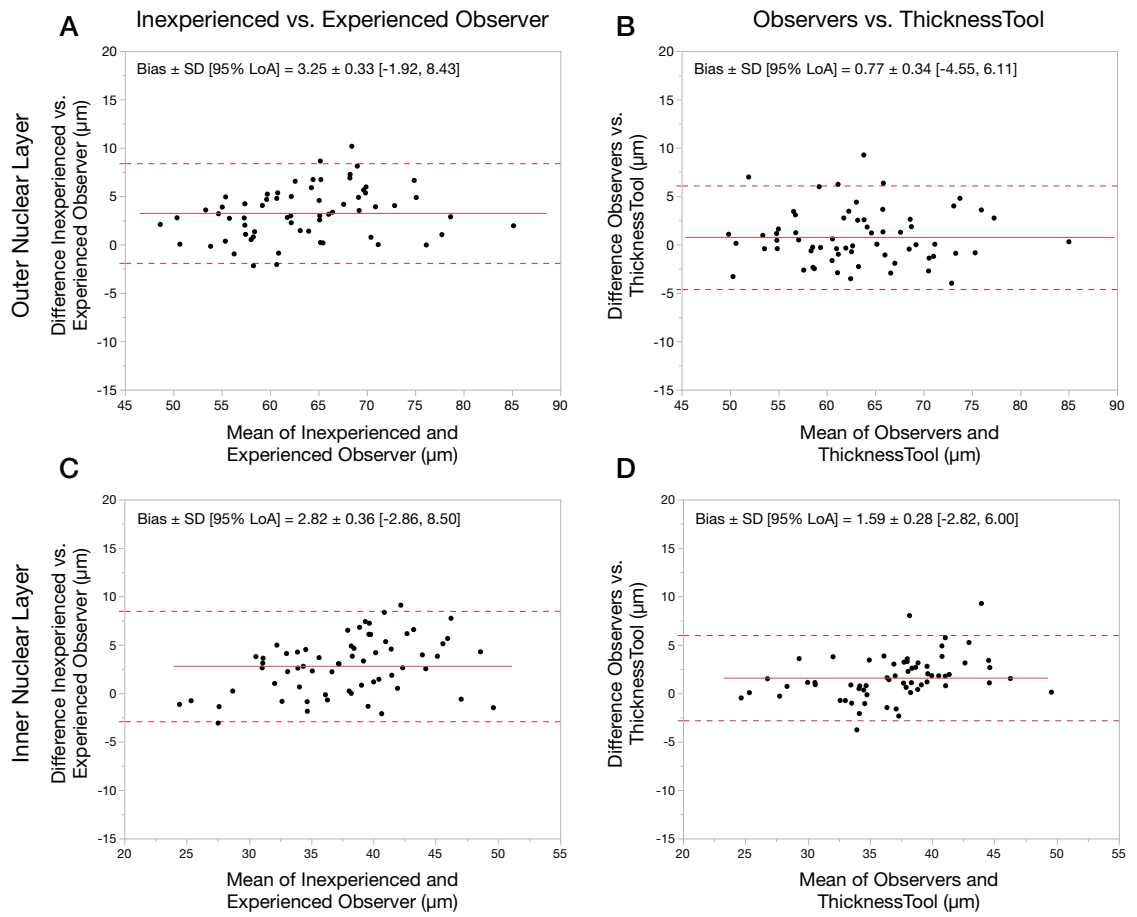
**Table 2.** Training dataset minimum and maximum thickness measured by observers and ThicknessTool. Values shown are means  $\pm$  standard deviations. *SD* standard deviation, *1st* first measurement, *2nd* second measurement. \*Comparisons for all pairs vs ThicknessTool using Tukey–Kramer HSD test.

	Bias $\pm$ SE	Lower 95% LoA	Upper 95% LoA	LoA interval	P value*
<b>Outer nuclear layer</b>					
Inexperienced 1st vs. inexperienced 2nd	0.86 $\pm$ 0.27	– 3.37	5.09	8.47	<b>0.002</b>
Experienced 1st vs. experienced 2nd	0.880 $\pm$ 0.36	– 4.76	6.52	11.29	<b>0.018</b>
Experienced mean vs. inexperienced mean	3.25 $\pm$ 0.33	– 1.92	8.42	10.35	< <b>0.001</b>
ThicknessTool vs. inexperienced mean	2.40 $\pm$ 0.42	– 4.19	8.99	13.17	< <b>0.001</b>
ThicknessTool vs. experienced mean	0.85 $\pm$ 0.34	– 4.48	6.18	10.66	<b>0.015</b>
ThicknessTool vs. observers mean	0.77 $\pm$ 0.34	– 4.56	6.10	10.66	<b>0.028</b>
<b>Inner nuclear layer</b>					
Inexperienced 1st vs. inexperienced 2nd	0.20 $\pm$ 0.27	– 4.03	4.43	8.47	0.465
Experienced 1st vs. experienced 2nd	0.38 $\pm$ 0.46	– 6.83	7.59	14.43	0.412
Experienced mean vs. inexperienced mean	2.82 $\pm$ 0.36	– 2.82	8.46	11.29	< <b>0.001</b>
ThicknessTool vs. inexperienced mean	3.00 $\pm$ 0.40	– 3.27	9.27	12.54	< <b>0.001</b>
ThicknessTool vs. experienced mean	0.18 $\pm$ 0.25	– 3.74	4.10	7.84	0.462
ThicknessTool vs. Observers mean	1.59 $\pm$ 0.28	– 2.80	5.98	8.78	< <b>0.001</b>

**Table 3.** Training dataset agreement analysis of thickness measurements between observers and ThicknessTool. *SE* standard error, *LoA* limit of agreement, *1st* first measurement, *2nd* second measurement. \*T-Test for matched pairs Values shown are means  $\pm$  standard errors.

and INL values without curve overlapping. These results suggest that experienced observers' measurements are not reproducible. In addition, there were no significant differences between TT and any observers' overall ONL minimum and maximum measurements (Table 2). However, we found significant differences across individual observer's minimum and maximum measurements and between the inexperienced overall minimum INL and TT measurements ( $p < 0.001$ ). In addition, the coefficient of variability between measurements was significantly different between inexperienced and observers' average measurements for both ONL and INL ( $p < 0.001$ ) and between the experienced observer and observers' average measurements for the ONL ( $p < 0.009$ ) (Supplementary Table 4). In summary, given the qualitative assessment of observers' callipers' lack of repeatability together with the quantitative analyses between manual callipers and a calibrated automated algorithm, results indicate that manual callipers measurements are not accurate nor reproducible.

**Agreement analysis between manual and thickness measurements.** We performed a bias analysis using Bland and Altman method<sup>13</sup>. For the ONL group, as seen in Table 3, we found statistically significant differences when comparing paired measurements, with a greater bias in inexperienced mean vs. experienced mean measurements (3.25  $\pm$  0.33  $\mu\text{m}$ ,  $p < 0.001$ ). In contrast, we found the lowest bias when comparing TT vs. observers' means (0.77  $\pm$  0.34  $\mu\text{m}$ ,  $p = 0.028$ ). For the INL group, we found statistically significant differences when comparing inexperienced mean vs. experienced mean ( $p < 0.001$ ), TT vs. inexperienced mean ( $p < 0.001$ ), and TT vs. observers' mean ( $p < 0.001$ ). Most importantly, the bias between TT vs. observers' INL mean (1.59  $\pm$  0.28  $\mu\text{m}$ ) was lower than the those between any observers mean against each other (2.82  $\pm$  0.36  $\mu\text{m}$ ). In addition, Bland–Altman plots displayed random variability within measurements (Fig. 4). In conclusion, these



**Figure 4.** Agreement analysis of observers and ThicknessTool by Bland–Altman plots. Agreement between inexperienced and experienced average mean (A) ONL and (C) INL thickness. Agreement between observers and ThicknessTool (B) ONL and (D) INL thickness ( $n = 64$ ). *SD* standard deviation, *LoA* limit of agreement.

results indicate that ThicknessTool can measure ONL and INL with lower bias than observers' average against each other for both ONL and INL.

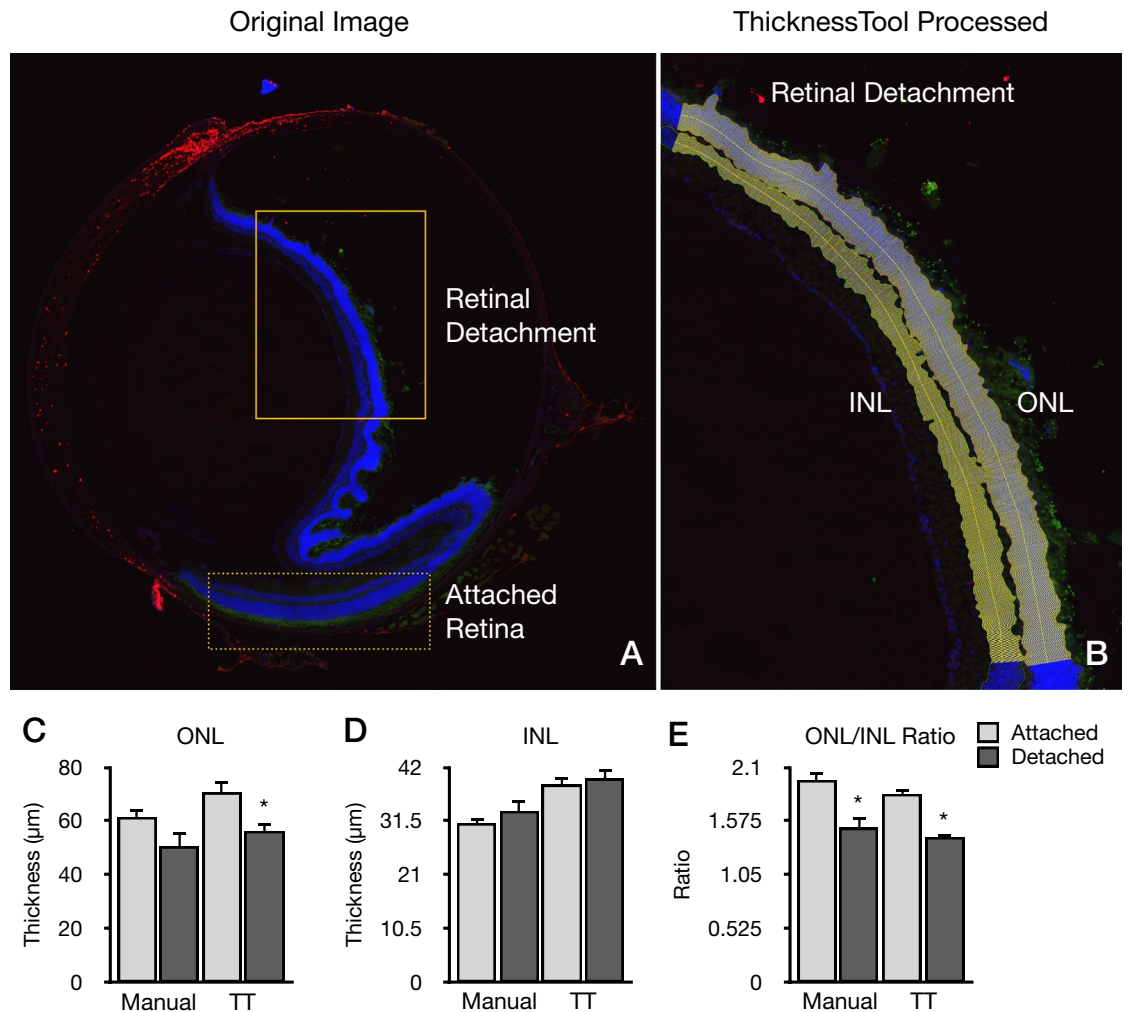
**Validation of thickness in a retinal detachment model.** To definitively validate the TT, we tested the tool in an experimental setting using a retinal detachment model, which causes photoreceptor cell death and subsequent ONL thinning and distortion (Fig. 5). We quantitated the ONL and INL thickness manually and using TT (Table 4). Interestingly, manual measurements failed to show significant ONL thinning in the detached retina ( $p = 0.069$ ), in comparison to TT ( $p = 0.006$ ). No differences were observed in the INL between the attached and detached retina by both methods as expected. Both methods showed significant differences in the ONL/INL ratio in the detached retina. These results suggest that TT can be more sensitive and detect significant true thinning beyond manual measurement capabilities. Collectively, these results suggest that TT is a sensitive, precise, and reliable tool to measure retinal nuclear layers in experimental models of disease.

**Application of thickness in digital images.** In order to examine the applicability of TT in different image modalities, we quantitated the thickness in various digital images. As seen in Fig. 6, TT was capable of quantitating the ONL thickness and corresponding profile in histology sections. Moreover, we challenged this tool in digital images of macular spectral-domain optical coherence tomography, with satisfactory results. Finally, we tested the performance in the quantitative assessment of retinal vessels in fluorescein angiogram images. The TT was able to draw accurate callipers with the correct vector in all cases. Collectively, these results indicate that ThicknessTool is a versatile tool with multi-platform capabilities.

## Discussion

In this study, we developed a sensitive, precise, and reliable tool to measure retinal layers. The performance of this tool was tested in calibration, training, and validation datasets with adequate and reliable performance. Collectively, results validate ThicknessTool for automated retinal thickness measurement. We speculate that this tool can aid in achieving precise, accurate, and fast measurements in multi-platform digital images.

In the development of new measurement tools, calibration images are mandatory to benchmark and finetune the performance of image processing algorithms to a standard<sup>14</sup>. We designed mock images with various known



**Figure 5.** Validation of ThicknessTool in a retinal detachment model. (A) Representative image of mouse retina section of a retinal detachment displaying the attached (dotted line box) and detached retina (solid line box). (B) Representative image of ThicknessTool measurement of detached retina inner nuclear layer (INL) and outer nuclear layer (ONL) at 1-pixel interval callipers. (C) Outer nuclear layer (ONL) and (D) Inner nuclear layer (INL) thickness measurements from observer manual callipers and ThicknessTool in the attached and detached retina. (E) ONL/ONL + INL ratio measurements of observer manual callipers and ThicknessTool (n = 8 per group, \*p ≤ .05).

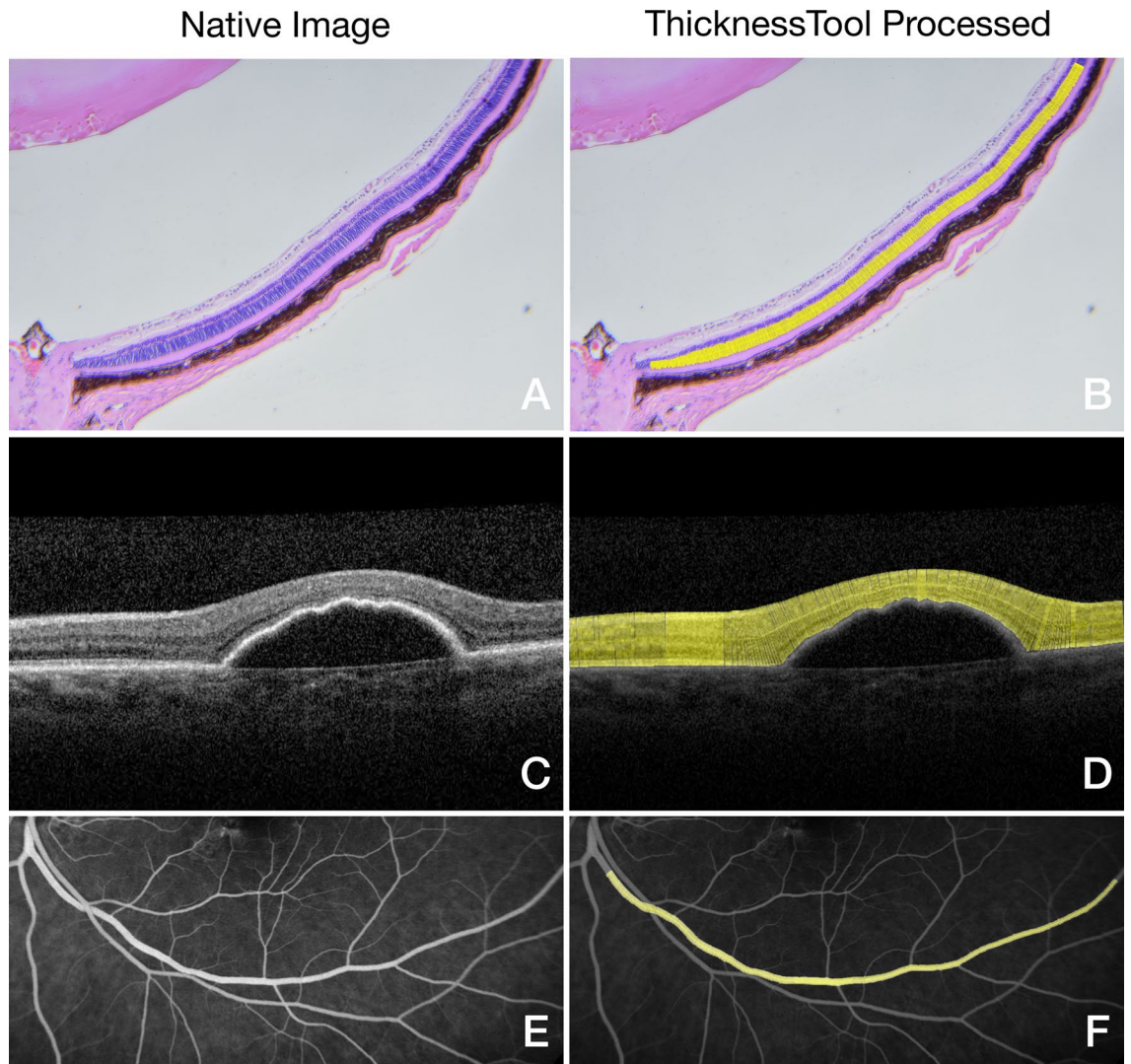
	Outer nuclear layer			Inner nuclear layer			ONL/ONL + INL Ratio		
	Attached	Detached	P value*	Attached	Detached	P value*	Attached	Detached	P value*
Manual	61.24 ± 8.39	50.47 ± 12.80	0.069	31.01 ± 2.58	33.43 ± 6.40	0.347	1.97 ± 0.22	1.51 ± 0.28	<b>0.003</b>
ThicknessTool	70.82 ± 10.04	56.14 ± 8.25	<b>0.006</b>	38.58 ± 4.34	39.79 ± 5.24	0.624	1.83 ± 0.16	1.41 ± 0.12	<b>&lt; 0.001</b>

**Table 4.** Validation dataset manual and ThicknessTool measurements in a retinal detachment model. Values shown are means ± standard deviations. ONL outer nuclear layer, INL inner nuclear layer. \*T-Test for independent samples.

thickness ranging from 10 pixels (given a single nuclei size in our images) to 500 pixels (twice the outer nuclear layer measured in our dataset). In addition, we created a dataset with multiple mask rotations to mimic the random mask orientations. This step is necessary as TT callipers are constructed in a vectorial manner, in number and angles as the user or region of interest requires. Moreover, we tested TT capability to detect true thinning and thickening in a mock image with decreasing and increasing thickness, respectively. We found that the TT can automatically and accurately measure thickness when challenged with different thicknesses and rotations.

Accurate assessment of photoreceptor cell death currently relies on manual counting by masked observers. Ideally, two masked two observers quantitate retina thickness, and their average is used to determine the overall outcome. We found that single observers' measurements were prone to over and underestimation, regardless of





**Figure 6.** ThicknessTool application in digital images. (A,B) Haematoxylin–eosin staining of retinal cryosection. (C,D) Macular spectral-domain optical coherence tomography scan showing a pigment epithelial detachment in age-related macular degeneration. (E,F) Fluorescein angiogram of retinal vessels.

their previous experience. In addition, observers draw incorrectly angled callipers, which can indirectly contribute to the abovementioned bias. Moreover, callipers were not located in corresponding locations in repeated same-image analysis, which dampens the reproducibility of the measurements. Altogether, this leads to a lack of accurate or reproducible retina thickness quantitation. This matter is of critical importance, as this bias can lead to  $\alpha$  or  $\beta$  errors. For example, finding no difference in retinal thickness in cases where a neuroprotective treatment is effective, or even worse, finding a significant difference in retinal thickness in an ineffective treatment<sup>15,16</sup>. Such errors are not only misleading to the scientific community but also high priced.

Regardless of the performance of manual measurements to date, it remains the gold standard. For purposes of method agreement, we refrained from relying on mean comparison or correlation, as Bland and Altman have previously reported in their seminal work<sup>12,17</sup>. As authors describe, such agreement must be quantified, and cannot be assessed by mean comparison or hypothesis testing, as it does not exist in a present or absent binary form<sup>13</sup>. In addition, high correlation does not necessarily imply a high agreement between two methods, as it quantifies the degree to which two methods are related<sup>18</sup>. Agreement analysis showed lower bias of TT against observers than observers' average against each other for both ONL and INL. Taken into consideration TT capability to detect true mask thinning and thickening, collectively, these results confer to TT a higher measurement agreement.

Moreover, the automated thickness analysis by TT yields a series of convenient metrics that remain hidden otherwise in manual measurements. On this note, Byun et al. have developed pioneering methods to estimate the ONL thickness exclusively<sup>9</sup>. However, their approach was not validated by thickness agreement analysis with manual observers. In their work, authors also described a method to quantify structural distortions in retinal tissue before and after injury, also known as distortion index. In essence, the authors stated that structural changes in the ONL are directly proportional to the distortion index. In our work, we estimated overall distortion in a similar approach using the index of dispersion (ID), calculated as the variance-to-mean ratio, as previously

defined by Cox and Lewis<sup>19</sup>. Similarly, the index of dispersion allows the assessment of the overall calliper variation, which can provide detailed layer architecture data. Briefly, an ID of 0 represents no dispersion; therefore, we used the ID for calibration purposes, as mock images of known and constant thickness were uniform rectangles (ID = 0). We speculate that the index of dispersion can be equally used for further morphometric assessment.

Several limitations to this work should be considered. First, despite the calibration analysis performed beforehand, the validation of TT can only be performed against manual measurements. This process is cumbersome, as observers' measurements carry an intrinsic bias. However, observers showed comparable performance, with overlapping confidence intervals. Second, TT analysis should be performed in areas distant from image edges, as skeleton are trimmed to avoid calliper-to-end image measurements. Despite the fact that TT allows skeleton and edge trimming, users should evaluate processed images for qualitative control. Third, TT measures thickness in binary segmented masks. Therefore, its performance is as strong as the segmentation algorithm used. In our case, we build this tool based on a previously validated algorithm for retina layer segmentation<sup>4</sup>. We recommend users to first tailor and validate their segmentation algorithm for optimal performance. Finally, regardless of the potential image modality application, this tool has been validated only for ONL and INL measurements in DAPI nuclei-stained murine retina cryosections. We believe future work can help expand the applicability of this tool for three-dimensional images, including upcoming imaging platforms.

In summary, we developed a fast, accurate, and precise retinal thickness measurement tool with multi-platform capabilities. In addition, the TT can be customized to user-defined preferences. This includes calliper number with a 1-pixel minimum distance between callipers, skeleton and edge trimming, and automated or manual layer segmentation. In addition, given the potential use with multi-platform capabilities, the source code can be easily modified to fit different applications. Finally, this tool is freely available to download and use. We expect this measurement tool could improve the outcomes and reduce bias in retina research.

Received: 19 March 2020; Accepted: 28 August 2020

Published online: 28 October 2020

## References

1. Cook, B., Lewis, G. P., Fisher, S. K. & Adler, R. Apoptotic photoreceptor degeneration in experimental retinal detachment. *Invest. Ophthalmol. Vis. Sci.* **36**, 990–996 (1995).
2. Arroyo, J. G., Yang, L., Bula, D. & Chen, D. F. Photoreceptor apoptosis in human retinal detachment. *Am. J. Ophthalmol.* **139**, 605–610 (2005).
3. Gavrieli, Y., Sherman, Y. & Ben-Sasson, S. A. Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *J. Cell Biol.* **119**, 493–501 (1992).
4. Maidana, D. E. *et al.* A novel ImageJ macro for automated cell death quantitation in the retina. *Invest. Ophthalmol. Vis. Sci.* **56**, 6701–6708 (2015).
5. Byun, J. *et al.* Automated tool for the detection of cell nuclei in digital microscopic images: Application to retinal images. *Mol. Vis.* **12**, 949–960 (2006).
6. Trichonas, G. *et al.* Receptor interacting protein kinases mediate retinal detachment-induced photoreceptor necrosis and compensate for inhibition of apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21695–21700 (2010).
7. Matsumoto, H. *et al.* Strain difference in photoreceptor cell death after retinal detachment in mice. *Invest. Ophthalmol. Vis. Sci.* **55**, 4165–4174 (2014).
8. Murakami, Y. *et al.* Receptor interacting protein kinase mediates necrotic cone but not rod cell death in a mouse model of inherited degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14598–14603 (2012).
9. Byun, J. *et al.* Quantifying structural distortions in retinal tissue before and after injury. in *Workshop on Multiscale Biological Imaging, Data Mining & Informatics, Santa Barbara, CA, USA 10–11* (2006).
10. Matsumoto, H., Miller, J. W. & Vavvas, D. G. Retinal detachment model in rodents by subretinal injection of sodium hyaluronate. *J. Vis. Exp.* <https://doi.org/10.3791/50660> (2013).
11. Arganda-Carreras, I., Fernández-González, R., Muñoz-Barrutia, A. & Ortiz-De-Solorzano, C. 3D reconstruction of histological sections: Application to mammary gland tissue. *Microsc. Res. Tech.* **73**, 1019–1029 (2010).
12. Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310 (1986).
13. Bland, J. M. & Altman, D. G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* **8**, 135–160 (1999).
14. Troscianko, J. & Stevens, M. Image calibration and analysis toolbox—a free software suite for objectively measuring reflectance, colour and pattern. *Methods Ecol. Evol.* **6**, 1320–1331 (2015).
15. Kim, H.-Y. Statistical notes for clinical researchers: Type I and type II errors in statistical decision. *Restor. Dent. Endod.* **40**, 249 (2015).
16. Rothman, K. J. Curbing type I and type II errors. *Eur. J. Epidemiol.* **25**, 223–224 (2010).
17. Bland, J. & Altman, D. Measuring agreement in method comparison studies with heteroscedastic measurements. *Stat. Methods Med. Res.* **8**, 135–160 (1999).
18. Giavarina, D. Understanding bland Altman analysis. *Biochem. Med.* **25**, 141–151 (2015).
19. Cox, D. & Lewis, P. *The Statistical Analysis of Series of Events* (Springer, Netherlands, 1966).

## Author contributions

D.E.M. and D.G.V. designed research. D.E.M. coded the ThicknessTool, performed the statistical analysis, and drafted the manuscript. S.N., T.U., T.Z., D.J., and C.K. measured retinal thickness from the digital image dataset and contributed to the manuscript. J.W.M., and D.G.V. reviewed the data and made substantial intellectual contributions to the research design and manuscript. Please visit <https://github.com/DanielMaidana/ThicknessTool> to download the ThicknessTool. ThicknessTool is licensed under the GNU General Public Licence.

## Funding

This work was supported by Bayer Healthcare Global Ophthalmology Awards Program (DEM); the Yeatts Family Foundation (DGV, JWM); Loefflers family foundation (JWM, DGV); a Macula Society Research Grant award

(DGV); a Physician Scientist Award (DGV) and an unrestricted grant (JWM) from the Research to Prevent Blindness Foundation; NEI Grant R21EY023079-01A1 (DGV); and NEI Grant EY014104 (MEEI Core Grant). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-75501-y>.

**Correspondence** and requests for materials should be addressed to D.G.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020