PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Automatic artery/vein classification in 2D-DSA images of stroke patients

Vivian van Asperen, Josefien van den Berg, Fleur Lycklama, Victoria Marting, Sandra Cornelissen, et al.

Vivian van Asperen, Josefien van den Berg, Fleur Lycklama, Victoria Marting, Sandra Cornelissen, Wim H. van Zwam, Jeanette Hofmeijer, Aad van der Lugt, Theo van Walsum, Matthijs van der Sluijs, Ruisheng Su, "Automatic artery/vein classification in 2D-DSA images of stroke patients," Proc. SPIE 12034, Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling, 120341N (4 April 2022); doi: 10.1117/12.2606412

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

Automatic Artery/Vein Classification in 2D-DSA Images of Stroke Patients

Vivian van Asperen[a], Josefien van den Berg[a], Fleur Lycklama[a], Victoria Marting[a], Sandra Cornelissen[b], Wim H. van Zwam[c], Jeanette Hofmeijer[d], Aad van der Lugt[b], Theo van Walsum[b], Matthijs van der Sluijs[b], and Ruisheng Su[b]

[a]Clinical Technology, Delft University of Technology, Delft, The Netherlands
[b]Department of Radiology & Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
[c]Department of Radiology & Nuclear Medicine, Maastricht University Medical Center, Maastricht, The Netherlands.
[d]Clinical Neurophysiology, University of Twente, Enschede, The Netherlands

ABSTRACT

To develop an objective system for perfusion assessment in digital subtraction angiography (DSA), artery-vein (A/V) classification is essential. In this study, an automated A/V classification system in 2D DSA images of stroke patients is proposed.

After preprocessing through vessel segmentation with a Frangi fitler and Gaussian smoothing, a time-intensity curve (TIC) of each vessel pixel was extracted and relevant parameters were calculated. Different combinations of input parameters were systematically tested to come to the optimal set of input parameters. The parameters formed the input for k-means (KM) and fuzzy c-means (FCM) clustering. Both algorithms were tested for clustering into 2 to 7 clusters. Cluster labeling was performed based on the average time to peak (TTP) of a cluster.

A reference standard consisted of manually annotated DSA images of the MR CLEAN registry. Outcome measures were accuracy, true artery rate (TAR) and true vein rate (TVR).

The optimal value for k was found to be 2 for both KM and FCM clustering. The optimal parameter set was: variance, standard deviation, maximal slope, peak width, time to peak, arrival time, maximal intensity and area under the TIC. No significant difference was found between FCM and KM clustering and. Both FCM and KM clustering yielded an average accuracy of 76%, average TAR of 74% and average TVR of 80%.

Keywords: Stroke, DSA, Vessels, Artery-vein classification

1. INTRODUCTION

Stroke is among the top causes of death worldwide,[1] and ischemic stroke accounts for 88% of all stroke cases.[2] A widely proven therapy for patients with acute ischemic stroke due to large vessel occlusion is endovascular thrombectomy (EVT).[3,4] Peri-procedural imaging and evaluation of treatment can be valuable for predicting treatment outcome, and subsequently treatment quality.[5] After an EVT procedure, digital subtraction angiography (DSA) imaging is performed to assess brain perfusion.[6] DSA images provide visualization of cerebral arteries and veins through intra-arterial contrast. By subtracting the background pixels from the images with the contrast in different phases, vessels are segmented and visualized.[7] A DSA sequence typically consists of a sequence of 20-30 frames, shot with a frequency of 0.5-4 frames per second.

Because a DSA consists of different images taken at different time points, it contains a time component that can be visualized in a time intensity curve (TIC) (fig. 1). In a TIC, a pixel's intensity is plotted in time. From this TIC, different characteristics can be identified which are descriptive for the blood flow in the DSA image.

Further author information:
V.v.A., J.v.d.B., F.L.a.N. and V.M. contributed equally to this manuscript.
(Send correspondence to R.S.,M.v.d.S.)
R.S.: E-mail: r.su@erasmusmc.nl
M.v.d.S.: E-mail: p.vandersluijs@erasmusmc.nl

Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling,
edited by Cristian A. Linte, Jeffrey H. Siewerdsen, Proc. of SPIE Vol. 12034, 120341N
 · 1605-7422 · doi: 10.1117/12.2606412

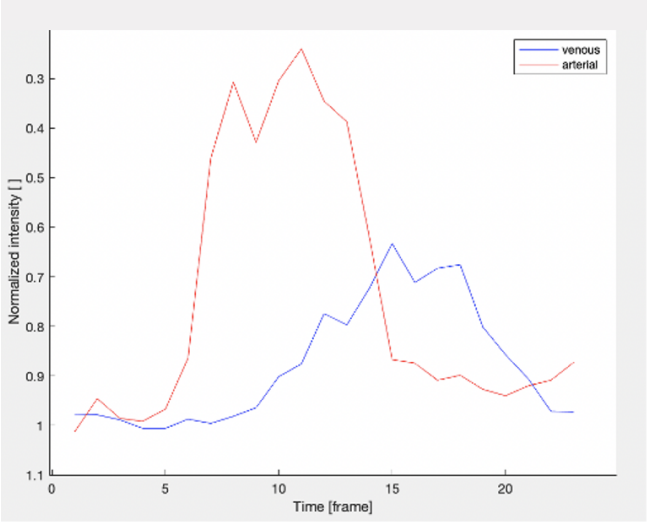


Figure 1. Example of a time intensity curve of an arterial pixel and a venous pixel.

Thrombolysis in cerebral infarction (TICI) score is a 5-point scale, and the most commonly used grading system to assess brain perfusion before and after EVT.[8] Assessment of DSA to grade TICI is now performed through manual inspection by a radiologist, which is time-consuming and error-prone.[6] It is affected by intra- and inter-observer variability, and (in)experience and (in)attentiveness of the observer. Furthermore, the scale is too coarse and the score is often overestimated,[6] which may lead to unfavorable results for the patient. A quantitative and standardized analysis of DSA images could be beneficial for efficient and accurate patient diagnosis and prognosis.

A step towards quantitative and standardized DSA analysis for perfusion assessment is an automated system for classification of arterial and venous structures. By excluding venous phase frames, the relevant vessels will become more visible and can be used for TICI assessment. Furthermore, other perfusion biomarkers may be found by separating the arterial from the venous phrase. Therefore, in this work, a fully automated artery/vein (A/V) classification system for 2-dimensional (2D) DSA images was proposed.

Multiple automatic classification systems have already been developed. One attempt has been made towards A/V classification of DSA images by Tache.[7] Tache exploited unsupervised machine learning for vessel clustering. However, Tache's system included visual inspection and manual classification of clustered vessels by a medical expert.

The rest of this paper is structured as follows: Section 2 provides an overview of the method. Sections 3 and 4 contain the experiments and results, respectively. Section 5 contains the discussion, and section 6 concludes the paper.

2. METHOD

The artery-vein classification method consists of four steps: pre-processing to segment the vessels, feature extraction from the TICs, an unsupervised clustering approach, and assignment of the correct label to each of the clusters.

2.1 Preprocessing

Before classifying vessel structures as arteries and veins, DSA images were pre-processed to obtain a vessel mask, to remove text and to smoothen the TICs.

Vessel segmentation is the separation of vessel structures from the background. Although it was not the principal goal of this work, it is essential to perform A/V classification, since this classification should be performed on vessels only. Vessel segmentation was performed by a Frangi filter. This filter can detect ridges like

vessel structures and fibers in image data. The filter uses eigenvectors of the Hessian matrix, which describes the second order derivatives of the image,[9] to compute the likeliness of a pixel to be a vessel.[10] The values given by the Frangi filter were binarized (value 0 for background, value 1 for vessels) and used to create a vessel mask. This mask served as an overlay for the DSA images.

The DSA images contained textual information, such as the patient number and the time point of the image. Segmentation by a Frangi filter would recognize these pixels as vessels and wrongly include them in the vessel mask. The pixels were excluded based on their extreme intensities (blacker or whiter than other pixels in the image). First, a matrix was created that assigns the value 1 to all black and white pixels, and value 0 to all other pixels. Then, morphological closing was performed. This created a new mask, which was combined with the original vessel mask.

Gaussian smoothing on the time dimension was applied to smooth the data and remove outliers. Gaussian smoothing applies a low pass filter to the data, which attenuates high frequencies. Features can be extracted more accurately after applying this filter because it smoothens the TICs.

2.2 Feature extraction

The TIC contains the pixel values over the time dimension. TICs are expected to be different for arteries compared to veins, considering that the contrast bolus first flows through arteries and later through veins.[7,11] Also, due to washout, intensity of contrast is generally lower in veins than in arteries. The different characteristics of the arterial and venous 'peaks' were used to identify multiple parameters to distinguish arteries from veins. Figure 2 shows a schematic representation of an arterial and a venous TIC with the extracted characteristics.

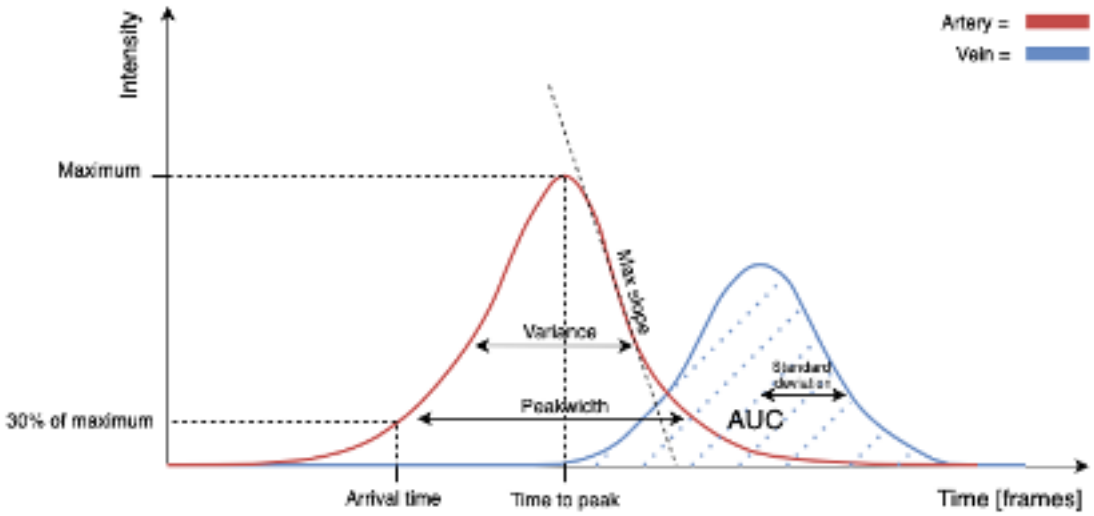


Figure 2. Schematic representation of arterial TIC (red) and venous TIC (blue) and extractable features.

Intensity-related features are expected to be suitable for distinguishing arteries from veins .[7,11] The following features were extracted from the Gaussian smoothed TIC of each individual pixel: area under the curve (AUC), peak intensity, time to peak (TTP), arrival time, peak width, variance, standard deviation, and maximum upward slope of the TIC. All features were normalized to a range from 0 to 1 after extraction, to avoid magnitude spread to play a role in clustering. The AUC of arterial TICs is larger in the first half of the curve, whereas venous TICs have a larger AUC in the second half of the curve.[11] Previous work on A/V classification also extracted the mean of the TIC.[7,11] Since AUC depicts the same as the mean intensity of a pixel, mean intensity was not included as a feature in this work. TTP depicts the time point of the peak intensity. Arrival time indicates the arrival of the contrast bolus in a vessel, and is the time point (frame) in which a pixel reaches 30% of the peak intensity. The exact percentage of the height is trivial, because the importance is in the ratio between all TICs. This value was also used for the peak width, which runs until the next moment the TIC has reached a value of 30% of the peak height again. Peaks below a visually determined threshold of 0.001 were excluded to avoid

including pixels of the parenchymal phase. The maximum slope was determined by calculating the maximum intensity difference between consecutive frames.

2.3 Clustering

Two unsupervised machine learning (UML) algorithms were applied to perform A/V classification: k-means (KM) clustering and fuzzy c-means clustering (FCM). In KM clustering, a number of clusters (k) are identified by choosing k random datapoints (prototypes) in a dataset. All datapoints are subsequently assigned to their closest prototype in an iterative process. Different distance metrics can be chosen to determine the closest prototype for each datapoint. In this work, Euclidean distance was implemented because data was normalized and not high-dimensional. Euclidean distance can be used in both KM and FCM clustering. FCM clustering is similar to KM, except this algorithm calculates the probability of each pixel belonging to one of k clusters. The class that is most likely for a pixel will be assigned.

2.4 Cluster labeling

Since UML uses unlabeled data as input, the output consists of unlabeled clusters. Clusters were labeled as arteries or veins based on prior knowledge of the circulation. Because of the direction of blood flow, the arterial peak always precedes the venous peak, and therefore the arterial TTP is expected to be lower. The average TTP of each cluster was calculated, after which the lowest and highest TTP were labeled as arterial and venous, respectively. If the system assigned pixels into two clusters (k=2), the cluster labeling was finished. If more than two clusters (k>2) were identified, two different approaches were explored.

The first approach was a k-nearest neighbour (KNN) algorithm, in which each cluster was assigned to a class based on the smallest difference between its average TTP and those of the outer two clusters. To correct for potential outliers, in a second approach, a KM algorithm was implemented that used the average TTP of the k clusters as input and the lowest and highest average TTPs as prototypes. To avoid confusion, we called this second approach 'labeled KM'.

3. EXPERIMENTS

3.1 Data

Data used in this research were obtained from the MR CLEAN Registry.[12] This registry contains images of all acute ischemic stroke patients that underwent EVT in the Netherlands since March 2014. The original dataset contains 50582 2D-DSA sequences from 3232 patients in anteroposterior (AP) view, lateral view or both. From this set, 984 sequences from 618 patients were selected for a previous study.[6] From this set, 1/3 of the sequences were selected for potential use in this study. Of these, 90 sequences remained after removing those with strong motion artifacts or incomplete venous phase, or because of too much overlap between arterial and venous phase (2 sequences). The 90 sequences were divided in a development set of 50 sequences and a test set of 40 sequences. A flowchart of the selection described above can be found in Appendix A. The development set was used in the experiments for finding the optimal combination of input parameters and amount of clusters. The test set was used to determine the final reported outcome measures.

3.2 Manual annotation

Manually annotated DSA image sets served as reference material to assess the performance of the proposed technique. Manual annotation of the DSA images was performed in MeVisLab software by the first four authors, working in pairs to minimize inter-user variability. An experienced radiologist, dr. S.A.P. Cornelissen, was available for a second opinion if doubt occurred. An extensive and detailed protocol on the annotation can be found in Appendix B. Pixels were either annotated as artery with value 1, or vein with value 2. The pixels in arterial and venous overlapping areas were given value 3.

3.3 Implementation

Finally, the algorithm that follows the method as described in 2 was tested on the test set consisting of 40 DSA sequences. A sigma of [1 1 2] was used for Gaussian smoothing and a fixed seed point was chosen to ensure consequent outcomes when rerunning the algorithm. The algorithm was run on the test set twice: once with KM clustering and once with FCM clustering. All experiments were performed in MATLAB.

3.4 Evaluation metrics

The resulting images obtained by the classification system were compared pixel by pixel with the reference images. For every image, the accuracy, true artery rate (TAR) and true vein rate (TVR) were calculated. Pixels that were classified as both arterial and venous in the reference standard, were classified correctly in both cases, if the system classified this pixel either arterial or venous. Equation 1 explains how accuracy is calculated. A_t means true artery, A_f means false artery, V_t is true vein, V_f indicates false vein. A false artery and false vein were defined as venous pixels that were recognized as arteries and arterial pixels recognized as veins, respectively.

$$Accuracy = \frac{A_t + V_t}{A_t + A_f + V_t + V_f} \tag{1}$$

The TAR (equation 2) and TVR (equation 3) are the probabilities that an actual artery or vein will be recognized as such.

$$TAR = \frac{A_t}{A_t + A_f} \tag{2}$$

$$TVR = \frac{V_t}{V_t + V_f} \tag{3}$$

The methods for calculating accuracy, TAR and TVR were previously described and used by Sun et al.[13] This way of quantifying classification performance explicitly addressed distinctiveness of arteries and veins, while background was not taken into account.

To gain insight into the performance of the proposed A/V classification system when using an optimal segmentation, the results of classification while using the Frangi filter-based segmentation and the segmentation from manual annotation were compared with a two-tailed, two-sample, paired t-test.

3.5 Determine amount of clusters

Different settings (amount of clusters, clustering method, cluster labeling method and included features) were tested on the development set of 50 DSA images. To determine the amount of clusters that was most appropriate, the performance of the system was evaluated for different values of k. 50 DSA images were classified with k=2, k=3, k=4, k=5, k=6 and k=7. An ANOVA F-test was performed to determine whether the average accuracies for different values of k are significantly different.

3.6 Feature selection

An ablation study was performed on the same development set to assess the influence of each feature on classification quality. Baseline performance was measured with all features. Then, features were excluded alternately. If performance increased, the feature was left out. If performance decreased or remained the same, the feature was inserted back into the function. A one-way ANOVA F-test was performed to calculate the statistical difference between different combinations of features used.

3.7 Method assessment

The A/V classification obtained from the results of the experiments described in Section 3.5 and Section 3.6 was tested on the test set twice. In the first run, KM clustering was used. In the second run, FCM clustering was used. Accuracy, TAR and TVR were reported for both KM and FCM.

4. RESULTS

4.1 Preprocessing

Figure 3 shows an example of a manually annotated DSA frame. The included DSA images were preprocessed. First, a vessel mask was applied (fig. 4) and text was removed (fig. 5). After this, Gaussian smoothing was applied to the TICs (fig. 6).

Figure 3. Example of manually annotated DSA image

Figure 4. Vessel mask of a DSA image made with a Frangi filter. White areas are included as vessels, black areas indicate background.

Figure 5. Vessel mask of a DSA image made with a Frangi filter after text removal was applied.

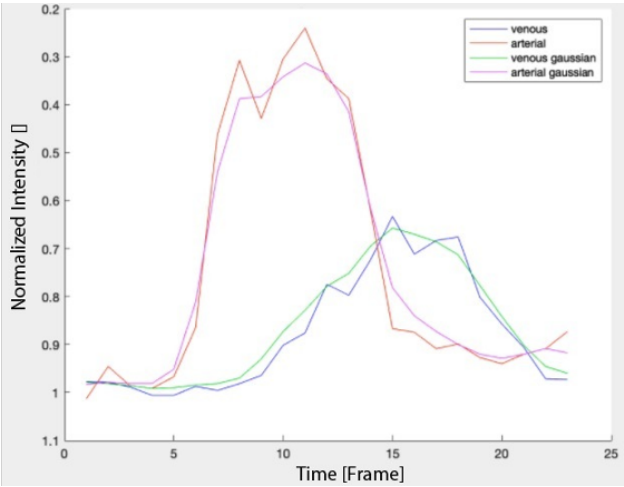


Figure 6. Gaussian smoothed time intensity curves of an arterial (red) and venous (green) pixel

4.2 Parameter optimization

All experiments were performed in labeled KM and KNN, combined with either KM or FCM clustering. Every combination of these settings was tested for clustering in 2, 3, 4, 5, 6 or 7 clusters. The results of the accuracy are reported in figures 7, 8, 9 and 10. Table 1 shows the average outcome measures of 50 DSA images for different approaches to determine the number of clusters and the method to label these clusters. The performance of KNN and labeled KM in labeling clusters as the right class were identical. KNN was selected randomly for the final A/V classification system. The amount of clusters k that yielded highest results was 2, for both FCM and KM clustering. For k=3, k=4, k=5 and k=6, outliers were found with accuracies as low as 0.55.

A one-way ANOVA F-test was performed to calculate the significance of the difference in classification performance when using different amounts of clusters and clustering-cluster labeling combinations. The only difference that was found to be significant is between using 2 or 3 clusters, when deploying FCM clustering. Therefore, we perform clustering in 2 clusters for A/V classification in the test set.

4.3 Feature selection

Statistical analysis of the ablation study revealed there was no difference in performance when using different input features in both KM clustering ($p = 0.60$) and FCM clustering ($p = 1.00$) (Figure 11). Therefore, all

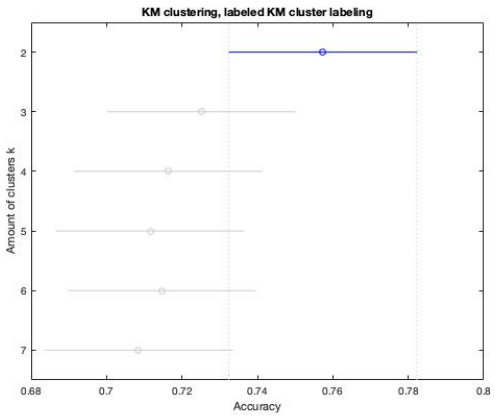


Figure 7. Forest plot of accuracies of artery/vein (A/V) classification in 2, 3, 4, 5, 6 and 7 clusters with k-means (KM) clustering and labeled KM cluster labeling.

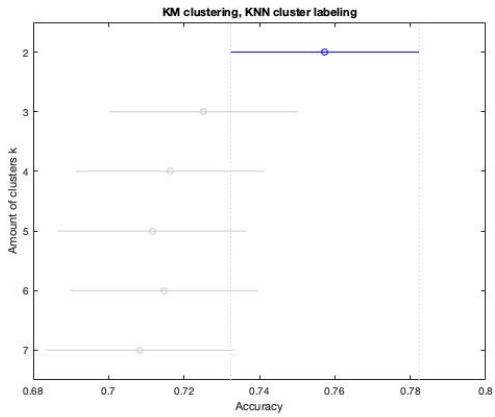


Figure 8. Forest plot of accuracies of A/V classification in 2, 3, 4, 5, 6 and 7 clusters with KM clustering and k-nearest neighbor (KNN) cluster labeling.

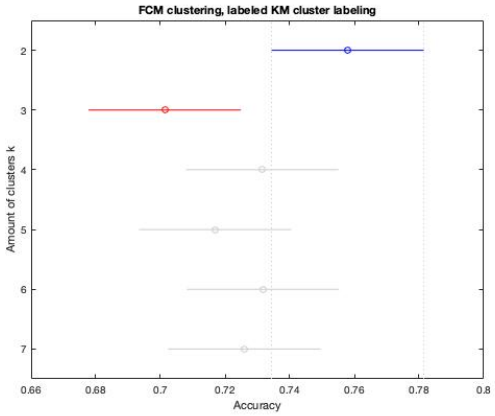


Figure 9. Forest plot of accuracies of A/V classification in 2, 3, 4, 5, 6 and 7 clusters with fuzzy c-means (FCM) clustering and K-means (KM) cluster labeling.

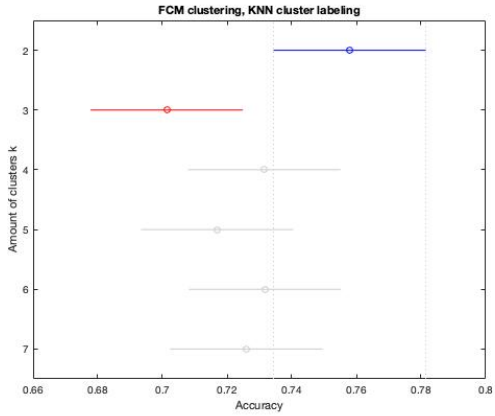


Figure 10. Forest plot of accuracies of A/V classification in 2, 3, 4, 5, 6 and 7 clusters with FCM clustering and KNN cluster labeling

parameters (arrival time, peak width, AUC, max slope, standard deviation, variance, peak intensity and TTP) were included in the clustering algorithm.

4.4 Method assessment

Based on the experiments described above, the final system performed clustering in 2 clusters with eight input features, after which class assignment was performed through a KNN algorithm. Both FCM and KM clustering with these settings were performed on the test set of 40 2D-DSA images. The mean outcome values of the accuracy, TAR and TVR for both clustering algorithms are listed in Table 2. A/V classification with FCM clustering and KM clustering yielded equal performances, with an accuracy of 0.76, TAR 0.74 and TVR 0.80. Figure 12 shows visual results of A/V classification using KM and FCM clustering.

Comparison of the A/V classification system with the highest performance on Frangi-segmented image and manually segmented image showed no significant difference in accuracy (0.76 (frangi) vs. 0.77 (manual), p=0.154). In contrast, the TAR was significantly higher for the Frangi segmentation versus manual segmentation (0.75 (frangi) vs. 0.74 (manual), p=0.0276), and the TVR was significantly lower (0.80 (frangi), vs 0.83 (manual) p=0.001).

Figure 13 shows the distribution of the accuracies of the DSA images, for the total and both views. The average accuracy of the total number of images is 0.76, the average accuracy of the images in AP view is 0.75

k	KNN labeling		KM labeling	
	KM	FCM	KM	FCM
2	0.76 ± 0.055	0.76 ± 0.056	0.76 ± 0.055	0.76 ± 0.056
3	0.73 ± 0.071	0.70 ± 0.078	0.73 ± 0.071	0.70 ± 0.078
4	0.72 ± 0.102	0.73 ± 0.099	0.72 ± 0.102	0.73 ± 0.099
5	0.71 ± 0.103	0.72 ± 0.098	0.71 ± 0.103	0.72 ± 0.098
6	0.71 ± 0.085	0.73 ± 0.076	0.71 ± 0.085	0.73 ± 0.076
7	0.71 ± 0.100	0.73 ± 0.080	0.71 ± 0.100	0.73 ± 0.080

Table 1. Accuracies of A/V classification with eight parameters for different values of k, KNN = k-nearest neigbour cluster labeling; KM=k-means; sd = standard deviation; FCM=fuzzy-c means

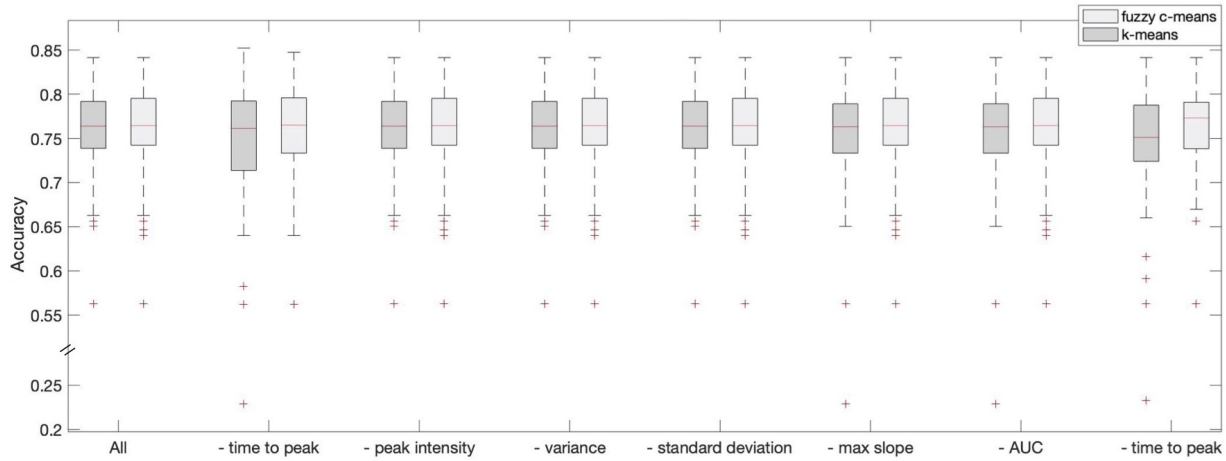
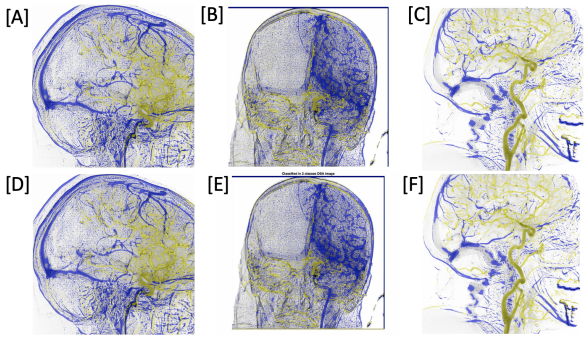


Figure 11. Performance of different feature sets as input for clustering. The first column shows performance including all features: time to peak, peak intensity, variance, standard deviation, maximal slope, area under the curve (AUC), peak width and arrival time. Other columns show the performance after leaving out features alternately.



[!h]
Figure 12. Examples of images with different outcome values for different clustering algorithms. Accuracies fuzzy-c means: [A] (0.77), [B] (0.55), [C] (0.90); Accuracies k-means: [D] (0.77), [E] (0.55), [F] (0.90)

and the average accuracy of the images in lateral view is 0.78. No significant difference was found between the AP and lateral images (p=0.156).

5. DISCUSSION

In this work, a fully automatic artery/vein classification system was presented. We have implemented unsupervised machine learning techniques to cluster vessel pixels based on different features and assign either an arterial or venous label to each cluster. Below, we will reflect on the findings following our methods. Next to this, limitations of this work will be discussed, as well as directions for future research.

An overall accuracy of 0.76 was found when using parameters that appeared to be optimal for classification

	k-means	fuzzy c-means
Acc	0.76 ± 0.074	0.76 ± 0.078
TAR	0.74 ± 0.106	0.74 ± 0.101
TVR	0.80 ± 0.111	0.80 ± 0.109

Table 2. Performance of A/V classification on the test set. Acc = accuracy; TAR = true artery rate; TVR = true vein rate.

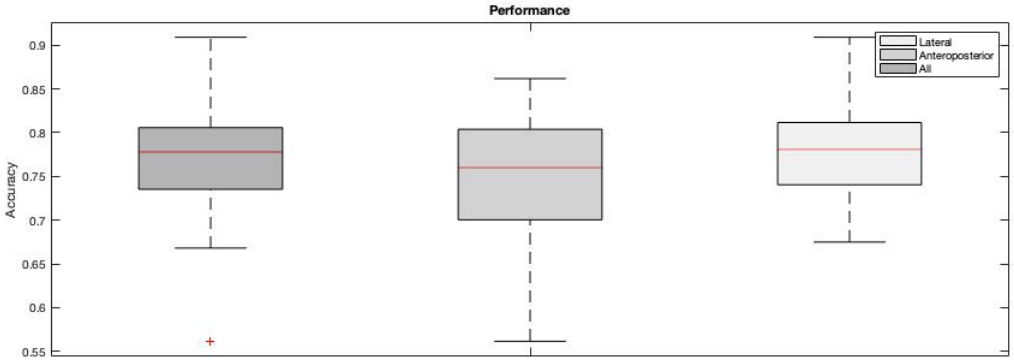


Figure 13. Performance of AP and lateral view images compared to the total performance

on the dataset used in this work. Previous studies performed by others performed accuracies that were in general higher. The only other study that was found that performed A/V classification on DSA images also used FCM and reported an accuracy of 0.91 and 0.86 for arteries and veins respectively, tested on data from 5 patients. Individual accuracies in this work ranged from 0.906 to 0.344 for arteries and from 0.936 to 0.491 for veins. For further improvement of the A/V classification system, it could be very valuable to look at the differences between images that were classified with high accuracy and images that were classified with low accuracy.

A repetitive superior outcome of images from the lateral point of view over those from the AP point of view was found for this dataset. It suggests that the proposed system works better for DSA images from a lateral point of view. A plausible explanation for this is that there is less overlap of structuring in these images than in images from an AP point of view. For some clinical settings, the system may be of less value.

To make well-founded decisions for choosing the optimal feature set and algorithm settings for the classification system, statistical tests were performed on the ablation study for feature selection and on the experiments to determine the optimal value for k. The different experiments performed did not have significantly different outcomes. Therefore, features and parameters were selected based only on their performance on the development dataset. There exists a chance that different feature combinations and algorithm settings may yield better results when tested on other datasets.

Labeling of clusters was now done based on TTP because of the knowledge on blood circulation direction, that is universal for each individual. Based on TTP, classification of two clusters yielded the highest overall accuracy, implying better methods and features may exist to decide what cluster belongs to what class when k>2. Seen from the low outliers for k=3, k=4, k=5 and k=6, in some cases, cluster labeling based on TTP yielded accuracies as low as 0.55. By systematic experiments, the added value of other parameters in this cluster-labeling can be examined, such as the time of bolus arrival. Also, other rule-based decision algorithms with varying thresholds may be explored to come to a cluster labeling system for more than two clusters.

The issue of class imbalance, in which an unjustified high performance due to included background pixels would be reported, is prevented by calculating the accuracy of pixels that are included in the vessel mask only. However, the approximation by the Frangi filter may have excluded vessel pixels that were labeled as such in the reference standard.[14] To show the potential contribution a perfect segmentation may have, the average accuracy with a manually annotated vessel mask was reported and yielded 0.77.

5.1 Limitations

The reference standard consisted of DSA images that were manually annotated by third year students, who in case of doubt, received help from an experienced radiologist. Even in the case of a classification system that is more accurate than human judgment, comparison will yield performances below 1. Wrongly classified vessels in the manual annotation, or missed vessels, will automatically and wrongly decrease the performance of a good classification system. In future research, this problem can be minimized by letting multiple experienced radiologists perform the manual annotation.

When selecting images, images with too many artefacts to perform manual annotation were left out of the dataset. This may give a distorted display of the performance of this A/V classification system, compared to a clinical situation.

The dataset used for developing and testing consisted of 90 DSA images. An amount higher than this was not feasible given the time as every annotated image had to be as truthful as possible and evaluated with care to obtain a trustworthy reference standard. The annotated set was split in a development and test set of 50 and 40, respectively. In supervised machine learning, a common distribution for training and testing data is 70% and 30%, respectively. Since UML originally does not require training and we use the development set to merely optimize the settings, we opted for a more equal distribution to make sure the results on the test dataset are representative, while still trying to avoid overfitting. However, a larger dataset for developing and testing may have yielded more trustworthy results.

The proposed system uses fixed seed points for the clustering to make sure the results are consistent for the same image when the algorithm is run multiple times. The influence of the location of the initial seed points on the final outcome has not been tested systematically. Therefore, it is unknown whether other initial seed point locations may have yielded different results.

5.2 Future research

In this study, eight different input parameters were evaluated on their contribution to A/V classification. One other study on this topic performed a fourier transform on the TIC, yielding even different parameters. A gamma fit with a gamma variate function was performed by the same study.[7] In the system proposed in this work, such fit was not proposed since parameters similar to the gamma fit parameter were already used as an input. Nevertheless, extracting these parameters from a gamma fit may yield different results.

Our work has performed preprocessing steps, such as text removal and Gaussian smoothing. More preprocessing of the images may improve performance of the classification. For example, a more precise vessel segmentation tool could ensure fewer artefacts are recognized as a vessel. Segmentation was beyond the scope of our research. In future studies, different segmentation techniques could be combined with the A/V classification results to enhance results by adequately removing artifacts and including vessels more precisely.

The value of k for the amount of clusters was now chosen to be 2. However, this decision was based on the highest overall accuracy. In individual cases, a k higher than 2 yielded the best classification result. In future studies, the feasibility of a system with a varying amount of k may be explored, to come to higher classification results. When more than two clusters are formed, multiple clusters need to be labeled as either artery or vein. Future studies may look into even more ways to divide clusters in an artery-vein ratio that differs per cluster, as the division between arterial and venous clusters is not expected to be universal for each image.

When looking at applications of A/V classification, follow-up research can be done on the relationship between patient outcome and the time between the arterial and venous peak in the time intensity curves. For example, this gives information on blood flow speed in the brain, which can in turn be linked to clinical outcome.

6. CONCLUSION

In this work, an automated classification system for A/V separation was proposed based on unsupervised machine learning. The feature set and algorithm settings were determined on a set of 50 development images and assessed on 40 test images. A maximal average accuracy of 0.76 was achieved when performing clustering through either FCM or KM clustering, with eight TIC-derived features, two clusters and cluster labeling based on TTP. The TAR and TVR scored 0.74 and 0.80, respectively, for A/V classification with either KM or FCM clustering.

APPENDIX A. RESULTS OF DATA ACQUISITION

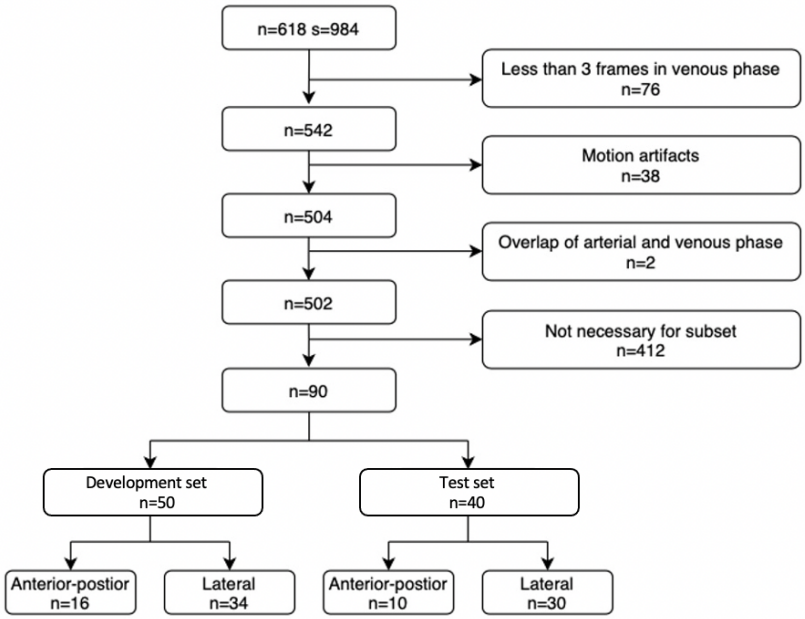


Figure 14. Flowchart of the selection process of 2D DSA images to obtain a development set and a test set.

APPENDIX B. ANNOTATION PROTOCOL

MeVisLab 2.7.1 (VC12-64) was used for semi-automatic annotation of DSA images. First, DICOM files of the images were loaded into the tool. Then, luminance was adjusted to visualize vessels. Image sequences were excluded when they contained prominent (motion) artefacts and/or less than three venous phase frames. The arterial and venous phase needed to be defined to support semi-automatic annotation. The phases were defined according to the following criteria:

- First arterial frame: appearance of contrast in image;
- Last arterial frame: no more arterial progression;
- First venous frame: appearance of contrast in venous structures and arterial contrast has washed out;
- Last venous frame: last frame with clearly visible contrast.

By selecting the frames of the different phases, the algorithm will analyze the pixels one by one and will give a suggestion of which pixels are arteries and which are veins. The parenchymal phase was not defined because this work only addressed the vessels that are stained in the arterial and venous phase. Microvasculature was excluded from the final DSA image by applying the above definitions.

Manually, the threshold for pixel intensity is determined for each DSA sequence. If the threshold is too high, relevant pixels will be missed. If the threshold is too low, background pixels will be falsely labeled as an artery of a vein. Next, a part of the artery or vein is selected manually by ticking a labeled region. All the pixels in this region are then recognized by the automatic algorithm as an artery or vein, and will be labeled. If necessary, pixels are excluded, that are incorrectly selected as an artery of vein. This can be relevant when falsely labeled background pixels and correct labeled pixels are connected as one region, or when arteries and veins overlap and are labeled as being one structure. Pixels that are wrongly defined as background can be added manually by selecting the region.

ACKNOWLEDGEMENTS

We thank the MR CLEAN Registry investigators for sharing the DSA image data. The MR CLEAN Registry was funded and carried out by the Erasmus University Medical Centre, Amsterdam UMC location AMC, and Maastricht University Medical Centre. The study was additionally funded by the Applied Scientific Institute for Neuromodulation (Toegepast Wetenschappelijk Instituut voor Neuromodulatie).

REFERENCES

[1] WHO, G., "Global health estimates 2016: deaths by cause, age, sex, by country and by region, 2000–2016," (2018).

[2] Higashida, R. T. and Furlan, A. J., "Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke," *stroke* **34**(8), e109–e137 (2003).

[3] Goyal, M., Menon, B. K., van Zwam, W. H., Dippel, D. W., Mitchell, P. J., Demchuk, A. M., Dávalos, A., Majoie, C. B., van der Lugt, A., De Miquel, M. A., et al., "Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials," *The Lancet* **387**(10029), 1723–1731 (2016).

[4] Berkhemer, O. A., Fransen, P. S., Beumer, D., Van Den Berg, L. A., Lingsma, H. F., Yoo, A. J., Schonewille, W. J., Vos, J. A., Nederkoorn, P. J., Wermer, M. J., et al., "A randomized trial of intraarterial treatment for acute ischemic stroke," *n Engl J Med* **372**, 11–20 (2015).

[5] Dargazanli, C., Consoli, A., Barral, M., Labreuche, J., Redjem, H., Ciccio, G., Smajda, S., Desilles, J.-P., Taylor, G., Preda, C., et al., "Impact of modified tici 3 versus modified tici 2b reperfusion score to predict good outcome following endovascular therapy," *American Journal of Neuroradiology* **38**(1), 90–96 (2017).

[6] Su, R., Cornelissen, S. A. P., van der Sluijs, M., van Es, A. C. G. M., van Zwam, W. H., Dippel, D. W. J., Lycklama, G., van Doormaal, P. J., Niessen, W. J., van der Lugt, A., and van Walsum, T., "autotici: Automatic brain tissue reperfusion scoring on 2d dsa images of acute ischemic stroke patients," *IEEE Transactions on Medical Imaging* **40**(9), 2380–2391 (2021).

[7] Tache, I.-A., "Blood vessels separation on angiograms," *University Politehnica of Bucharest Scientific Bulletin Series C-electrical Engineering and Computer Science* **77**(4), 143–154 (2015).

[8] Higashida, R. T. and Furlan, A. J., "Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke," *Stroke* **34**(8), e109–e137 (2003).

[9] Rudzki, M., "Vessel detection method based on eigenvalues of the hessian matrix and its applicability to airway tree segmentation," in [*Proceedings of the 11th International PhD Workshop OWD*], 100–105 (2009).

[10] Barra, B., El Hadji, S., De Momi, E., Ferrigno, G., Cardinale, F., and Baselli, G., "The introduction of capillary structures in 4d simulated vascular tree for art 3.5 d algorithm further validation," in [*Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*], **10135**, 101350G, International Society for Optics and Photonics (2017).

[11] Laue, H. O., Oei, M. T., Chen, L., Kompan, I., Hahn, H. K., Prokop, M., and Manniesing, R., "Automated artery and vein detection in 4d-ct data with an unsupervised classification algorithm of the time intensity curves," in [*Medical Imaging 2013: Image Processing*], **8669**, 86691W, International Society for Optics and Photonics (2013).

[12] Jansen, I. G., Mulder, M. J., and Goldhoorn, R.-J. B., "Endovascular treatment for acute ischaemic stroke in routine clinical practice: prospective, observational cohort study (mr clean registry)," *bmj* **360** (2018).

[13] Sun, G., Liu, X., Gong, J., and Gao, L., "Artery-venous classification in fluorescein angiograms based on region growing with sequential and structural features," *Computer methods and programs in biomedicine* **190**, 105340 (2020).

[14] Tiu, E., "Metrics to evaluate your semantic segmentation model," *Towards datascience* (2019).