

## 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 magnification and 0.2528  $\mu\text{M}/\text{pixel}$  resolution. The Stereo Investigator Imaging system used the Virtual Slice module to acquire a montage of the entire slide at 400x 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.<sup>1</sup> 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.<sup>2</sup> 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 magnification and 0.2528  $\mu\text{M}/\text{pixel}$  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.