

X-LMC: Cross-View Spatiotemporal Collateral Circulation Scoring from DSA

Maedeh Hafezi Moghadas¹, Hakim Baazaoui², Lukas Bastian Otto²,
Susanne Wegener², Björn Menze³, and Ezequiel De la Rosa³

¹ Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany
maedeh.hafezi@fau.de

² University Hospital Zurich, Zurich, Switzerland

³ University of Zurich, Zurich, Switzerland

Abstract. Digital subtraction angiography (DSA) is the reference standard for leptomeningeal collateral (LMC) assessment, providing critical prognostic insights to guide secondary treatment strategies, neurorehabilitation planning, and retrospective stroke research. However, clinical LMC grading via the ASITN/SIR scale relies on manual, highly variable visual inspection. We introduce X-LMC, a spatiotemporal framework for automated collateral scoring from time-resolved biplane DSA. The proposed architecture encodes spatial frame representations through a DINOv2 backbone, fuses orthogonal projections via a token-level cross-view attention module, and models representations of contrast bolus dynamics using a recurrent network architecture. We evaluate our framework on a multicenter dataset of 134 patients with M1-segment occlusions. In a 5-fold cross-validation setting, X-LMC yields higher point estimates than static architectures and spatiotemporal baselines adapted from related angiographic tasks, achieving a Quadratic Weighted Kappa (QWK) of 0.398 (vs. 0.322) and a dichotomized macro-F1 score of 0.711 (vs. 0.663) against the best-performing baseline. X-LMC performance also aligns with the observed clinical inter-rater agreement (QWK: 0.314). As the first DSA study attempting to automate LMC scoring, we demonstrate that multi-view temporal deep learning can capture collateral-specific contrast kinetics. Ultimately, these benchmarks delineate the clinical ambiguities and achievable performance boundaries of automated ASITN/SIR grading, establishing a reproducible foundation for objective hemodynamic phenotyping in stroke cohorts. Code is available at <https://github.com/maedehafezi/X-LMC>.

Keywords: Deep learning · cross-view attention · digital subtraction angiography · ischemic stroke.

1 Introduction

Acute ischemic stroke (AIS) remains a leading cause of mortality and disability worldwide [6]. Therapeutic approaches focus on rapid recanalization of an

M. Hafezi Moghadas and H. Baazaoui—Equal contribution.

B. Menze and E. De la Rosa—Shared senior authorship.

occluded blood vessel through endovascular treatment or intravenous thrombolysis. The viability of brain tissue critically depends on collateral flow, and especially on leptomeningeal collaterals (LMCs) at the pial surface of the brain that are “recruited” following an arterial obstruction. They expand due to pressure gradient changes and provide retrograde blood flow, thereby delaying the progression from penumbra to irreversible infarction [9]. Despite their prognostic significance, LMCs fall below the spatial resolution of conventional neuroimaging due to their small caliber (< 1 mm) and pial location. Consequently, clinicians must rely on indirect surrogates for LMC assessment, such as retrograde filling in computed tomography angiography (CTA) or perfusion delay maps [11, 14].

Intra-procedural digital subtraction angiography (DSA) is the reference standard for evaluating LMCs [12]. It provides high spatial resolution (0.08×0.08 to 0.2×0.2 mm² pixels), enabling precise visualization of the collateral microvasculature. Beyond supporting stroke etiology identification [20], collateral status supports secondary treatment and rehabilitation decision making. However, clinical grading via the reference American Society of Interventional and Therapeutic Neuroradiology/Society of Interventional Radiology (ASITN/SIR) scale [8] remains manual, time-intensive, and dependent on specialized expertise. Prior work reported poor interobserver agreement for angiographic ASITN/SIR collateral assessment, with $\kappa = 0.16 \pm 0.0065$ overall and $\kappa = 0.27 \pm 0.014$ after dichotomization into poor versus good collaterals [7], limiting reproducible large-scale cohort analysis. Automated collateral scoring has been mostly studied in (single/multi-phase) CTA and CT perfusion [18, 21]. However, these modalities provide *indirect* LMC surrogates, with limited spatial resolution compared to DSA. In DSA, deep learning has been applied to diverse tasks, including artery-vein segmentation [13, 23] and thrombus detection [15]. For reperfusion grading, frameworks automate (modified) Thrombolysis in Cerebral Infarction ((m)TICI) scoring via spatiotemporal recurrent networks (*DeepTICI*) [16], leveraging intensity-projection perfusion maps [22], or using cross-view fusion modules in *CVFSNet* [27]. While graph neural networks were recently introduced to segment collateral branches [4], direct grading of LMCs in DSA remains unaddressed.

Herein, we introduce X-LMC, a cross-view spatiotemporal framework for automated pre-interventional ASITN/SIR collateral scoring from biplane DSA series. By incorporating a bidirectional cross-view attention (X-Attn) module, the architecture dynamically exchanges token-level information between synchronized anteroposterior (AP) and lateral (Lat) projections, capturing complex contrast bolus kinetics within the microvasculature. As a first proof-of-concept for automated collateral assessment on standard DSA, this work establishes a reproducible foundation for large-scale hemodynamic phenotyping. Our main contributions are threefold: (i) the first deep learning framework to derive ordinal ASITN/SIR collateral scores directly from raw 2D+time biplane DSA series; (ii) a multi-view architecture employing a DINOv2 [17] vision backbone with bidirectional X-Attn modules to integrate orthogonal angiographic projections prior to spatiotemporal modeling; and (iii) a comprehensive evaluation against

static and spatiotemporal baselines, benchmarked directly against human inter-rater agreement in this clinically complex task.

2 Methods & Experimental Setup

We formulate the LMC scoring task as a multi-view spatiotemporal learning problem. Let $\mathcal{X} \subset \mathbb{R}^{T \times H \times W \times C}$ denote the space of time-resolved DSA sequences, where T , H , W , and C denote the number of temporal frames, spatial height, spatial width, and input channels, respectively. For each patient, we consider a paired observation $\mathbf{I} = \{I_{\text{AP}}, I_{\text{Lat}}\} \in \mathcal{X}^2$, corresponding to simultaneous biplane projections in the AP and Lat planes, providing orthogonal perspectives of the cerebral vasculature. Our objective is to learn a mapping $f_\theta : \mathcal{X}^2 \rightarrow \mathcal{Y}$, parameterized by weights θ , mapping angiographic dynamics to a discrete collateral score $y \in \mathcal{Y}$. The target labels follow the ASITN/SIR scale, the most widely adopted clinical scale for grading collateral flow in DSA based on the completeness and speed of retrograde contrast arrival to the ischemic territory [8], treated as an ordered label space $\mathcal{Y} = \{0, 1, \dots, 4\}$, where $y = 0$ represents the absence of collateral flow and $y = 4$ denotes complete and rapid contrast arrival. The proposed framework f_θ jointly optimizes spatial feature extraction from multi-view inputs and the modeling of contrast bolus kinetics over time, ensuring that the predicted score $\hat{y}$ captures both the anatomical extent of the collateral network and the temporal blood flow dynamics within the ischemic territory.

2.1 Spatiotemporal Framework

The mapping f_θ is implemented as an end-to-end framework designed to process the paired biplane observation $\mathbf{I} = \{I_{\text{AP}}, I_{\text{Lat}}\}$ of variable temporal length T . The architecture consists of two symmetric branches for parallel feature extraction from the sequences in $\mathcal{X}$, followed by a temporal reasoning module operating on the derived cross-view fused representation. The pipeline proceeds in three stages: (i) frame-wise spatial encoding, where features are extracted independently from each view I_v ; (ii) per-timepoint cross-view feature interaction, where information is exchanged between orthogonal projections to produce view-enhanced feature representations, which are subsequently aggregated into a unified representation h_t ; and (iii) bidirectional temporal modeling, to capture the temporal contrast-flow dynamics across the sequence length T (Fig. 1).

Multi-view spatial encoding. To obtain robust descriptors of the cerebral vasculature, we utilize DINOv2, a self-supervised Vision Transformer (ViT-B/14), as the spatial encoder for each angiographic view. Each frame $f_t^{(v)} \in I_v$ is processed independently to extract token-level representations. Specifically, the frame is partitioned into non-overlapping patches and prepended with a learnable class token, which, after processing through the transformer layers, yields a sequence of latent token embeddings $Z_t^{(v)} \in \mathbb{R}^{N \times D}$, where N denotes the number

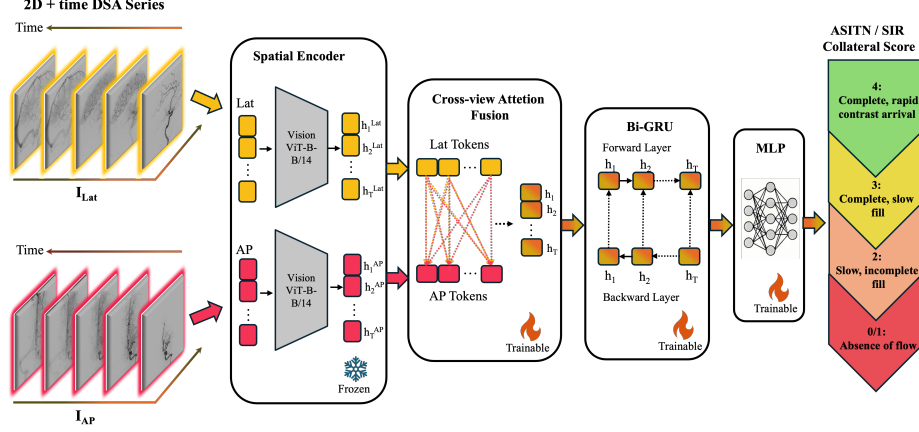


Fig. 1. Overview of the proposed X-LMC framework for collateral scoring.

of tokens including the class token and $D = 768$ is the embedding dimension. These token-level representations are retained for subsequent cross-view feature interaction, rather than reducing each frame to a single global descriptor before fusion. Due to the limited size of our cohort and the high capacity of the ViT-B/14 backbone, the encoder weights are kept frozen to serve as a fixed feature extractor, thereby preventing overfitting to the training set.

Feature fusion. To integrate complementary information from orthogonal angiographic projections while preserving temporal alignment, we employ bidirectional cross-view attention (X-Attn) [26]. For each timepoint $t \in \{1, \dots, T\}$, X-Attn is applied at the token level, allowing AP and Lat embeddings to mutually attend to one another. Given query $Q = Z_q W_Q$, key $K = Z_{kv} W_K$, and value $V = Z_{kv} W_V$ projections with learnable matrices W_Q, W_K, W_V , the attention output is computed as $\text{Attention}(Q, K, V) = \text{softmax}(QK^\top / \sqrt{d_k}) V$, and is applied bidirectionally as $\tilde{Z}_t^{(\text{AP})} = \text{X-Attn}(Z_t^{(\text{AP})}, Z_t^{(\text{Lat})})$ and $\tilde{Z}_t^{(\text{Lat})} = \text{X-Attn}(Z_t^{(\text{Lat})}, Z_t^{(\text{AP})})$. The updated class-token embeddings are summed to obtain the fused frame-level representation $h_t = \tilde{z}_{t,0}^{(\text{AP})} + \tilde{z}_{t,0}^{(\text{Lat})} \in \mathbb{R}^D$. The resulting sequence $\mathcal{H} = \{h_1, h_2, \dots, h_T\}$ serves directly as input to the temporal modeling module.

Perfusion dynamics via temporal learning. To model the temporal evolution of contrast dynamics across the sequence length T , we process the fused frame-level representations $\mathcal{H}$ using a bidirectional gated recurrent unit (Bi-GRU) [5], integrating contextual information from both earlier and later time-steps to capture the overall hemodynamic progression of contrast flow. For a GRU with hidden dimension H_d , the final forward ($\vec{h}_T$) and backward ($\overleftarrow{h}_1$)

hidden states are concatenated to form the global spatiotemporal representation $h_{\text{seq}} = [\vec{h}_T \parallel \overleftarrow{h}_1] \in \mathbb{R}^{2H_d}$. This representation is passed to an MLP classifier for collateral grade prediction.

2.2 Dataset and preprocessing

Due to the scarcity of public DSA datasets with collateral grading, we used data from the multicenter acute stroke MAGIC repository [1], including anonymized data of AIS patients. We included pre-interventional DSA imaging of 134 patients undergoing endovascular treatment for AIS. Each case consists of simultaneous biplane DSA series acquired via a single contrast injection. Collateral circulation was graded using the ASITN/SIR scale (0–4) by a fellow in interventional neuro-radiology and a neurology resident specializing in stroke. For supervised training and evaluation, the score assigned by the interventional neuroradiology fellow was used as the reference label. A subset of cases (N=98) was independently graded by both raters and used to quantify inter-rater agreement.

To minimize variability, we restricted the dataset to M1-segment middle cerebral artery (MCA) occlusions. To harmonize acquisitions, series were temporally standardized to 2 frames/s via linear interpolation and truncated at 30 frames. Pre-contrast baseline frames were removed to prioritize arterial and early venous phases, preserving hemodynamics critical for collateral assessment. Finally, images were resized to 224×224 pixels and normalized to $[0, 1]$ for compatibility with standard deep learning backbones.

2.3 Experiments and baselines

We evaluate the model on two tasks: multiclass ASITN/SIR grading (0–4) and binary classification of *good* ($n = 66$) versus *poor* (ASITN/SIR ≤ 2 , $n = 68$) collateralization. Due to its low prevalence (2.24%), grade 0 was merged with grade 1. Model performance was contextualized with inter-rater agreement and compared against: (i) a mean-guess baseline; and (ii) an adapted TICI-style spatiotemporal baseline inspired by Nielsen et al. [16]. This baseline adopts the high-level DeepTICI-inspired architecture, consisting of dual-branch frame-wise feature extraction followed by recurrent temporal modeling. Specifically, it uses 1×1 convolutional fusion and a unidirectional GRU (Uni-GRU), but is adapted without DeepTICI’s frame-wise pseudo-perfusion supervision or TICI-specific loss formulation. All models were evaluated via 5-fold cross-validation, using four folds for training/validation and one fold for independent testing.

Metrics. Multiclass performance is evaluated using mean absolute error (MAE) and accuracy within a ± 1 grade tolerance ($\text{ACC}_{\pm 1}$) [16], reflecting clinical acceptability of minor discrepancies. To account for the ordinal nature of the ASITN/SIR scale, we report quadratic weighted kappa (QWK) with weights $w_{ij} = (i - j)^2 / (N - 1)^2$. For the binary task (*good* vs. *poor*), we report macro-averaged F1-score (mF1) and Cohen’s kappa (Kappa). Both QWK and Kappa

are calculated as $1 - \sum w_{ij}O_{ij} / \sum w_{ij}E_{ij}$, where O_{ij} and E_{ij} are the observed and expected agreements, respectively.

Implementation. We utilized a DINOv2 ViT-B/14 backbone as the spatial feature extractor of our model, initialized with pretrained weights and kept frozen during training. Temporal dependencies were modeled using a single-layer BiGRU with 256 hidden units. Data augmentation included horizontal flipping, random shifts, scaling, rotation, and contrast perturbations. Optimization was performed using Adam with a batch size of 1. The model was trained using a hybrid loss combining cross-entropy with an auxiliary MAE term on the predicted class probabilities, $\mathcal{L} = (1 - \alpha)\mathcal{L}_{\text{CE}} + \alpha\mathcal{L}_{\text{MAE}}$, with $\alpha = 0.3$ selected empirically on the validation set. Models were trained for a maximum of 1000 epochs with early stopping on an NVIDIA Tesla V100 SXM2 (32 GB) GPU.

3 Results

Comparative results across all evaluated metrics are summarized in Fig. 2. In the multiclass task (Fig. 2a–c), X-LMC achieves the highest agreement with reference labels across all models (QWK = 0.398), with a higher point estimate than the adapted TICI-style baseline (QWK = 0.322), while also yielding higher $\text{ACC}_{\pm 1}$ (84.33% vs. 76.12%) and lower MAE (0.746 vs. 0.866). Observed inter-rater agreement in this cohort was comparably modest (QWK = 0.314), placing both the model and the human raters within the same fair-to-moderate agreement range; the model’s point estimate is thus better interpreted as comparable to, rather than decisively exceeding, a single pair of expert raters, particularly given the overlapping confidence intervals at this cohort size (Fig. 2a). In the binary task (Fig. 2d–e), X-LMC achieves the highest Kappa (0.430) and mF1 = 0.711 among evaluated methods, followed by the TICI-style baseline (Kappa=0.327, mF1 = 0.663). Furthermore, all optimized deep learning models substantially outperform the mean-guess baseline across both ordinal and dichotomized settings. This performance gap demonstrates that the networks successfully capture task-specific hemodynamic representations rather than merely exploiting the structural target statistics of the cohort.

To qualitatively evaluate model behavior, frame-averaged Grad-CAM [19] activations were computed (Fig. 3). Saliency maps for good collateral cases (Fig. 3a–b) were frequently localized to distal cortical territories, consistent with the anatomical regions utilized during clinical ASITN/SIR assessment, while poor collateral cases (Fig. 3c–d) showed relatively more central vascular activation, reflecting the absence of distal filling. These activation profiles indicate that the framework attends to physiologically relevant perfusion cues rather than relying on uninformative image regions.

Ablations. We conducted ablations following an 80/20 train/test split scheme (Table 1). We report QWK and Kappa as the primary agreement metrics for

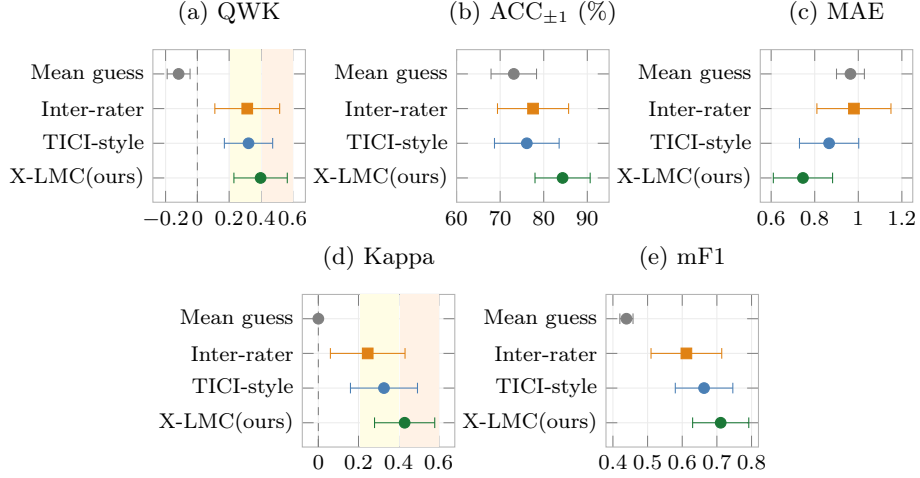


Fig. 2. Multiclass (top) and binary (bottom) performance. Shaded κ bands show fair [0.21 - 0.40] / moderate [0.41 - 0.60] agreement. Whiskers show 95% confidence intervals.

the ordinal and dichotomized tasks, respectively. Among the static spatial encoders, DINOv2-ViT-B achieved the best ordinal agreement ($\text{QWK} = 0.364$), outperforming EfficientNet-B0 [24] and EfficientNetV2-S [25] despite the EfficientNet variants being fine-tuned rather than frozen, supporting the use of transformer-based representations for capturing fine-grained collateral features. Adding temporal modeling improved agreement across ablation variants. When keeping the fusion strategy fixed, replacing Uni-GRU with Bi-GRU increased QWK from 0.397 to 0.478 and Kappa from 0.250 to 0.429. When the temporal module was fixed to Bi-GRU, replacing concatenation with X-Attn yielded smaller but consistent relative gains in QWK and Kappa (8.0% and 19.3%, respectively), suggesting that cross-view attention provides an additional benefit beyond bidirectional temporal modeling alone.

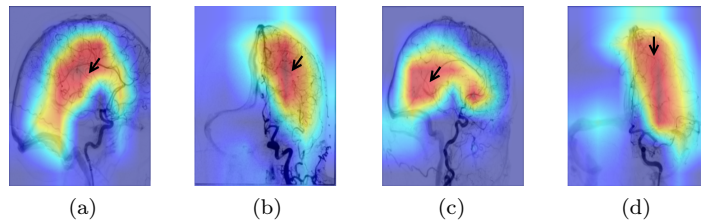


Fig. 3. Heatmap activations for good (a,b) and poor (c,d) ASITN/SIR scores. Arrows point to collaterals.

Table 1. Ablation analysis in a 80/20 train/test scheme.

Ablation	Spatial Encoder	GRU	Fusion	QWK $\uparrow$	Kappa $\uparrow$
Spatial Encoder	EfficientNet-B0	–	–	0.332	0.462
	EfficientNetV2-S	–	–	0.341	0.318
	DINOv2-ViT-B	–	–	0.364	0.456
Temporal & Fusion	DINOv2-ViT-B	Uni	Concat	0.397	0.250
	DINOv2-ViT-B	Bi	Concat	0.478	0.429
	DINOv2-ViT-B	Bi	X-Attn	0.516	0.512

4 Discussion & Conclusion

This work presents, to our knowledge, the first dedicated framework for automated ASITN/SIR collateral scoring from biplane DSA, establishing a foundation for exploring the feasibility and challenges of temporal deep learning for this task. Compared with spatiotemporal baselines adapted from related angiographic scoring tasks, X-LMC achieved higher performance across the evaluated metrics. The frozen DINOv2 backbone provided a stable, data-efficient representation for our limited cohort, while bidirectional X-Attn enabled synchronized AP and Lat projections to be jointly exploited across the angiographic sequence. The model’s agreement was comparable to inter-rater agreement, suggesting that performance is partly bounded by label ambiguity rather than architecture alone. Discrimination of adjacent ASITN/SIR grades, particularly the 2 vs. 3 boundary, proved most challenging, reflecting subtle differences in flow completeness and perfusion delay that are difficult to operationalize even for expert raters. The subjectivity inherent in the ASITN/SIR scale therefore represents a shared ceiling for both human and automated assessment, and adjudicated consensus annotation would be a meaningful step toward more reliable evaluation. This work should be read as an opening contribution mapping the feasibility and current boundaries of the task, rather than a mature clinical tool.

The clinical significance of automating this specific task extends beyond thrombectomy decision support. Because DSA provides direct, high-fidelity visualization of collaterals, DSA-derived scores offer a more reliable ground-truth reference [12]. Automating this assessment at scale therefore unlocks retrospective cohort analysis that was previously infeasible due to the manual grading burden, enabling systematic association of collateral status with functional outcomes and imaging biomarkers, and paving the way toward cross-modal label propagation to pre-interventional modalities (e.g. CTA/CTP). Reliable collateral estimates also carry direct value for secondary treatment planning, neurorehabilitation scheduling, and hospital resource allocation, clinical insights that remain highly desired even after thrombectomy [2, 3, 10].

To conclude, X-LMC demonstrates that automated architectures can reliably capture specialized microvascular kinetics. This work bridges the gap between raw biplane angiographic sequences and standardized collateralization indices,

establishing a promising and reproducible foundation for automated hemodynamic profiling in stroke research.

Limitations. Supervised training relied on a single expert reference label rather than adjudicated consensus. The cohort was further limited to occlusions of the M1 segment, reducing generalizability to other vascular territories. Future work should therefore prioritize consensus-labeled, multi-territory cohorts and external validation.

Acknowledgments. EdIR and BM are supported by the Helmut Horten Foundation.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Baazaoui, H., Engelter, S.T., Gensicke, H., Enz, L.S., Psychogios, M., Mutke, M., Michel, P., Strambo, D., Salerno, A., Marquering, H.A., et al.: The multicentre acute ischemic stroke imaging and clinical data (MAGIC) repository: rationale and blueprint. *Frontiers in neuroinformatics* **18**, 1508161 (2025)
2. Binder, N.F., El Amki, M., Glück, C., Middleham, W., Reuss, A.M., Bertolo, A., Thurner, P., Deffieux, T., Lambride, C., Epp, R., et al.: Leptomeningeal collaterals regulate reperfusion in ischemic stroke and rescue the brain from futile recanalization. *Neuron* **112**(9), 1456–1472 (2024)
3. Campbell, B.C., Christensen, S., Tress, B.M., Churilov, L., Desmond, P.M., Parsons, M.W., Barber, P.A., Levi, C.R., Bladin, C., Donnan, G.A., et al.: Failure of collateral blood flow is associated with infarct growth in ischemic stroke. *Journal of Cerebral Blood Flow & Metabolism* **33**(8), 1168–1172 (2013). <https://doi.org/10.1038/jcbfm.2013.77>
4. Cao, J., Baazaoui, H., Prabhakar, C., Shit, S., Otto, L.B., Wegener, S., Menze, B., de la Rosa, E.: Leptomeningeal collateral detection on dsa via vessel-graph neural networks. *arXiv preprint arXiv:2606.14828* (2026)
5. Cho, K., Van Merriënboer, B., Gulçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder–decoder for statistical machine translation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. pp. 1724–1734 (2014)
6. Collaborators, G.S., et al.: Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet. Neurology* **20**(10), 795 (2021)
7. Hassen, W.B., Malley, C., Boulouis, G., Clarençon, F., Bartolini, B., Bourcier, R., Régent, C.R., Bricout, N., Labeyrie, M.A., Gentric, J.C., et al.: Inter-and intraobserver reliability for angiographic leptomeningeal collateral flow assessment by the american society of interventional and therapeutic neuroradiology/society of interventional radiology (ASITN/SIR) scale. *Journal of neurointerventional surgery* **11**(4), 338–341 (2019)
8. Higashida, R.T., Furlan, A.J.: Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. *stroke* **34**(8), e109–e137 (2003)
9. Liebeskind, D.S.: Collateral circulation. *Stroke* **34**(9), 2279–2284 (2003)

10. Liebeskind, D.S., Luff, M.K., Bracard, S., Guillemin, F., Jahan, R., Jovin, T.G., Majoie, C.B., Mitchell, P.J., van der Lugt, A., Menon, B.K., et al.: Collaterals at angiography guide clinical outcomes after endovascular stroke therapy in hermes. *Journal of NeuroInterventional Surgery* **17**(8), 811–816 (2025)
11. Lin, L., Chen, C., Tian, H., Bivard, A., Spratt, N., Levi, C.R., Parsons, M.W.: Perfusion computed tomography accurately quantifies collateral flow after acute ischemic stroke. *Stroke* **51**(3), 1006–1009 (2020)
12. Liu, L., Ding, J., Leng, X., Pu, Y., Huang, L.A., Xu, A., Wong, K.S.L., Wang, X., Wang, Y., Cai, J.m., et al.: Guidelines for evaluation and management of cerebral collateral circulation in ischaemic stroke 2017. *Stroke and vascular neurology* **3**(3) (2018)
13. Liu, W., Tian, T., Wang, L., Xu, W., Li, L., Li, H., Zhao, W., Tian, S., Pan, X., Deng, Y., et al.: DIAS: A dataset and benchmark for intracranial artery segmentation in dsa sequences. *Medical Image Analysis* **97**, 103247 (2024)
14. Menon, B., Smith, E., Modi, J., Patel, S., Bhatia, R., Watson, T., Hill, M., Demchuk, A., Goyal, M.: Regional leptomeningeal score on CT angiography predicts clinical and imaging outcomes in patients with acute anterior circulation occlusions. *American journal of neuroradiology* **32**(9), 1640–1645 (2011)
15. Mittmann, B.J., Braun, M., Runck, F., Schmitz, B., Tran, T.N., Yamlahi, A., Maier-Hein, L., Franz, A.M.: Deep learning-based classification of DSA image sequences of patients with acute ischemic stroke. *International journal of computer assisted radiology and surgery* **17**(9), 1633–1641 (2022)
16. Nielsen, M., Waldmann, M., Frölich, A.M., Flottmann, F., Hristova, E., Bendszus, M., Seker, F., Fiehler, J., Sentker, T., Werner, R.: Deep learning-based automated thrombolysis in cerebral infarction scoring: a timely proof-of-principle study. *Stroke* **52**(11), 3497–3504 (2021)
17. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023)
18. Potreck, A., Scheidecker, E., Weyland, C., Neuberger, U., Herweh, C., Möhlenbruch, M., Chen, M., Nagel, S., Bendszus, M., Seker, F.: RAPID CT perfusion-based relative CBF identifies good collateral status better than hypoperfusion intensity ratio, CBV-index, and time-to-maximum in anterior circulation stroke. *American Journal of Neuroradiology* **43**(7), 960–965 (2022)
19. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. *International journal of computer vision* **128**(2), 336–359 (2020)
20. Shi, F., Zeng, Q., Gong, X., Zhong, W., Chen, Z., Yan, S., Lou, M.: Quantitative collateral assessment on CTP in the prediction of stroke etiology. *American Journal of Neuroradiology* **43**(7), 966–971 (2022)
21. Su, J., Wolff, L., van Es, A.C.M., Van Zwam, W., Majoie, C., Dippel, D.W., Van Der Lugt, A., Niessen, W.J., Van Walsum, T.: Automatic collateral scoring from 3D CTA images. *IEEE transactions on medical imaging* **39**(6), 2190–2200 (2020)
22. Su, R., Cornelissen, S.A., Van Der Sluijs, M., Van Es, A.C., Van Zwam, W.H., Dippel, D.W., Lycklama, G., Van Doormaal, P.J., Niessen, W.J., Van Der Lugt, A., et al.: 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)
23. Su, R., van der Sluijs, P.M., Chen, Y., Cornelissen, S., van den Broek, R., van Zwam, W.H., van der Lugt, A., Niessen, W.J., Ruijters, D., van Walsum, T.: CAVE:

- Cerebral artery–vein segmentation in digital subtraction angiography. *Computerized Medical Imaging and Graphics* **115**, 102392 (2024)
24. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International conference on machine learning*. pp. 6105–6114. PMLR (2019)
 25. Tan, M., Le, Q.: Efficientnetv2: Smaller models and faster training. *arxiv* 2021. *arXiv preprint arXiv:2104.00298* **5** (2021)
 26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
 27. Xu, W., Tan, T., Yang, H., Liu, W., Chen, Y., Zhang, L., Pan, X., Gao, F., Deng, Y., van Walsum, T., et al.: Cvfsnet: A cross view fusion scoring network for end-to-end mtici scoring. *Medical Image Analysis* **102**, 103508 (2025)