Towards joint representation learning for multimodal patient similarity networks for ovarian cancer

Emilia Agasi[1], Charlie Gourley[2], Robb Hollis[2], Ian Simpson[1]

[1]School of Informatics, University of Edinburgh [2]Institute of Genetics and Cancer, University of Edinburgh

Ovarian cancer is a heterogeneous disease.

Ovarian cancer is the eighth most common malignancy among women globally with notably poor survival outcomes[1]. Around 90% of cases fall into five carcinoma subtypes with distinct prognoses and treatment responses[2] but heterogeneity remains within each subtype: across histology, molecular profile and clinical course. This intra-subtype variation is a major barrier to precision medicine and characterising it requires methods that can integrate information across multiple modalities.

Contribution.

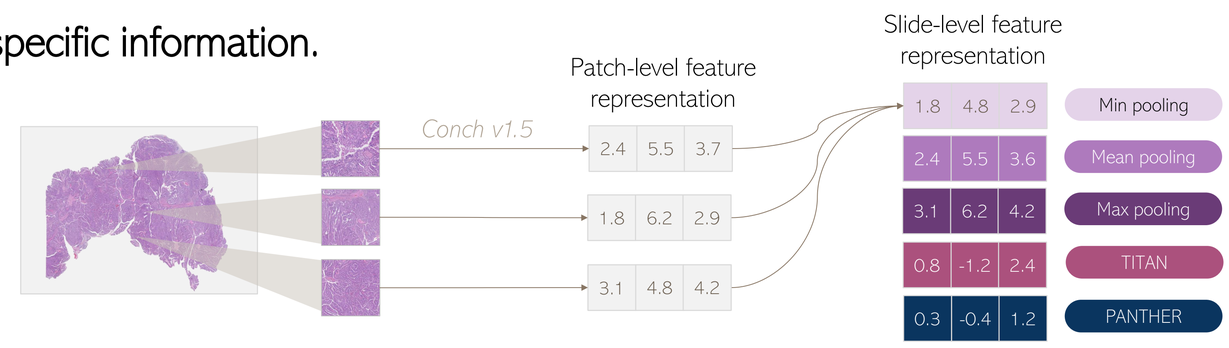
This work motivates the need for interpretable, outcome-aware slide-level representations in ovarian cancer. We show that:

i. pre-trained pathology foundation models transfer to high-grade serous ovarian carcinoma (HGSOC) without fine-tuning, capturing morphological information
ii. five existing slide aggregation strategies carry some survival signal but none provide interpretable, patient-specific predictions and
iii. we propose a multimodal patient similarity network framework that integrates histology with genomic and clinical data as a path toward interpretable outcome prediction.

Histopathology encodes valuable disease-specific information.

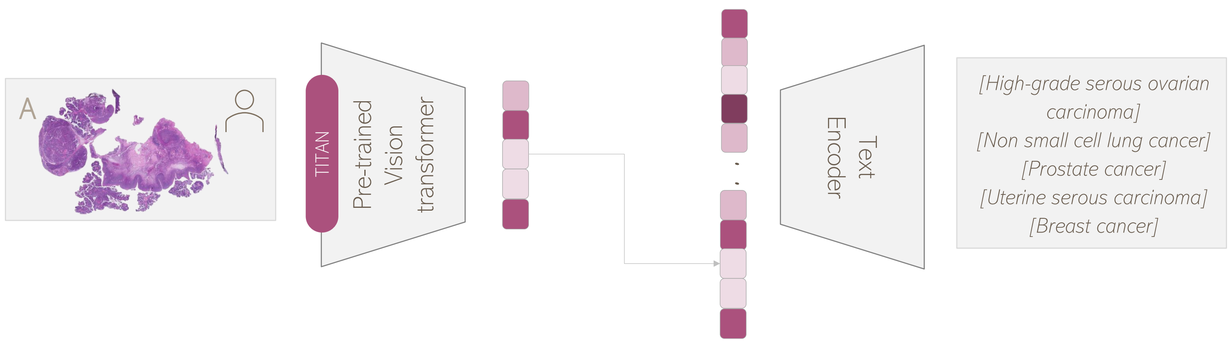
We extract 512x512 px patches at 20x from whole-slide images, encode them with CONCH v1.5 to produce 768-dim patch embeddings, then aggregate to a single slide-level representation using one of five methods:

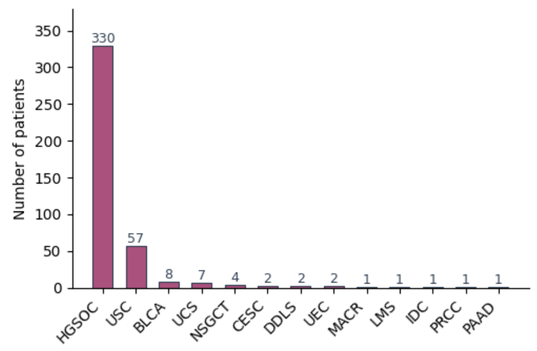
- Static: min, max, and mean pooling
- Contrastively pre-trained slide encoder (TITAN)
- Gaussian-mixture-based aggregator (PANTHER)



Zero-shot classification: Pre-trained embeddings capture morphological information.

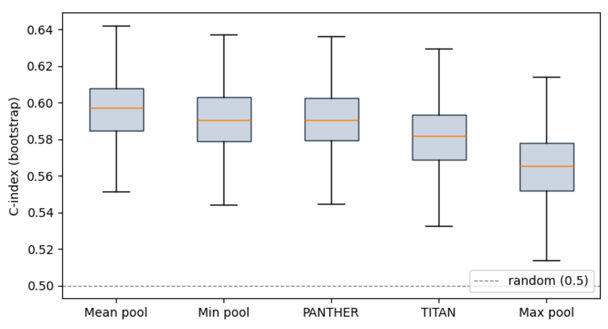
We tested whether TITAN[2], contrastively pre-trained on pathology reports, encodes ovarian-specific morphology by performing cross-modal zero-shot classification: each slide embedding was matched to the most similar of 46 cancer-type text embeddings using cosine similarity. 78% of slides were correctly classified as HGSOC. The most frequent confusion was with uterine serous carcinoma, a morphologically similar tumour, suggesting the embeddings capture meaningful tissue features

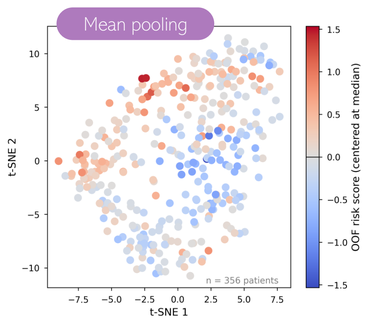


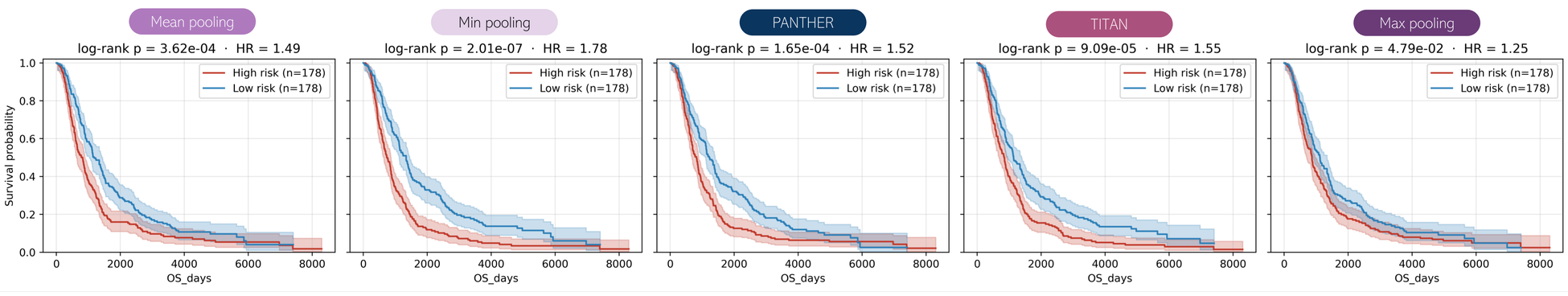


Slide-level aggregators encode survival signal but lack interpretability.

Using penalised Cox regression with median-split risk stratification, every aggregation method produced significant separation between high- and low-risk groups (log-rank $p < 0.05$; HR 1.25–1.78). Differences between methods were modest and none of the resulting models offer insight into which tissue features drive the prediction.

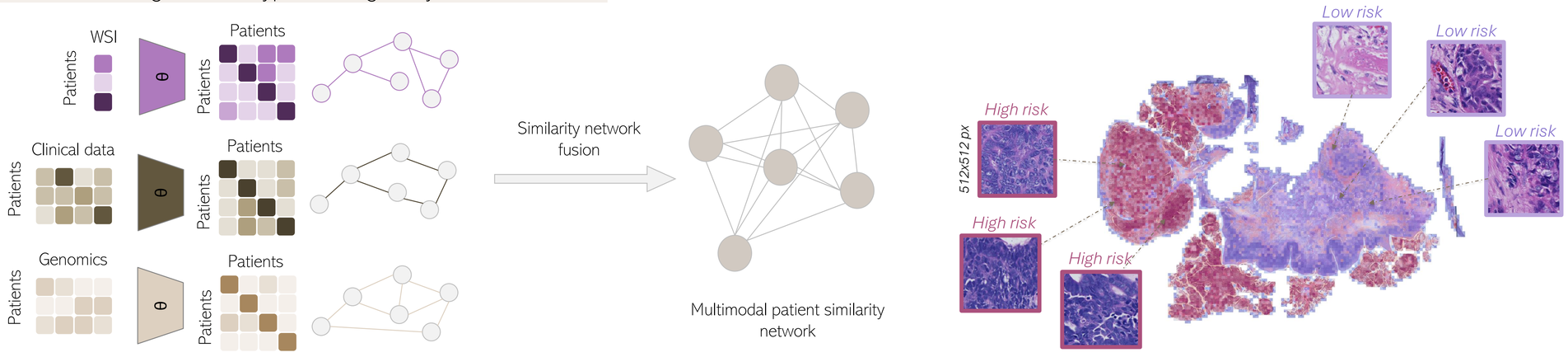




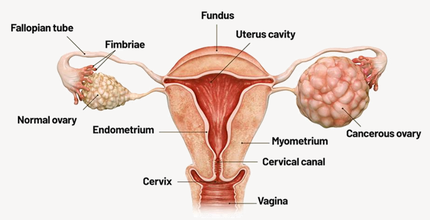


Multimodal patient similarity networks for joint representation learning.

To move beyond opaque slide-level embeddings, we represent patients as nodes in a similarity graph, with edges derived separately from histology, genomics, and clinical features. Fusing these unimodal networks into a single multimodal PSN provides a structured representation in which patient relationships are defined jointly across modalities. This graph is intended as the input for joint representation learning, where node embeddings are learned to capture shared structure across data types while remaining interpretable at the patient and feature level. Together, we hope this approach enables accurate outcome prediction while elucidating the mechanisms driving intra-subtype heterogeneity in ovarian cancer.

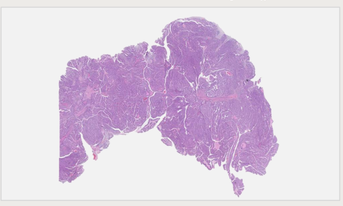


At a glance…



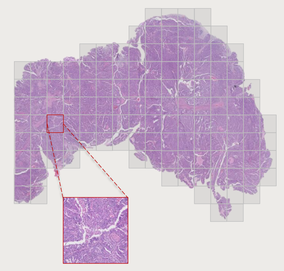
High-grade serous ovarian cancer

Tumour biopsy



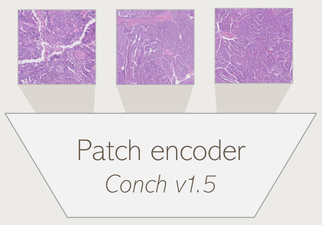
H&E Whole slide image

Preprocessing



Tissue segmentation and tiling (512px x512px, 20x)

Feature extraction



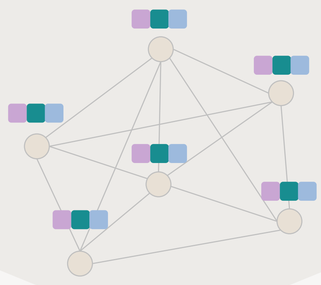
Patch encoder
Conch v1.5

Patch embeddings (768-dim vectors)

Slide aggregation

Min pooling | Mean pooling | Max pooling | TITAN | PANTHER

Multimodal graph construction



Survival prediction

Future work

i. Joint representation learning. Train GNNs on the multimodal PSN to learn embeddings that integrate histology, genomics and clinical data.
ii. Interpretability. Combine attention-based MIL with graph attention to trace predictions back to tissue regions, genes and clinical variables.
iii. Generalisation. Extend the framework to other cancer types to assess transferability.

[1] Bray, F. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries (2024).
[2] Ding, T., Wagner, S.J., Song, A.H. et al. A multimodal whole-slide foundation model for pathology (2025).
[3] Kilim, O., Olar, A., Biricz, A. et al. Histopathology and proteomics are synergistic for high-grade serous ovarian cancer platinum response prediction (2025).

Any questions left unanswered? Get in touch!