ARTICLE TEMPLATE

SPD Matrices Representing Artery Anatomy for First-pass Effect Prediction by Aggregated Networks with Multi-scale Attentions

Jianhai Zhang[a], Fouzi Bala[a], Petra Cimflova[a], Nishita Singh[b], Faysal Benali[a,c], Michael D. Hill[a,b], Bijoy K. Menon[a,b], Wu Qiu[d]

[a]University of Calgary, Calgary, Canada
[b]Foothills Medical Centre, Calgary, Canada
[c]Maastricht University Medical Center+, Maastricht, Netherlands
[d]Huazhong University of Science and Technology, Wuhan, China

ARTICLE HISTORY
Compiled September 13, 2022

ABSTRACT
First-pass Effect(FPE) for Endovascular Therapy(EVT) is associated with good clinical outcome (mRS$\leq$2) of patients with Acute Ischemic Stroke (AIS). A rapid and accurate prediction of FPE before EVT can help neurointerventionists plan the procedure and avoid delays in restoration of cerebral blood flow. However, there are rare studies focused on FPE prediction on arterial vessel anatomy immediately. The intractable difficulty lies in extracting discriminative features to represent a wide variety of vessels with irregular vessel shapes. In this paper, we propose a Symmetric Positive Definite(SPD) matrix-based feature representation extracted from the centerline of arterial vessels, encoding the global discriminative information over the artery for predicting FPE. Subsequently, a collaborative network of multi-scale Convolutional Neural Network(CNN) and Multiple Layer Perception (MLP) with specific attentions is developed. Specifically, the CNN is used for capturing the features among the mutli-scale local neighbors of the curve. MLPs are used for capturing more prominent global discriminative features at different scales. The attention mechanism is used to better filter and extract the useful information for feature fusion. Quantitative experimental results demonstrate that our proposed method is able to predict FPE accurately, outperforming the manually defined features and traditional machine learning-based methods in this task, regarding the metrics of AUC, precision, sensitivity, specificity and accuracy.

KEYWORDS
First-pass Effect; SPD Matrices; collaborative network

1. Introduction

Endovascular Therapy (EVT) is the current standard care of Acute Ischemic Stroke(AIS) with large vessel occlusion Flottmann et al. (2021). Previous studies Goyal et al. (2016); Zaidat et al. (2018a); Aubertin et al. (2021); Abdullayev et al. (2021) have shown that rapid and effective reperfusion of the occluded artery during stroke have a significant impact on clinical outcomes, and therefore excellent reperfusion after one pass, i.e., First-pass Effect(FPE) should be the aim of every EVT procedure. Several factors are linked to FPE He et al. (2022); Zaidat et al. (2018b), such as occlusion

Corresponding Author: Wu Qiu, Email: Qiu.wu.ch@gmail.com

location on baseline imaging, operator experience, and device type used for thrombus retrieval. However, the impact of artery anatomy on FPE and other procedural outcomes is still understudied. A significant part of the EVT procedure is represented by the time-cost on establishing intracranial access, which is influenced by the degree of tortuosity Morris et al. (2011) of extracranial and intracranial arteries Mokin et al. (2020). Therefore, difficult vessel anatomy can result in failure or delay in accessing intracranial vessels and retrieval of intracranial thrombi, and subsequently lower rates of FPE. Identification of unfavorable vascular anatomy on baseline imaging and accurate estimation of the likelihood of FPE before EVT can help neurointerventionists plan the procedure and avoid delays in restoration of cerebral blood flow Chen et al. (2018).

The intractable difficulty for predicting FPE using arterial vessels lies in shape variability and representation of arterial vessels. Previous methods Di Maria et al. (2021); Jadhav et al. (2021) mainly used shallow learning techniques with handcrafted features on CTA images, such as vessel length, vessel angles at certain anatomical points etc,. No existing methods were well established for this task. Prvious methods suffered from three facts: 1) manually placing anatomical points on the vessels is time-consuming and observer dependent; 2) the representation for vessel anatomy is not sufficient and suitable for representing the vessels; 3) the used off-the-shelf classification algorithms may not be optimal. To address these issues, we propose a deep learning approach to predict FPE while using a novel representation on curves representing vessels. Specifically, the whole vessel shape can be described and embedded into a Symmetric Positive Definite(SPD) matrix Huang et al. (2014) which is located on a Riemanian manifold. The obtained SPD matrices can be mapped on a Euclidean space defined by Riemannian metric for better feature extraction. A collaborative network of CNN and MLP with multi-scale attention mechanism Vaswani et al. (2017) is then proposed to exploit efficient features using both global and local dependencies of the feature descriptions of the curves for better FPE prediction.

Our main contributions are summarized as follows:

- A novel curve representation based on SPD matrices on manifold is proposed, which can be processed and integrated into a deep learning framework.
- An efficient collaborative neural network with multi-scale attention MLP is developed.
- Several loss functions are proposed to handle the imbalanced distributions of the training data.

To the best of our knowledge, there were few studies on predicting FPE using deep learning. This paper reports the first study on FPE prediction using a collaborative neural network of deep learning with global artery curve descriptions.

2. Proposed Methodology

2.1. Feature Representation for a Centerline Curve

The artery centerline from the origin of brachiocephalic truncus of the aortic arch to MCA bifurcation can be denoted as a directed curve $\mathbf{C} \in \mathbb{R}^{N\times 3}$ in 3-dimensional coordinates, where N is the number of points on the curve. In order to standardize the extracted features of all the curves in different image acquisition space, each curve $\mathbf{C}$ needs to be mapped or transmitted into a feature space independent on coordinates

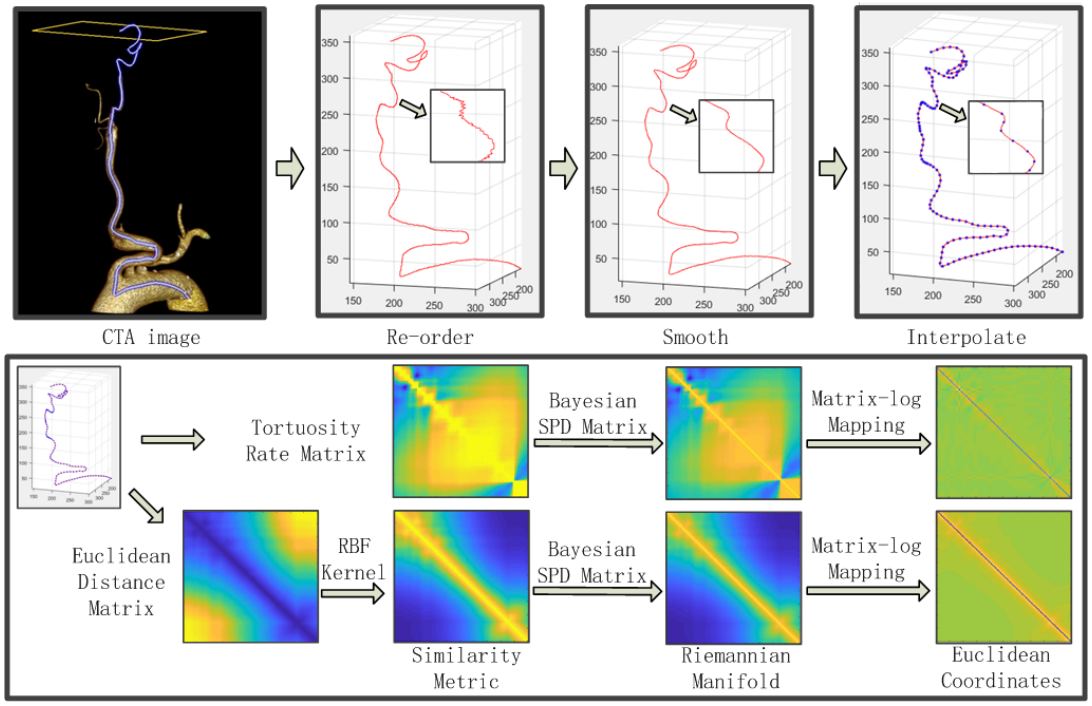

Figure 1. The overview of our proposed feature representation

Vemulapalli et al. (2014). The degree of tortuosity Togay-Isikay et al. (2005) is relevant to the FPE in EVT procedure. In our work, we define the tortuosity rate between two points p_i and p_j in a curve as:

$$k(p_i, p_j) = 1 - \mathrm{L_E}/\mathrm{L_C} \in [0, 1) \tag{1}$$

where $\mathrm{L_E}$ is the Euclidean distance and $\mathrm{L_C}$ is the curve length or geodesic distance between the two points p_i and p_j. The larger value of k means more tortuous for the curve with two end-points, whereas zero of k means the straight line. To describe the curve at scale both locally and globally, the tortuosity rates between arbitrary two points in a curve are calculated to form a matrix $\mathbf{K}_\mathcal{M} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ with the diagonal elements equal to 1. It should be noted that $\mathbf{K}_\mathcal{M}$ is a Symmetric Positive Definite(SPD) Matrix. To guarantee that $\mathbf{K}_\mathcal{M}$ is a strict SPD matrix for the valid numerical operation, an additional Bayesian regularization Zhang et al. (2021) is conducted on $\mathbf{K}_\mathcal{M}$ by the formula:

$$\tilde{\mathbf{K}}_\mathcal{M} = \pi\mathbf{K}_0 + (1-\pi)\mathbf{K}_\mathcal{M} \tag{2}$$

where $\pi \in [0, 1]$ controls the intensity of the prior knowledge and the prior is set as $\mathbf{K}_0 = \mathbf{I} \odot \mathbf{K}_\mathcal{M}$, $\odot$ means matrix Hadamard product. Since SPD matrices are the elements on a Riemannian manifold Huang et al. (2014), there exists a homeomorphic mapping $\varphi(\cdot) : \mathcal{M} \mapsto \mathbb{E}$ which is able to map matrix $\mathbf{K}_\mathcal{M}$ into a Euclidean space as a vector representation.

Since Euclidean distance denoted by $\mathrm{L_E}$ between two points on a curve can be thought of as a non-normalized version of tortuosity measurement in (1), it is also able to offer additional discriminative information for FPE prediction. Thus, a matrix $\mathbf{K}_\mathrm{E}$ of

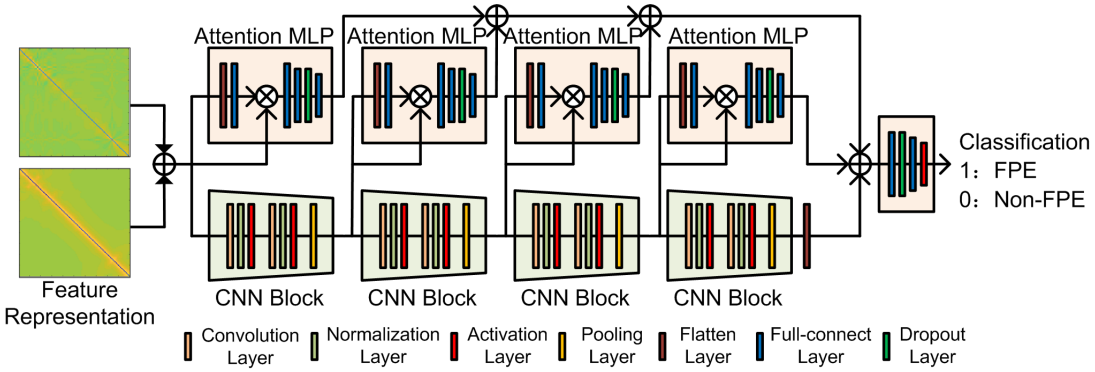

Figure 2. Illustration of the proposed collaborative neural network

measuring Euclidean distance $\mathrm{L_E}$ needs to be aggregated into the overall representation of a curve. In order to align the range of two matrices $\mathbf{K}_{\mathcal{M}} \in [0,1]$ and $\mathbf{K}_{\mathrm{E}} \in [0,T]$, a Radial Basis Function (RBF) Hofmann et al. (2008) $\kappa(d) = exp(-\sigma^2 * d)$ is applied to map all of the entries of $\mathbf{K}_{\mathrm{E}}$ to the range of $[0,1]$ with a bandwidth σ, where T is the maximum of the Euclidean distance due to the physical body restriction. The mapped $\mathbf{K}_{\mathrm{E}}$ by RBF is an analogous similarity-based matrix and is on the Riemannian manifold as well. Thus, the two regularized matrices by (2) can be combined to form a composite representation $\tilde{\mathbf{K}}_{\mathcal{M}} \times \tilde{\mathbf{K}}_{\mathrm{E}}$ on a product manifold. This composite representation is subsequently mapped and aggregated into the Euclidean coordinates $\mathbf{K}_{\mathrm{R}} \in \mathbb{R}^{2\times \mathrm{N}\times \mathrm{N}}$ as the final representation for a curve by using a Riemannian metric of matrix-log mapping $\mathrm{Log}(\mathbf{K}) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{(\mathbf{K}-\mathbf{I})^k}{k}$. Since any SPD matrix $\mathbf{K}$ can be diagonalized into $\mathbf{K} = \mathrm{U}\mathbf{\Lambda}\mathrm{U}^T$ Huang et al. (2015), where the columns of U are eigenvectors and the diagonal elements of $\mathbf{\Lambda}$ are the related eigenvalues, the matrix-log mapping can be calculated as follows:

$$\mathrm{Log}(\mathbf{K}) = \mathrm{U}\big((log(\mathbf{\Lambda 1})\mathbf{1}^T \odot \mathbf{I})\big)\mathrm{U}^T \tag{3}$$

where $\mathbf{1} \in \mathbb{R}^{\mathrm{N}\times 1}$ is a vector with all elements of 1, italic *log* means the logarithm operator for each element in the vector.

2.2. Collaborative Learning of multi-scale CNN and MLP for Classification

To better extract efficient discriminative features on the directed curves, one of the widely used ideas is to align all the curves on a generated mean curve by the alignment techniques, such as Dynamic Time Warping(DTW) Folgado et al. (2018); Vemulapalli et al. (2014). However, a wide variety of different vessel anatomies make alignment intractable and time-consuming. Alternatively, convolutional neural network (CNN) is fit into our situation where the combination of SPD matrices $\tilde{\mathbf{K}}_{\mathcal{M}} \times \tilde{\mathbf{K}}_{\mathrm{E}}$ on the product manifold is viewed as the two-dimensional image input with the channel of 2, i.e., each SPD matrix is viewed as one channel. Thus, the introduction of CNN is utilized to extract effective features based on the relationships among adjacent pixels at multiple scales. Unfortunately, traditional CNN has a weakness of being short of extracting the global information at different scales and aggregating them together. In our work, in order to mitigate the difficulty in capturing global dependencies along

the curves, we introduce a Flattening Gated-Attention (FGA), which is illustrated in upper part of the network in Fig. 2, to capture the global information at different scales. As stated in many successful applications Vaswani et al. (2017), attention mechanism could induce more benefits of rapid training and improvement of performance. Though FGA in our work is relatively simple, it is efficient for improving the performance of predicting FPE. Finally, all the information at multiple scales from CNN and FGA are aggregated into a fully connected layer followed by a Multi-layer Perceptron (MLP) to obtain the final classification results. The architecture of the network is illustrated in Fig.2.

2.3. Loss Functions of Neural Network

Binary Cross-entropy(BCE) loss is used for this classification task, where 1 means presence of FPE and 0 means absence of FPE. Since the positive and negative samples are imbalanced, a mixed loss function combined sensitivity $\mathcal{L}_{\mathbf{S}}$, precision $\mathcal{L}_{\mathbf{P}}$ and specified true-positive $\mathcal{L}_{\mathbf{T}}$ is proposed to improve classification accuracy. A collaborative learning framework is therefore introduced for minimizing the holistic loss function:

$$\mathcal{L} = \sigma_1 \mathcal{L}_{\mathbf{B}} + \sigma_2 \mathcal{L}_{\mathbf{S}} + \sigma_3 \mathcal{L}_{\mathbf{P}} + \sigma_4 \mathcal{L}_{\mathbf{T}} \tag{4}$$

where $\mathbf{\Sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ is the hyper-parameters of the related weighting coefficients for the sub-losses. The holistic loss function is differentiable so it could be optimized by the gradient-based methods, such as SGD, Adam etc. The detailed expressions of loss functions are shown as follows:

Classification Loss $\mathcal{L}_{\mathbf{B}}$ for the Holistic Network For the output $\hat{y}$ of the network and data label y, the BCE loss is defined as:

$$\mathcal{L}_{\mathbf{B}} = \frac{1}{\mathrm{M}} \sum_{n=1}^{\mathrm{M}} y_n \mathrm{Log}\hat{y}_n + (1 - y_n)\mathrm{Log}(1 - \hat{y}_n) \tag{5}$$

where M is the number of the min-batch.

Enhanced Losses $\mathcal{L}_{\mathbf{S}}$, $\mathcal{L}_{\mathbf{P}}$, $\mathcal{L}_{\mathbf{T}}$ for Positive Sample Learning The losses of sensitivity $\mathcal{L}_{\mathbf{S}}$, precision $\mathcal{L}_{\mathbf{P}}$ and true-positive $\mathcal{L}_{\mathbf{T}}$ are defined as:

$$\mathcal{L}_{\mathbf{S}} = \frac{1}{\mathrm{M}} \sum_{n=1}^{\mathrm{M}} \frac{y_n \odot \hat{y}_n}{y_n \odot \hat{y}_n + y_n \odot (1 - \hat{y}_n) + \epsilon} \tag{6}$$

$$\mathcal{L}_{\mathbf{P}} = \frac{1}{\mathrm{M}} \sum_{n=1}^{\mathrm{M}} \frac{y_n \odot \hat{y}_n}{y_n \odot \hat{y}_n + (1 - y_n) \odot \hat{y}_n + \epsilon} \tag{7}$$

$$\mathcal{L}_{\mathbf{T}} = \frac{1}{\mathrm{M}} \sum_{n=1}^{\mathrm{M}} y_n \odot \hat{y}_n \tag{8}$$

3. Experiments

3.1. Dataset

The dataset contains 539 patients in total, where the derivation set has 429 training samples and testing set has 110 testing samples. The FPE ratios of positive samples in training and testing sets are 31.7% and 31.82%, respectively. The thin-slice ($\leq$2.5 mm) neck CTA images in patients with AIS were included. Arterial centerlines from the aortic arch to the intracranial occlusion site were semi-automatically traced by three experienced radiologists and neuroradiologists on CTA images in order to eliminate the segmentation errors introduced by automated segmentation algorithms. One example is shown in upper left image in Fig.1.

3.2. Pre-processing

Since the points in the curves are stored disorderly, a re-arrangement is needed for adjusting the order of the curve. Subsequently, the curves have analogous zig-zag shapes which cause the unreal measurements of longer curve distance, so a smooth operation is conducted on the curve with moving average of 5 points. To accurately represent the curves, the last process is to adjust the intervals between two adjacent points in a curve. A interpolating operation based on splines is conducted to make the intervals between two adjacent points consistent. This process is illustrated in the upper layer in Fig.1.

Each centerline curve is discretized and denoted by N points with the identical intervals, where N could be divided by 2 evenly for at least 4 times in order to fit into the design of CNN. The larger N means the higher resolution(or spacing) for the representation of the curve. The higher spacing may improve the performance of the neural network, but accordingly increase the computational cost. As a tradeoff, $N = 96$ is applied in our work. Therefore, the generated representation of curve $\mathbf{K}_{\mathrm{R}}$ has the dimension of $dim = 2 \times 96 \times 96$, where 2 is the number of two matrices represented as the channels for CNN, 96×96 is the matrix dimension. For MLP, $\mathbf{K}_{\mathrm{R}}$ is flattened into a vector as the input. Though we used a fixed number of points($N = 96$) to represent a curve, they don't appear the obvious variations on curve lengths. Thus, we empirically applied an currently optimal number of 96. Subsequently, a neural network is able to adaptively learn the required and discriminative features on the varied data. Also, in order to improve the generalization of the model on the curves with different spacings, we conducted the data augments which will introduced in the next section in details.

In addition, since the elements in two matrices of $\mathbf{K}_{\mathrm{R}}$ are under the different Gaussian distributions, they are required to be in the same distribution for further feeding into the network. Thus, each matrix after matrix-log mapping is processed with whitening of subtracting mean value followed by dividing standard variation.

3.3. Implementation Details

Data Augmentation To improve the generalization and performance of the network, we conduct two types of data augmentation in the experiments.

i) Noise Augmentation Since the data is already whitening in the stage of preprocessing and follows a standard normal distribution $\mathcal{N}(0, 1)$. We add a disturbing noise matrix with the same shape of $\mathbf{K}_{\mathrm{R}}$. Each element of the noise matrix is generated by the rules: With a occurrence probability of $p_1 = 0.8$, a small value is sampled from

a Gaussian distribution $\mathcal{N}(0, 0.02a)$. With a occurrence probability of $p_2 = 0.15$, a bit larger value is sampled from a Gaussian distribution $\mathcal{N}(0, 0.05b)$. With the last probability interval of $p_3 = 0.05$, a large value is sampled from the most oscillated Gaussian distribution $\mathcal{N}(0, 0.1c)$, where the scalars a, b, c follow a uniform distribution of range $[0, 1]$, i.e., $a, b, c \sim \mathbf{U}(0, 1)$. For each element in $\mathbf{K}_{\mathrm{R}}$, we set a probability $p_o = 0.5$ controlling the occurrence of doing the noise augmentation. To guarantee the symmetric property for two sub-matrices in the noise matrix, a upper-triangle matrix with diagonal elements is generated first using the above noise generation mechanism. Then, the transposed lower-triangle matrix is added to form the symmetric matrices for the final noise augmentation.

ii) Scaling Augmentation Some parts from two sides of the curve are randomly cut off, then the trimmed part is resized into the consistent shape with the training samples using interpolating. Specifically, we set two occurrence probabilities $\alpha = 0.2$ and $\beta = 0.8$. The reason why β is larger than α is that the most of the redundancy length exists on one side because of the body anatomy and the data acquisition. Accordingly, the cutting percentages are set $\left[0, \lceil 10g \rceil\right]$ and $\left[0, \lceil 20h \rceil\right]$ for two sides, respectively, where the scalars $g, h \sim \mathbf{U}(0, 1)$. Then, the remaining sub-matrix along the diagonal-axis is kept, guaranteeing the symmetric property of matrix for the final scaling augmentation.

Hyper-parameters Settings The parameter σ is set to 0.1 in the RBF kernel function for mapping $\mathbf{K}_{\mathrm{E}}$. The elements in hyper-parameter set in (4) are identical $\mathbf{\Sigma} = (1, 1, 1, 1)$.

Metrics for Assessment The metric of Accuracy(Acc.) is used to assess the classification performance for the models. Since the positive samples are relatively less, the metric index of F1-score or accuracy may bias the results. For example, if 70% samples are negative and the classifier assigns all the testing samples as negative class, the accuracy will retain high up to 70%, but the sensitivity and precision both will be 0. Thus, four more metrics for more tangible evaluation, i.e., AUC, Sensitivity(Sen.), Specificity(Spe.) and Precision(Pre.) are additionally used to comprehensively assess the model performance. The corresponding numbers of prediction(n/N) are also offered below four metrics of precision, sensitivity, specificity, and accuracy.

3.4. Experimental Results

Quantitative results in Table.1 show that our proposed feature representation matrix based on tortuosities and Euclidean distance measurements with the collaborative neural network (CNN+MLP-Attention+Augmentation) obtained the best performance with AUC of 0.61, Precision of 0.46, Sensitivity of 0.49, Specificity of 0.73, and Accuracy of 0.65. The proposed method outperforms the baseline models based on the individual CNN or MLP components and the traditional SVM classifier with hand crafted features. Ablation studies show that several proposed contributing techniques are helpful to improve the model performance, including the collaborative neural network of CNN and MLP(CNN+MLP), MLP with attention mechanism(MLP-Attention) to learn the global dependencies and data augmentation method(Augmentation) for the generalization of the model.

Table 1. Quantitative performance on our FPE prediction dataset, where MKL: Multi-kernel Learning, MLP: Multi-layer Perceptron, CNN: Convolution Neural Netork, ATT.: Attention, Aug.:Augmentation

Features	Methods	AUC	Pre.	Sen.	Spe.	Acc.
Manual	MKL+SVM	0.54	0.42 8/19	0.23 8/35	**0.85** 64/75	**0.65** 72/110
$\mathbf{K}_{\mathrm{R}}$	MKL+SVM	0.58	0.45 13/29	0.37 13/35	0.79 59/75	**0.65** 72/110
$\mathbf{K}_{\mathrm{R}}$	MLP	0.57	40 17/42	**0.49** 17/35	0.67 50/75	0.61 50/75
$\mathbf{K}_{\mathrm{R}}$	CNN	0.57	0.41 14/34	0.40 14/35	0.73 55/75	0.63 69/110
$\mathbf{K}_{\mathrm{R}}$	CNN+MLP	0.59	0.44 15/34	0.43 15/35	0.75 56/75	0.65 71/110
$\mathbf{K}_{\mathrm{R}}$	CNN+MLP-Att.	0.59	0.42 17/40	**0.49** 17/35	0.69 52/75	0.63 69/110
$\mathbf{K}_{\mathrm{R}}$	CNN+MLP-Att.+Aug.	**0.61**	**0.46** 17/37	**0.49** 17/35	0.73 55/75	**0.65** 72/110

4. Discussion and Conclusion

Predicting FPE before EVT can help physicians plan the procedure and avoid delays in treatments. The current solutions for the FPE prediction is based on the occlusion location on baseline imaging, operator experience, and type of device. Though there has been attempts to predict FPE based on the CTA images by using the manually annotated features with off-the-self machine learning methods, the prediction accuracy was limited. In order to reduce the time-consuming annotation work and improve the prediction accuracy, we proposed a curve descriptor over the full arterial vessels without manual intervention(marking the land-markers, like end-points of MCA etc., for the further analysis) and related collaborative neural network for automatically predict the FPE on CTA images. The quantitative evaluations demonstrated the efficacy of the proposed method in Table.1. This work provides a new solution for predicting FPE using deep learning and a global descriptor of arterial vessels.

Predicting FPE from arterial vessels on CTA images is challenging. Previous methods required to manually place a few landmarks on the vessel centerlines to calculate curve tortuosity. In our experiments, four landmarkers along the arterial vessel from aorta to MCA are annotated by three radiologists and neuroradiologists with $\geq$3 year experiences. The tortuosity rate among the landmarkers and the angulations at the landmarkers are calculated as the features for machine learning models. This traditional machine learning methods with manually annotated features obtained the worst results of AUC of 0.54 and sensitivity of 0.23 in our experiments in Table.1. When the proposed feature representation matrix $\mathbf{K}_{\mathrm{R}}$ is introduced, the AUC was increased to 0.58, and the sensitivity was increased to 0.37 as well. All the ablation experiments in Table.1 showed that several neural network techniques aggregated with our proposed feature representation $\mathbf{K}_{\mathrm{R}}$ can jointly improve the prediction performance. The best model(CNN+MLP-Attention+Augmentation) has reached the highest AUC of 0.61, precision of 0.46, sensitivity of 0.49 and accuracy of 0.65, significantly outperforming the baseline models based on manually annotated features and traditional machine learning methods.

However, this study has several limitations. 1) Our dataset is limited, much more

data is required to mitigate the issue of data imbalance between positive and negative samples, and to assess the validity and generalizability of the trained model more comprehensively. 2)In fact, arteries were semi-automatically segmented using region growing in our in-house software, followed by manual annotations of centerline points. We should mention that the centerline extraction could be automated in the future with the off-the-shelf or self-developed methods, such as nnU-net Isensee et al. (2021), Medical Transformer Valanarasu et al. (2021), etc. These tasks can be aggregated into an end-to-end holistic processing method. 3) The feature representation just includes two sources of curve descriptors, future work will involve more useful features into the feature representations.

In conclusion, a SPD-matrix-based feature representation of arterial vessels and a related collaborative network with attention mechanism have been proposed for predicting FPE on CTA images. Experimental results demonstrate promising results, suggesting its potentials of being used in clinical practice.

References

Abdullayev N, Maus V, Behme D, Barnikol UB, Kutschke S, Stockero A, Goertz L, Celik E, Zaeske C, Borggrefe J, et al. 2021. True first-pass effect in basilar artery occlusions first-pass complete reperfusion improves clinical outcome in stroke thrombectomy patients. Journal of Clinical Neuroscience. 89:33–38.

Aubertin M, Weisenburger-Lile D, Gory B, Richard S, Blanc R, Ducroux C, Piotin M, Labreuche J, Lucas L, Dargazanli C, et al. 2021. First-pass effect in basilar artery occlusions: Insights from the endovascular treatment of ischemic stroke registry. Stroke. 52(12):3777–3785.

Chen Z, Shi F, Gong X, Zhang R, Zhong W, Zhou Y, Lou M. 2018. Thrombus permeability on dynamic cta predicts good outcome after reperfusion therapy. American Journal of Neuroradiology. 39(10):1854–1859.

Di Maria F, Kyheng M, Consoli A, Desilles JP, Gory B, Richard S, Rodesch G, Labreuche J, Girot JB, Dargazanli C, et al. 2021. Identifying the predictors of first-pass effect and its influence on clinical outcome in the setting of endovascular thrombectomy for acute ischemic stroke: results from a multicentric prospective registry. International Journal of Stroke. 16(1):20–28.

Flottmann F, Brekenfeld C, Broocks G, Leischner H, McDonough R, Faizy TD, Deb-Chatterji M, Alegiani A, Thomalla G, Mpotsaris A, et al. 2021. Good clinical outcome decreases with number of retrieval attempts in stroke thrombectomy: beyond the first-pass effect. Stroke. 52(2):482–490.

Folgado D, Barandas M, Matias R, Martins R, Carvalho M, Gamboa H. 2018. Time alignment measurement for time series. Pattern Recognition. 81:268–279.

Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, Dávalos A, Majoie CB, van der Lugt A, De Miquel MA, et al. 2016. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. The Lancet. 387(10029):1723–1731.

He G, Deng J, Lu H, Wei L, Zhao Y, Zhu Y, Li Y. 2022. Thrombus enhancement sign on ct angiography is associated with the first pass effect of stent retrievers. Journal of NeuroInterventional Surgery.

Hofmann T, Schölkopf B, Smola AJ. 2008. Kernel methods in machine learning. The annals of statistics. 36(3):1171–1220.

Huang Z, Wang R, Shan S, Chen X. 2014. Learning euclidean-to-riemannian metric for point-to-set classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. p. 1677–1684.

Huang Z, Wang R, Shan S, Li X, Chen X. 2015. Log-euclidean metric learning on symmet-

ric positive definite manifold with application to image set classification. In: International conference on machine learning. PMLR. p. 720–729.

Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. 2021. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods. 18(2):203–211.

Jadhav AP, Desai SM, Budzik RF, Gupta R, Baxter B, English JD, Bartolini BM, Krajina A, Haussen DC, Nogueira RG, et al. 2021. First pass effect in patients with large vessel occlusion strokes undergoing neurothrombectomy: insights from the trevo retriever registry. Journal of NeuroInterventional Surgery. 13(7):619–622.

Mokin M, Primiani CT, Castonguay AC, Nogueira RG, Haussen DC, English JD, Satti SR, Chen J, Farid H, Borders C, et al. 2020. First pass effect in patients treated with the trevo stent-retriever: a track registry study analysis. Frontiers in Neurology. 11:83.

Morris SA, Orbach DB, Geva T, Singh MN, Gauvreau K, Lacro RV. 2011. Increased vertebral artery tortuosity index is associated with adverse outcomes in children and young adults with connective tissue disorders. Circulation. 124(4):388–396.

Togay-Isikay C, Kim J, Betterman K, Andrews C, Meads D, Tesh P, Tegeler C, Oztuna D. 2005. Carotid artery tortuosity, kinking, coiling: stroke risk factor, marker, or curiosity? Acta neurologica belgica. 105(2):68.

Valanarasu JMJ, Oza P, Hacihaliloglu I, Patel VM. 2021. Medical transformer: Gated axial-attention for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer. p. 36–46.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017. Attention is all you need. Advances in neural information processing systems. 30.

Vemulapalli R, Arrate F, Chellappa R. 2014. Human action recognition by representing 3d skeletons as points in a lie group. In: Proceedings of the IEEE conference on computer vision and pattern recognition. p. 588–595.

Zaidat OO, Castonguay AC, Linfante I, Gupta R, Martin CO, Holloway WE, Mueller-Kronast N, English JD, Dabus G, Malisch TW, et al. 2018a. First pass effect: a new measure for stroke thrombectomy devices. Stroke. 49(3):660–666.

Zaidat OO, Castonguay AC, Linfante I, Gupta R, Martin CO, Holloway WE, Mueller-Kronast N, English JD, Dabus G, Malisch TW, et al. 2018b. First pass effect a new measure for stroke thrombectomy devices. Stroke. 49(3):660–666.

Zhang J, Feng Z, Su Y, Xing M. 2021. Bayesian covariance representation with global informative prior for 3d action recognition. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM). 17(4):1–22.