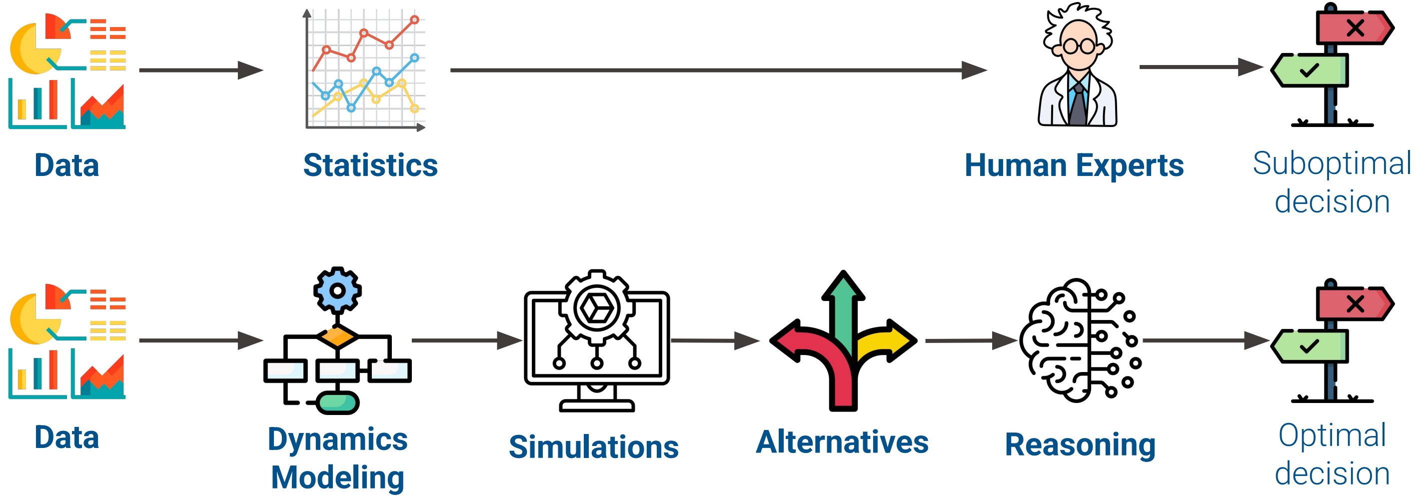
Combining GNNs and Mamba to capture local and global tissue spatial relationships in whole slide images

Ruiwen Ding[1] *et al.*, Nature Scientific Reports, 2025

[1]Medical and Imaging Informatics, University of California, USA

Presented by Amaury WEI

Personal Introduction

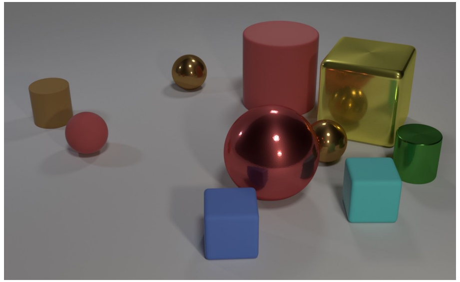


Amaury WEI

2nd year PhD candidate

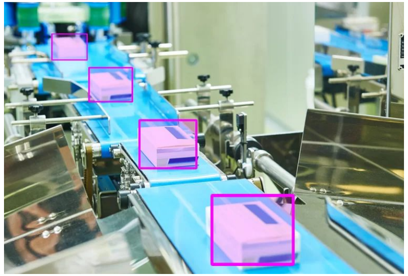
IMOS laboratory
Supervisor: Prof. Olga FINK

Background: Robotics, Mixed Reality, Software



Rigid body dynamics

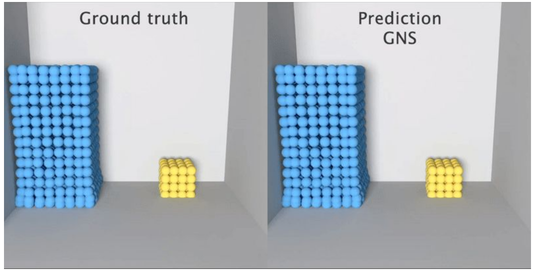
Robotic manipulation



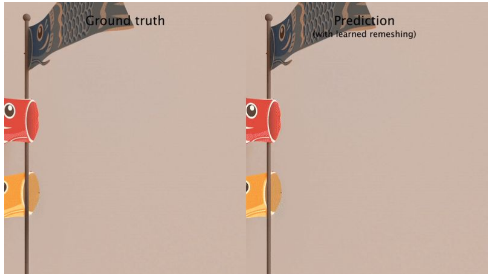
Manufacturing tracking

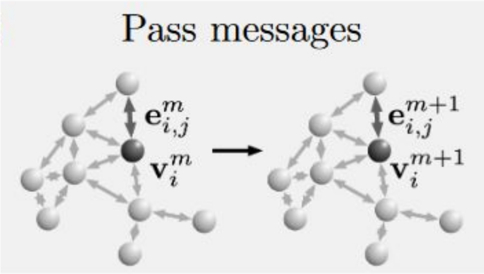
Physics-Informed Surrogate Models

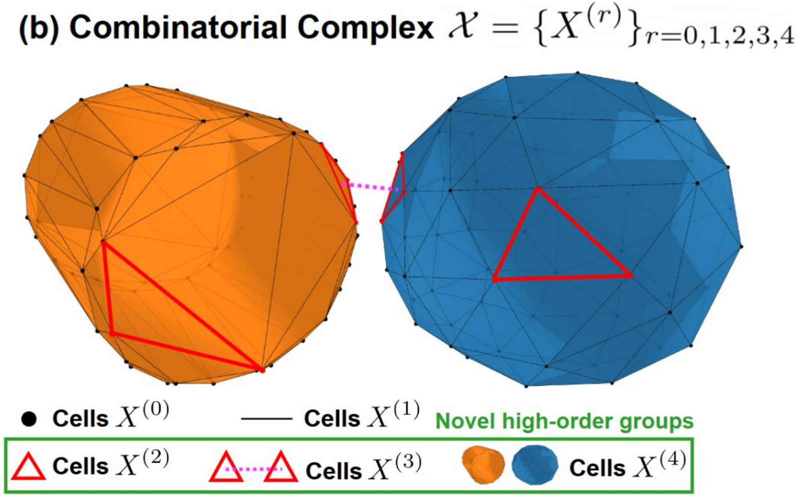
Particle-based (GNS)

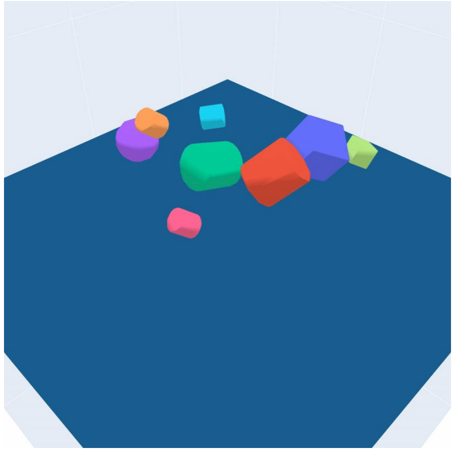


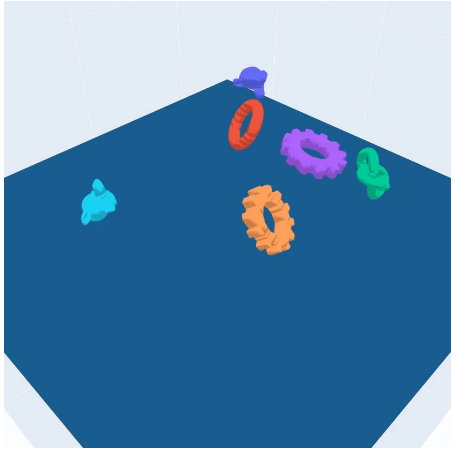
Mesh-based (MGN, FIGNet)











Amaury WEI, Olga FINK, "*Integrating physics and topology in neural networks for learning rigid body dynamics*," Nature Communications (2025)

Combining GNNs and Mamba to capture local and global tissue spatial relationships in whole slide images

Ruiwen Ding[1] *et al.*, Nature Scientific Reports, 2025

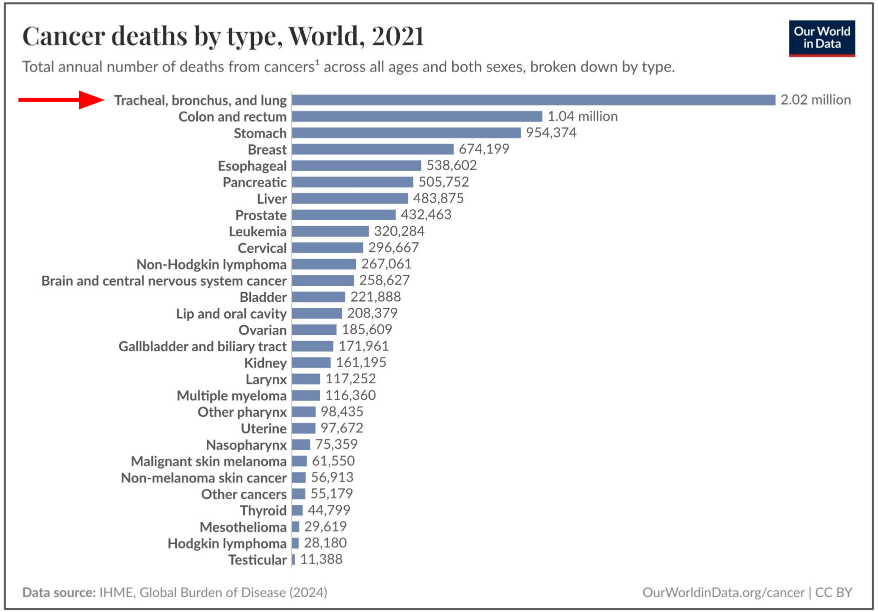
[1]Medical and Imaging Informatics, University of California, USA

Presented by Amaury WEI

Motivation

Lung Adenocarcinoma (LUAD)

- Around 40% of all lung cancers
- Leading cause of global cancer mortality



Progression-Free Survival (PFS)

- Duration in time from a starting point (diagnosis, surgery, treatment start, …) to either disease progression or death.
- Reflects how treatment keeps cancer under control

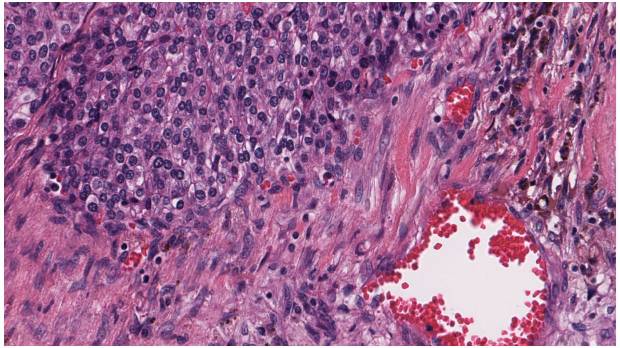
Task: Predict per-patient the PFS risk (*i.e.* how likely the patient is to progress sooner)

Whole Slides Images (WSIs)

Tissue-level view of the tumor carry strong prognostic signals.
(tumor architecture, growth, metastatic lesions, local recurrence after surgery, …)

Whole Slide Images (WSIs)

- Thin slices (3-5 μm) of tumor tissue
- Stained (Hematoxylin and Eosin = H&E)
- Gigapixel images (up to 100’000 x 100’000 pixels)
- Images divided into many tiles



Example of a whole slide image

Datasets

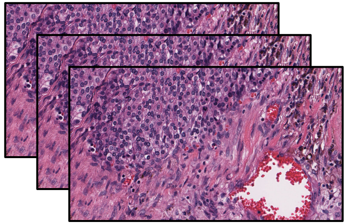
- National Lung Screening Trial (NLST)
  - 132 patients, 243 WSIs
- Cancer Genome Atlas (TCGA)
  - 312 patients, 343 WSIs

444 patients (all had surgery), **568 WSIs**
177 patients with progression (39.9%)

Challenge Overview

Input

Patient WSIs



Proposed Method

GAT+Mamba

Output

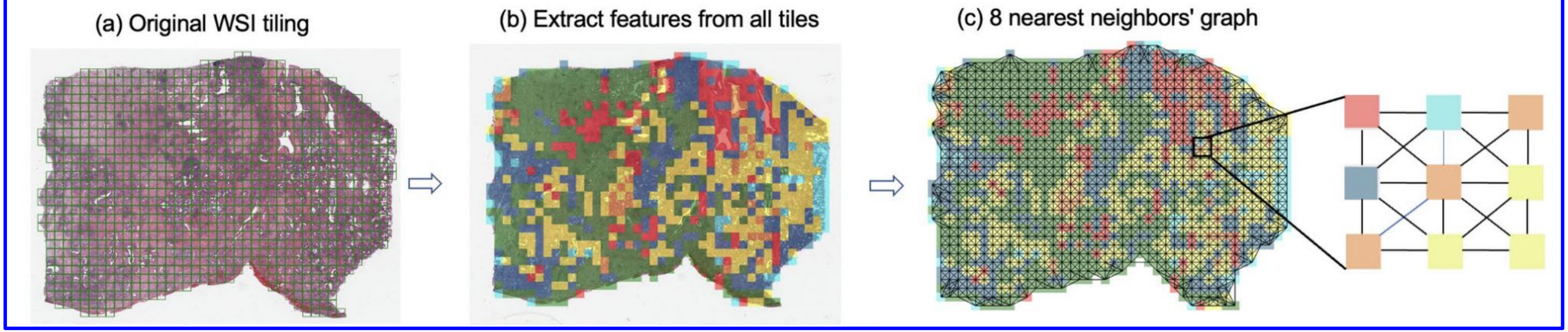
Patient-level risk score

Challenges

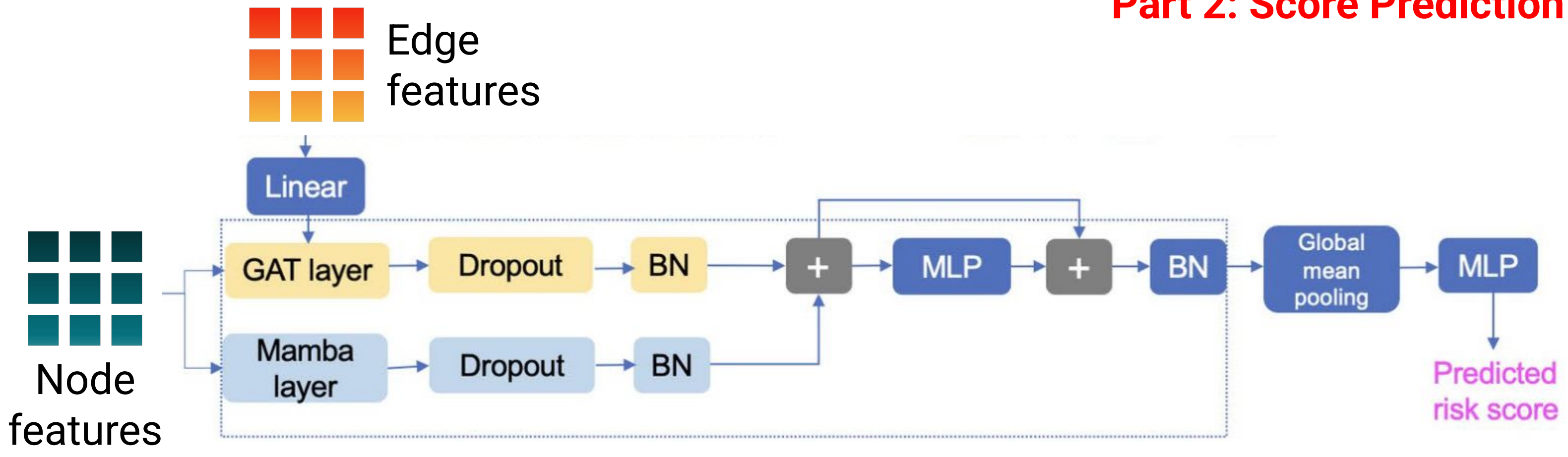
- Scale: 10^2 - 10^3 tiles per WSI, global relationships are hard to compute
- Local vs. global: message-passing GNNs capture local info. but not global
- Sampling: using all tiles is costly, need to sample some of them
- Feature quality: prognostic power depends heavily on feature extractor

Overview of the Proposed Method

Part 1: Graph generation

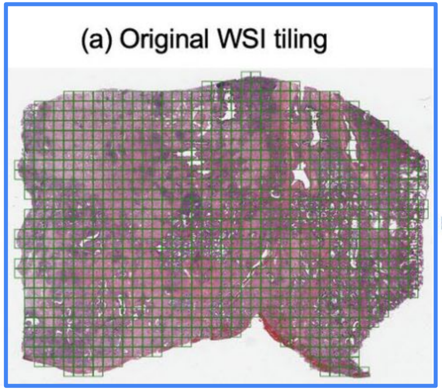


Part 2: Score Prediction



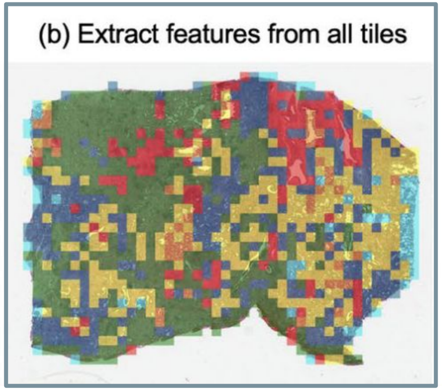
Part 1 – Graph Generation

Input



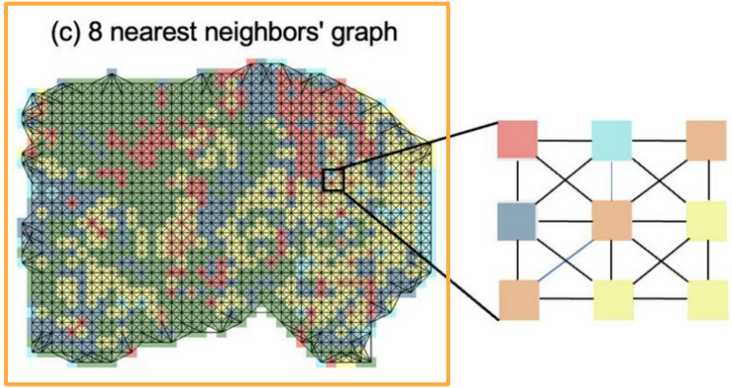
1. Tile WSI
2. Filter tiles

Node Features



1. UNI* features [1024, N]
2. Pos encoding [16, N]

Edge Creation



1. Compute Euclidean distances
2. Create edges using k-NN (k=8)
3. Compute edge features ( )

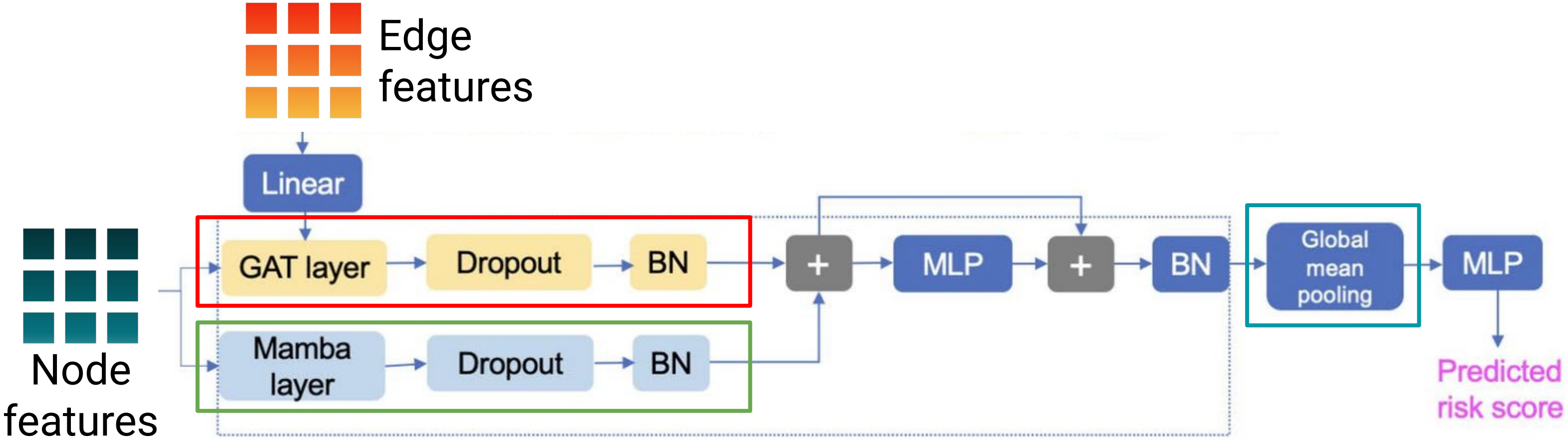
- Subtype-subtype† connection (21 possibilities) $\boldsymbol{E_{cat}}$
- Cosine similarity between node UNI features
- Euclidean distance between nodes

$$\left.\begin{array}{l}\text{Cosine similarity between node UNI features}\\ \text{Euclidean distance between nodes}\end{array}\right\} \boldsymbol{E_{cont}}$$

* Chen, R. J. *et al.*, "*Towards a general-purpose foundation model for computational pathology*," Nature Medicine (2024)

† Ding, R. *et al.*, "*Tailoring pretext tasks to improve self-supervised learning in histopathologic subtype classification of lung adenocarcinomas*", Comput. Biol. Med. (2023)

Part 2 – GAT + Mamba



Graph Path (GAT)

Uses node + edge features

Classic GAT architecture (LeakyReLU)

Focuses on **local** representation

$$\alpha_{i,j} = \frac{exp(LeakyReLU(\boldsymbol{a}^T[\boldsymbol{W}\boldsymbol{X}_i \parallel \boldsymbol{W}\boldsymbol{X}_j \parallel \boldsymbol{W}\boldsymbol{E}_{i,j}]))}{\sum_{k \in \mathcal{N}_i} exp(LeakyReLU(\boldsymbol{a}^T[\boldsymbol{W}\boldsymbol{X}_i \parallel \boldsymbol{W}\boldsymbol{X}_k \parallel \boldsymbol{W}\boldsymbol{E}_{i,k}]))}$$

State Space Model Path (MAMBA)

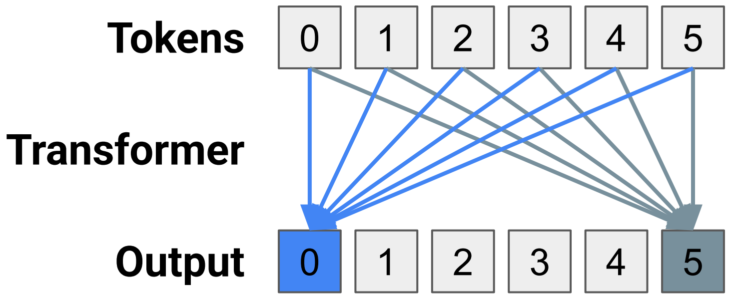
Uses node features only

Treats graph nodes as a sequence of tokens

Can handle long **global** contexts

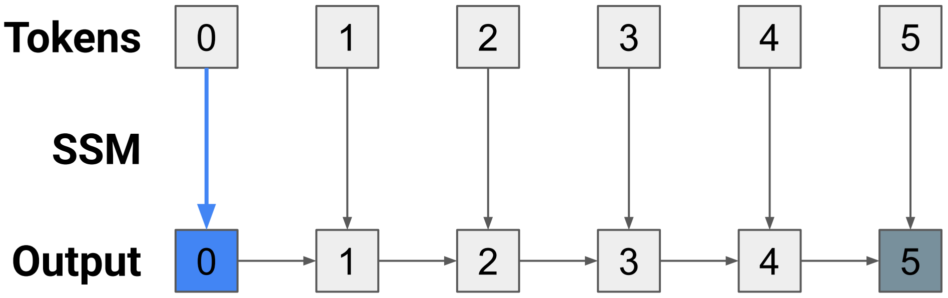
$$h'(t) = \boldsymbol{A}h(t) + \boldsymbol{B}x(t),$$
$$y(t) = \boldsymbol{C}h(t),$$

State Space Models (SSMs) vs. Transformer



All tokens attend to all tokens

Parallel computation



Notion of past / present / future

Token *i* depends only on past and current information

Sequential computation

Results

Baselines

Not graph based

1. Clinical (regression model using age, gender, …) ← Traditional model
2. MLP (statistics from tile-level UNI* features fed to an MLP) ← Simplest NN baseline
3. AttentionMIL (Multi-Instance Learning with Attention)
4. TransMIL (Multi-Instance Learning with Transformer)
6. SCMIL (similar to TransMIL with more sparsity enforced)

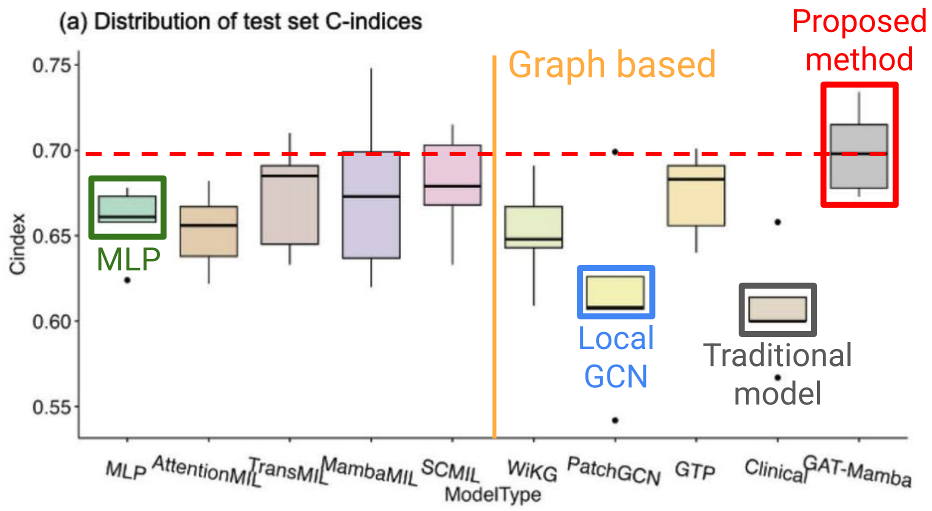
Transformer variations

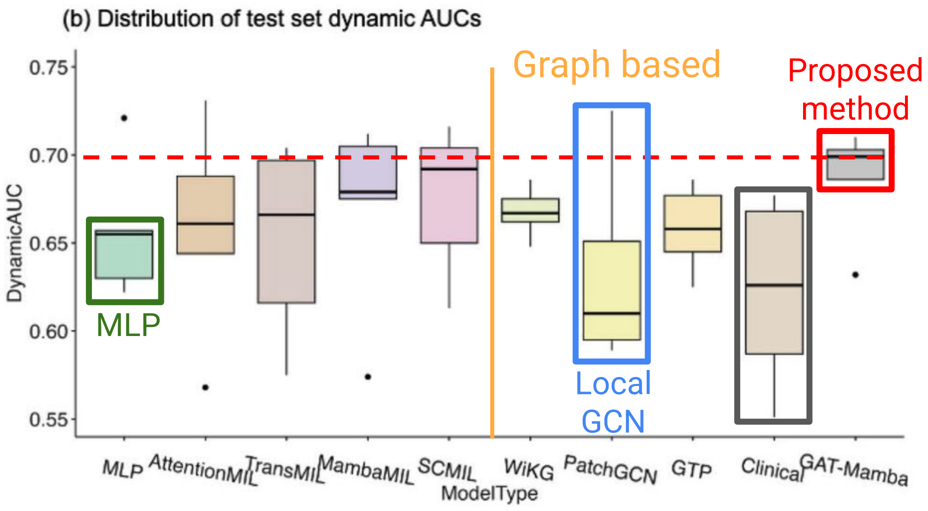
5. MambaMIL (Multi-Instance Learning where Mamba calculates tile-importance)

Graph based

7. WiKG (learnable connectivity based on node features)
8. PatchGCN (Graph Convolution Network between WSI patches) ← Local only
9. GTP (transformer combining GCN then ViT to classify WSIs)

Results





Concordance Index (C-index)

Measures if the model orders pairs of patients by risk in agreement with their PFS outcomes

0.5 = random ranking; 1.0 = perfect ranking

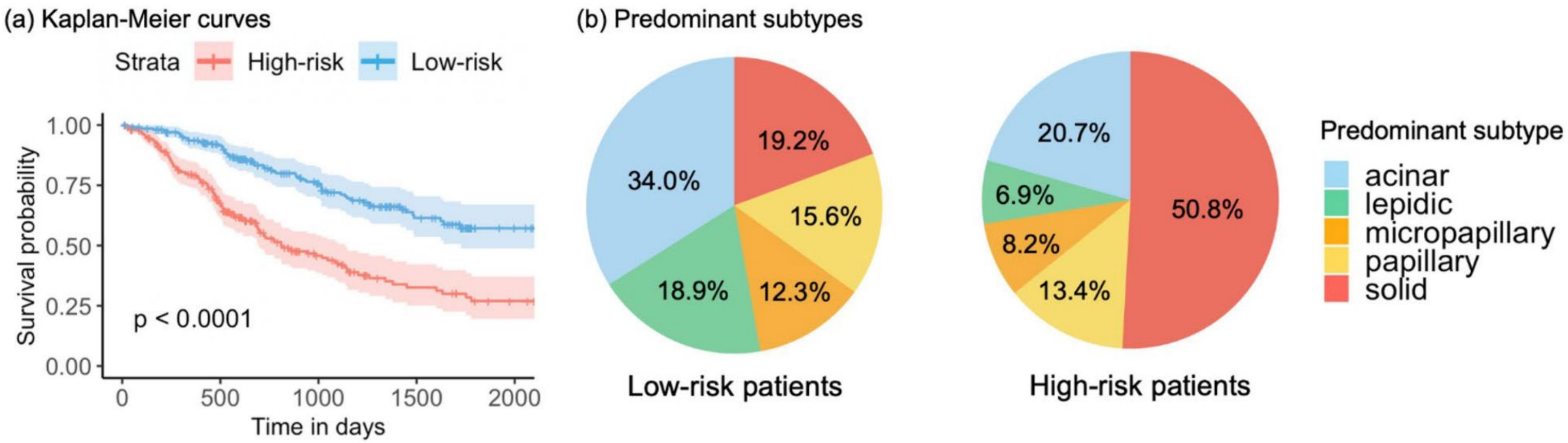
Dynamic AUC

Measures classification performance at a certain time horizon (1-year, 3-year, 5-year)

0.5 = random class; 1.0 = perfect classification

Results

GAT+Mamba splits all patients into low or high risk category.



Kaplan-Meier curve

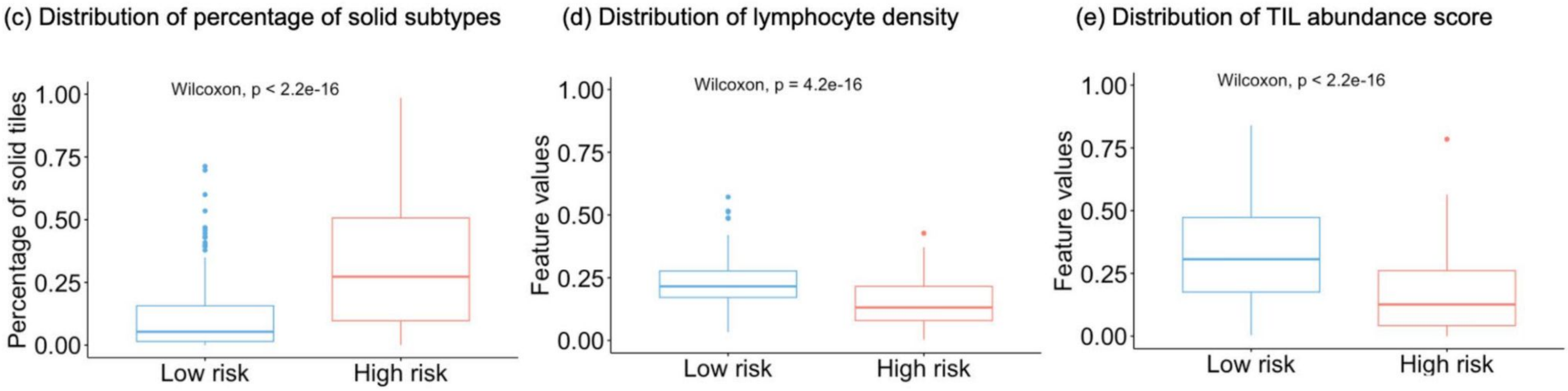
Probability of being “progression-free” at time *t*

Downward step of the curve on progression or death

Findings consistent with well-known prognostic patterns

Results

GAT+Mamba splits all patients into low or high risk category.



TIL = Tumor Infiltrating Lymphocyte

Ablation Studies

Architecture Ablations

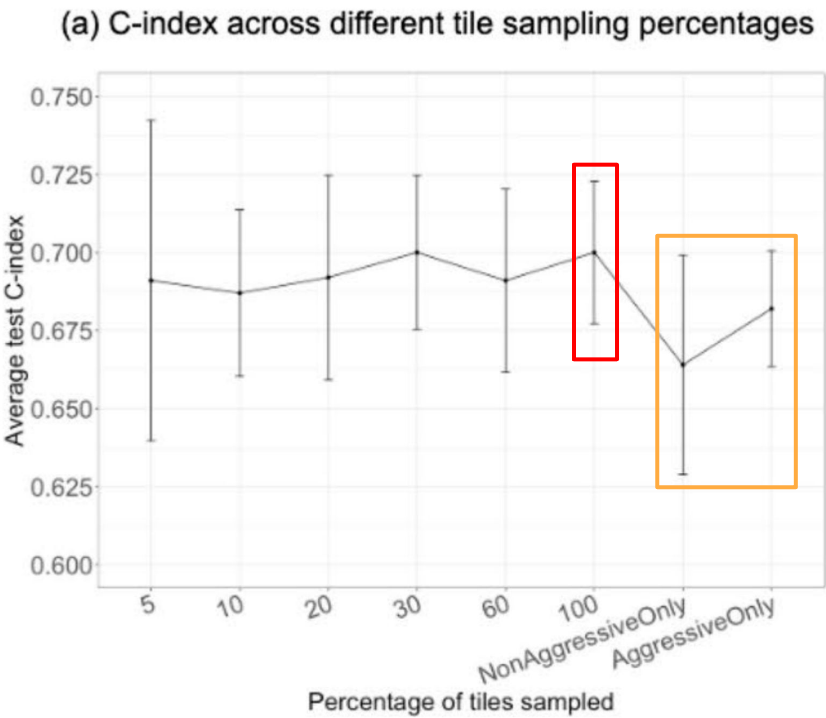
Model	Parameters	TrainTime	InfTime	C-index	Dynamic AUC
GAT	89,666	0.775 ± 0.0660	0.246 ± 0.1193	0.675 ± 0.0304	0.663 ± 0.0329
Mamba	120,401	0.774 ± 0.0817	0.250 ± 0.1051	0.680 ± 0.0267	0.673 ± 0.0395
GAT-Transformer	128,945	1.316 ± 0.0749	0.404 ± 0.1954	0.676 ± 0.0430	0.678 ± 0.0636
GAT-MambaNo$\boldsymbol{E}$	127,041	0.801 ± 0.0985	0.264 ± 0.1405	0.691 ± 0.0311	0.699 ± 0.0255
GAT-MambaNo$\boldsymbol{X_{PE}}$	106,177	0.782 ± 0.0820	0.280 ± 0.1380	0.687 ± 0.0447	0.680 ± 0.0684
GAT-Mamba	127,425	0.819 ± 0.0736	0.286 ± 0.1551	0.700 ± 0.0228	0.686 ± 0.0281

Node Features Ablations

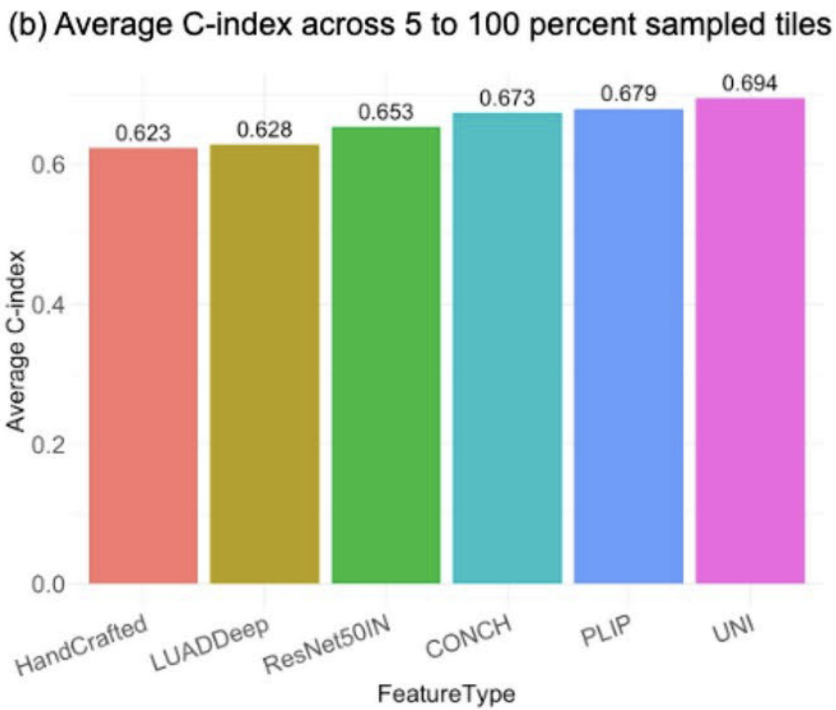
Features	Raw features	C-index	Dynamic AUC
Hand-crafted	57	0.629 ± 0.0427	0.633 ± 0.0663
LUADDeep	512	0.635 ± 0.0357	0.638 ± 0.0376
ResNet50IN	1024	0.669 ± 0.0297	0.668 ± 0.0373
CONCH	512	0.683 ± 0.0377	0.663 ± 0.0463
PLIP	512	0.684 ± 0.0353	0.671 ± 0.0396
UNI	1024	0.700 ± 0.0228	0.686 ± 0.0281

Ablation Studies

Tile Sampling Strategies



Robustness to Sampling



Conclusion

PFS risk prediction using Whole Slide Images (WSIs)

0. Graph creation using UNI model (nodes) + tile subtype classes (edges)

1. Local information using Graph Attention model (GAT)

2. Global information using state-space model (Mamba)

3. Combine local + global with sum aggregation

4. Risk prediction using global mean pooling

Robust to tile sampling, **Discovers** feature trends for low & high risk groups

Discussion Starters

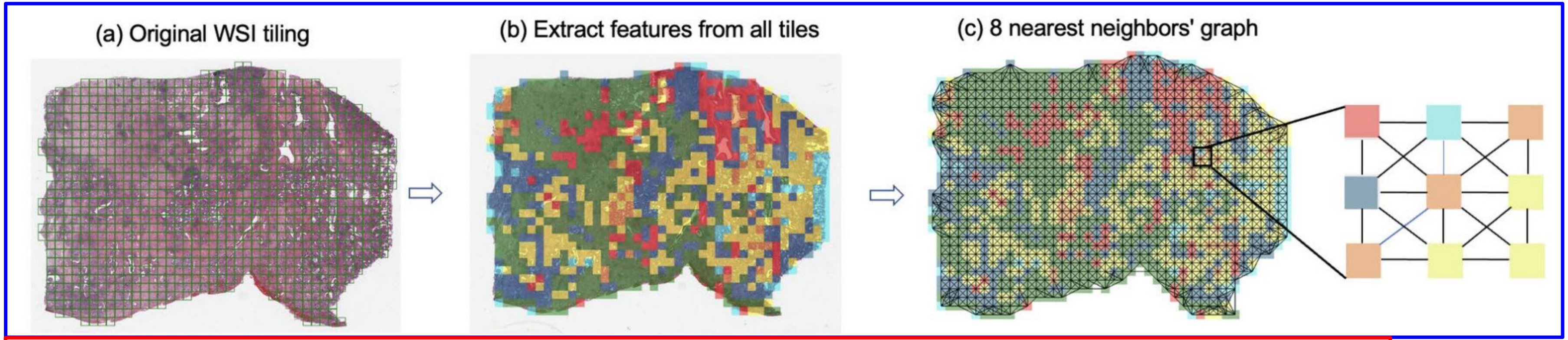
Positive Aspects

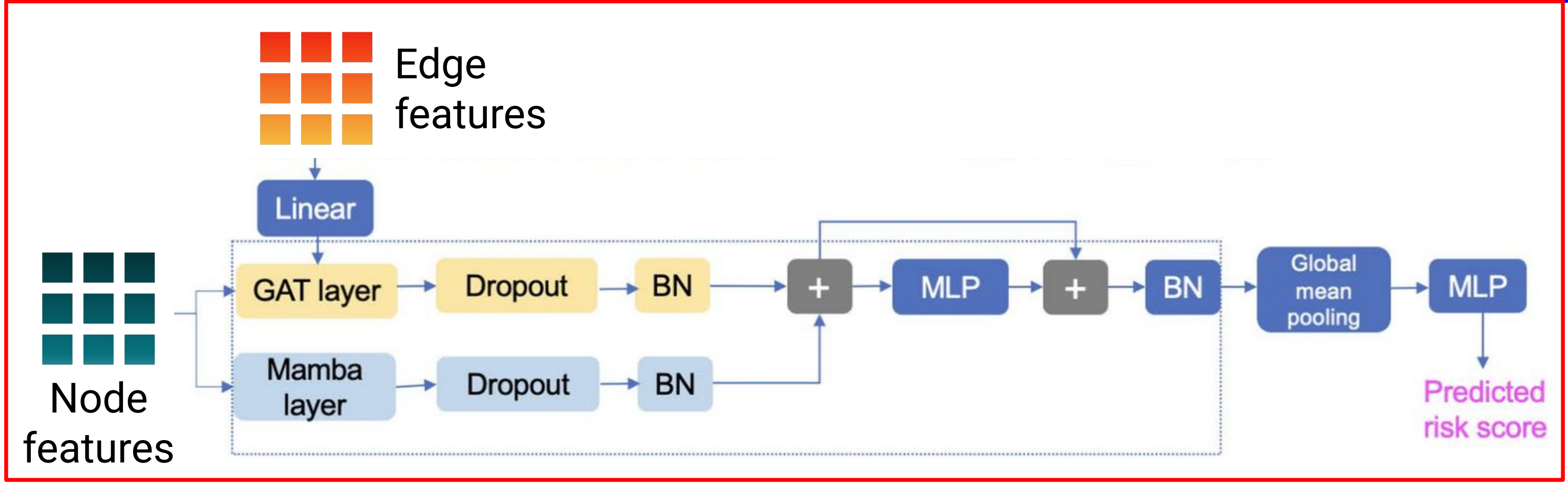
+ Interesting way to combine both **local** and **global** information
+ Classify **low-risk** and **high-risk** patients
+ Correlates with well-known medical patterns
+ Robust to **sampling** (even with 30% of the total tiles)

Negative Aspects

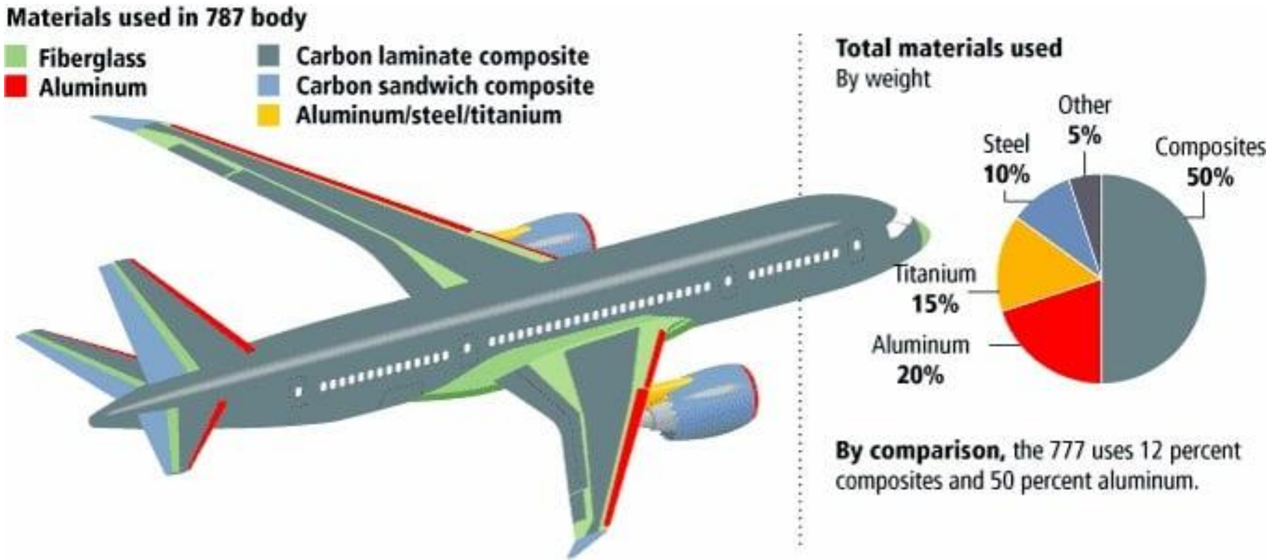
- Choice of a **State Space Model** for global information
- Robustness analysis of the SSM to flipping, cropping, sequence inversion, …
- Global mean pooling as aggregation
- Lack of comparison against pure vision-based models
- Computational cost compared to a basic MLP

Summary

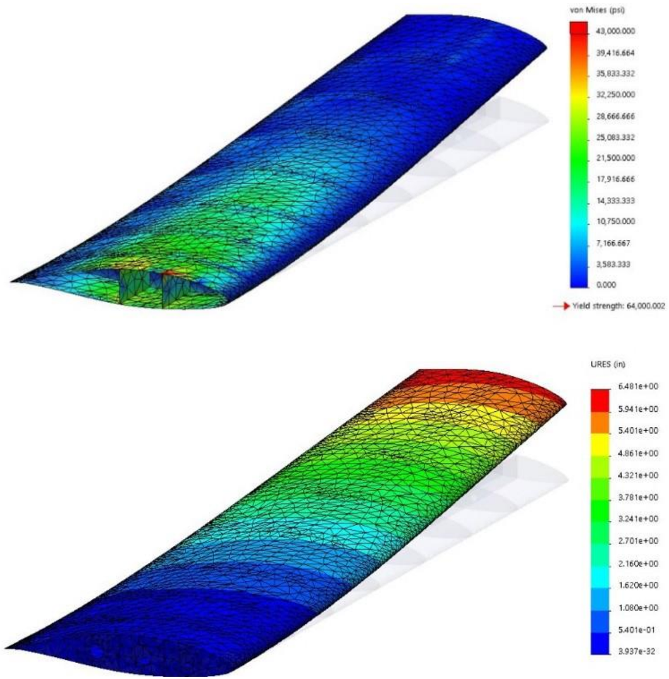

(a) Original WSI tiling
(b) Extract features from all tiles
(c) 8 nearest neighbors' graph

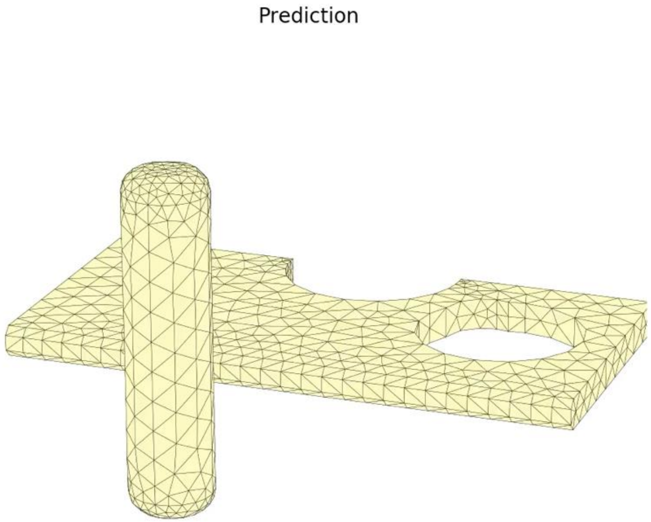

Edge features
Linear
Node features
GAT layer
Dropout
BN
Mamba layer
Dropout
BN
+
MLP
+
BN
Global mean pooling
MLP
Predicted risk score

Physics Informed GNNs for Composite Materials



GNNs with Material Mesh as Nodes





www.nature.com/scientificreports

scientific reports

Check for updates

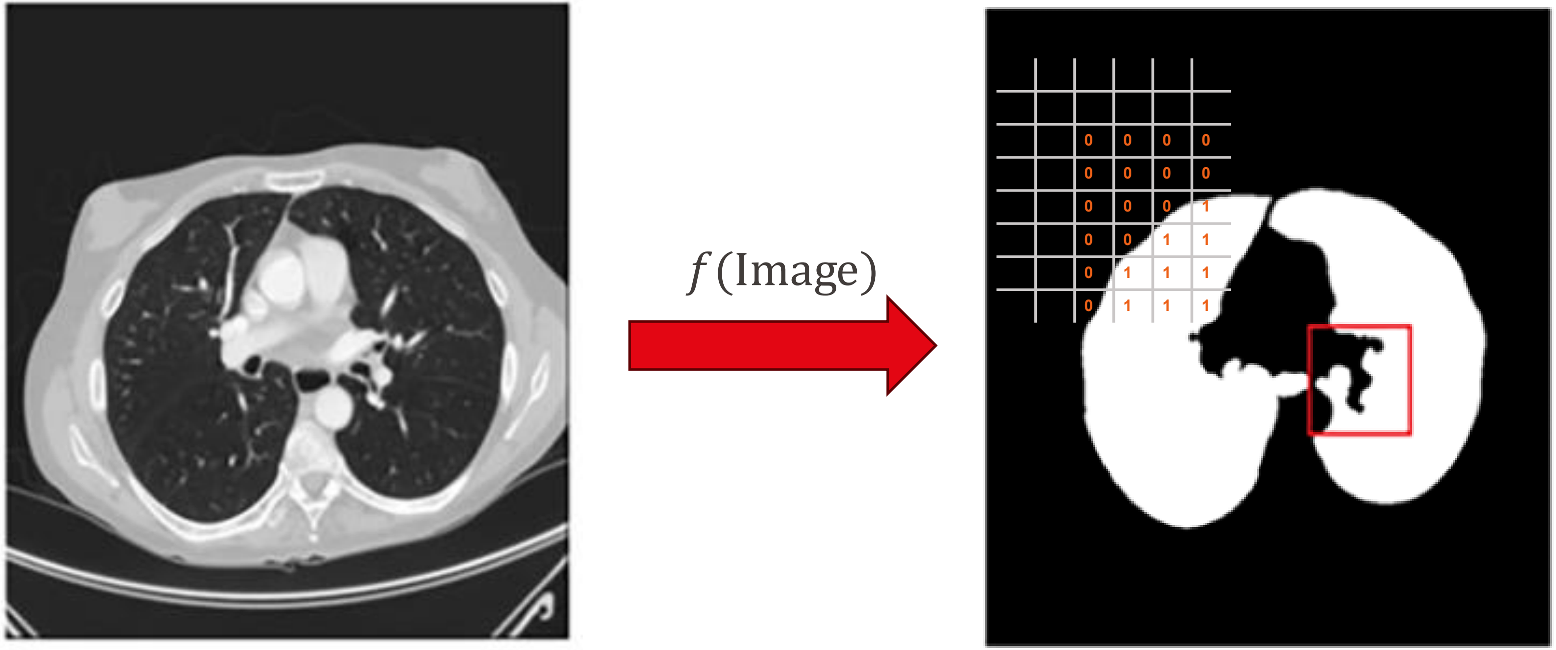
OPEN

Graph neural network model using radiomics for lung CT image segmentation

Mohammad Khalid Faizi[1,6]✉, Yan Qiang[1,5,6], Md Masum Billa Shagar[1,6], Yangyang Wei[2,6], Ying Qiao[2,6], Juanjuan Zhao[1,3,4]✉ & Zia Urrehman[1,6]

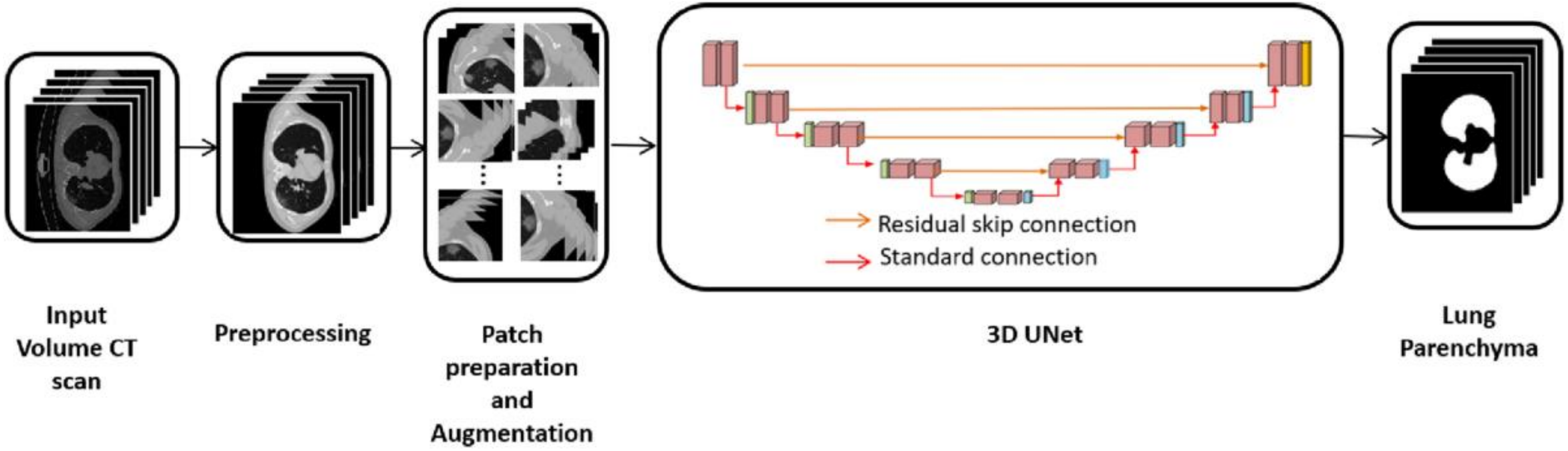
Lung CT Image Segmentation

- Goal is to have a Neural Network f, that learns a segmentation grid



Current State of the Art

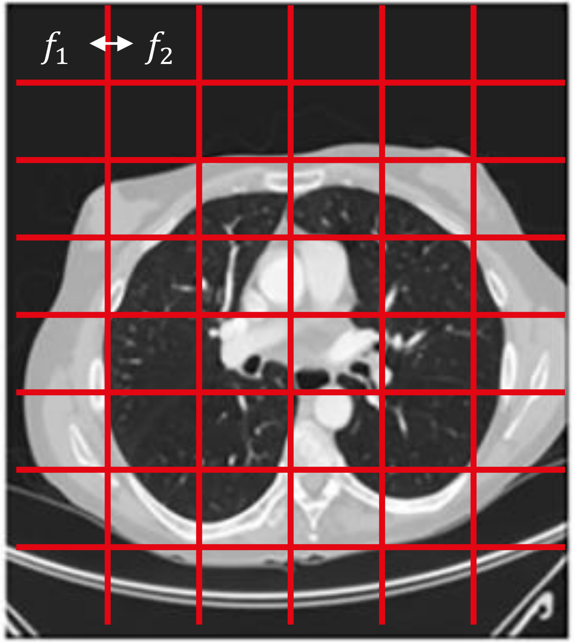
- Current state-of-the-art are Unet like architectures

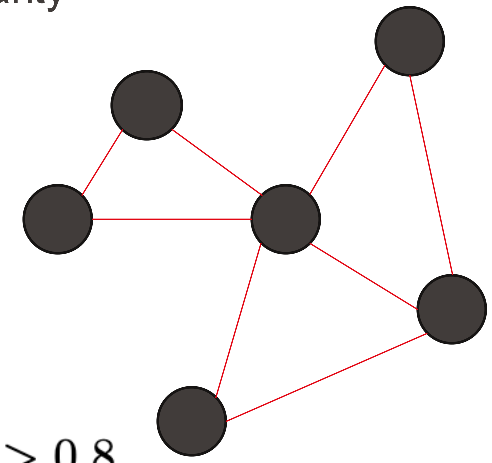


L., A.A., S., V.C.S. Cascaded 3D UNet architecture for segmenting the COVID-19 infection from lung CT volume. *Sci Rep* **12**, 3090 (2022). https://doi.org/10.1038/s41598-022-06931-z

Extract Graph Representation

- Divide image into patches
- Assign radiomics features f_i
- Build graph based on
  - Connectivity and
  - Cosine similarity





$$A_{ij} = 1 \text{ if } \frac{f_i \cdot f_j}{\|f_i\| \|f_j\|} > 0.8.$$

Radiomics

Radiomic Features

This section contains the definitions of the various features that can be extracted using PyRadiomics. They are subdivided into the following classes:

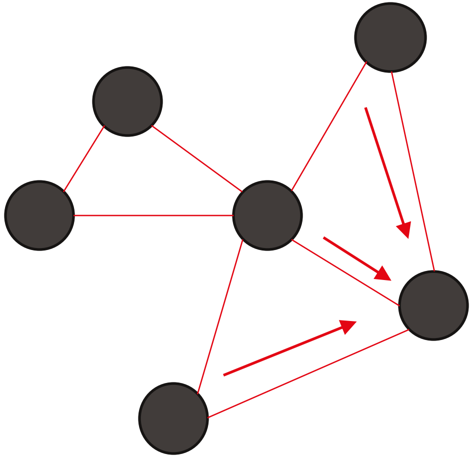
- `First Order Statistics` (19 features)
- `Shape-based (3D)` (16 features)
- `Shape-based (2D)` (10 features)
- `Gray Level Co-occurrence Matrix` (24 features)
- `Gray Level Run Length Matrix` (16 features)
- `Gray Level Size Zone Matrix` (16 features)
- `Neighbouring Gray Tone Difference Matrix` (5 features)
- `Gray Level Dependence Matrix` (14 features)

$$energy = \sum_{i=1}^{N_p} (\mathbf{X}(i) + c)^2$$

$$\begin{bmatrix} 0.2 & 0.9 \\ 0.3 & 0.01 \end{bmatrix}$$

$$E = 0.235, \text{ for } c = 0$$

Extract Graph Embedding

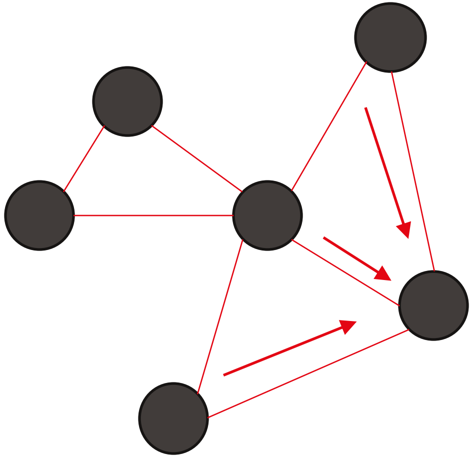


Graph Attention Network

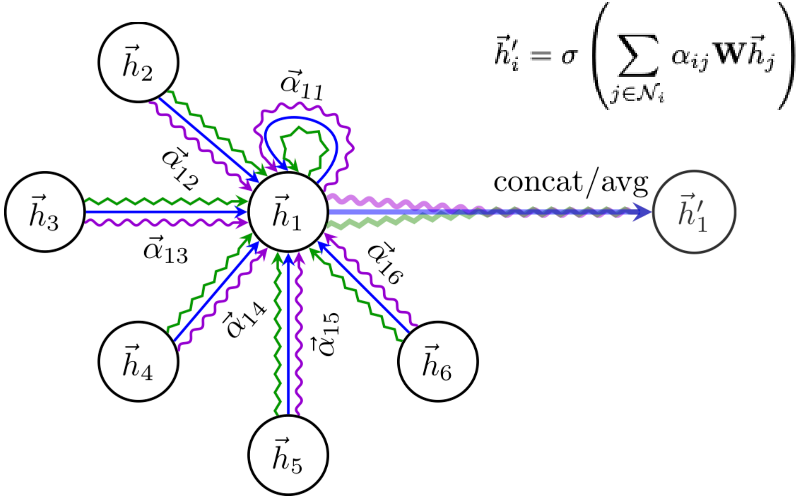
- Graph Attention Layers extract meaningful node embeddings
- Attention adds weight to connections

Extract Graph Embedding

- One attention layer

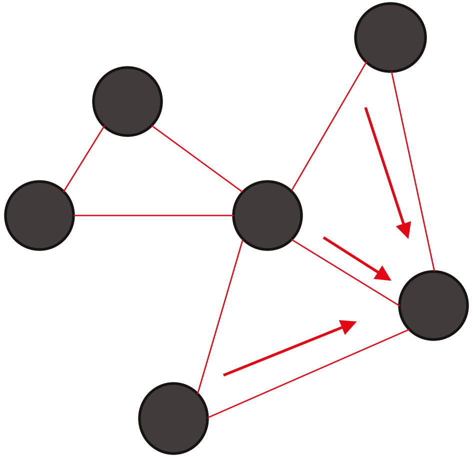


Graph Attention Network



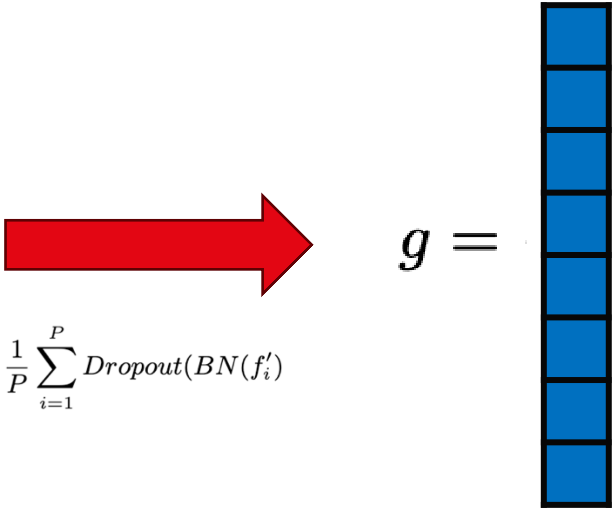
Extract Graph Embedding

- Aggregate node features into graph embedding
- Mean Aggregation

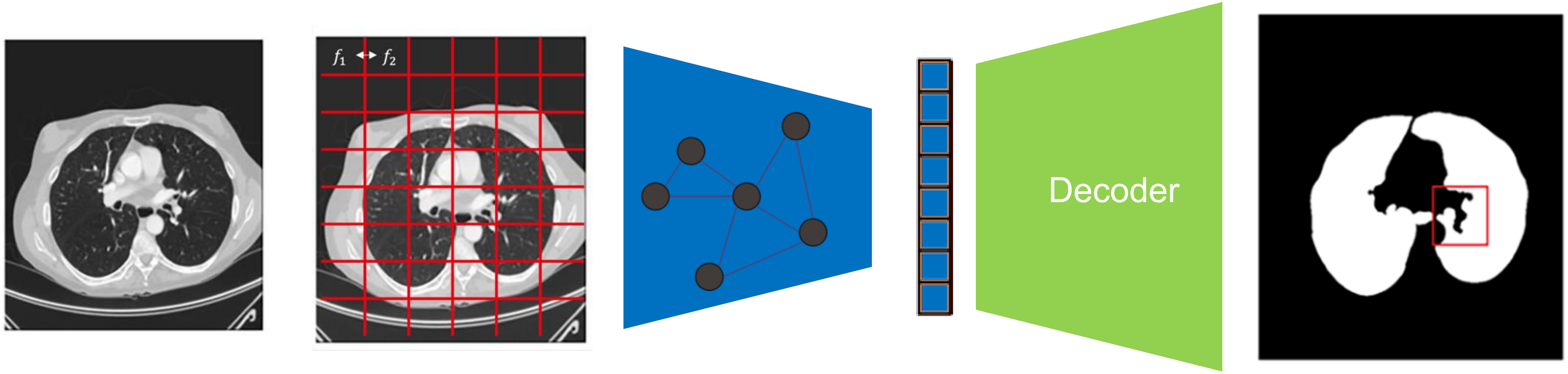


Graph Attention (GAT)
3x GAT Block

$$g = \frac{1}{P}\sum_{i=1}^{P} Dropout(BN(f_i'))$$



Not the Architecture Overview



Original Image

Patched Image

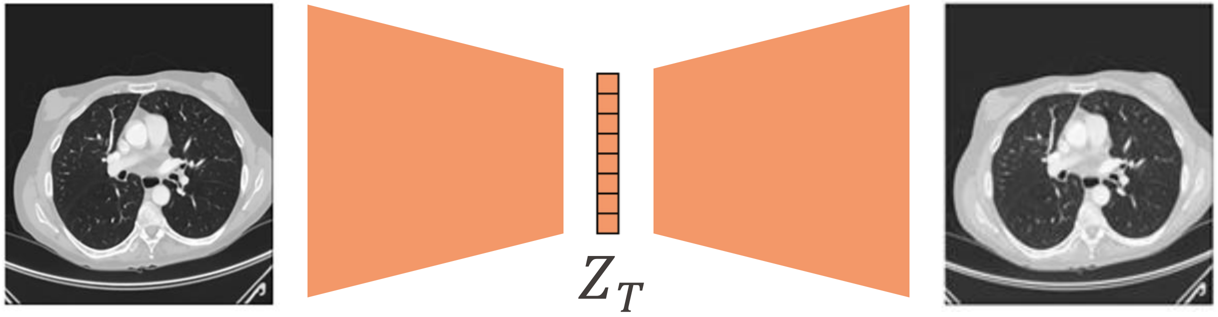
GNN Encoder

CNN Decorder

Segmentation Map

AutoEncoder

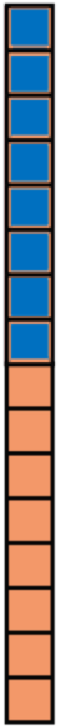
- Additionally capture image relationships with features from image encoder



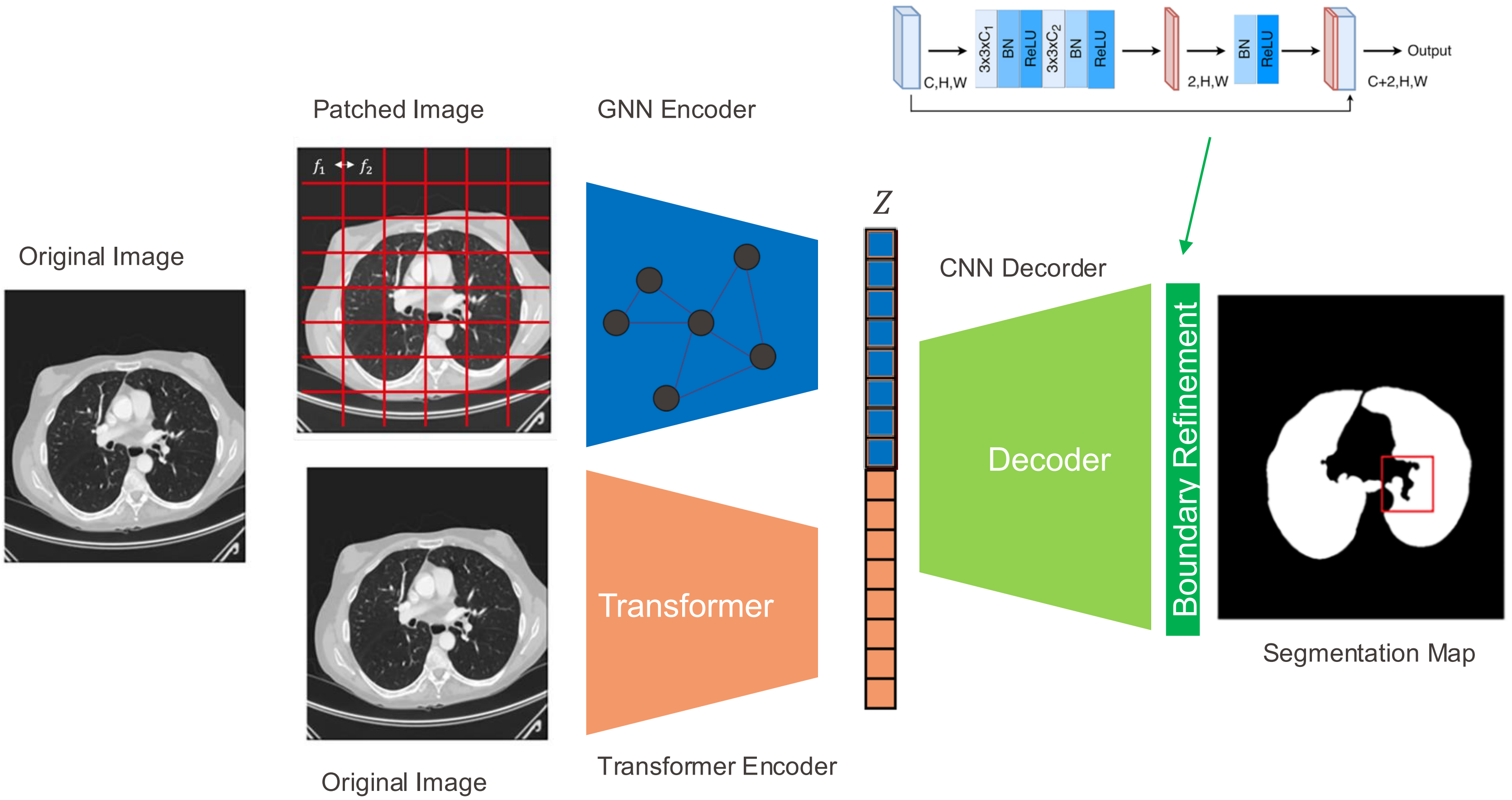
Feature Fusion

- Feature fusion to capture topological and image relationships
- Concatenate embeddings
- Simple fusion reduces parameters and inference time

$$Z = Concat(g, Z_T) =$$



Architecture Overview

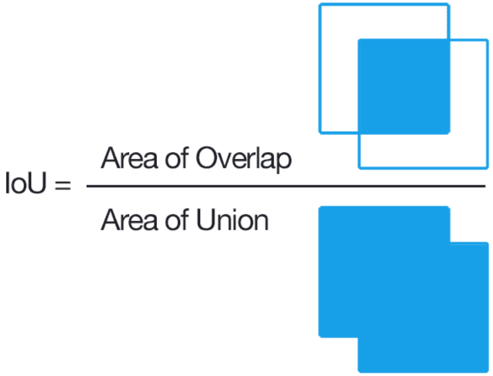


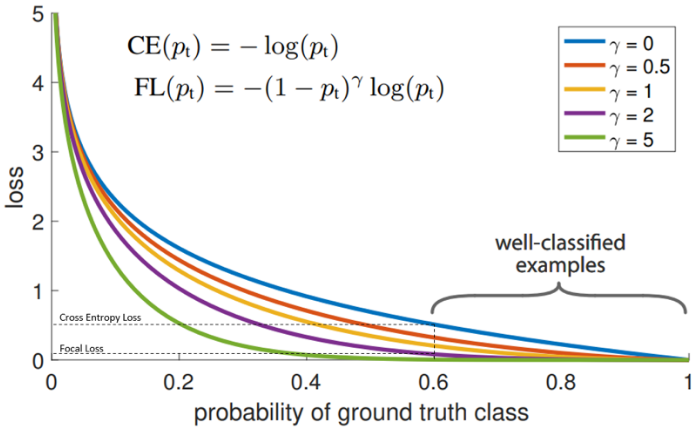
Special Loss

$$Loss = \lambda_1 Loss_{focal} + \lambda_2 Loss_{focal} + \lambda_3 Loss_{iou}$$

$$IoU = \frac{TP}{TP + FP + FN}$$

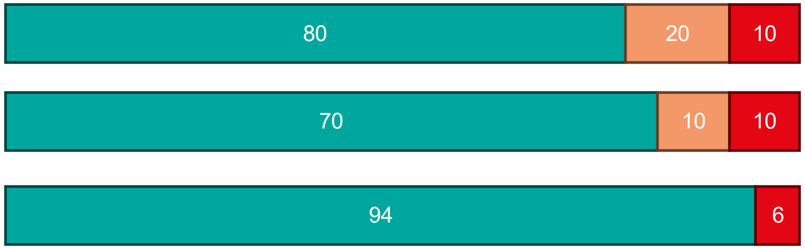
$$FL = -\alpha\,(1 - p_t)^{\gamma}\log(p_t)$$





Datasets

- LIDC-IDRI
  - 1800 Images
  - 512x512
  - DICOM format
- Chest CT
  - 700 Images
  - JPG, PNG format
- Interesting Data Split
  - Capture anatomical deviation
  - Maximize training data
  - Test set selected manually to be diverse and similar to training ($p>0.05$)
  - 5-fold cross-validation

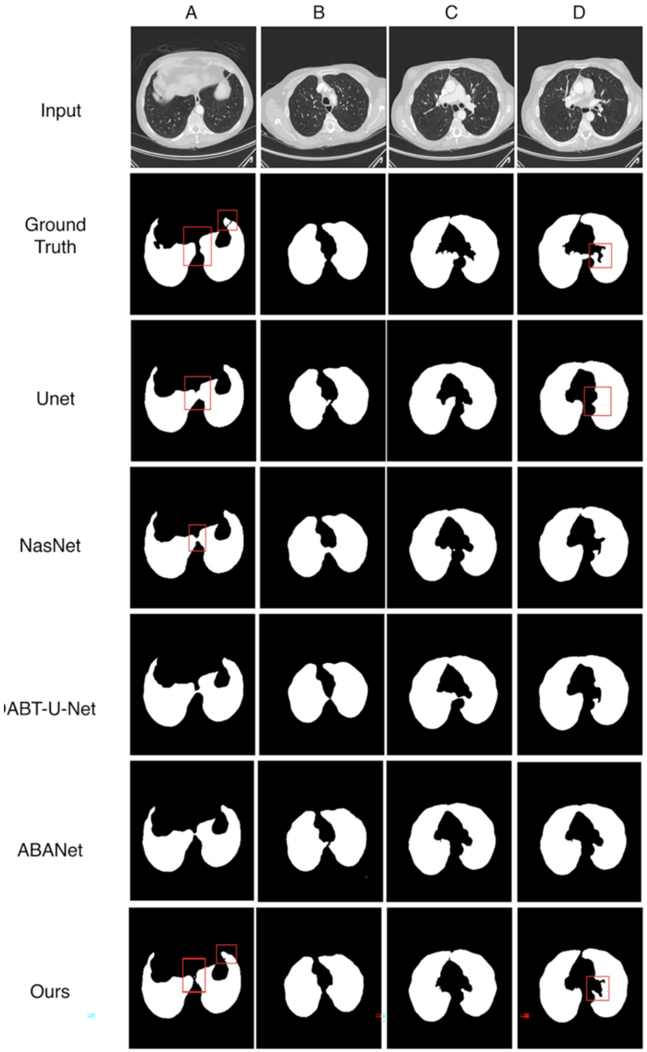


Performance

- Outperforms other methods

Method	IoU	F1-score	Precision	Recall	Accuracy	GPU Memory (GB)
U-Net	94.07	95.03	98.15	99.45	95.85	6.5
RU-Net	94.67	96.33	95.65	97.45	97.10	6.8
ResNet34-Unet	95.38	97.83	97.29	98.25	96.85	7.0
BCDU-Net	96.10	98.51	99.00	98.01	97.20	7.2
ResBCDUnet	97.15	98.03	99.09	97.58	97.06	7.4
NasNet	97.07	98.53	98.05	97.45	98.85	8.0
DABT-U-Net	96.65	97.69	98.21	98.15	98.35	7.6
ABANet	96.57	97.70	98.25	98.35	99.15	7.5
TransUNet	96.80	97.80	97.90	98.50	98.90	9.5
UKAN	97.20	98.10	98.30	98.70	99.20	7.8
GEANet	**98.07**	**99.03**	**98.65**	**99.45**	**99.85**	8.2

Table 1. Performance comparison on Dataset Version 1 (LIDC-IDRI) test set, including IoU, F1-score, Precision, Recall, Accuracy, and GPU memory. Significant values are in bold.

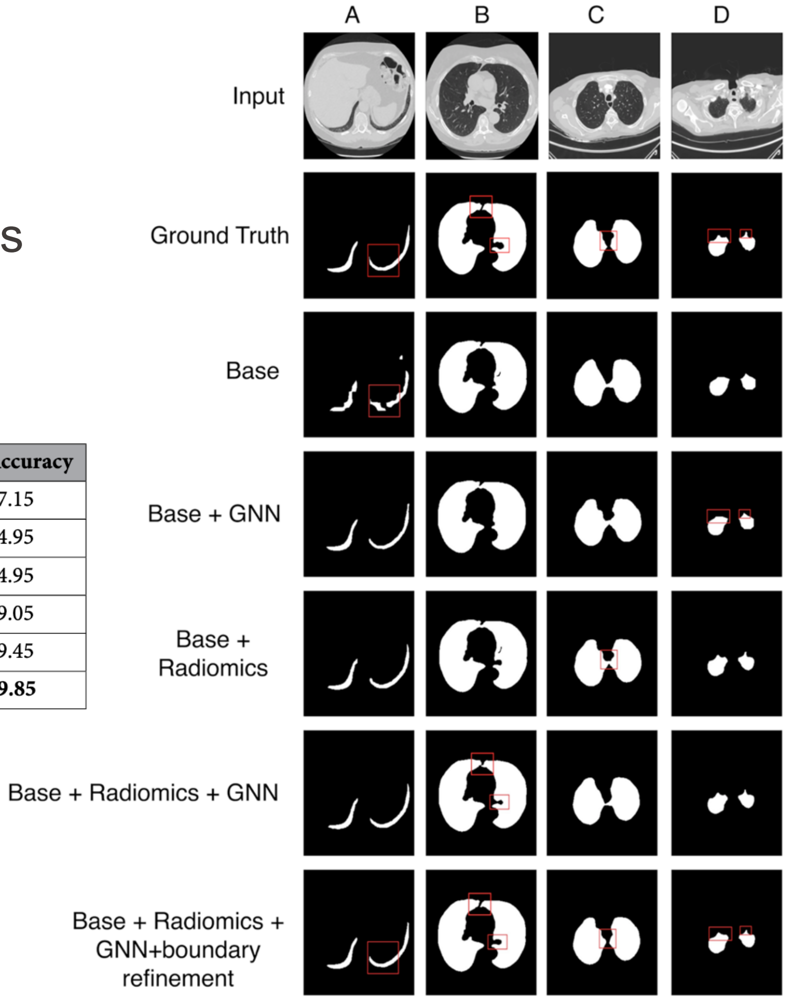


Ablation

- Ablation shows that GNN+Radiomics work together to achieve superior accuracy

Method	IoU	F1-score	Precision	Recall	Accuracy
Baseline	95.52	96.63	97.46	95.82	97.15
Baseline + GNN	95.64	96.25	94.46	94.82	94.95
Baseline + Radiomics	95.21	96.25	94.46	94.82	94.95
Baseline + GNN + Radiomics	97.15	98.05	98.46	98.82	99.05
Baseline + GNN + Radiomics + Concatenation	97.65	98.53	98.65	99.05	99.45
GEANet Model	**98.07**	**99.03**	**98.65**	**99.45**	**99.85**

Table 4. Ablation study results on the test set, evaluating the contribution of each component using IoU, F1score, Precision, Recall, and Accuracy. Significant values are in bold.



Summary

Positive

- Accurate segmentation of Lungs from CT images
- Better Performance through inclusion of
  - topological features with GNN
  - domain knowledge via radiomics
- Augmentation of current architecture (Transformer) through feature fusion
  - Approach transferable to other segmentation problems

Negative

- Exact architecture unclear
- Overclaimed small memory usage