

# Pulmonary Embolism Risk Stratification from CTPA and Medical Records: Vascular Graphs Are Not All You Need

Nathan Painchaud<sup>1</sup>(✉), Tristan Habémont<sup>1</sup>, Morgane des Ligneris<sup>1</sup>, Allan Serva<sup>2,3</sup>, Pierre Croisille<sup>1,4</sup>, Laurent Bertolotti<sup>2</sup>, Thomas Lampert<sup>5</sup>, Johannes F. Lutzeyer<sup>6</sup>, and Odysée Merveille<sup>1</sup>

<sup>1</sup> INSA-Lyon, Université Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Lyon, France

<sup>2</sup> Université Jean Monnet Saint-Étienne, CHU Saint-Étienne, Mines Saint-Étienne, INSERM, SAINBIOSE U1059, CIC 1408, Département de Médecine Vasculaire et Thérapeutique, F-CRIN INNOVTE network, all in F-42055, Saint-Étienne, France

<sup>3</sup> Department of Pneumology, CHU Saint-Etienne, UJM Saint-Etienne, France

<sup>4</sup> Department of Radiology, CHU Saint-Etienne, UJM Saint-Etienne, France

<sup>5</sup> ICube, University of Strasbourg, France

<sup>6</sup> LIX, CNRS, École Polytechnique, Institut Polytechnique de Paris, Palaiseau, France

`nathan.painchaud@insa-lyon.fr`

**Abstract.** Risk stratification for pulmonary embolism (PE) is critical for clinical decision-making. Stratification guidelines are based on patient medical records, parameters measured from computed tomography pulmonary angiography (CTPA), and blood tests. However, blood tests are often missing in routine practice. This work studies whether state-of-the-art models can accurately classify risk stratification from only medical records and biomarkers extracted from CTPA images. We benchmark different approaches to combine medical records and cardiac biomarkers with rich pulmonary vascular information; we add vascular biomarkers to tabular models and apply graph neural networks (GNNs) on the vascular tree’s intrinsic graph representation. We use a private dataset (n=353) with uniquely complete data for PE risk stratification. Our results show that, among global features, medical records and cardiac biomarkers are the most significant predictors, while vascular biomarkers do not further improve stratification. Even more surprising, even GNNs on vascular graphs fail to outperform strong tabular baseline on global features. We consider hypotheses, on both models and data, that could explain this suboptimal performance. Our investigation suggests that, counter-intuitively, vascular graphs may not help predict PE guidelines risk better than simpler clinical biomarkers. Code is available from <https://github.com/creatis-myriad/GENESIS>.

**Keywords:** multimodal fusion · tabular data · foundation model · graph neural networks · graph classification.

## 1 Introduction

Risk stratification for pulmonary embolism (PE) determines how patients are treated, e.g., whether they receive intensive care, thrombolytic therapy, or anticoagulants. Current stratification guidelines rely on a combination of medical records, visual evaluation of Computed Tomography Pulmonary Angiography (CTPA), and blood tests to measure troponin and NT-proBNP [19]. CTPA is the standard diagnostic modality and is almost universally available, but blood tests are frequently not performed for the initial diagnosis [23]. This gap between guidelines and clinical practice motivates research on maximising the diagnostic accuracy from CTPA and medical records only.

While CTPA is used to confirm PE diagnosis by the presence of thrombi, its current use for risk stratification is limited to measuring right ventricle (RV) heart dysfunction, rather than thrombus burden. Clinical scores were proposed to globally assess thrombus burden from CTPA, e.g., Qanadli [28] and Mastora [26] vascular obstruction scores, and were correlated to PE risk factors. However, they require to locate thrombi’s positions in the pulmonary arterial tree, a labour too time-intensive to perform manually in routine clinical settings.

Deep neural networks have been successfully applied to automate part of this process by detecting [30] and segmenting [8] PE in CTPA. They have also been used to segment the pulmonary arterial tree [6, 21]. Recently, an automated pipeline was introduced to segment both the arterial tree and thrombi, and to extract a feature-rich patient-specific graph representation of the arterial tree [20]. Still, since thrombi and arteries are thin and sparse in CTPA, the spatial priors of deep vision models make them ill-suited for patient-level PE risk stratification.

By contrast, the new patient-specific vascular graphs [20] and the rising popularity of graph neural networks (GNN) for medical applications [1] present the opportunity to use a framework intrinsically adapted to the application. However, global features, i.e., patient-level data and biomarkers, are important for PE risk stratification [19], and comparatively little research has explored multi-modal learning with GNNs, all applications considered [3, 32, 25]. Furthermore, no specific GNN model dominates across tasks and datasets [22], and even graph foundation models (GFMs) for healthcare currently target specific tasks, mostly knowledge discovery and brain networks [17].

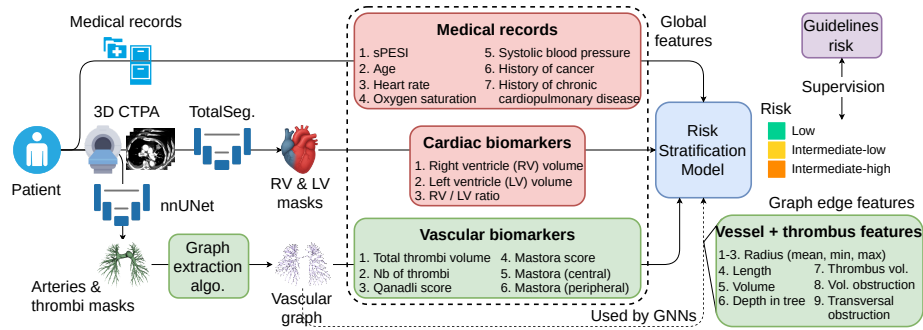
In this context, specifying an optimal PE risk stratification model is not trivial. Using a private dataset with uniquely complete PE risk stratification data, we propose the first study of the usefulness of pulmonary vascular graphs for PE risk stratification, structured into three main contributions:

1. We introduce a pipeline for automatic PE risk stratification from routine clinical data, i.e., structured medical records and CTPA images;
2. We benchmark state-of-the-art tabular models and GNNs to compare two approaches for adding vascular information: global biomarkers vs. full graphs. We test several backbones, notably two new proposed methods to combine global features in GNNs;

3. We empirically investigate hypotheses to explain counter-intuitive negative results, concluding that vascular graphs may be more relevant to predict patient outcome, or other pathologies, rather than PE guidelines risk.

## 2 Methods

We propose a pipeline, detailed in Fig. 1, to categorize PE risk in three classes by combining routine clinical data, i.e., structured medical records and CTPA images. Clinical guidelines were used to determine reference PE risk, using blood tests results not provided to our pipeline.



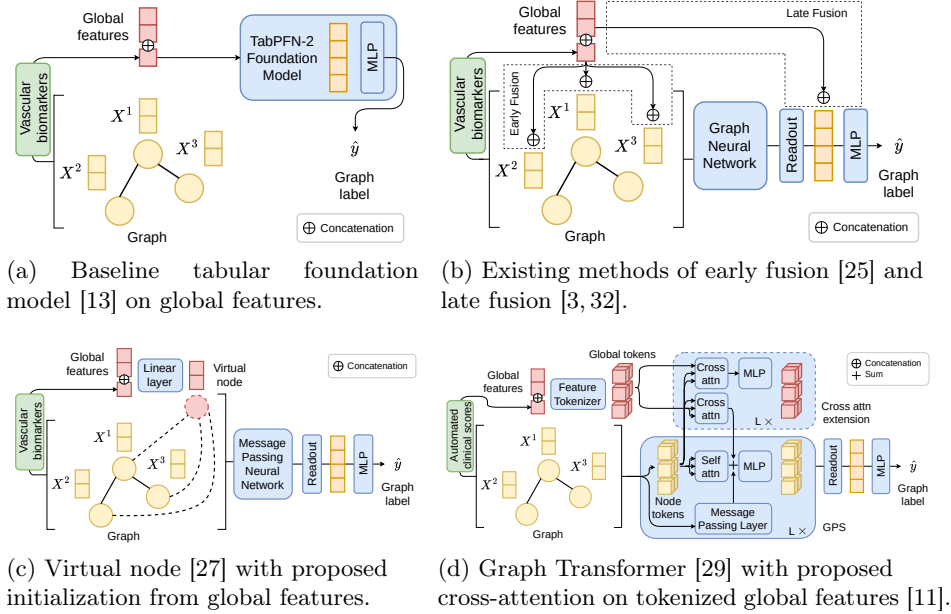
**Fig. 1.** Proposed pipeline for PE risk stratification, including CTPA preprocessing.

We first preprocess high-dimensional 3D CTPA by segmenting structures of interest for PE, i.e., ventricles and pulmonary vasculature. The latter is then converted into a graph. Finally, cardiac and vascular biomarkers are computed from the ventricle masks and vascular graph, respectively (see Section 2.1).

The stratification model then combines structured medical records and extracted image features. Since medical records and CTPA are complementary, we focus on exploring two fusion strategies, which tend to perform better in this context [12]. We test *tabular models* where image information is provided by concatenating cardiac and vascular biomarkers to medical records, as shown in Fig. 2a, motivated by similar applications having shown tabular models to be strong baselines [31]. We also test *GNNs*, extended to support global features (see Section 2.2), to make use of the rich vascular graph representation.

### 2.1 CTPA Images Preprocessing & Vascular Graph Representation

To segment the heart’s left and right ventricles (LV & RV), we use TotalSegmentator [33], a reference nnUNet [15] pretrained to segment important anatomical structures in CT. From these masks, we compute 3 cardiac biomarkers related to RV function (middle branch in Fig. 1).



**Fig. 2.** Schemas summarizing benchmarked methods for combining global features and vascular graphs for patient risk classification. Some layers (e.g., residual connections, norm, dropout) are not shown for readability.

To extract the pulmonary arterial graph, we use the pipeline developed in [20]. An mUNet was iteratively trained and its predictions semi-automatically checked to obtain anatomically and topologically consistent masks of arteries and thrombi (e.g., ensuring a single connected artery component, thrombi within arteries). The arteries masks were converted to graphs by representing junctions or terminal vessels as nodes and vessels as edges, with automatic quality checks and corrections to ensure topological consistency (e.g., removing artifact branches and cycles). Each vessel, i.e., edge in the graph, is described using 6 features computed from the masks (bottom right in Fig. 1). Thrombi are described by concatenating 3 thrombus features, i.e., thrombus volume and obstructions, equal to 0 in the absence of thrombi, to the 6 vessel features.

Finally, the vascular graph enables the automatic computation of 6 patient-level vascular biomarkers defined by clinicians to summarise thrombus burden (bottom centre in Fig. 1), previously too time-consuming to measure manually.

## 2.2 Graph Neural Networks with Global Features

Compared to the extensive research on multimodal fusion in computer vision, few works have studied how to combine global features with graphs. Previous works have proposed two fusion methods (see Fig. 2b), namely *early fusion* by concatenating global features on each node [25] and *late fusion* by concatenating

global features to the graph embedding of a GNN encoder [3, 32]. Considering computer vision literature has shown *intermediate fusion* typically performs better than early or late fusion when data modalities are structured differently [12], we propose two new intermediate fusion methods designed for GNNs.

**Virtual Node from Global Features.** Previous works have proposed to facilitate interactions between faraway nodes by adding virtual nodes (VN), i.e., nodes not originally in the graph, connected to all other nodes [27]. VNs have primarily been used as they tend to improve performance on graph-level tasks. Prior works also suggested to use VNs to integrate global features and store a global state [27], but did not test this in practice. Thus, we perform the first empirical evaluation of VNs initialised from global features, as shown in Fig. 2c.

**Cross-attention in Graph Transformers.** Graph Transformers (GT) have become a state-of-the-art alternative to message passing neural networks (MPNN), demonstrating high performance on graph-level tasks [29]. Additionally, Transformers’ ability to adapt to data with different structures and levels of details helped them succeed in multimodal applications [35]. These successes motivated us to propose to use cross-attention in GTs to fuse global features.

We use the state-of-the-art GPS [29] as the backbone. It augments the graph by adding a node’s positional encoding to its features. Node features are updated in parallel by an MPNN layer and self-attention between nodes treated as a sequence of tokens. On top of GPS, we propose to use a Feature Tokeniser [11] to map each global feature to a token, and to perform bidirectional cross-attention between global tokens and node tokens, as shown in Fig. 2d.

### 3 Experiments & Results

#### 3.1 Experimental Settings

**Dataset.** Data were collected from 353 PE patients, diagnosed by CTPA (see [20] for acquisition details) at the CHU Saint-Étienne emergency department between 04/2014 and 06/2020. The retrospective study was approved by the local ethics committee (IRBN262017/CHUSTE). Compared to other PE datasets, additional information to stratify risk using ESC guidelines [19] was collected: troponin & BNP measurements, sPESI score and its 6 factors [16]. Risk groups are: low (n=88), intermediate-low (n=138), and intermediate-high (n=127). To evaluate models, 10-fold cross-validation with an 80/10/10 split was used.

**Tabular Models.** Since tree-based models have historically outperformed deep learning [11] we compare XGBoost [5], a reference machine learning method, against TabPFN-2 [13], a state-of-the-art Transformed-based foundation model.

**Graph Neural Networks and Architecture Search.** All GNNs were implemented using PyTorch Geometric [10]. We conducted an exhaustive hyperparameter search across the options detailed in Table 1 using Optuna [2] with the Tree-structured Parzen Estimator (TPE) and a budget of 250 cross-validation trials for each GNN backbone. For each backbone, we chose the configuration with the best validation average across folds.

**Table 1.** Configuration space for automated hyperparameter tuning of GNNs.

Hyperparameter	Tested configurations	
	Message passing GNN (MPNN)	Graph Transformer (GT)
Backbone	GCN [18], GATv2 [4], GIN [34] with/without Virtual Node (VN)	GPS [29]
Attention heads	–	1, 2, 4
Graph representation (readout)	Global pooling: sum, mean, virtual node	Global pooling: sum, mean, [CLS] token
Layer depth	2...10	
Hidden channels	16, 32, 64	
Normalization	BatchNorm, InstanceNorm, GraphNorm	
Dropout	[0, 0.6]	
Prediction head layers	1...5	
L2 regularization	[ $10^{-5}$ , $10^{-3}$ ]	

Additionally, we ran an ablation study of the best GNN configurations’ key components, not detailed because of space constraints. Results showed that using the line graphs, i.e., inverting edges and vertices, improves performance, explained by the graphs’ features being exclusively on edges. They also confirmed that embedding global features using a Feature Tokenizer [11] (see Fig. 2d) outperforms both linear projection and TabPFN-2’s internal embedding.

### 3.2 Key Results

In this section, we present key results on the relevance of pulmonary vascular graphs for PE risk stratification. We first study whether clinicians’ vascular biomarkers improve tabular models. Then, we test if GNNs applied to the intrinsic vascular graph manage to use the richer data representation to improve upon the tabular baseline. Finally, we investigate whether data and models limitations explain the unexpected lack of improvement from vascular graphs.

Unless otherwise specified, scores in Tables 2 to 4 correspond to cross-validation test mean  $\pm$  s.d. Results in **bold** are not statistically different from the best ( $p > 0.05$  using a one-sided Wilcoxon signed-rank test).

**Can tabular models learn from vascular biomarkers?** Our first experiment studies the predictive power of global features. Cardiac biomarkers should be important, considering the link between RV dysfunction and PE risk [19]. Similarly, vascular biomarkers extracted from the graph should be predictive, given correlations between obstruction scores [26, 28] and PE risk factors.

We tested tabular models on combinations of global features, reporting the results in Table 2. Cardiac biomarkers are more predictive than vascular biomarkers (+6.6 on average for rows 2-3 and 4-5). Still, neither are informative on their own ( $\leq 52.7$ ), and they benefit from taking into account medical records (+35 on average for rows 4-2 and 5-3). Surprisingly, adding vascular biomarkers to cardiac

**Table 2.** Ablation study on the types of global features considered, using tabular models. Only the best model for each feature combination is presented.

Model	Global features			F1 (%) $\uparrow$
	Best backbone	Medical records	Cardiac biomarkers	
TabPFN-2	✓			75.2 $\pm$ 6.1
XGBoost			✓	52.7 $\pm$ 10.3
XGBoost				44.9 $\pm$ 8.5
TabPFN-2	✓		✓	<b>86.8 <math>\pm</math> 6.6</b>
TabPFN-2	✓			81.4 $\pm$ 6.3
TabPFN-2	✓		✓	<b>86.3 <math>\pm</math> 6.2</b>

**Table 3.** Benchmark of GNNs, including models integrating global features, compared to tabular baseline. Rows without *fusion method* indicate monomodal models.

Type	Model			Global features	Vascular graph	F1 (%) $\uparrow$	
	Fusion method	Best backbone	# params			Validation	Test
Tabular	–	TabPFN-2	7.2M	✓		–	<b>86.3 <math>\pm</math> 6.2</b>
MPNN	–	GCN+VN	22.3K		✓	65.5 $\pm$ 9.0	55.9 $\pm$ 8.3
	Early fusion	GCN+VN	4.3K	✓	✓	88.1 $\pm$ 4.4	<b>83.4 <math>\pm</math> 4.4</b>
	Late fusion	GIN+VN	89K	✓	✓	88.1 $\pm$ 4.6	83.3 $\pm$ 5.1
	Virtual node	GIN+VN	8.4K	✓	✓	87.6 $\pm$ 3.5	81.4 $\pm$ 7.6
GT	–	GPS	23K		✓	65.6 $\pm$ 7.2	50.2 $\pm$ 8.6
	Early fusion	GPS	35.3K	✓	✓	88.3 $\pm$ 4.6	81.3 $\pm$ 4.8
	Late fusion	GPS	97.4K	✓	✓	87.5 $\pm$ 5.3	80.8 $\pm$ 6.4
	Cross-attention	GPS	932K	✓	✓	89.2 $\pm$ 5.4	<b>83.8 <math>\pm</math> 4.9</b>

biomarkers and medical records does not significantly change results ( $-0.5$  for rows 6-4). This suggests that discriminative information for PE risk stratification may be lost when summarising vascular graphs to a few biomarkers.

**Can GNNs leverage rich vascular graphs?** To explore the hypothesis that vascular biomarkers discard important details from vascular graphs, we evaluated GNNs directly on the graphs, with and without global features. Table 3 benchmarks them against the best tabular model. GNNs without global features improve over tabular models on vascular biomarkers (+8 on average for rows 2 & 6 vs. row 3 in Table 2). Still, global features are necessary to reach competitive results. Compared to the impact of global features, differences between fusion methods across MPNN/GT are minor (test means all between 80.8 and 83.8). Still, the top two methods end up within statistical significance of TabPFN-2, even if they do not outperform it.

**Table 4.** Results for the regression of vascular biomarkers.

Model		Global features		Vascular graph	R <sup>2</sup> (%) $\uparrow$	
Type	Best backbone	Medical records	Cardiac biomarkers		Total Embolism Volume	Qanadli Score
Baseline Tabular	TabPFN-2	✓	✓		35.3 $\pm$ 20.3	32.7 $\pm$ 5.6
GNN	MPNN	GCN+VN		✓	95.7 $\pm$ 3.0	<b>83.1 <math>\pm</math> 4.2</b>
	GT	GPS		✓	<b>98.2 <math>\pm</math> 0.8</b>	<b>82.0 <math>\pm</math> 4.4</b>

**What explains suboptimal results with vascular graphs?** As both vascular biomarkers and learning directly on vascular graphs counter-intuitively fail to improve risk predictions, we explore three limitations (overfitting, failure to learn, and biased labels) that could explain suboptimal performance.

First of all, given our sample size ( $n=353$ ), GNNs may be overfitting, even if some are as small as 4.3K parameters. Table 3 does show a generalisation gap between validation and test of  $-5.8$  on average for multimodal GNNs. However, this is in line with generalisation gaps reported on much larger datasets: 0.319/0.387 (+21%) train/test MAE on ZINC with 12K graphs [9], and 84.79/77.07 ( $-7.72$ ) val/test AUROC on ogbg-molhiv with 41K+ graphs [14]. Thus, GNNs do not overfit our dataset more than datasets two orders of magnitude larger.

Alternatively, the tree structure of vascular graphs may be fundamentally challenging for GNNs. We tested this by regressing vascular biomarkers more directly linked to local graph features than PE risk. Table 4 compares TabPFN-2 to GNNs for this, where *Total Embolism Volume* and *Qanadli score* are the simplest and most complex biomarkers to regress, respectively. Improvements over the tabular baseline (+56 on average for rows 2 & 3 vs. 1) confirm that GNNs can learn patterns found only in vascular graphs and not in global features.

As for the labels, guideline risk groups might not exactly reflect real PE risk, and erroneous predictions from our models could be closer to real risk. This hypothesis is harder to test, as real PE risk in populations is typically quantified from 30-days mortality risk, too rare in our case for significant analysis (11 deaths for 353 patients). We therefore tried to predict another clinically meaningful endpoint, i.e., elevated troponin or BNP, but we observed the same ranking between methods. Alternatively, we looked at ambiguous cases, which we define as patients whose stratification changed between guidelines versions (2014 vs 2019), to check if they correlate with our models’ errors. A large overlap could suggest that our models align with the guidelines overall, but disagree on difficult cases. However, 0 out of 64 errors from our best GNN (GPS with cross-attention) match the 39 ambiguous patients, refuting the initial hypothesis.

After ruling out the explanations above, we conclude that vascular graphs are not more discriminative than medical records and cardiac biomarkers for predicting PE guidelines risk. Other endpoints, e.g., mortality or complications, may need to be studied on larger cohorts to reveal vascular graph potential.

## 4 Limitations & Conclusions

**Limitations.** The focus on one dataset may raise concerns about the cohort’s diversity. However, no other dataset provides similarly comprehensive data. Public datasets like PARSE (n=203) [24] and AirRC (n=254) [21] provide pulmonary vessel annotations but lack thrombi annotations, medical records and disease information. RSNA (n=12 195) [7] targets PE and covers more cases, but lacks segmentations and complete medical records. Compared to these, our dataset provides a unique combination of vessel and thrombi segmentations and medical records, at a larger scale (n=353) than pixel-level annotated public datasets.

**Conclusions.** This work presented the first study of the relevance of vascular graphs for pulmonary embolism risk stratification, enabled by an automatic CTPA processing pipeline. We extensively benchmarked methods to combine rich vascular graphs and medical records, from tabular models on global features to multimodal GNNs. The lack of improvements from vascular graphs lead us to investigate and rule out models and data limitations. Instead, our results refute clinical intuitions and suggest that vascular graphs do not contribute discriminative information for PE severity on top of simpler clinical biomarkers.

**Acknowledgments.** This work was supported by the French National Research Agency through the PERSEVERE (ANR-22-CE45-0018) and HistoGraph (ANR-23-CE45-0038) projects, and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007).

**Disclosure of Interests.** The authors have no relevant competing interests to declare.

## References

1. Ahmedt-Aristizabal, D., et al.: Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future. *Sensors* **21**(14), 4758 (Jan 2021). <https://doi.org/10.3390/s21144758>
2. Akiba, T., et al.: Optuna: A Next-generation Hyperparameter Optimization Framework. In: *Proc. KDD*. pp. 2623–2631 (2019). <https://doi.org/10.1145/3292500.3330701>
3. Brasoveanu, A.D., et al.: Extending Graph Neural Networks with Global Features. In: *Learning on Graphs (LoG)* (2023), <https://openreview.net/forum?id=aisVQy6R2k>
4. Brody, S., Alon, U., Yahav, E.: How Attentive are Graph Attention Networks? In: *Proc. ICLR* (2021), <https://openreview.net/forum?id=F72ximsx7C1>
5. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In: *Proc. KDD*. pp. 785–794 (2016). <https://doi.org/10.1145/2939672.2939785>
6. Chu, Y., et al.: Deep learning-driven pulmonary artery and vein segmentation reveals demography-associated vasculature anatomical differences. *Nat Commun* **16**(1), 2262 (Mar 2025). <https://doi.org/10.1038/s41467-025-56505-6>
7. Colak, Errol et al., F.: The RSNA Pulmonary Embolism CT Dataset. *Radiology: Artificial Intelligence* **3**(2), e200254 (Mar 2021). <https://doi.org/10.1148/ryai.2021200254>

8. Djahnine, A., et al.: Detection and severity quantification of pulmonary embolism with 3D CT data using an automated deep learning-based artificial solution. *Diagnostic and Interventional Imaging* **105**(3), 97–103 (Mar 2024). <https://doi.org/10.1016/j.diii.2023.09.006>
9. Dwivedi, V.P., et al.: Benchmarking Graph Neural Networks. *Journal of Machine Learning Research* **24**(43), 1–48 (2023)
10. Fey, M., et al.: PyG 2.0: Scalable Learning on Real World Graphs. In: *Proc. TGL Workshop @ KDD (2025)*, <https://openreview.net/forum?id=DHHLkQvWqs>
11. Gorishniy, Y., et al.: Revisiting Deep Learning Models for Tabular Data. In: *Proc. NeurIPS*. pp. 18932–18943 (2021)
12. Guarrasi, V., et al.: A systematic review of intermediate fusion in multimodal deep learning for biomedical applications. *Image Vis Comput* **158**, 105509 (May 2025). <https://doi.org/10.1016/j.imavis.2025.105509>
13. Hollmann, N., et al.: Accurate predictions on small data with a tabular foundation model. *Nature* **637**(8045), 319–326 (Jan 2025). <https://doi.org/10.1038/s41586-024-08328-6>
14. Hu, W., et al.: Open Graph Benchmark: Datasets for Machine Learning on Graphs. In: *Proc. NeurIPS*. pp. 22118–22133 (2020)
15. Isensee, F., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* **18**(2), 203–211 (Feb 2021). <https://doi.org/10.1038/s41592-020-01008-z>
16. Jiménez, D., RIETE Investigators, et al.: Simplification of the Pulmonary Embolism Severity Index for Prognostication in Patients With Acute Symptomatic Pulmonary Embolism. *Arch Intern Med* **170**(15), 1383–1389 (Aug 2010). <https://doi.org/10.1001/archinternmed.2010.199>
17. Khan, W., et al.: A Comprehensive Survey of Foundation Models in Medicine. *IEEE Rev Biomed Eng* pp. 1–20 (2025). <https://doi.org/10.1109/RBME.2025.3531360>
18. Kipf, T.N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks. In: *Proc. ICLR (2017)*, <https://openreview.net/forum?id=SJU4ayYg1>
19. Konstantinides, S.V., ESC Scientific Document Group, et al.: 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). *European Heart Journal* **41**(4), 543–603 (Jan 2020). <https://doi.org/10.1093/eurheartj/ehz405>
20. des Ligneris, M., et al.: A Patient-Specific Pulmonary Arterial Tree Digital Twin to Extract Pulmonary Embolism Biomarkers (May 2026), <https://arxiv.org/abs/2605.28217v1>, Under review.
21. Liu, J., et al.: A Custom Annotated Dataset for Segmentation of Pulmonary Veins, Arteries, and Airways. *Sci Data* **12**(1), 1806 (Nov 2025). <https://doi.org/10.1038/s41597-025-06074-6>
22. Liu, J., et al.: Graph Foundation Models: Concepts, Opportunities and Challenges. *IEEE Trans Pattern Anal Mach Intell* **47**(6), 5023–5044 (Jun 2025). <https://doi.org/10.1109/TPAMI.2025.3548729>
23. Liu, X., et al.: Joint analysis of D-dimer, N-terminal pro b-type natriuretic peptide, and cardiac troponin I on predicting acute pulmonary embolism relapse and mortality. *Sci Rep* **11**(1), 14909 (Jul 2021). <https://doi.org/10.1038/s41598-021-94346-7>
24. Luo, G., et al.: Efficient automatic segmentation for multi-level pulmonary arteries: The PARSE challenge (Aug 2024). <https://doi.org/10.48550/arXiv.2304.03708>

25. Lutchyn, T., Mardal, M., Ricaud, B.: Efficient Learning of Molecular Properties Using Graph Neural Networks Enhanced with Chemistry Knowledge. *ACS Omega* (Nov 2025). <https://doi.org/10.1021/acsomega.5c07178>
26. Mastora, I., et al.: Severity of acute pulmonary embolism: evaluation of a new spiral CT angiographic score in correlation with echocardiographic data. *Eur Radiol* **13**(1), 29–35 (Jan 2003). <https://doi.org/10.1007/s00330-002-1515-y>
27. Pham, T., et al.: Graph Classification via Deep Learning with Virtual Nodes (Aug 2017). <https://doi.org/10.48550/arXiv.1708.04357>
28. Qanadli, S.D., et al.: New CT Index to Quantify Arterial Obstruction in Pulmonary Embolism. *American Journal of Roentgenology* **176**(6), 1415–1420 (Jun 2001). <https://doi.org/10.2214/ajr.176.6.1761415>
29. Rampášek, L., et al.: Recipe for a General, Powerful, Scalable Graph Transformer. In: *Proc. NeurIPS*. pp. 14501–14515 (2022)
30. Soffer, S., et al.: Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis. *Sci Rep* **11**(1), 15814 (Aug 2021). <https://doi.org/10.1038/s41598-021-95249-3>
31. Stym-Popper, J., et al.: DAFTEd: Decoupled Asymmetric Fusion of Tabular and Echocardiographic Data for Cardiac Hypertension Diagnosis. In: *Proc. MIDL* (2025), <https://openreview.net/forum?id=ghhGIwv07>
32. Sun, X., et al.: Graph Neural Network based Future Clinical Events Prediction from Invasive Coronary Angiography. In: *Proc. ISBI*. pp. 1–5 (2024). <https://doi.org/10.1109/ISBI56570.2024.10635813>
33. Wasserthal, J., et al.: TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. *Radiology: Artificial Intelligence* **5**(5), e230024 (Sep 2023). <https://doi.org/10.1148/ryai.230024>
34. Xu, K., et al.: How Powerful are Graph Neural Networks? In: *Proc. ICLR* (2018), <https://openreview.net/forum?id=ryGs6iA5Km>
35. Xu, P., Zhu, X., Clifton, D.A.: Multimodal Learning With Transformers: A Survey. *IEEE Trans Pattern Anal Mach Intell* **45**(10), 12113–12132 (Oct 2023). <https://doi.org/10.1109/TPAMI.2023.3275156>