JoS JOURNAL OF STROKE

Journal of Stroke 2022;24(3):433-435
https://doi.org/10.5853/jos.2022.02054

Letter to the Editor

Automated Composition Analysis of Thrombus from Endovascular Treatment in Acute Ischemic Stroke Using Computer Vision

JoonNyung Heo,[a] Young Seog,[a] Hyungwoo Lee,[a] Il Hyung Lee,[a] Sungeun Kim,[a] Jang-Hyun Baek,[b] Hyungjong Park,[c] Kwon-Duk Seo,[d] Gyu Sik Kim,[d] Han-Jin Cho,[e] Minyoul Baik,[f] Joonsang Yoo,[f] Jinkwon Kim,[f] Jun Lee,[g] Yoon-Kyung Chang,[h] Tae-Jin Song,[i] Jung Hwa Seo,[j] Seong Hwan Ahn,[k] Heow Won Lee,[a] Il Kwon,[a,l] Eunjeong Park,[l] Young Dae Kim,[a,l] Hyo Suk Nam[a,l]

[a]Department of Neurology, Yonsei University College of Medicine, Seoul, Korea
[b]Department of Neurology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Korea
[c]Department of Neurology, Brain Research Institute, Keimyung University School of Medicine, Daegu, Korea
[d]Department of Neurology, National Health Insurance Service Ilsan Hospital, Goyang, Korea
[e]Department of Neurology, Pusan National University School of Medicine, Busan, Korea
[f]Department of Neurology, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Korea
[g]Department of Neurology, Yeungnam University College of Medicine, Daegu, Korea
[h]Department of Neurology, Ewha Womans University Mokdong Hospital, Ewha Womans University School of Medicine, Seoul, Korea
[i]Department of Neurology, Ewha Womans University Seoul Hospital, Ewha Womans University School of Medicine, Seoul, Korea
[j]Department of Neurology, Inje University Busan Paik Hospital, College of Medicine, Inje University, Busan, Korea
[k]Department of Neurology, Chosun University Hospital, Chosun University College of Medicine, Gwangju, Korea
[l]Integrative Research Center for Cerebrovascular and Cardiovascular Diseases, Seoul, Korea

Dear Sir:

Histological analysis of thrombi acquired from endovascular thrombectomy for ischemic stroke provides an unprecedented understanding of thrombus formation in stroke.[1] Composition analysis is currently the mainstream of research for stroke thrombi, and immunohistochemistry (IHC) staining is frequently used.[2,3] Most studies have used general purpose software, including ImageJ and Photoshop (Adobe, San Jose, CA, USA).[4,5] However, for composition analysis, these tools require manual drawing for area measurement, which is labor-intensive and prone to bias, and the results are irreproducible. This study aimed to develop an open-sourced software named Automated Region-of-interest based Image Analysis (ARIA) for automated composition analysis of IHC-stained thrombi.

This was a retrospective study using a nationwide multicenter prospective registry. The Specialized Multi-center Attributed Registry of sTroke–Clot (SMART-Clot) is a prospective registry that enrolls consecutive patients with acute ischemic stroke from 10 centers in Korea who underwent endovascular thrombectomy. Forty stained slides from 10 randomly selected patients were included in this study. Detailed methodology is described in the Supplementary Methods and Supplementary Figure 1. Screen recordings using both ARIA and traditional method for thrombus image analysis are presented in the Supplementary Video. The accuracy and time needed for analysis were compared between the traditional analysis method and ARIA. Four analysts with varying experiences measured the same 40 slides using both the traditional method and ARIA. External validation was performed for two datasets: (1) anti-CD42b IHC stained slides of stroke thrombi from a previously published study[6] and (2) an open dataset of IHC slides from breast tissue.[7] This study was approved by the Institutional Review Board of Yonsei University College of Medicine (Approval number: 4-2017-0426), and informed consent was obtained from each patient. ARIA is publicly available as an open-sourced project and can be installed in all major operating systems (https://github.com/jnheo-md/aria).

The median age of the patients included in this study was 71.5 years (interquartile range [IQR], 66.2 to 79.8) and six (60.0%) were male (Supplementary Table 1). The results obtained using ARIA by all analysts showed highly accurate results compared to the results from the professional analyst using the traditional method. The Spearman's correlation coefficient of the stained ratio, defined as the stained area divided by the total area, ranged between 0.913 and 0.916 (all P<0.001) (Table 1). The Bland-Altman analysis showed 95% limits of agreement between 0.11 and 0.14 (Supplementary Figure 2). Agreement of the results from each analyst were significantly higher using ARIA than the traditional method for thrombus area (P=0.005), stained area (P<0.001), and stained ratio (P<0.001) (Table 2 and Supplementary Figure 3). The focused and total times needed for analysis were significantly shorter when ARIA was used than when the traditional method was used (Figure 1 and Supplementary Table 2). The median focused time needed while using ARIA was 7 seconds (IQR, 3.0 to 11.0), whereas the traditional method required a median of 231 seconds (IQR, 182.0 to 286.0; P<0.001). There was disagreement on one sample with homogeneous staining pattern due to differences in thresholding algorithm. There was high correlation, except for that sample with tissue factor staining, and significant analysis time difference between the two methods across each staining method and composition (Supplementary Tables 3 and 4, Supplementary Figure 4). External validation showed high correlation for both stroke thrombi and breast tissue (Spearman's correlation coefficient 0.929 and 0.875 respectively, both P<0.001) with significant focused time reduction (median 7.0 seconds [IQR, 7.0 to 10.0] vs. 198.0 seconds [IQR, 184.0 to 313.0] in stroke thrombi; and 11.0 seconds [IQR, 8.0 to 13.0] vs. 185.0 seconds [IQR, 165.5 to 217.5] in

Table 1. Comparison between stained ratio of each analyst using ARIA and those of the professional analyst using the traditional method

	Spearman's correlation coefficient*	Mean difference
Professional	0.913	−0.010 (0.064)
Trained	0.913	−0.010 (0.065)
Untrained 1	0.915	−0.012 (0.063)
Untrained 2	0.916	−0.010 (0.063)

Values are presented as difference (standard deviation).
ARIA, Automated Region-of-interest based Image Analysis.
*All P-values of the Spearman's correlation coefficient was <0.001.

Table 2. Differences between analysts for ARIA and traditional method

	ARIA	Traditional method	P
Total thrombus area pixel	3.72×10^5 (5.65×10^4 to 1.20×10^6)	3.93×10^5 (1.07×10^5 to 1.70×10^6)	0.005
Stained area pixel	1.26×10^5 (2.35×10^4 to 6.34×10^5)	2.87×10^5 (4.87×10^4 to 1.20×10^6)	<0.001
Stained ratio (%)	0.179 (0.050 to 0.544)	0.612 (0.163 to 1.90)	<0.001

Values are presented as median (interquartile range).
ARIA, Automated Region-of-interest based Image Analysis.

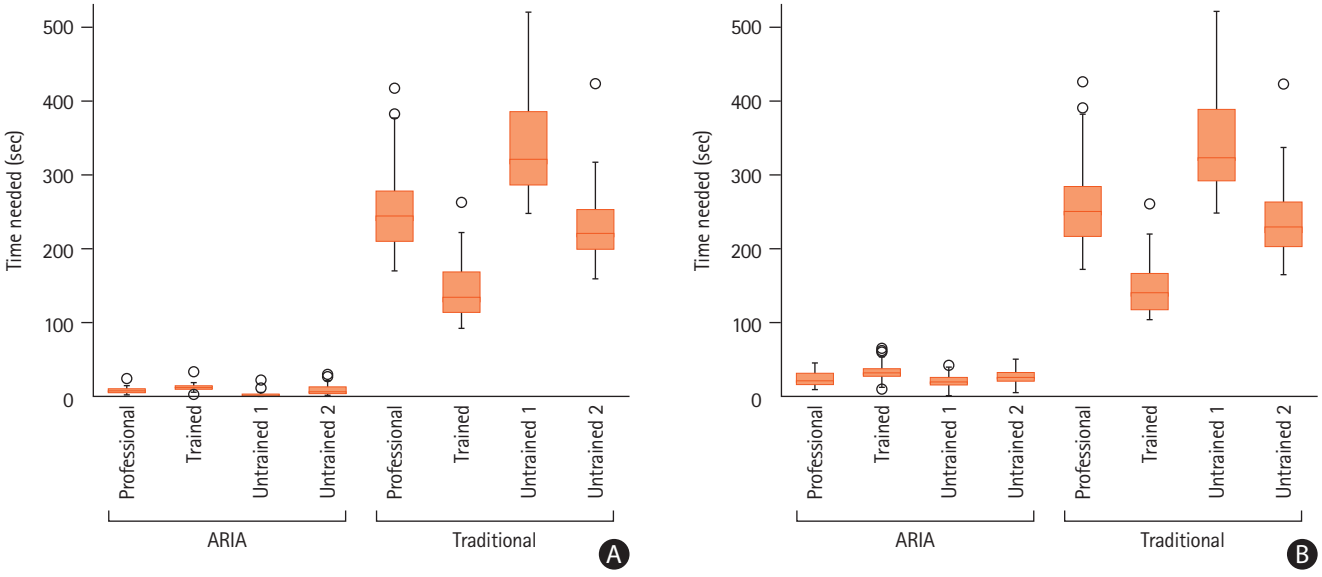


Figure 1. Comparison of (A) focused time and (B) total time needed for analysis using Automated Region-of-interest based Image Analysis (ARIA) and traditional methodology.

https://doi.org/10.5853/jos.2022.02054

breast tissue, both $P<0.001$).

This study validated a custom-built software named ARIA developed with computer vision for automated stroke thrombus composition analysis. ARIA (1) was highly accurate; (2) provided consistent results even for analysts without previous experience; and (3) was 30 times faster than the traditional method. Additionally, ARIA produced the same results when the input image and analysis parameters were equal, which is impossible using traditional methods.

Traditional method of composition analysis requires manual demarcation of the thrombus border, which is strenuous and inevitably introduces potential error due to inconsistencies in freehand drawing. ARIA mainly uses the Canny Edge Detection algorithm to replace freehand drawing, which is a widely used technique.[8] Additionally, slide image processing and automatic thresholding have been used in ARIA to reduce unnecessarily burden on the researchers.[9,10] Furthermore, this study demonstrated that the four analysts with varying experiences produced more consistent results when they used ARIA. Considering that the reproducibility and consistency between analysts are critical factors for a research methodology, ARIA has advantages for thrombus analysis over traditional methodologies.

This study has several limitations. The number of samples included in this study was small and originated form a single laboratory. There is no gold standard in measuring the composition of thrombi and the assessment of the accuracy of the novel software is limited. The expert's analysis result using traditional method does not necessarily represent the true composition of the thrombus. As seen in one sample with exceptional disagreement between methods, thresholding algorithm may have great effect on the results, especially in images with homogeneous texture.

In conclusion, ARIA may be used as an efficient and accurate tool that provides reproducible results in thrombus analysis.

Supplementary materials

Supplementary materials related to this article can be found online at https://doi.org/10.5853/jos.2022.02054.

References

1. Heo JH, Nam HS, Kim YD, Choi JK, Kim BM, Kim DJ, et al. Pathophysiologic and therapeutic perspectives based on thrombus histology in stroke. *J Stroke* 2020;22:64-75.
2. Schuhmann MK, Gunreben I, Kleinschnitz C, Kraft P. Immunohistochemical analysis of cerebral thrombi retrieved by mechanical thrombectomy from patients with acute ischemic stroke. *Int J Mol Sci* 2016;17:298.
3. Simons N, Mitchell P, Dowling R, Gonzales M, Yan B. Thrombus composition in acute ischemic stroke: a histopathological study of thrombus extracted by endovascular retrieval. *J Neuroradiol* 2015;42:86-92.
4. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;9:671-675.
5. Office Adobe Photoshop. Photo and design software. Adobe. https://www.adobe.com/products/photoshop.html. Accessed September 7, 2022.
6. Ahn SH, Hong R, Choo IS, Heo JH, Nam HS, Kang HG, et al. Histologic features of acute thrombi retrieved from stroke patients during mechanical reperfusion therapy. *Int J Stroke* 2016;11:1036-1044.
7. Borovec J, Kybic J, Arganda-Carreras I, Sorokin DV, Bueno G, Khvostikov AV, et al. ANHIR: automatic non-rigid histological image registration challenge. *IEEE Trans Med Imaging* 2020;39:3042-3052.
8. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell* 1986;8:679-698.
9. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 1979;9:62-66.
10. Goode A, Gilbert B, Harkes J, Jukic D, Satyanarayanan M. OpenSlide: a vendor-neutral software foundation for digital pathology. *J Pathol Inform* 2013;4:27.

Correspondence: Hyo Suk Nam
Department of Neurology, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Tel: +82-2-2228-1617
Fax: +82-2-393-0705
E-mail: hsnam@yuhs.ac
https://orcid.org/0000-0002-4415-3995

Received: June 22, 2022
Revised: August 7, 2022
Accepted: September 5, 2022

This research was supported by a grant of Patient-Centered Clinical Research Coordinating Center (PACEN) funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HC19C0028) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (grant number : 2022R1I1A1A01063439).

We would like to thank the students (S.Y. Kang and N.Y. Baik) who participated in this study.

The authors have no financial conflicts of interest.

Supplementary Methods

Slide preparation and image acquisition

Fresh thrombi were immediately fixed using 4% paraformaldehyde solution and sent to the central laboratory for further analysis. Thrombus samples were immunohistochemistry (IHC) stained with rabbit monoclonal anti-CD42b (ab134087, 1:100, Abcam, Cambridge, UK) for platelets, rabbit polyclonal anti-fibrinogen (ab34269, 1:200, Abcam) for fibrin/fibrinogen, rabbit monoclonal anti-glycophorin A (ab129024, 1:400, Abcam) for erythrocytes, and rabbit polyclonal anti-CD142 (PA5-27278, 1:100, Invitrogen, Waltham, MA, USA) for tissue factors. Except for anti-CD42b, antigen retrieval was performed using IHC-Tek epitope retrieval solution and a steamer. Overnight incubation was performed with primary antibodies at 4°C. An avidin/biotin/horse-radish peroxidase complex (Vector Laboratories Ltd., Peterborough, UK) was used for secondary antibody reaction. Incubation with 3,3'-diaminobenzidine solution was performed for color development. Hematoxylin counter staining was also performed, and slides were mounted with Permount Mounting Medium (Fischer Scientific, Fair Lawn, NJ, USA). The slides were scanned using either a whole-slide scanner (Leica Biosystems, Richmond, IL, USA) or a Stereo Investigator Imaging system (MBF Bioscience, Williston, VT, USA) equipped with a light microscope (Axio Imager D2, Carl Zeiss Co. Ltd., Jena, Germany). The whole-slide scanner captured the image in a 40x magnitude and 0.2528 μM/pixel resolution. The Stereo Investigator Imaging system used the Virtual Slice module to acquire a montage of the entire slide at 400× magnification.

Development of Automated Region-of-interest based Image Analysis (ARIA)

ARIA was developed using Python (Python Software Foundation, Wilmington, DE, USA). Python libraries, including "OpenSlide," "scikit-image," and "OpenCV," were used. The software receives an image file and supports most major slide formats. ARIA processes the image in the following order: (1) ARIA provides an option to enable cropping of the original image (cropping); (2) it automatically draws a contour that demarcates the area of the thrombus (contouring), with adjustable handles to enable customization; (3) color deconvolution is initiated to separate colors into the IHC color space for quantitative analysis; (4) ARIA performs automatic and manual thresholding to define the stained area (thresholding); and (5) it outputs a comma-separated value file (.csv), including the results from the analysis (Supplementary Figure 1). The parameters used during analysis, including the thresholds for contouring and definition of the stained area, were automatically set during analysis within the software. If these parameters are exactly equal, the software outputs the same results for the same image.

Analysis process

ImageJ was used as a traditional method for analysis: (1) the slide image file was opened in ImageJ; (2) color deconvolution was performed using the "Colour Deconvolution" plugin with the "H DAB" option selected; (3) the contour of the entire thrombus was drawn using the lasso tool; (4) the entire thrombus area was measured using the "Measure" menu; (5) automatic thresholding was performed using the "Threshold" menu; and (6) the area after thresholding was measured using the "Measure" menu.

Four analysts with varying experiences measured the same 40 slides using both the traditional method and ARIA. The first analyst (professional, H.L.) was a certified stroke neurologist with a Bachelor's Degree in Chemical and Biological Engineering, who is well-experienced in composition analysis of IHC-stained samples. The second analyst (trained analyst, Y.S.) was a medical student who had a thorough understanding of the study, with modest experience in composition analysis and additional training. Two other analysts were college students (untrained analysts) without any experience or understanding of this study. These untrained students were provided a specific set of instructions for analysis prepared as a screen recording movie and a detailed document with screenshots. For each analysis, all analysts were instructed to measure the time needed for analysis, which was divided into focused time and waiting time. Waiting time is defined as the time consumed by the computer to process the image. Focused time is calculated by subtracting the waiting time from the total time needed for analysis. All analyses were performed on the same computer (2017 iMac Pro, 3.2GHz 8-Core Intel Xeon W processor, 32GB memory, Radeon Pro Vega 56 8 GB graphics processor, Apple, Cupertino, CA, USA). All analysts were blinded to the clinical information of the patients included in this study. Each analyst performed the analysis on different dates to ensure that the analysis was conducted independently.

External validation

External validation was performed on two datasets. The first dataset was from a previously published study which included thrombi from endovascular thrombectomy of ischemic stroke from a single center in Korea.[1] A total of 13 images were used which were anti-CD42b IHC stained. Digital images were obtained using virtual slide microscopy scanning system (VS 120, Olympus, Japan). The second dataset was an open dataset of

IHC slides.[2] The dataset included total of 15 slides from five samples of human breast tissue stained against three antibodies (estrogen receptor, progesterone receptor, and Her2-neu). The images were whole-slide images with acquired in a 40x magnitude and 0.2528 μM/pixel resolution. The images were analyzed by the professional analyst. Time needed for analysis and analysis results were obtained with the same method as the internal dataset.

Statistical analysis

The accuracy of ARIA was determined by comparing the results from each analyst with those obtained using the traditional method by the professional analyst. Spearman's correlation coefficient (rho) was used for assessing correlation of total area, stained area, and stained ratio. The mean difference was calculated by obtaining the mean difference value between two results of the same sample. The Bland–Altman analysis was performed to obtain 95% limits of agreement. The consistency of the results between analysts was assessed by obtaining the difference in results between two analysts for the same thrombus image. The differences between all six possible combinations of two analysts were calculated. The absolute values of the differences were compared between ARIA and the traditional method using Mann-Whitney U-test. The time needed for analysis was compared between the two analysis methods using Mann-Whitney U-test as well. Samples were grouped into red blood cell (RBC)-rich thrombi or platelet+fibrin rich thrombi. A thrombus was classified as RBC-rich when anti-Glycophorin A stained ratio was higher than the sum of anti-CD42b and anti-fibrinogen stained ratio. Platelet+fibrin rich thrombi was defined as thrombi with the sum of anti-CD42b and anti-fibrinogen stained ratio higher than the anti-Glycophorin A stained ratio. The results used for classification were those obtained by the professional analyst with traditional method. All statistical analyses were performed using Python with "tableone" and "SciPy" packages.

Supplementary References

1. Ahn SH, Hong R, Choo IS, Heo JH, Nam HS, Kang HG, et al. Histologic features of acute thrombi retrieved from stroke patients during mechanical reperfusion therapy. *Int J Stroke* 2016;11:1036-1044.
2. Borovec J, Kybic J, Arganda-Carreras I, Sorokin DV, Bueno G, Khvostikov AV, et al. ANHIR: automatic non-rigid histological image registration challenge. *IEEE Trans Med Imaging* 2020; 39:3042-3052.

https://doi.org/10.5853/jos.2022.02054

Supplementary Table 1. Characteristics of the patients included in the study

Variable	Patients (n=10)
Age (yr)	71.5 (66.2–79.8)
Male sex	6 (60.0)
Hypertension	7 (70.0)
Diabetes	2 (20.0)
Dyslipidemia	2 (20.0)
TOAST classification	
CE	6 (60.0)
LAA	1 (10.0)
UT	3 (30.0)
Occlusion site	
ICA	4 (40.0)
MCA	6 (60.0)
Composition ratio (%)	
RBC	47.4 (31.4–54.4)
Platelet	16.9 (14.2–21.0)
Fibrinogen	31.8 (27.8–41.5)
Tissue factor	30.7 (26.3–39.8)

Values are presented as median (interquartile range) or number (%).
TOAST, Trial of Org 10172 in Acute Stroke Treatment; CE, cardioembolic; LAA, large artery atherosclerotic; UT, more than two causes; ICA, internal carotid artery; MCA, middle cerebral artery; RBC, red blood cell.

Supplementary Table 2. Time needed for analysis using ARIA and traditional methodology for each analyst

Analyst	ARIA	Traditional method	*P*
Professional			
Focused time	7.0 (5.0–10.0)	244.0 (209.5–278.0)	<0.001
Total time	24.5 (19.8–34.8)	254.0 (219.8–287.0)	<0.001
Trained			
Focused time	12.0 (10.0–14.2)	134.5 (113.0–168.8)	<0.001
Total time	35.5 (30.8–41.2)	143.5 (120.8–169.8)	<0.001
Untrained 1			
Focused time	1.0 (1.0–3.0)	320.5 (285.0–385.5)	<0.001
Total time	24.0 (19.8–29.0)	326.5 (294.8–391.0)	<0.001
Untrained 2			
Focused time	7.0 (4.0–13.0)	220.5 (199.0–253.0)	<0.001
Total time	29.5 (24.0–36.0)	232.5 (206.2–266.2)	<0.001
All analysts			
Focused time	7.0 (3.0–11.0)	231.0 (182.0–286.0)	<0.001
Total time	29.0 (21.0–37.0)	242.0 (183.5–296.5)	<0.001

Values are presented as median second (interquartile range).
ARIA, Automated Region-of-interest based Image Analysis.

Supplementary Table 3. Comparison of results grouped by dominant component of the thrombus of each patient

Variable	RBC dominant	PLT+Fibrin dominant
Spearman's correlation coefficient (rho)	0.954	0.767
P	<0.001	<0.001
Differences between analysts (stained ratio, %)		
ARIA	0.4 (0.1–1.4)	0.1 (0.0–0.5)
Traditional	0.7 (0.3–2.5)	0.4 (0.2–1.7)
P	0.001	<0.001
Time needed for analysis (focused time, sec)		
ARIA	7.0 (3.0–11.2)	6.5 (2.8–11.0)
Traditional	238.5 (140.2–293.2)	229.0 (177.8–282.8)
P	<0.001	<0.001

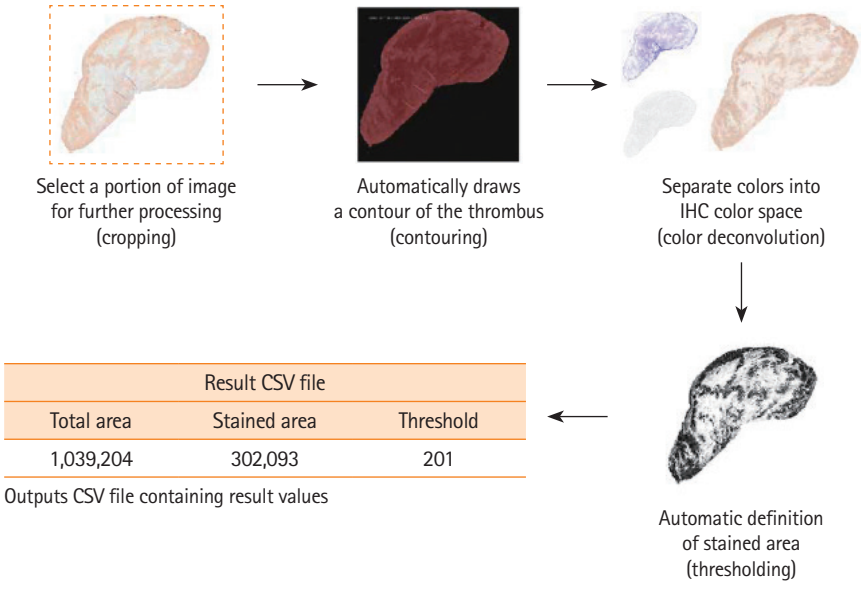
Values are presented as median (interquartile range) for differences between analysts and time needed for analysis.
RBC, red blood cell; PLT, platelet; ARIA, Automated Region-of-interest based Image Analysis.

Supplementary Table 4. Comparison of analyst results grouped by stained antibody

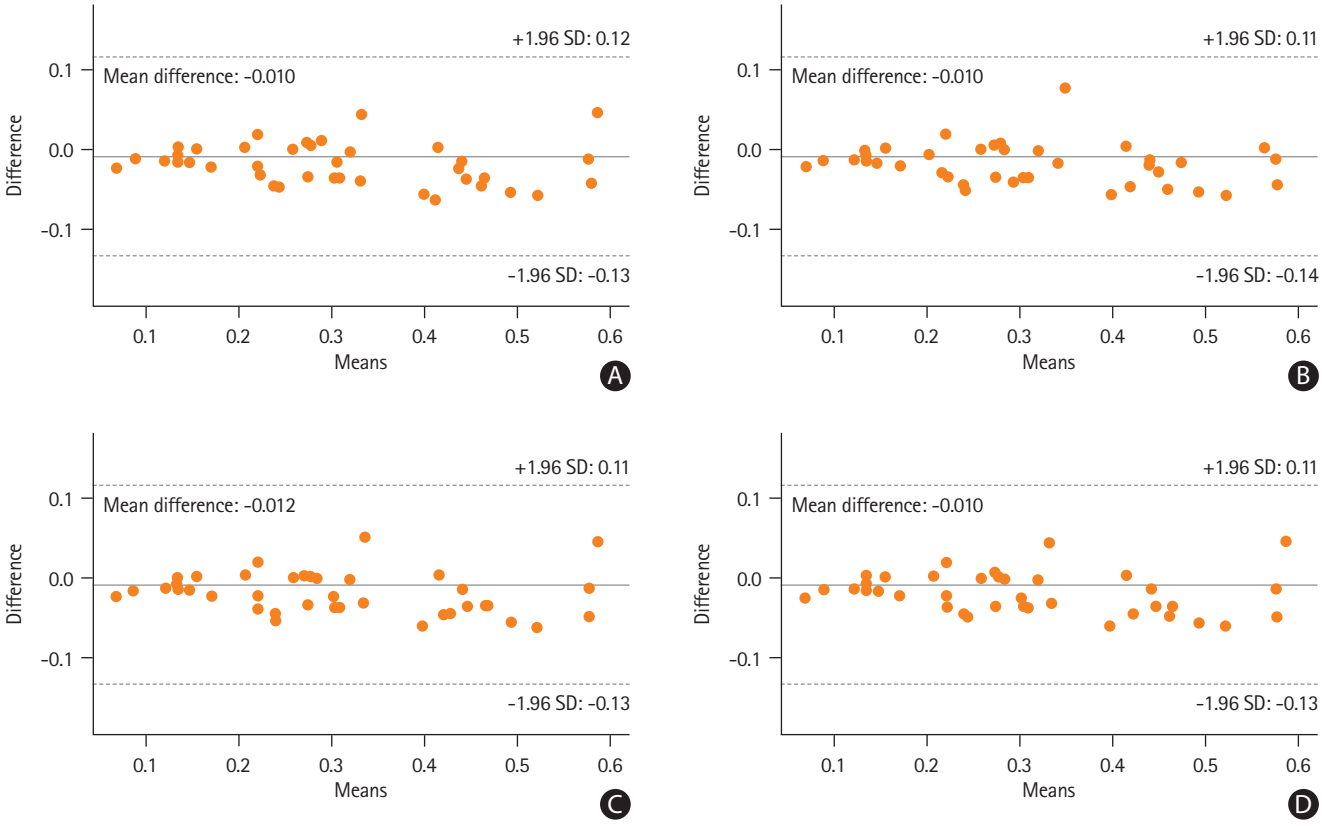
Variable	CD42b	Glycophorin A	Fibrinogen	Tissue factor	Tissue factor*
Spearman's correlation coefficient (rho)	0.888	0.922	0.767	0.441	0.820
P	<0.001	<0.001	<0.001	<0.001	<0.001
Differences between analysts (stained ratio, %)					
ARIA	0.2 (0.1–0.9)	0.4 (0.1–0.8)	0.3 (0.0–1.6)	0.1 (0.0–0.4)	0.1 (0.0–0.4)
Traditional	0.6 (0.3–0.7)	0.2 (0.0–0.7)	1.3 (0.6–4.3)	0.6 (0.1–2.2)	0.2 (0.1–2.2)
P	<0.001	0.185	<0.001	0.001	0.011
Time needed for analysis (focused time, sec)					
ARIA	7.0 (2.8–10.0)	4.5 (2.0–11.0)	7.0 (3.8–12.2)	8.0 (1.8–12.0)	8.0 (1.8–11.2)
Traditional	243.5 (183.2–283.8)	216.0 (164.8–291.2)	225.5 (184.5–282.8)	243.0 (169.5–295.5)	243.0 (165.8–295.5)
P	<0.001	<0.001	<0.001	<0.001	<0.001

Values are presented as median (interquartile range) for differences between analysts and time needed for analysis.
ARIA, Automated Region-of-interest based Image Analysis.
*Without one sample with exceptional disagreement between methods due to differences in thresholding algorithm.

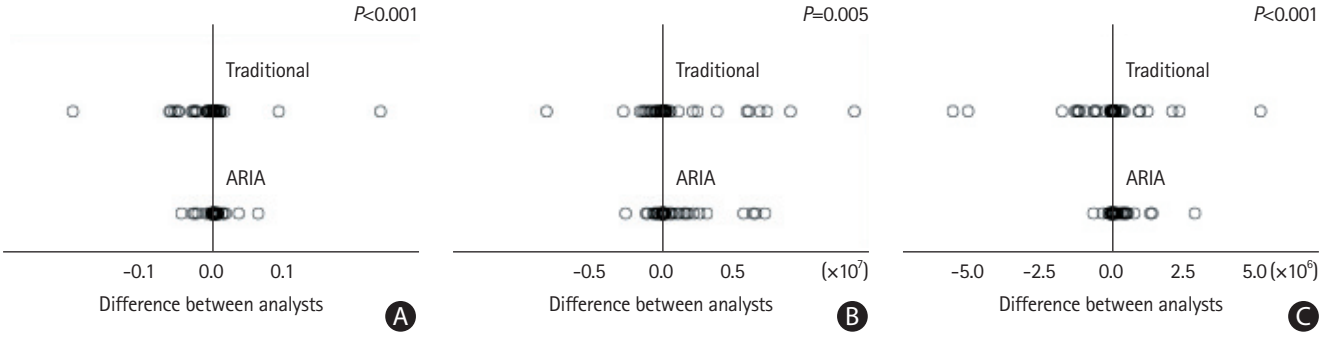
https://doi.org/10.5853/jos.2022.02054



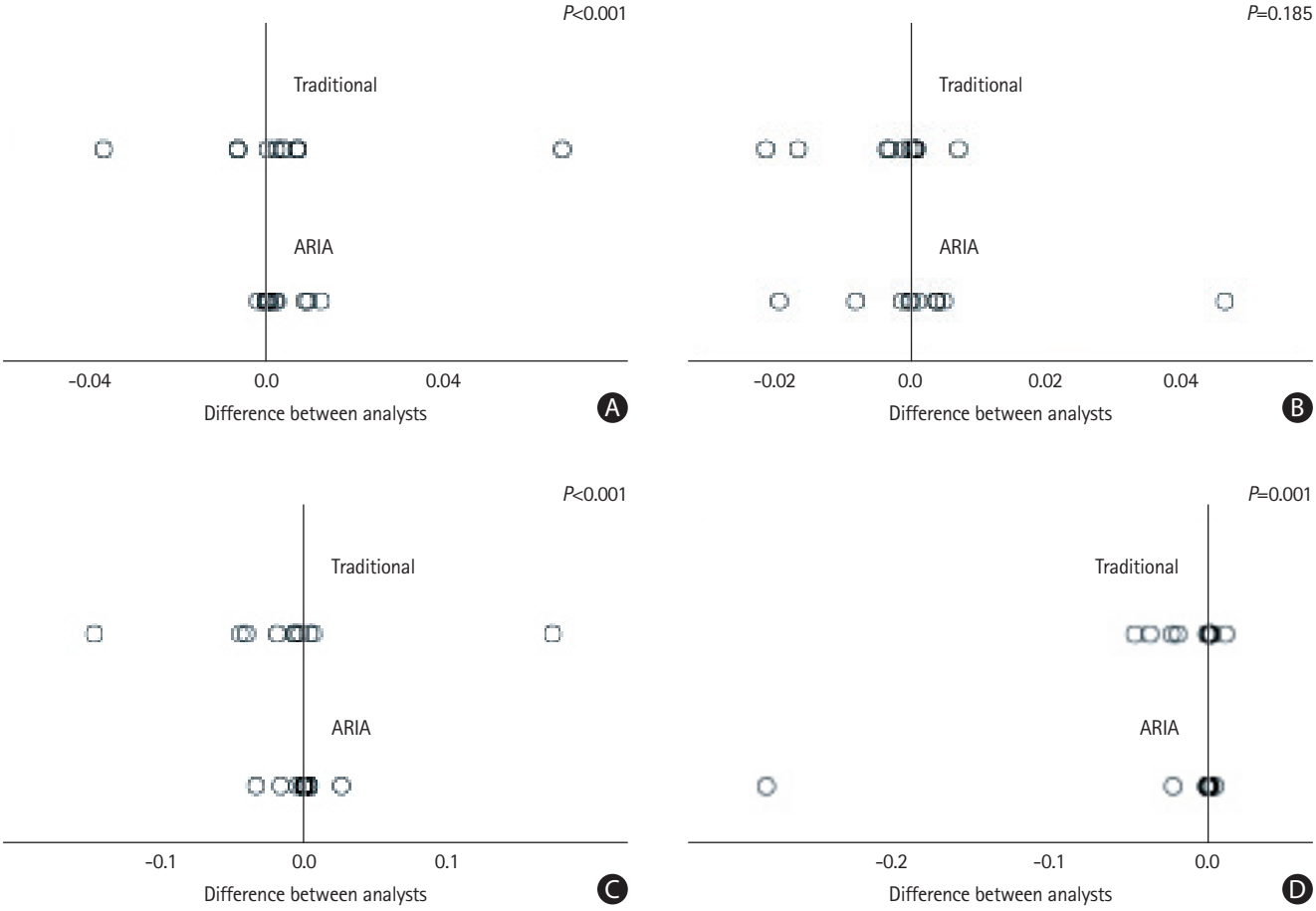
Supplementary Figure 1. Imaging analysis process performed by Automated Region-of-interest based Image Analysis (ARIA). IHC, immunohistochemistry; CSV, comma separated value.



Supplementary Figure 2. Bland-Altman plots of stained ratio results from each analyst using Automated Region-of-interest based Image Analysis (ARIA) compared to those of the professional analyst using the traditional method: (A) the professional analyst, (B) trained analyst, (C) untrained analyst 1, and (D) untrained analyst 2. SD, standard deviation.



Supplementary Figure 3. Scatter plot showing distribution of the result differences between analysts of (A) stained ratio, (B) total thrombus area, and (C) stained area for the same slide image using Automated Region-of-interest based Image Analysis (ARIA) and traditional methodology. Differences from all possible combination of two analysts were plotted.



Supplementary Figure 4. Scatter plot showing distribution of the result differences between analysts of (A) anti-CD42b, (B) anti-glycophorin A, (C) anti-fibrinogen, and (D) anti-tissue factor-stained slides using Automated Region-of-interest based Image Analysis (ARIA) and traditional methodology. Differences from all possible combination of two analysts were plotted.

https://doi.org/10.5853/jos.2022.02054