Multimodal Colon Cancer Prognosis with Inferred Spatial Transcriptomics and Biologically-Informed LLM loss

Navneet Prakash, Saatvik Kesarwani, Arjun Chitla, Ali Usman, Joshua Levy
Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center

Dartmouth Health

ABSTRACT

- Cancer prognosis is important because it informs treatment decisions. It remains a critical challenge due to the complexity of tumor biology and patient heterogeneity.
- We developed a multimodal pipeline integrating WSIs, inferred ST, bulk RNA-seq, and DNA methylation for cancer prognosis on colorectal adenocarcinoma. We use crossmodally pretrained modality-specific encoders to embed each modality, and combine the embeddings using gated attention and basic concatenation.
- In order to increase the generalization and performance of the model, especially to compensate our limited data and use of inferred instead of ground-truth ST, we developed a novel bioligically-informed loss function using an LLM

INTRODUCTION

- Cancer prognosis is difficult because outcomes are shaped by both tumor morphology and molecular complexity
- Currently, most models often use single data types such as histology or bulk sequencing, which fail to capture the full biological complexity
- Histopathology slides provide rich spatial and morphological context but lack direct molecular information
- Bulk RNA-seq and DNA methylation (DNAm) capture global molecular states but lack information on tissue architecture
- Spatial transcriptomics (ST) links gene activity to tissue structure, but is highly expensive and therefore largely inaccessible
- Approach: develop a multimodal pipeline that integrates whole-slide images (WSI), inferred ST, bulk RNA, and DNA methylation trained with the aid of a powerful custom LLM loss.
- Objective: reveal patterns linking tumor morphology and molecular activity that separate patients by survival risk and improve prediction accuracy

METHODS

Whole-Slide Images (WSI):
- Extracted tissue patches with the Gigapath tile encoder, then used a Top-K selector with diversity and entropy regularizations to choose 256 informative embeddings.
- Passed selected embeddings into the Gigapath slide encoder to generate a representative [CLS] token.
- Performed crossmodal pretraining by reconstructing DNA methylation (DNAm) embeddings, then finetuned on survival with the pretrained slide encoder, a new Top-K selector, and an MLP head.

DNA Methylation (DNAm):
- Generated embeddings with MethylGPT, then aligned them through crossmodal pretraining to reconstruct raw RNA.
- Finetuned on survival with a modality-specific MLP head.

Bulk RNA-seq:
- Encoded bulk RNA-seq data with BulkRNA-BERT, and finetuned on survival with an MLP head.

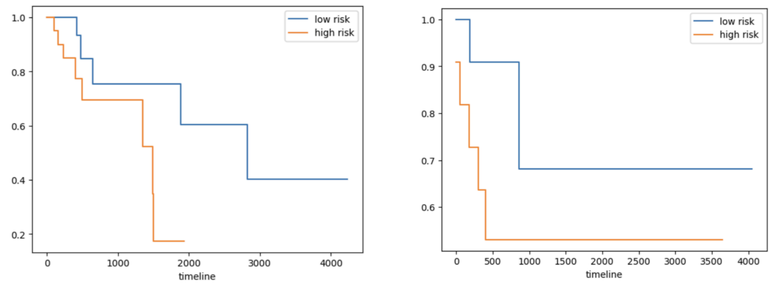
Inferred Spatial Transcriptomics (ST):
- Predicted ST from WSI patches using a UNI-based VRI model, then encoded gene expression patterns with a Graph Attention Network (GAT) and trianed on survival with an MLP head.

Multimodal Fusion:
- Projected all modality embeddings into a matched latent space using projection MLPs, ensuring each projected to the same dimension.
- Combined embeddings across modalities using gated attention and basic concatenation, followed by an MLP head for final survival prediction.

LLM Loss:
- Used GPT-5 API to generate discriptions of whole slide images
- Created a custom LLM loss function that prompts a local LLM with predictions from our model as well as bioligical info (including WSI descriptions) for each patient, and has the LLM assess the accuracy of the model's predictions considering the information given

RESULTS

WSI masking prior to encoding



Kaplan Meier from model without LLM and with respectively

Final test C-index of 0.69, val C-index of 0.72

CONCLUSION & NEXT STEPS

Conclusions
- Built a multimodal pipeline integrating WSI, inferred ST, bulk RNA, and DNAm for survival prediction in colon adenocarcinoma and related cancers.
- Crossmodal pretraining aligned heterogeneous data types and improved survival modeling.
- Multimodal fusion outperformed single-modality models, highlighting the value of integration.
- LLM-powered loss function increased performance all-round

Limitations & Next Steps
- Current evaluation is limited by small cohort size and lack of external validation.
- Fusion via basic concatenation may not fully capture cross-modality interactions.
- Plan to validate on larger, independent datasets and across additional cancer types.
- Explore other fusion methods and improved interpretability methods.
- Implement modality reconstruction for DNAm and bulk RNA
- Conduct further validation on our LLM loss

edit.