A GRAPH-DRIVEN HYBRID CNN-GNN APPROACH FOR VESSEL CLASSIFICATION IN ADAPTIVE OPTICS IMAGES

Abir Aissa[1], *Guillaume Lachaud*[2], *Florence Rossant*[1], *Hélène Urien*[1], *Michel Paques*[3]

[1]{abir.aissa, florence.rossant, helene.urien}@isep.fr, ISEP, Issy-les-Moulineaux, France
[2]guillaume.lachaud@polytechnique.edu, École polytechnique, Palaiseau, France
[3]mpaques@15-20.fr, CIC 1423, Paris, France

ABSTRACT

Adaptive Optics Ophthalmoscopy (AOO) enables high-resolution retinal imaging for clinical biomarker analysis. While existing methods automatically segment retinal bifurcations and extract vascular biomarkers, they require manual selection of high-quality arterial regions, limiting scalability and efficiency. We propose a fully automated patch-based CNN-GNN framework for vessel classification in AOO images. Our method extracts deep features from image patches, constructs graphs with cost-weighted edges capturing vessel connectivity, then applies Graph Attention Networks (GATs) for classification. This hybrid approach integrates local features with global topological information through graph processing, eliminating manual region selection requirements. We evaluate our pipeline on a dataset of 4,258 vessel patches across four CNN/Transformer backbones (ResNet-18, EfficientNet-B2, DenseNet-121, TinyViT). The backbone+GAT combination consistently outperforms backbone-only baselines, achieving up to 86.5% accuracy with improvements ranging from 1.9% to 4.4% across architectures. Our approach provides a robust, scalable foundation for automated vascular biomarker analysis in AOO imaging, supporting more efficient clinical workflows.

Index Terms— Arteriovenous classification, Adaptive Optics Ophthalmoscopy, Deep learning, Graph Neural Network

1. INTRODUCTION

Adaptive Optics Ophthalmoscopy (AOO) is a non-invasive imaging technique that produces high-resolution images of the retinal microvasculature. As illustrated in Fig. 1, it enables direct visualization of fine structures such as vessel lumen and vascular walls. Hence, AOO images have been used to quantify clinically relevant biomarkers, most notably the wall-to-lumen ratio (WLR). These measures, computed mainly from the arterial tree, i.e., the network of arteries and bifurcations, capture vascular morphology and have proven essential in the study and characterization of pathologies such as diabetic retinopathy (DR), vascular ageing, and hypertension [1, 2, 3]. AOO images display a variety of vascular configurations, encompassing arteries, arterial bifurcations (composed of a mother branch and two daughter branches), veins, venous confluences (vein junctions), and arteriovenous crossings. While veins are generally darker with thinner, less discernible walls, excepting anomalous cases, arteries exhibit more visible vessel walls. Nevertheless, variability in image quality often renders vessel differentiation ambiguous.

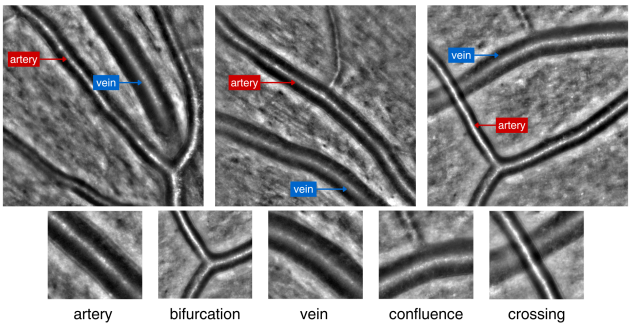

Fig. 1. Examples of vessel configurations in AOO images.

Prior work [4, 5, 6], has addressed automatic segmentation of AOO images, enabling the calculation of vessel diameter measurements used in biomarker computation. However, clinicians must still manually delineate arterial segments and bifurcations within high-quality regions (sharp image areas, with good contrast, and sufficient visibility of vessel walls), a process that is both time-consuming and resource-intensive. In our previous work, we automated the complementary task of quality assessment using a patch-based deep learning approach [7], enabling the reliable identification of such high-quality regions in AOO images. Nevertheless, vessel type identification remains manual, limiting the automation of biomarker computation.

In this context, we propose an automated vessel classification framework that leverages deep neural networks (CNNs and Vision Transformers) for local feature extraction and graph neural networks (GNNs) to incorporate global

topological context. We adopt a patch-based approach that overcomes pixel-wise labeling limitations: patches are easier to annotate, computationally efficient, and naturally serve as graph nodes for backbone-GNN integration while preserving both local and global vessel structure.

This method supports automatic identification of arterial segments and bifurcations, streamlining biomarker extraction workflows.

Our contributions are threefold: (i) a lightweight method for constructing graph edges that encodes contextual relationships between patches; (ii) a hybrid backbone-GNN pipeline for patch-wise arteriovenous classification; (iii) ablation studies demonstrating superior performance over backbone-only baselines across multiple architectures.

2. RELATED WORKS

Prior AOO-based studies have largely focused on automating biomarker extraction through vessel segmentation and morphometric analysis [5, 6, 8]. In contrast, automated vessel classification has been more extensively studied in other imaging modalities, most notably in fundus photography, OCT and OCTA [9], while comparable approaches for AOO images remain limited.

In recent years, deep learning and graph-based methods have emerged as the dominant paradigms for artery/vein classification [9]. Many of these approaches integrate vessel segmentation and classification into a unified framework, often leveraging U-Net-based architectures [10, 11, 12].

More recently, GNNs have been integrated with CNNs to incorporate vessel topology information. Mishra et al. [13] extract vessel features using a CNN, construct a graph from the segmentation mask, and apply a GCN to classify vessel segments, fusing outputs from both networks. Similarly, Xu et al. [14] use a CNN for initial segmentation and a GNN to refine the classification by enforcing structural coherence.

In contrast, our method adopts a patch-wise approach, extracting features and constructing graph nodes from image patches. This design avoids exhaustive pixel-level labeling while enabling robust, context-aware vessel identification.

3. METHODOLOGY

3.1. Patch-based Dataset

Our dataset is composed of AOO eye fundus images acquired at our partner Hospital, using the Rtx1 Adaptive Optics Retinal camera [15]. It includes 253 grayscale images with a resolution of 1300×1300 pixels, obtained from 45 healthy subjects, 27 patients with diabetes, and 33 patients with CADASIL[1] syndrome, a hereditary pathology that affects small vessels and causes strokes. The dataset includes various vessel configurations and calibres. For each image, the binary vessel mask is also provided, generated semi-automatically under the supervision of clinical experts using the AOV software tool in supervised mode described in [8].

[1]Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy

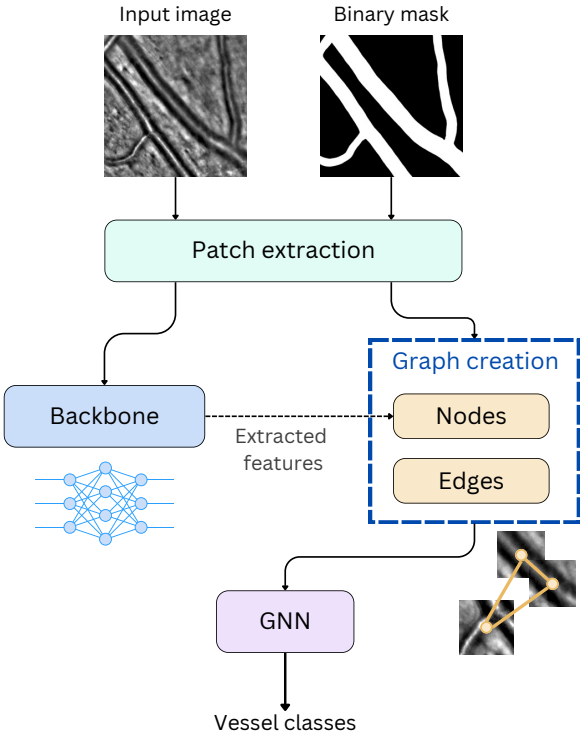

Fig. 2. Overview of the class prediction model

We extract overlapping image patches of size 256×256 pixels with stride 84 pixels across the full image. We retain only patches overlapping the morphological skeleton of the segmentation mask. A patch qualifies if the centroid of its largest vessel component lies within 10% of the patch center. This filtering ensures vessel-centered patches for classification.

Each patch belongs to one of five vessel categories: artery, vein, bifurcation, confluence, or crossing. Three annotators with varying expertise independently label patches using a custom annotation tool, considering both vessel wall visibility and surrounding context. Distinguishing arteries from veins and bifurcations from confluences proves particularly challenging, as wall thickness alone can be misleading. These labeled patches serve as input for graph construction.

3.2. Graph Construction

Each image is converted to a graph, where the nodes are the patches and the edges represent the vessel connectivity. Each node has a multidimensional feature vector that is extracted from the backbone network. Two nodes are connected by an

edge if the line between the two centroids overlaps with the vessel binary mask. Let p_1 and p_2 be the centroids of two patches. We define the cost $c(p_1, p_2)$ as

$$c(p_1, p_2) = \frac{1}{len([p_1, p_2])} \sum_{p \in [p_1, p_2]} \mathbb{I}(p \in \text{Vessel mask}), \quad (1)$$

where $len([p_1, p_2])$ is the number of pixels in the segment, and $\mathbb{I}(x)$ equals 1 if x is true, 0 otherwise. In other words, the cost function measures the normalized fraction of pixels in the segment $[p_1, p_2]$ that lies within the vessel mask. The set of edges E comprises all nodes pairs whose cost is greater than a threshold τ, i.e., $E = \{(i, j) \mid c(p_i, p_j) \geq \tau\}$. In practice, the pixels in the segment $[p_1, p_2]$ can be computed in linear time using Bresenham's algorithm (or `skimage.draw.line` from scikit-image). Intuitively, the threshold τ measures the selectivity of our edge filtering: the higher the threshold, the fewer edges will be created. Figure 3 shows an example of the extracted graph, using different threshold levels.

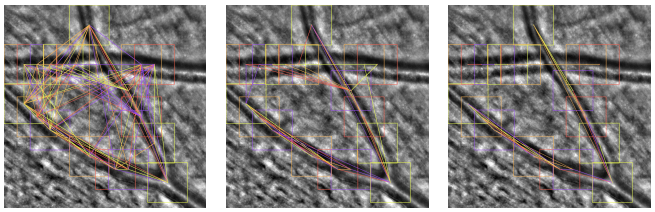

Fig. 3. Extracted graph structure for $\tau \in 0.3, 0.5, 0.7$ (from left to right)

3.3. Vessel classification

The classification pipeline decouples feature extraction from graph processing, enabling flexible combinations of backbone and GNN architectures. This modular design allows the approach to be adapted across different network types and medical imaging tasks. An overview of the full pipeline is shown in Figure 2.

The feature extractor processes individual patches to produce multidimensional feature vectors. The graph processor combines these vectors with the patch connectivity to generate the final classification, leveraging both local patch information and global vessel topology. In our implementation, patch features are extracted from intermediate backbone layers, and the graph is fed to a GNN.

The GNN processing adds minimal computational overhead to the backbone-only baseline. GNN complexity depends on the number of edges and feature dimensions, both of which remain small in our application. For example, with 256×256 patches and 1,024-dimensional feature vectors, the GNN operates on feature spaces that are $64\times$ smaller per patch than the pixel spaces processed by the backbone. Additionally, vessel connectivity creates naturally sparse graphs rather than fully-connected structures, further reducing GNN processing requirements.

4. EXPERIMENTS

4.1. Experimental setup

We evaluate our pipeline on a dataset of 4,258 vessel patches using 5-fold cross-validation with a 10% held-out test set. Patient-level splitting prevents data leakage, and class distributions remain balanced across splits.

We test four backbone architectures—three CNNs (ResNet-18, EfficientNet-B2, DenseNet-121) and one Vision Transformer (TinyViT) [16, 17, 18, 19]—as feature extractors, and two GNN types (GCN[20], GAT [21]) as graph processors. All backbones use ImageNet pretraining with standard data augmentation (flips, random rotations within the ±15° range). Training employs AdamW optimization (learning rate of 1e-3, with cosine scheduling) for 100 epochs. We report mean and standard deviation of test set performance across the five folds using accuracy, precision, recall, and F1-score.

4.2. Pipeline effectiveness

Table 1. Pipeline effectiveness: backbone-only baselines vs backbone+GAT across four architectures.

Model	Accuracy	Precision	Recall	F1
DenseNet-121	81.8±1.3	81.4±0.9	76.2±1.6	78.1±1.0
+ GAT	**85.4**±1.3	**86.5**±3.1	**79.7**±2.6	**82.2**±2.7
EfficientNet-B2	83.8±1.2	84.3±2.9	77.1±1.6	80.0±1.6
+ GAT	**85.7**±1.4	**87.5**±2.0	**79.1**±0.4	**82.3**±0.7
ResNet-18	80.9±1.5	81.1±1.6	74.3±1.4	76.9±1.0
+ GAT	**83.5**±1.9	**83.7**±1.4	**78.1**±3.2	**80.0**±2.3
TinyViT	82.1±2.1	83.4±1.7	76.9±2.4	79.3±2.0
+ GAT	**86.5**±0.9	**86.3**±0.9	**80.4**±1.1	**82.7**±0.7

We evaluate our backbone+GAT pipeline across four diverse architectures using threshold $\tau = 0.5$. Table 1 demonstrates that our backbone+GAT pipeline consistently outperforms backbone-only baselines across all four architectures. GAT integration provides substantial improvements: +3.6% accuracy for DenseNet-121, +1.9% for EfficientNet-B2, +2.6% for ResNet-18, and +4.4% for TinyViT. These consistent gains across diverse architectural families, from traditional CNNs to modern transformers, validate the broad applicability of our graph-based approach.

The improvements are particularly notable for precision and recall, indicating that GAT effectively leverages vessel connectivity information to reduce both false positives and false negatives. TinyViT+GAT achieves the highest overall performance (86.5% accuracy, 82.7% F1), while even the

lowest-performing combination (ResNet-18+GAT) surpasses the best backbone-only baseline. This architecture-agnostic effectiveness demonstrates that our pipeline's value stems from the graph construction methodology rather than specific backbone-GNN pairings.

4.3. Ablation study

We explore the impact of GNN architecture and connectivity threshold on our pipeline performance. We compare GCN and GAT performance across three threshold values (0.3, 0.5, 0.7) using EfficientNet-B2 as the backbone, selected for its highest baseline performance among backbone-only models. Both GNNs use 3 layers and are trained on backbone-extracted features with identical cross-validation splits.

Table 2 shows that GAT consistently outperforms GCN across all thresholds and metrics. Despite the noisy graph construction from our fast connectivity approach, GAT's attention mechanism effectively leverages topological information while GCN struggles with the sparse, imperfect connectivity. GAT demonstrates robustness to threshold variation, maintaining stable performance across different sparsity levels, with $\tau = 0.5$ providing the best balance of connectivity and performance.

Table 2. Ablation study comparing GCN and GAT performance across connectivity thresholds using EfficientNet-B2 features.

Model	τ	Accuracy	Precision	Recall	F1
GCN	0.3	50.6±0.9	27.6±0.7	29.9±1.1	28.0±1.0
GCN	0.5	59.1±1.6	46.8±6.9	37.4±3.2	37.1±4.1
GCN	0.7	66.1±1.1	50.2±3.3	42.3±2.3	41.9±2.9
GAT	0.3	84.0±1.0	87.0±1.4	74.9±1.6	78.3±1.8
GAT	0.5	**86.0±0.9**	**88.2±1.2**	**77.1±3.2**	**80.5±2.8**
GAT	0.7	84.8±1.7	86.2±1.0	74.6±4.5	77.1±5.1

5. CONCLUSION

In this article, we introduced a fully automated, patch-based framework for vessel classification in AOO images, integrating deep networks (CNNs and Vision Transformers) for local visual features with GNNs for global structural information. Our framework consistently outperforms backbone-only baselines across all architectures, achieving improvements of 1.9% to 4.4% in accuracy with minimal computational overhead. The ablation study demonstrates that GAT's attention mechanism effectively leverages vessel connectivity information, even with our simple graph construction approach.

By enabling patch-level modeling of vascular topology, this work provides a scalable foundation for automated arteriovenous classification and lays the groundwork for future fully automated reconstruction of the vascular tree in wide-field AOO images. The architecture-agnostic effectiveness of our approach suggests broad applicability to other medical imaging domains requiring spatial relationship modeling.

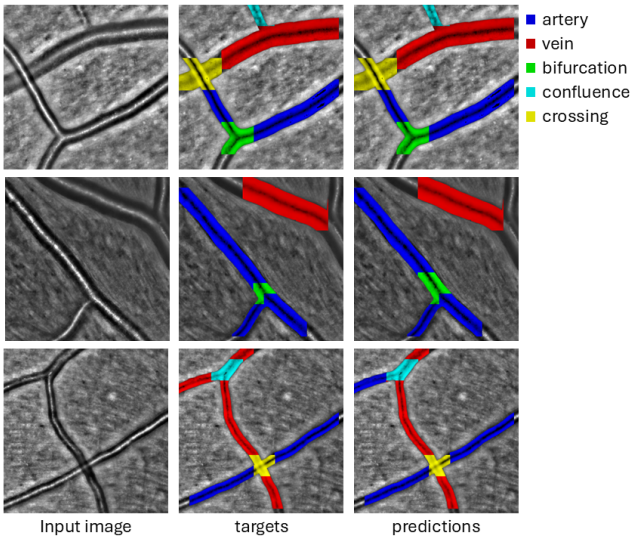

Fig. 4. Vessel classification results from the EfficientNet+GAT model

6. COMPLIANCE WITH ETHICAL STANDARDS

The study was an observational study and was not declared as a trial on clinicaltrials.gov. The study was approved by an independent ethics committee (updated agreement CEEI-IRB-17/388) from INSERM–France and conducted in accordance with the Declaration of Helsinki and guidelines for Good Clinical Practice and General Data Protection Regulation (GDPR) in Europa.

7. REFERENCES

[1] Edouard Koch, David Rosenbaum, Aurélie Brolly, José-Alain Sahel, Philippe Chaumet-Riffaud, Xavier Girerd, Florence Rossant, and Michel Paques, "Morphometric analysis of small arteries in the human retina using adaptive optics imaging: Relationship with blood pressure and focal vascular changes," *Journal of hypertension*, vol. 32, pp. 890–898, 2014.

[2] Ting Luo, Thomas J. Gast, Tyler J. Vermeer, and Stephen A. Burns, "Retinal vascular branching in healthy and diabetic subjects," *Investigative Ophthalmology & Visual Science*, vol. 58, pp. 2685 – 2694, 2017.

[3] David Rosenbaum, Alessandro Mattina, Edouard Koch, Florence Rossant, Antonio Gallo, Nadjia Kachenoura,

Michel Paques, Alban Redheuil, and Xavier Girerd, "Effects of age, blood pressure and antihypertensive treatments on retinal arterioles remodeling assessed by adaptive optics," *Journal of Hypertension*, vol. 34, pp. 1115–1122, 2016.

[4] Iyed Trimeche, Florence Rossant, Isabelle Bloch, and Michel Paques, "Segmentation of retinal arterial bifurcations in 2d adaptive optics ophthalmoscopy images," *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1490–1494, 2019.

[5] Iyed Trimeche, Florence Rossant, Isabelle Bloch, and Michel Pâques, "Fully automatic cnn-based segmentation of retinal bifurcations in 2d adaptive optics ophthalmoscopy images," *2020 International Conference on Image Processing Theory (IPTA)*, pp. 1–6, 2020.

[6] Abir Aissa, Florence Rossant, Hélène Urien, and Michel Pâques, "Deep learning-based segmentation of retinal vessels in adaptive optics ophthalmoscopy images," *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, 2024.

[7] Abir Aissa, Florence Rossant, Hélène Urien, and Michel Pâques, "Deep-learning based quality assessment in adaptive optics ophthalmoscopy images," *2025 IEEE International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6, 2024.

[8] Florence Rossant, Isabelle Bloch, Iyed Trimeche, J.-B. de Regnault de Bellescize, Daniela Castro Farias, Valerie Krivosic, Hugues Chabriat, and Michel Pâques, "Characterization of retinal arteries by adaptive optics ophthalmoscopy image analysis," *IEEE Transactions on Biomedical Engineering*, vol. 71, pp. 3085–3097, 2024.

[9] Qihan Chen, Jianqing Peng, Shen Zhao, and Wanquan Liu, "Automatic artery/vein classification methods for retinal blood vessel: A review," *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, vol. 113, pp. 102355, 2024.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[11] Ruben Hemelings, Bart Elen, Ingeborg Stalmans, Karel van Keer, Patrick De Boever, and Matthew B. Blaschko, "Artery-vein segmentation in fundus images using a fully convolutional network," *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, vol. 76, pp. 101636, 2019.

[12] Jingfei Hu, Hua Wang, Zhaohui Cao, Guang Wu, Jost Bruno Jonas, Ya Xing Wang, and Jicong Zhang, "Automatic artery/vein classification using a vessel-constraint network for multicenter fundus images," *Frontiers in Cell and Developmental Biology*, vol. 9, 2021.

[13] Qingyu Chen, Suraj Mishra, Ya Xing Wang, Chuan Chuan Wei, Danny Ziyi Chen, and Sharon Hu, "VTG-Net: A CNN Based Vessel Topology Graph Network for Retinal Artery/Vein Classification," *Frontiers in Medicine*, vol. 8, 2021.

[14] Xiayu Xu, Peiwei Yang, Hong Wang, Zhanfeng Xiao, Gang Xing, Xiulan Zhang, Wei Wang, Feng Xu, Jiong Zhang, and Jianqin Lei, "Av-casNet: Fully automatic arteriole-venule segmentation and differentiation in OCT angiography," *IEEE Transactions on Medical Imaging*, vol. 42, pp. 481–492, 2022.

[15] Imagine Eyes, "Rtx1 adaptive optics retinal camera," .

[16] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015.

[17] Mingxing Tan and Quoc Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[18] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2016.

[19] Kan Wu, Jinnian Zhang, Houwen Peng, Mengchen Liu, Bin Xiao, Jianlong Fu, and Lu Yuan, "Tinyvit: Fast pretraining distillation for small vision transformers," in *European conference on computer vision*. Springer, 2022, pp. 68–85.

[20] Thomas N. Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. 2017, OpenReview.net.

[21] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio, "Graph attention networks," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. 2018, OpenReview.net.