AI-Driven Characterisation of Ovarian Cancer Heterogeneity

Emilia Agasi, Ian Simpson, Charlie Gourley
School of Informatics, 10 Crichton Street, University of Edinburgh, EH8 9AB, UK.

Background

Ovarian cancer ranks as the eighth most common malignancy among women globally, with notably poor survival outcomes[1]. Five carcinoma subtypes comprise approximately 90% of cases, characterised by distinct prognoses and therapeutic responses[2].

However, **accurate subtype classification and characterisation of intra-subtype heterogeneity remain major barriers in precision medicine**. This heterogeneity is found across histological variation, molecular profiles, and clinical outcomes.

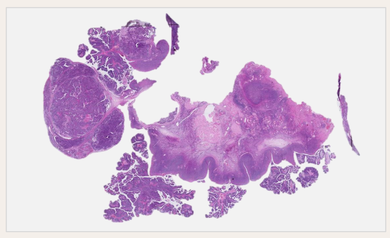
This PhD project addresses these challenges by:

1) Developing AI-based methods to accurately predict ovarian cancer subtypes
2) Characterising molecular and clinical heterogeneity within subtypes, and
3) Linking these patterns to treatment responses and patient outcomes.

Histopathology: A data treasure chest

Histopathology images are key to diagnosing many cancers.

AI-based methods allow us to systematically **encode high-dimensional features** into quantitative **embeddings**. Such embeddings encode information on tissue architecture and cellular composition into a quantitative representation that enables large-scale analysis of heterogeneity.

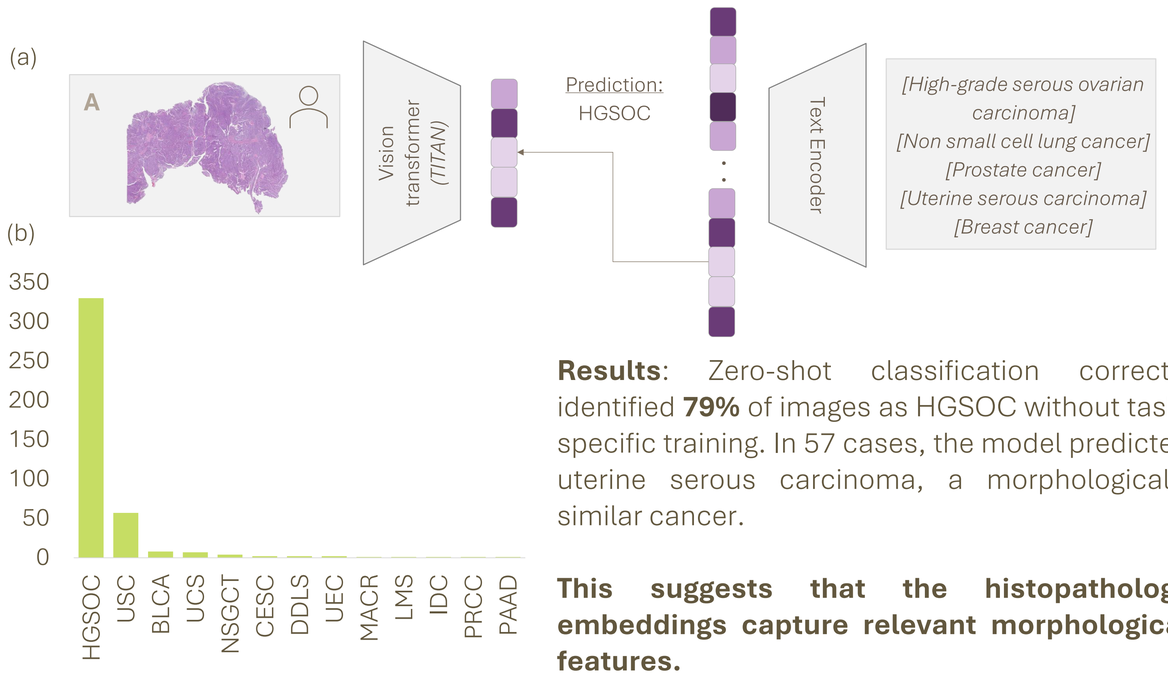


Histopathology embeddings capture cell type information

What information is encoded within histopathology image embeddings?

Data: Edinburgh houses over 1,000 H&E-stained whole slide images (WSI) from ovarian cancer patients with matched molecular data. For this experiment, a HGSOC cohort (n=417) was analysed.

Method: TITAN[3], a contrastively pre-trained vision encoder, generated embeddings from WSI. We performed cross-modal zero-shot classification by encoding each WSI and matching it to the most similar text embedding from 46 cancer type descriptions using cosine similarity.



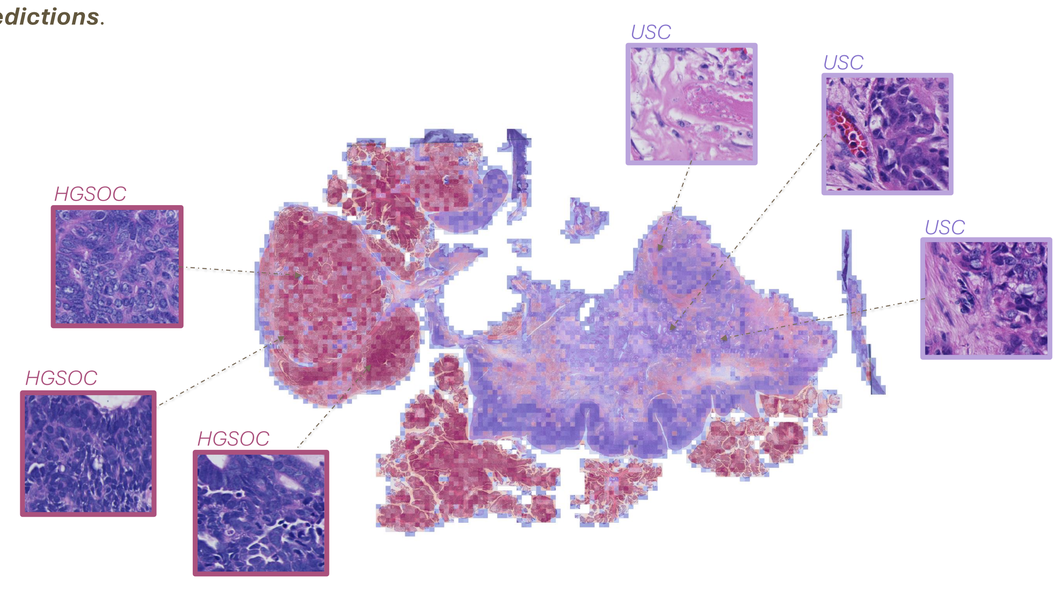
Results: Zero-shot classification correctly identified **79%** of images as HGSOC without task-specific training. In 57 cases, the model predicted uterine serous carcinoma, a morphologically similar cancer.

This suggests that the histopathology embeddings capture relevant morphological features.

Fig 1 *– (a) Cross-modal zero-shot experiment set-up. A query WSI is encoded using TITAN and is classified matching it to the most similar text embedding out of 46 cancer types. (b) Zero-shot classification results.*

Interpretable AI models for outcome prediction

While embeddings capture morphological features, their utility for precision medicine depends on our ability to predict patient-specific outcomes and understand ***which tissue characteristics drive these predictions***.



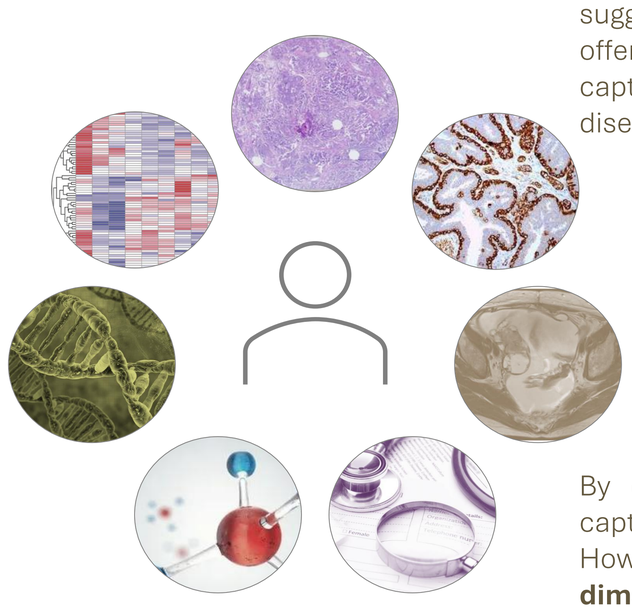
Attention-based multiple instance learning (ABMIL)[3] enables both objectives through its attention mechanism.

How it works: WSI are divided into smaller patches, each representing a localised tissue region. Each patch is encoded into an embedding, creating a 'bag' of patch embeddings per slide. An attention mechanism learns to weight each patch contribution to the slide-level prediction.

The result: An interpretable predictive model based on histopathology.

Fig 2 *– Preliminary results from an ABMIL model showing multi-head patch attention visualised on a WSI.*

Patient data is complex



Beyond histopathology, emerging evidence suggests that integrating additional modalities offers improved predictive performance by capturing complementary aspects of disease biology[4, 5].

Molecular data: Gene expression profiles, mutational status, copy number alterations, transcriptomic subtypes

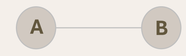
Clinical data: Patient demographics, treatment history, stage, survival outcomes, treatment response

By combining these data, we can better capture the complexity of ovarian cancer. However, **analysing these complex, high-dimensional datasets requires frameworks that can model patient relationships across multiple biological scales**.

What are patient similarity networks?

Patient similarity networks (PSNs) offer an analytical framework for modelling complex patient relationships.

PSNs represent patients as nodes connected by edges that reflect their similarity.



The distance between nodes (e.g., patients A and B) quantifies how similar patients are to each other based on their features, enabling identification of patient subgroups with shared characteristics.

Multimodal patient similarity networks

Multimodal PSNs provide a framework for modelling patient relationships on different biological scales. Using methods such as graph neural networks and unsupervised clustering, we can **identify clinically meaningful patient subgroups** and elucidate mechanisms driving heterogeneity.

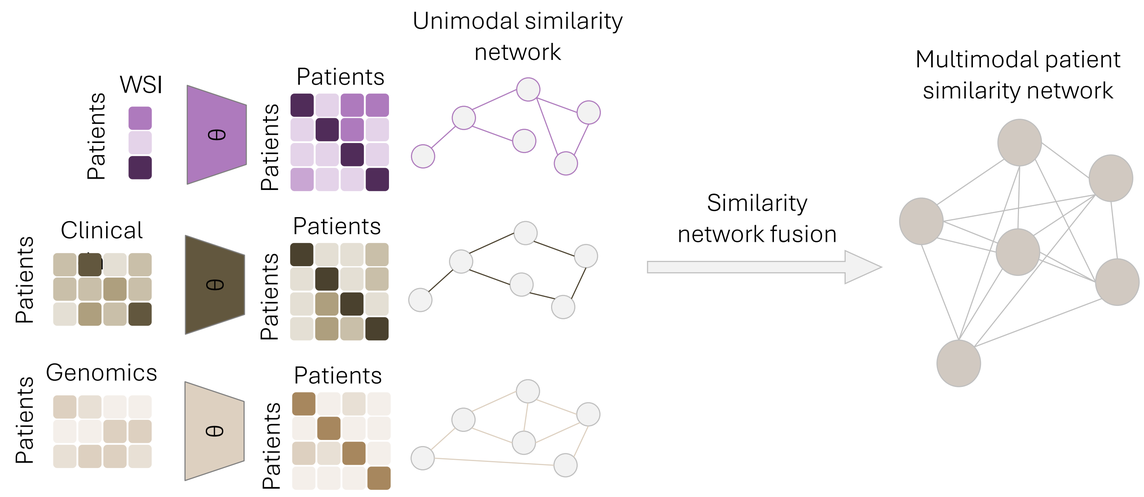


Fig 3 *– Process of extracting patient features from modalities and merging the resulting unimodal patient similarity networks into a multimodal network.*

Future work and impact

1. Deploy attention-based models for clinical subtype classification, prognosis, and biomarker discovery.
2. Apply graph neural networks to multimodal PSNs for stratification and treatment response prediction.
3. Generalise multimodal approach across cancer types to establish precision medicine approach.

Sources

[1] Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 74, 229–263 (2024).
[2] Köbel, M. et al. Ovarian carcinoma subtypes are different diseases: implications for biomarker studies. PLoS Med. 5, e232 (2008).
[3] Ilse, M., Tomczak, J. M. & Welling, M. Attention-based deep multiple instance learning. In Proc. 35th International Conference on Machine Learning (eds Dy, J. & Krause, A.) 2127–2136 (PMLR, 2018).
[4] Ding, T., Wagner, S.J., Song, A.H. et al. A multimodal whole-slide foundation model for pathology. Nat Med (2025).
[5] Kilim, O., Olar, A., Biricz, A. et al. Histopathology and proteomics are synergistic for high-grade serous ovarian cancer platinum response prediction. npj Precis. Onc. 9, 27 (2025).

Any questions left unanswered? Get in touch!