

RESEARCH ARTICLE

Automated thresholding algorithms outperform manual thresholding in macular optical coherence tomography angiography image analysis

Jan Henrik Terheyden¹ [✉], Maximilian W. M. Wintergerst¹ [✉], Peyman Falahat¹, Moritz Berger², Frank G. Holz¹, Robert P. Finger¹ ^{*}

1 Department of Ophthalmology, University of Bonn, Bonn, Germany, **2** Department of Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany

 These authors contributed equally to this work.

* Robert.Finger@ukbonn.de



OPEN ACCESS

Citation: Terheyden JH, Wintergerst MWM, Falahat P, Berger M, Holz FG, Finger RP (2020) Automated thresholding algorithms outperform manual thresholding in macular optical coherence tomography angiography image analysis. PLoS ONE 15(3): e0230260. <https://doi.org/10.1371/journal.pone.0230260>

Editor: Ireneusz Grulkowski, Nicolaus Copernicus University, POLAND

Received: September 10, 2019

Accepted: February 25, 2020

Published: March 20, 2020

Copyright: © 2020 Terheyden et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data proving the main findings of the study are contained within the manuscript. The Optical Coherence Tomography Angiography images of all participants cannot be made publicly available as public availability of the data might compromise participant privacy. All participants consented to scientific analysis of the data by the researchers involved but not to public data sharing (according to the consent form approved by the Ethics committee, University of Bonn). As subjects might be identifiable by the

Abstract

Introduction

For quantification of Optical Coherence Tomography Angiography (OCTA) images, Vessel Density (VD) and Vessel Skeleton Density (VSD) are well established parameters and different algorithms are in use for their calculation. However, comparability, reliability and ability to discriminate healthy and impaired macular perfusion of different algorithms are unclear, yet, of potential high clinical relevance. Hence, we assessed comparability and test-retest reliability of the most common approaches.

Materials and methods

Two consecutive 3×3mm OCTA en face images of the superficial and deep retinal layer were acquired with swept-source OCTA. VD and VSD were calculated with manual thresholding and six automated thresholding algorithms (Huang, Li, Otsu, Moments, Mean, Percentile) using ImageJ and compared in terms of intra-class correlation coefficients, measurement differences and repeatability coefficients. Receiver operating characteristic analyses (healthy vs. macular pathology) were performed and Area Under the Curve (AUC) values were calculated.

Results

Twenty-six eyes (8 female, mean age: 47 years) of 15 patients were included (thereof 15 eyes with macular pathology). Binarization thresholds, VD and VSD differed significantly between the algorithms and compared to manual thresholding ($p < 0.0001$). Inter-measurement differences did not differ significantly between patients with healthy versus pathologic maculae ($p \geq 0.685$). Reproducibility was higher for the automated algorithms compared to manual thresholding on all measures of reproducibility assessed. AUC was significantly

high-quality OCTA images in the future, the image dataset cannot be made publically available. However, anonymized image data sharing with qualified researchers who meet the criteria for access to confidential data will be possible upon request from the University of Bonn, Department of Ophthalmology at phone number +4922828715505

Funding: This research was supported by Carl Zeiss Meditec AG, Jena (www.zeiss.com) which provided the Optical Coherence Tomography Angiography device used in this study, funding by the German Scholars Organization/Else Kröner Fresenius Stiftung (GSO/EKFS 16, www.ekfs.de) and the BONFOR GEROK Program, Faculty of Medicine, University of Bonn, (Grant No O-137.0028, MWMW). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: The Department of Ophthalmology, University of Bonn received imaging devices from Heidelberg Engineering, Optos, Carl Zeiss Meditec, and CenterVue. JHT: No additional conflicts of interest. MWMW: travel grant from DigiSight Technologies, grant from Heine Optotechnik, imaging devices from D-EYE, consultant of Heine Optotechnik GmbH. PF: No additional conflicts of interest. MB: No additional conflicts of interest. FGH: consultant of Acucela, Bayer, Bioeq, Boehringer-Ingelheim, Genentech/Roche, Heidelberg Engineering, Novartis, and Thea, grants from Acucela, Allergan, Bayer, Bioeq, Genentech/Roche, Merz, NightstarX, and Novartis, honoraria from Allergan, Bayer, Carl Zeiss Meditec, Genentech/Roche, Heidelberg Engineering, and Novartis. RPF: consultant of Bayer, Novartis, Santen, Opthea, Novelon, Santhera, Inositec, Alimera and RetinalImplant, honoraria from Bayer, Ellex, Alimera. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

higher for the Mean algorithm compared to the manual approach with respect to the superficial retinal layer.

Conclusions

Automated thresholding algorithms yield a higher reproducibility of OCTA parameters and allow for a more sensitive diagnosis of macular pathology. However, different algorithms are not interchangeable nor results readily comparable. Especially the Mean algorithm should be investigated in further detail. Automated thresholding algorithms are preferable but more standardization is needed for clinical use.

Introduction

Optical coherence tomography angiography (OCTA) provides depth resolved high resolution images of the retinal and choroidal blood flow. [1,2] A number of different approaches are available to quantify OCTA image data, but to date both their reproducibility as well as their comparability are unclear.

Image processing is a crucial step when generating comparable and reliable quantitative data from retinal images. For the calculation of global vessel density (VD) from OCTA images, definition of a threshold for image binarization is essential. The three most common solutions are manual binarization methods, automated binarization methods using open source software and automated binarization using commercial software. [3–9] Manual and semi-automated methods often gain a threshold for binarization based on the signal within the vessel-free foveal avascular zone (FAZ). Automated algorithms use e.g. the histogram of a complete OCTA image or local clusters to obtain a threshold. Advantages of manual/semiautomated and automated binarization methods, as compared to commercial software, include high transparency for research purposes. Besides these two options, commercial software is available from various device manufacturers as well as other sources, mostly using proprietary image processing algorithms not publicly available. Additionally, fixed threshold and machine learning approaches are available. [10,11]

A recent study by Rabiolo and colleagues [12] found significant differences in VD calculations between manual and automated approaches but used arbitrary cut-offs for their manual binarization and did not assess test-retest reliability of two consecutive examinations. Thus, in this study we assessed both repeatability and comparability of manual binarization based on the FAZ and six automated algorithms for OCTA image binarization in patients with and without macular disease.

Materials and methods

Subject recruitment

Participants both healthy and with any macular pathology impairing the vasculature were consecutively recruited at the Department of Ophthalmology, University of Bonn, Germany, between April and August 2018. Ethical approval was obtained from the ethics committee of the University of Bonn (approval ID 089/08) and informed consent was obtained from all study participants prior to study initiation after explanation of the nature and possible consequences of the study. The study was conducted in adherence to the tenets of the Declaration of

Helsinki. Exclusion criteria were artefacts diminishing image quality, OCTA signal strength index < 7, inability to fixate and clinically relevant media opacities.

Image acquisition

Two OCTA images per eye were consecutively obtained using a swept-source OCTA device (Zeiss PLEX Elite 9000, Carl Zeiss Meditec, Dublin, California, USA) with 100,000 A-scans per second (central wavelength: 1040–1060 nm). The scan size of all scans was 3×3mm with focus on the macula. B-scans were segmented automatically using the automated algorithm of the device and manually reviewed for segmentation errors. The proprietary general sliding slab method of the device was used to remove decorrelation tail artefacts within the OCTA volume (Bagherinia H, et al., IOVS 2017;58:ARVO E-Abstract 643). According to the current OCTA nomenclature, en face images of the superficial retinal layer (encompassing nerve fibre layer, ganglion cell layer, inner plexiform layer) and deep retinal layer (inner nuclear layer, outer plexiform layer, Henle fibre layers) were exported [13].

Image analysis

Fiji [14], an open-source image processing software based on ImageJ [15] (version 1.51w) was used for image analysis. Per eye, two 8-bit grey scale en face bitmap images of the superficial and deep retinal layers were binarized by manual thresholding and six previously published automated algorithms [16–21] implemented in Fiji, named as followed, based on the abbreviations used in the software [15]: Huang, Li, Otsu, Moments, Mean and Percentile. As previously described, the manual approach was based on delineating the FAZ by selecting its outer borders with a free-hand tool on the superficial retinal layer and using the maximum grey value in this area as the threshold for binarization. [3,8,22–28] A second measurement was performed after at least 1 week by the same examiner, in case of a grey value threshold difference between the 2 measurements ≥ 5 , a third measurement was performed immediately and the median of these 3 measurements used as the threshold for the manual method. It has been shown that this methodology has high interrater reliability. [3] For the deep retinal layer, the FAZ selection of the superficial retinal layer was applied to the respective image area of the deep layer for binarization threshold determination. This approach combines best practice manual thresholding approaches published in the literature. [3,22] The threshold determination of the six automated algorithms has been previously published elsewhere. [16–21] VD was calculated based on the binarized images according to the formula $VD = \frac{n(\text{white pixels in binarized image})^2}{n(\text{all pixels in binarized image})^2}$. [22] For calculation of Vessel Skeleton Density (VSD) images were skeletonized by ImageJ and VSD was calculated according to the formula $VSD = \frac{n(\text{white pixels in skeletonized image})}{n(\text{all pixels in skeletonized image})^2}$. [22]

Statistical analyses

Statistical analyses were performed with SPSS Statistics for Windows, version 25 (IBM Corporation, Armonk, New York). Mean values of all VD and VSD measurements per layer and per eye were calculated and tested for associations with signal strength index of the OCTA image, intraocular pressure and patient age. The relative differences between test and retest VD and VSD were calculated. Linear regression analysis was performed to adjust for age, including relative differences between test and retest VD and VSD values as dependent variables and age as well as seven binary variables for the respective algorithm used as independent variables. Intra-class correlation coefficients (ICCs) between the two OCTA images of each eye were determined. Additionally, the Repeatability Coefficient (RC) was calculated according to the

formula $RC = 1.96 \times \sqrt{\frac{\sum(\text{measurement } 2 - \text{measurement } 1)^2}{n}}$. [29–31] The Mann-Whitney-U test, the Friedman test and the Kruskal Wallis test were used as indicated. To measure discriminatory ability (healthy versus macular pathology), differences between the different approaches were assessed using Receiver Operating Characteristic (ROC) curves and area under the curve (AUC) values. For age adjustment, we performed a binary logistic regression analysis to discriminate between healthy eyes and eyes with a macula based on VD or VSD and age (per algorithm and per retinal layer). The resulting probabilities were then used for ROC analysis. A p-value of <0.05 was considered statistically significant. Correction for multiple testing was done using the Holm-Bonferroni method [32]. Corrected p-values are reported as p_c.

Results

Twenty-six eyes (8 female, 18 male; mean age: 47 years) of 15 patients were included, resulting in 104 images (two consecutive images of two layers per eye) of the superficial and deep retinal layer. These included 11 healthy eyes and 15 eyes with an impaired macular vasculature (due to diabetic changes, previous central venous occlusions or other maculopathies). Some eyes had to be excluded due to insufficient image quality. Sphere ranged between ± 2.0 dpt in all eyes. Intraocular pressure was within a normal range in all patients (mean ± standard deviation: 17 mmHg ± 4 mmHg). In all included images, the OCTA en face image quality was high with a minimum signal strength index of 9/10.

The average threshold values differed significantly between different binarization approaches (p < 0.0001, Table 1). Overall, the Moments algorithm produced the highest and the Percentile and Li algorithms the lowest mean thresholds. Overall binarization thresholds did not differ significantly between subjects with healthy maculae and subjects with macular vessel pathology (p = 0.447). Exemplary unprocessed and binarized images are displayed in Fig 1.

VD and VSD differed significantly between different binarization algorithms (p < 0.0001, Table 1). Both values were significantly lower in individuals with macular pathology compared to those without (p < 0.0001). For instance, mean VD of the superficial layer was 0.22 ± 0.01 for healthy eyes and 0.19 ± 0.03 for diseased eyes while VSD of the superficial layer was 6.9×10⁻⁸ ± 0.3×10⁻⁸ for healthy eyes and 6.1×10⁻⁸ ± 0.8×10⁻⁸ for diseased eyes according to the Mean algorithm. VD and VSD did not differ significantly between different image signal strength indices (p = 0.157 and p = 0.079, respectively) or intraocular pressures (p = 0.271 and

Table 1. Mean ± standard deviation of binarization thresholds, vessel densities and skeleton densities obtained using the different algorithms for images of the superficial and deep retinal layers.

Parameter	Layer	Binarization approach						
		Manual	Huang	Li	Otsu	Moments	Mean	Percentile
Binarization threshold	Superf.	63 ± 12	49 ± 4	49 ± 4	66 ± 8	74 ± 10	58 ± 7	52 ± 9
	Deep	56 ± 13	30 ± 5	30 ± 4	41 ± 7	45 ± 9	32 ± 4	26 ± 5
Vessel Density	Superf.	0.196 ± 0.089	0.276 ± 0.058	0.149 ± 0.024	0.163 ± 0.035	0.120 ± 0.025	0.206 ± 0.025	0.250 ± 0.003
	Deep	0.052 ± 0.037	0.190 ± 0.034	0.197 ± 0.038	0.102 ± 0.034	0.079 ± 0.027	0.174 ± 0.024	0.249 ± 0.005
Skeleton Density	Superf.	6.1×10 ⁻⁸ ± 1.4×10 ⁻⁸	7.0×10 ⁻⁸ ± 1.0×10 ⁻⁸	7.1×10 ⁻⁸ ± 1.0×10 ⁻⁸	5.9×10 ⁻⁸ ± 0.9×10 ⁻⁸	5.3×10 ⁻⁸ ± 0.8×10 ⁻⁸	6.4×10 ⁻⁸ ± 0.7×10 ⁻⁸	6.9×10 ⁻⁸ ± 0.5×10 ⁻⁸
	Deep	3.5×10 ⁻⁸ ± 1.4×10 ⁻⁸	6.6×10 ⁻⁸ ± 0.7×10 ⁻⁸	6.7×10 ⁻⁸ ± 0.7×10 ⁻⁸	4.9×10 ⁻⁸ ± 0.9×10 ⁻⁸	4.4×10 ⁻⁸ ± 0.9×10 ⁻⁸	6.4×10 ⁻⁸ ± 0.5×10 ⁻⁸	7.5×10 ⁻⁸ ± 0.3×10 ⁻⁸

Superf. = superficial

<https://doi.org/10.1371/journal.pone.0230260.t001>

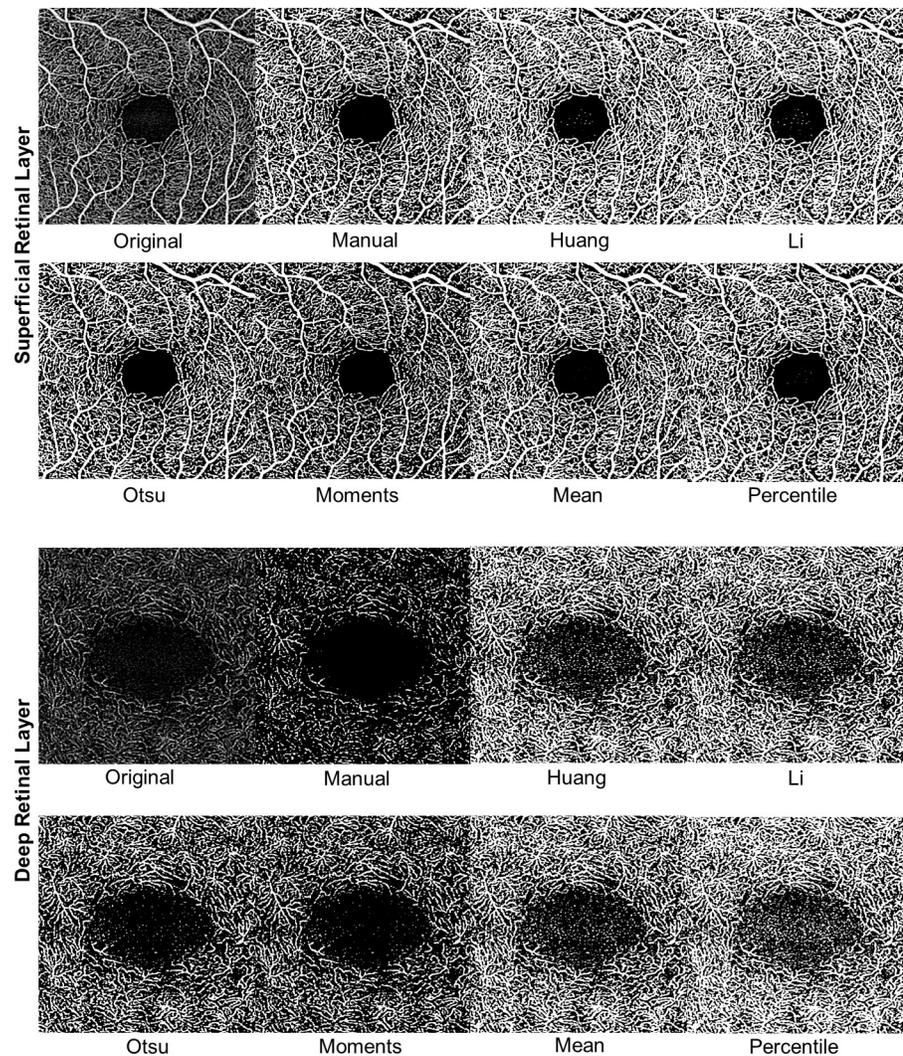


Fig 1. OCTA en face images of the superficial and deep retinal layers of one healthy eye. The image order from left to right is: Original image, binarized images using the manual thresholding and the Huang, Li, Otsu, Moments, Mean, Percentile algorithms.

<https://doi.org/10.1371/journal.pone.0230260.g001>

$p = 0.462$, respectively). Age was negatively correlated with VD ($r = -0.405$, $p = 0.040$) and VSD ($r = -0.406$, $p = 0.039$).

Relative test-retest measurement differences of VD and VSD obtained from two consecutive OCTA examinations varied significantly between the algorithms when comparing all approaches using the Friedman test both in the superficial and deep retinal layers ($p < 0.0001$). The manual thresholding approach had a significantly lower VD and VSD repeatability in the superficial and deep retinal layer compared to most automated algorithms in pair-wise comparisons between the different binarization approaches ($p_c < 0.0084$ in paired algorithm comparisons, **Fig 2**). The relative inter-measurement differences of the Huang algorithm (VD, deep retinal layer: $p_c = 0.126$; VSD, superficial and deep retinal layer: $p_c \geq 0.064$) and the Li algorithm (VSD, superficial retinal layer: $p_c = 0.183$) did not differ significantly from the repeatability values of the manual method. The relative inter-measurement differences, however, did not differ significantly when comparing automated algorithms with one

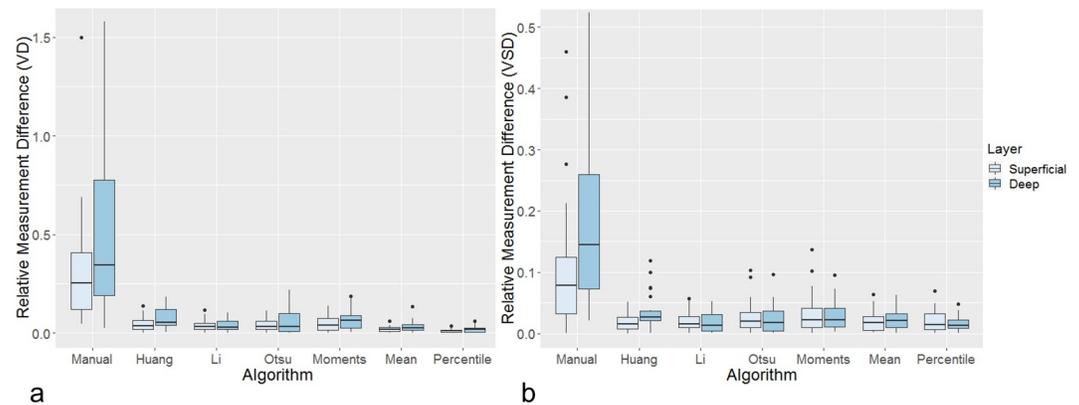


Fig 2. Inter-algorithm repeatability: Relative Vessel Density (a) and Vessel Skeleton Density (b) measurement differences obtained using a manual approach and six automated algorithms to binarize images of the superficial and deep retinal layers. Some outliers are not shown due to the scaling of the diagram. Values over 1.5 interquartile range below the first quartile or above the third quartile were defined as outliers; the horizontal lines indicate the 1st quartile, median and 3rd quartile. VD = Vessel Density, VSD = Vessel Skeleton Density.

<https://doi.org/10.1371/journal.pone.0230260.g002>

another ($p_c > 0.05$). To adjust for age, linear regression analysis was performed. For both VD and VSD, the relative test-retest differences significantly depended on use of the manual approach compared to not using it ($\beta = 0.661$, $p < 0.0001$ and $\beta = 0.163$, $p < 0.0001$, respectively). The VD and VSD relative test-retest differences did not significantly depend on age or use of any of the automated algorithms. Inter-examination VD and VSD differences between algorithms were not significantly different between healthy eyes and eyes with macular vessel pathologies (VD, $p = 0.685$; VSD, $p = 0.770$), indicating that the results of both groups can be interpreted in total.

A post-hoc power analysis for the Friedman test using a 10000-fold simulation with data generated from a normal distribution according to our sample characteristics revealed that a sample size of 6 eyes is sufficient to detect significant differences at 5% level with 80% power between VD relative differences of different algorithms in the superficial and in the deep retinal layers. Therefore, the sample size available in our study was considered appropriate from a statistical standpoint.

Except for the percentile algorithm, ICCs of VD and VSD measurements between the two consecutive examinations per eye were noticeably higher for all the automated algorithms when compared to the manual approach (Table 2). Repeatability Coefficients of VD and VSD of the superficial and deep retinal layers were noticeably higher (i.e. poorer repeatability) for the manual approach compared to the six automated algorithms investigated (Table 2). ROC analysis (healthy versus macular pathology) based on binary logistic regression models to adjust for age revealed AUC values between 0.838 and 0.997. The manual method was less sensitive to pathologic change than most automated algorithms (Fig 3) and the four AUC values of the manual approach were lower than almost all (23/24, 96%) values of automated algorithms. The AUC value for the manual method of VD acquisition of the superficial retinal layer was significantly lower (indicating lower sensitivity and specificity values) than the AUC value for the Mean algorithm of VD acquisition of the same layer (Table 3).

Discussion

To our knowledge this is the first quantitative study on test-retest reliability of manual versus automated thresholding approaches for OCTA images. Automated algorithms outperformed

Table 2. Additional repeatability parameters.

Parameter	Layer	Repeatability value per Algorithm							
		Manual	Huang	Li	Otsu	Moments	Mean	Percentile	
ICC [95% CI]	Vessel Density	Superf.	0.625 [0.163–0.832]	0.907 [0.793–0.958]	0.829 [0.618–0.923]	0.974 [0.942–0.988]	0.949 [0.885–0.977]	0.949 [0.886–0.977]	-0.070[-1.386–0.520]
		Deep	0.735 [0.410–0.881]	0.812 [0.581–0.916]	0.929 [0.842–0.968]	0.942 [0.870–0.974]	0.957 [0.904–0.981]	0.966 [0.924–0.985]	0.096 [-1.017–0.595]
	Skeleton Density	Superf.	0.802 [0.559–0.911]	0.981 [0.957–0.991]	0.979 [0.953–0.991]	0.986 [0.969–0.994]	0.982 [0.961–0.992]	0.984 [0.965–0.993]	0.884 [0.742–0.948]
		Deep	0.643 [0.203–0.840]	0.854 [0.675–0.935]	0.977 [0.948–0.990]	0.942 [0.871–0.974]	0.947 [0.881–0.976]	0.945 [0.878–0.976]	0.305 [-0.550–0.688]
RC	Vessel Density	Superf.	0.191	0.026	0.024	0.016	0.016	0.009	0.007
		Deep	0.062	0.033	0.018	0.016	0.012	0.014	0.012
	Skeleton Density	Superf.	2.2×10 ⁻⁸	3.1×10 ⁻⁹	3.4×10 ⁻⁹	3.9×10 ⁻⁹	4.3×10 ⁻⁹	3.0×10 ⁻⁹	3.4×10 ⁻⁹
		Deep	1.9×10 ⁻⁸	5.4×10 ⁻⁹	3.1×10 ⁻⁹	3.2×10 ⁻⁹	3.1×10 ⁻⁹	3.4×10 ⁻⁹	3.0×10 ⁻⁹

CI = confidence interval; ICC = Intra-class correlation coefficient; p_c = corrected p-value; RC = Repeatability Coefficient; Superf. = superficial

<https://doi.org/10.1371/journal.pone.0230260.t002>

manual thresholding, lead to more reproducible results and, therefore, allow for a more sensitive discrimination of healthy maculae from maculae with pathology. Thus, automated binarization algorithms should be preferred over manual approaches for OCTA image analysis. However, the different algorithms are not interchangeable and results of the algorithms differed significantly. Therefore, a better – ideally international – standardization of algorithms is needed to increase comparability of studies.

Our study supports existing data showing low inter-method agreement for image binarization. [12,33] The thresholds between the algorithms varied significantly as did the VD and VSD values, highlighting a lack of comparability. Rabiolo and colleagues also compared different methods to quantify perfusion in the macular region, however, they only assessed VD, used lower resolution 6×6mm en face OCTA images and did not investigate test-retest reliability of the different approaches. [12] Mehta and colleagues recently applied five automated binarization algorithms to OCTA images and found significant differences between the detected VD values. However, they investigated neither the repeatability of VD values based on the algorithms nor their ability to detect pathology. [33] Shoji and colleagues compared different automated thresholding algorithms but they did not assess any manual methods. [34] In this study, we used high-detail 3×3mm en face OCTA images, provide data on VD as well as VSD and evaluated test-retest reliability of both manual and automated algorithms on consecutive images.

Reproducibility was excellent for five of the automated algorithms according to the scale proposed by Chan. [35] This is in keeping with the current literature where automated algorithms tend to outperform manual image analysis in terms of reproducibility. [36–38] The inter-measurement differences were significantly higher after image binarization using the manual algorithm compared to all automated algorithms. This effect was independent of age. Due to this, automated binarization algorithms should be preferred over manual approaches in OCTA image binarization.

Intra-class correlation coefficients between the two consecutive OCTA measurements were significantly higher when using the Otsu, Moments or Mean algorithms compared to the manual approach in at least three out of four categories investigated (VD and VSD of the superficial and deep retinal layers, respectively). For this reason, these three algorithms should be

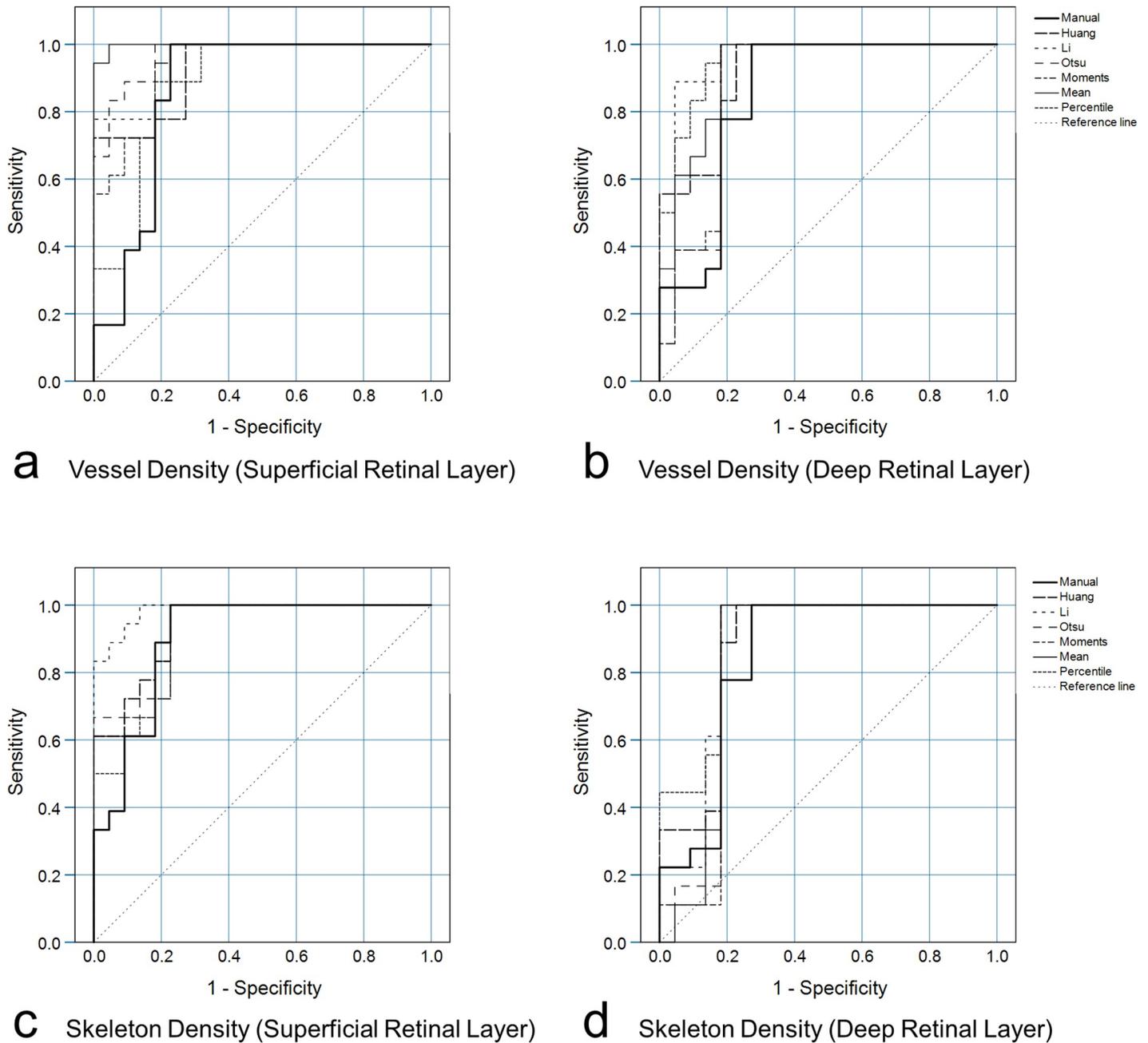


Fig 3. ROC curves (healthy versus macular pathology) of the binary logistic regression formulae based on the respective Vessel Density and Vessel Skeleton Density as well as age obtained by the manual approach and the automated algorithms in the superficial (a, c) and deep retinal layer (b, d).

<https://doi.org/10.1371/journal.pone.0230260.g003>

investigated in further detail. The automated Percentile algorithm proved not to be appropriate for analysis of VD or VSD due to inconsistent results. It measures the grey intensity closest to a percentile which limits the use of this algorithm when applied to OCTA (Table 4). We therefore do not recommend the Percentile Algorithm for future investigation on this topic.

Ability to discriminate between healthy and pathological maculae was good for almost all automated algorithms except the Percentile algorithm. Rabiolo et al found no significant differences between the algorithms tested in their study. [12] Interestingly, the manual approach

Table 3. Results of ROC analysis of binary logistic regression formulae based on the respective Vessel Density and Vessel Skeleton Density as well as age in the superficial and deep retinal layers per algorithm.

Variable	Layer	Area Under the Curve per Algorithm [95% Confidence Interval]						
		Manual	Huang	Li	Otsu	Moments	Mean	Percentile
Vessel Density	Superf.	0.864 [0.741;0.986]	0.929 [0.855;1.0]	0.957 [0.903;1.0]	0.967 [0.922;1.0]	0.934 [0.863; 1.0]	0.997 [0.989;1.0]	0.884 [0.778;0.989]
	Deep	0.851 [0.727;0.975]	0.917 [0.833;1.0]	0.965 [0.914;1.0]	0.876 [0.758;0.994]	0.879 [0.763;0.995]	0.927 [0.846;1.0]	0.955 [0.896;1.0]
Skeleton Density	Superf.	0.902 [0.806;0.998]	0.934 [0.864;1.0]	0.985 [0.958;1.0]	0.929 [0.855;1.0]	0.922 [0.842;1.0]	0.922 [0.842;1.0]	0.909 [0.821;0.997]
	Deep	0.843 [0.714;0.973]	0.876 [0.762;0.990]	0.876 [0.757;0.996]	0.864 [0.706;0.986]	0.838 [0.692;0.985]	0.843 [0.701;0.986]	0.884 [0.782;0.986]

Superf. = superficial

<https://doi.org/10.1371/journal.pone.0230260.t003>

used by Rabiolo and colleagues discriminated healthy individuals from those with macular disease noticeably better than our manual approach. However, information on reliability of their manual approach is missing. Our method for manual thresholding was standardized against the FAZ whereas Rabiolo et al used arbitrary binarization thresholds. Thus it is doubtful whether their results can be reproduced with other images and manual thresholding cannot be recommended. Shoji et al compared different automated global and local thresholding methods across two OCTA devices [34]. They also included the Otsu and the Mean algorithms in their analysis but reported significant lower intra-class correlations compared to our data (based on the 95% confidence intervals). Our work additionally included age-adjusted statistical comparisons of different algorithms, showing that the Mean algorithm can be used to detect pathology significantly better than the manual approach. Such data have not been available in the previous literature. According to our results, specifically the Mean algorithm might be a good option to use for binarization of en face OCTA images of the superficial retinal layer with the PLEX Elite device. It tends to enhance the retinal vasculature compared to other algorithms as reflected in relatively high mean VD and VSD values. For this reason, it might be more prone to image artefacts than algorithms that overall set higher binarization thresholds like the Otsu or Moments algorithms.

The strengths of our study include a comprehensive evaluation of reliability and comparability of seven different approaches for OCTA image binarization, all previously published, including four measures of repeatability. We also assessed ability to discriminate healthy from diseased eyes for all approaches. For image acquisition, well established protocols were used, allowing for easy replication of our study. In the literature, more comprehensive preprocessing steps for image processing have been proposed, including the use of filters such as the Frangi

Table 4. Principles of the binarization methods used in this study.

Binarization method	Short description [3–9]
Manual	Determines maximum gray value within the marked foveal avascular zone
Huang	Minimizes fuzziness of pixel-wise fuzzy membership functions
Li	Minimizes cross entropy between unedited and binarized images
Otsu	Minimizes variance between foreground and background structures in the image histogram
Moments	Preserves gray-level moments of the input image
Mean	Calculates mean of grey levels in the original image
Percentile	Measures the grey intensity closest to a percentile

<https://doi.org/10.1371/journal.pone.0230260.t004>

filter to enhance image contrasts. We have omitted such steps in our analysis on purpose, since these methods have several disadvantages including generation of image artifacts resembling vessel structures and different results for vessels that are not equally distributed in size [39,40]. Limitations include the relatively small number of subjects (therefore limited generalizability of the comparisons between different automated algorithms), having only one repeat measurement per eye and limited comparability of vessel density and skeleton density values because of different calculation approaches in the literature. We did not assess comparability and reproducibility across different OCTA devices and evaluated OCTA scans of the macula only. We did not compare the algorithms to commercial tools for quantification of vessel parameters because for scientific purposes, understanding as many of the image processing steps as possible is warranted. Future research should also focus on binarization of OCTA images of the optic disc with different algorithms.

In conclusion, because of higher repeatability and improved discrimination, automated binarization algorithms should be preferred over manual approaches. Better standardization of algorithms is needed to improve comparability of studies.

Author Contributions

Conceptualization: Jan Henrik Terheyden, Maximilian W. M. Wintergerst, Frank G. Holz, Robert P. Finger.

Data curation: Jan Henrik Terheyden, Maximilian W. M. Wintergerst.

Formal analysis: Jan Henrik Terheyden, Maximilian W. M. Wintergerst, Moritz Berger, Robert P. Finger.

Investigation: Jan Henrik Terheyden, Maximilian W. M. Wintergerst, Peyman Falahat.

Methodology: Jan Henrik Terheyden, Maximilian W. M. Wintergerst, Moritz Berger.

Project administration: Jan Henrik Terheyden, Maximilian W. M. Wintergerst.

Supervision: Frank G. Holz, Robert P. Finger.

Validation: Frank G. Holz, Robert P. Finger.

Visualization: Jan Henrik Terheyden, Maximilian W. M. Wintergerst.

Writing – original draft: Jan Henrik Terheyden, Maximilian W. M. Wintergerst.

Writing – review & editing: Peyman Falahat, Moritz Berger, Frank G. Holz, Robert P. Finger.

References

1. Jia Y, Tan O, Tokayer J, Potsaid B, Wang Y, Liu JJ, et al. Split-spectrum amplitude-decorrelation angiography with optical coherence tomography. *Opt Express*. 2012; 20: 4710–4725. <https://doi.org/10.1364/OE.20.004710> PMID: 22418228
2. Spaide RF, Klancnik JM Jr, Cooney MJ. Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography. *JAMA Ophthalmol*. 2015; 133: 45–50. <https://doi.org/10.1001/jamaophthalmol.2014.3616> PMID: 25317632
3. Wintergerst MWM, Pfau M, Muller PL, Berger M, de Sisternes L, Holz FG, et al. Optical Coherence Tomography Angiography in Intermediate Uveitis. *Am J Ophthalmol*. 2018; 194: 35–45. S0002-9394(18)30335-0. <https://doi.org/10.1016/j.ajo.2018.06.023> PMID: 30026083
4. BahadarKhan K, A Khaliq A, Shahid M. A Morphological Hessian Based Approach for Retinal Blood Vessels Segmentation and Denoising Using Region Based Otsu Thresholding. *PLoS One*. 2016; 11: e0158996. <https://doi.org/10.1371/journal.pone.0158996> PMID: 27441646
5. Lei J, Yi E, Suo Y, Chen C, Xu X, Ding W, et al. Distinctive Analysis of Macular Superficial Capillaries and Large Vessels Using Optical Coherence Tomographic Angiography in Healthy and Diabetic Eyes.

- Invest Ophthalmol Vis Sci. 2018; 59: 1937–1943. <https://doi.org/10.1167/iovs.17-23676> PMID: 29677360
6. Shahlaee A, Samara WA, Hsu J, Say EA, Khan MA, Sridhar J, et al. In Vivo Assessment of Macular Vascular Density in Healthy Human Eyes Using Optical Coherence Tomography Angiography. *Am J Ophthalmol*. 2016; 165: 39–46. <https://doi.org/10.1016/j.ajo.2016.02.018> PMID: 26921803
 7. Romano F, Arrigo A, Ch'ng SW, Battaglia Parodi M, Manitto MP, Martina E, et al. Capillary Network Alterations in X-Linked Retinoschisis Imaged on Optical Coherence Tomography Angiography. *Retina*. 2018; 0: 1–7. <https://doi.org/10.1097/IAE.0000000000002222> PMID: 29877903
 8. Kim AY, Chu Z, Shahidzadeh A, Wang RK, Puliafito CA, Kashani AH. Quantifying Microvascular Density and Morphology in Diabetic Retinopathy Using Spectral-Domain Optical Coherence Tomography Angiography. *Invest Ophthalmol Vis Sci*. 2016; 57: OCT362–70. <https://doi.org/10.1167/iovs.15-18904> PMID: 27409494
 9. Nesper PL, Roberts PK, Onishi AC, Chai H, Liu L, Jampol LM, et al. Quantifying Microvascular Abnormalities With Increasing Severity of Diabetic Retinopathy Using Optical Coherence Tomography Angiography. *Invest Ophthalmol Vis Sci*. 2017; 58: BIO307–BIO315. <https://doi.org/10.1167/iovs.17-21787> PMID: 29059262
 10. Zahid S, Dolz-Marco R, Freund KB, Balaratnasingam C, Dansingani K, Gilani F, et al. Fractal Dimensional Analysis of Optical Coherence Tomography Angiography in Eyes With Diabetic Retinopathy. *Invest Ophthalmol Vis Sci*. 2016; 57: 4940–4947. <https://doi.org/10.1167/iovs.16-19656> PMID: 27654421
 11. Lee H, Lee M, Chung H, Kim HC. Quantification of Retinal Vessel Tortuosity in Diabetic Retinopathy using Optical Coherence Tomography Angiography. *Retina*. 2018; 38: 976–985. <https://doi.org/10.1097/IAE.0000000000001618> PMID: 28333883
 12. Rabiolo A, Gelormini F, Sacconi R, Cicinelli MV, Triolo G, Bettin P, et al. Comparison of methods to quantify macular and peripapillary vessel density in optical coherence tomography angiography. *PLoS One*. 2018; 13: e0205773. <https://doi.org/10.1371/journal.pone.0205773> PMID: 30335815
 13. Campbell JP, Zhang M, Hwang TS, Bailey ST, Wilson DJ, Jia Y, et al. Detailed Vascular Anatomy of the Human Retina by Projection-Resolved Optical Coherence Tomography Angiography. *Sci Rep*. 2017; 7: 42201. <https://doi.org/10.1038/srep42201> PMID: 28186181
 14. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012; 9: 676–682. <https://doi.org/10.1038/nmeth.2019> PMID: 22743772
 15. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012; 9: 671–675. <https://doi.org/10.1038/nmeth.2089> PMID: 22930834
 16. Huang L-, Wang M-J. Image thresholding by minimizing the measure of fuzziness. *Pattern Recognit*. 1995; 28: 41–51.
 17. Li CH, Tam PKS. An Iterative Algorithm for Minimum Cross Entropy Thresholding. *Pattern Recog Lett*. 1998; 18: 771–776.
 18. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Sys, Man, Cyber*. 1979; 9: 62–66.
 19. Tsai W. Moment-preserving thresholding: A new approach. *Computer Vision Graphics and Image Processing*. 1985; 29: 377–393.
 20. Glasbey CA. An analysis of histogram-based thresholding algorithms. *CVGIP: Graphical Models and Image Processing*. 1993; 55: 532–537.
 21. Doyle W. Operation useful for similarity-invariant pattern recognition. *Journal of the Association for Computing Machinery*. 1962; 9: 259–267.
 22. Kim AY, Rodger DC, Shahidzadeh A, Chu Z, Koullis N, Burkemper B, et al. Quantifying Retinal Microvascular Changes in Uveitis Using Spectral-Domain Optical Coherence Tomography Angiography. *Am J Ophthalmol*. 2016; 171: 101–112. S0002-9394(16)30424-X. <https://doi.org/10.1016/j.ajo.2016.08.035> PMID: 27594138
 23. Cheng D, Chen Q, Wu Y, Yu X, Shen M, Zhuang X, et al. Deep perifoveal vessel density as an indicator of capillary loss in high myopia. *Eye (Lond)*. 2019. <https://doi.org/10.1038/s41433-019-0573-1> PMID: 31527763
 24. Chu Z, Lin J, Gao C, Xin C, Zhang Q, Chen CL, et al. Quantitative assessment of the retinal microvasculature using optical coherence tomography angiography. *J Biomed Opt*. 2016; 21: 66008. <https://doi.org/10.1117/1.JBO.21.6.066008> PMID: 27286188
 25. Richter GM, Madi I, Chu Z, Burkemper B, Chang R, Zaman A, et al. Structural and Functional Associations of Macular Microcirculation in the Ganglion Cell-Inner Plexiform Layer in Glaucoma Using Optical

- Coherence Tomography Angiography. *J Glaucoma*. 2018; 27: 281–290. <https://doi.org/10.1097/IJG.0000000000000888> PMID: 29394201
26. Hwang TS, Gao SS, Liu L, Lauer AK, Bailey ST, Flaxel CJ, et al. Automated Quantification of Capillary Nonperfusion Using Optical Coherence Tomography Angiography in Diabetic Retinopathy. *JAMA Ophthalmol*. 2016; 134: 367–373. <https://doi.org/10.1001/jamaophthalmol.2015.5658> PMID: 26795548
 27. Kuehlewein L, Bansal M, Lenis TL, Iafe NA, Sadda SR, Bonini Filho MA, et al. Optical Coherence Tomography Angiography of Type 1 Neovascularization in Age-Related Macular Degeneration. *Am J Ophthalmol*. 2015; 160: 739–48.e2. <https://doi.org/10.1016/j.ajo.2015.06.030> PMID: 26164826
 28. Jia Y, Bailey ST, Wilson DJ, Tan O, Klein ML, Flaxel CJ, et al. Quantitative optical coherence tomography angiography of choroidal neovascularization in age-related macular degeneration. *Ophthalmology*. 2014; 121: 1435–1444. <https://doi.org/10.1016/j.ophtha.2014.01.034> PMID: 24679442
 29. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1: 307–310. doi: S0140-6736(86)90837-8 PMID: 2868172
 30. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol*. 2003; 22: 85–93. <https://doi.org/10.1002/uog.122> PMID: 12858311
 31. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One*. 2013; 8: e73990. <https://doi.org/10.1371/journal.pone.0073990> PMID: 24040139
 32. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Statist*. 1979; 6: 65–70.
 33. Mehta N, Liu K, Alibhai AY, Gendelman I, Braun P, Ishibazawa A, et al. Impact of Binarization Thresholding and Brightness/Contrast Adjustment Methodology on Optical Coherence Tomography Angiography Image Quantification. *Am J Ophthalmol*. 2019. doi: S0002-9394(19)30106-0.
 34. Shoji T, Yoshikawa Y, Kanno J, Ishii H, Ibuki H, Ozaki K, et al. Reproducibility of Macular Vessel Density Calculations Via Imaging With Two Different Swept-Source Optical Coherence Tomography Angiography Systems. *Transl Vis Sci Technol*. 2018; 7: 31. <https://doi.org/10.1167/tvst.7.6.31> PMID: 30619651
 35. Chan YH. *Biostatistics 104: correlational analysis*. Singapore Med J. 2003; 44: 614–619. PMID: 14770254
 36. Ashton EA, Takahashi C, Berg MJ, Goodman A, Totterman S, Ekholm S. Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. *J Magn Reson Imaging*. 2003; 17: 300–308. <https://doi.org/10.1002/jmri.10258> PMID: 12594719
 37. Hsu YY, Schuff N, Du AT, Mark K, Zhu X, Hardin D, et al. Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia. *J Magn Reson Imaging*. 2002; 16: 305–310. <https://doi.org/10.1002/jmri.10163> PMID: 12205587
 38. Stalhammar G, Fuentes Martinez N, Lippert M, Tobin NP, Molholm I, Kis L, et al. Digital image analysis outperforms manual biomarker assessment in breast cancer. *Mod Pathol*. 2016; 29: 318–329. <https://doi.org/10.1038/modpathol.2016.34> PMID: 26916072
 39. Tan B, Wong A, Bizheva K. Enhancement of morphological and vascular features in OCT images using a modified Bayesian residual transform. *Biomed Opt Express*. 2018; 9: 2394–2406. <https://doi.org/10.1364/BOE.9.002394> PMID: 29760996
 40. Jerman T, Pernuš F, Likar B, Špiclin Ž. Beyond Frangi: an improved multiscale vesselness filter. 2015; 9413: 94132A.