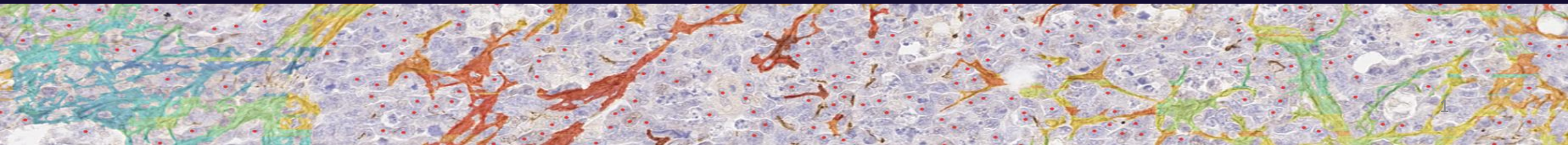
TIA Centre
Tissue Image Analytics

Multimodal Approaches to Computational Pathology

Exploring if radiology-histology fusion leads to enhanced patient outcome predictions

Muhammad Amjad
Seth Alain Chang

24/09/2025



Motivations

- Increase efficiency in pathology workflows
- Allow for more individualised care
- Smarter resource allocation
- Increase survival chance through earlier intervention [1]

[1] Nwaimo CS, Adegbola AE, Adegbola MD. Transforming healthcare with data analytics: Predictive models for patient outcomes. GSC Biol Pharm Sci. 2024;27(3):25-35.

Unimodal Approaches

- Paige Prostate
  - FDA and CE-IVDR clearance
  - Based off work by Campanella et al. which achieved >0.98 AUC and 100% sensitivity [1]
- Virchow
  - **0.95** specimen-level AUROC across 9 common and 7 rare cancers [2]

[1] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Silva VWK, Busam K, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine. 2019 Aug;25(8):1301-9
[2] Vorontsov E, Bozkurt A, Casson A, Shaikovski G, Zelechowski M, Severson K, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine. 2024 Jul 22

Multimodal Approaches

Multimodal approaches often outperform unimodal ones, reaching clinical grade performance in tasks unimodal models struggle to do so in [1].

- Detection and extension in prostate cancer [2]
- Diagnosis of chest based pathological conditions [3]
- Diagnosis of non-pigmented skin cancer
  - Work by Tschandl et al. was described as being “on par with human experts” [4]

[1] Yang H, Yang M, Chen J, Yao G, Zou Q, Jia L. Multimodal deep learning approaches for precision oncology: a comprehensive review. Briefings in Bioinformatics. 2024 Nov 22;26(1).
[2] Sonni I, Felker ER, Lenis AT, Sisk AE, Bahri S, Allen-Auerbach M, et al. Head-to-Head Comparison of ^{68}Ga-PSMA-11 PET/CT and mpMRI with a Histopathology Gold Standard in the Detection, Intraprostatic Localization, and Determination of Local Extension of Primary Prostate Cancer: Results from a Prospective Single-Center Imaging Trial. Journal of Nuclear Medicine. 2021 Oct 14;63(6):847–54.
[3] Khader F, Gustav Müller-Franzes, Wang TS, Han T, Soroosh Tayebi Arasteh, Christoph Haarburger, et al. Multimodal Deep Learning for Integrating Chest Radiographs and Clinical Parameters: A Case for Transformers. Radiology. 2023 Oct 1;309(1).
[4] Tschandl P, Rosendahl C, Akay BN, Argenziano G, Blum A, Braun RP, et al. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. JAMA Dermatology [Internet]. 2019 Jan 1;155(1):58–65. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6439580/

Trends in Existing Work [1]

Multimodal models exhibit an average **6.2% increase in AUC**

Radiology and pathology were the **least combined modalities**

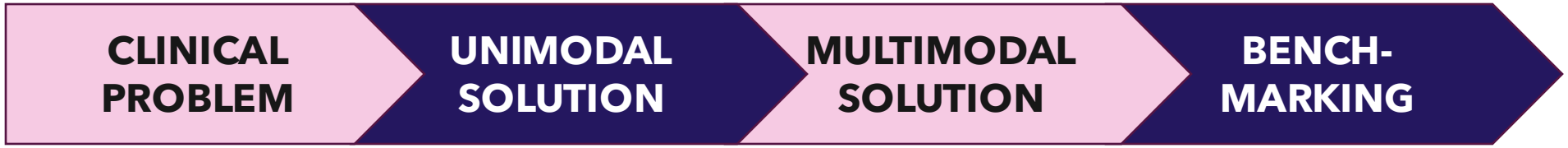
Diagnosis has been the most common task, followed by **survival prediction**

Attention-based fusion and **transformer** methods are on the rise

[1] Schouten D, Nicoletti G, Dille B, Chia C, Vendittelli P, Schuurmans M, et al. Navigating the landscape of multimodal AI in medicine: A scoping review on technical challenges and clinical applications. Medical image analysis [Internet]. 2025;105:103621. Available from: https://pubmed.ncbi.nlm.nih.gov/40482561/

Project Overview

To what extent do radiology-histology multimodal models perform better than unimodal approaches for patient outcome prediction?



Challenges with Data

- Modality incompleteness
- Ethics and privacy concerns
- Lack of fully-annotated datasets

Datasets

[✗] TCGA-GBM

~200 matched cases

New NIH controlled access policy meant it was unavailable

[✗] TCGA-LUAD

Had a large amount of individual histology and radiology data

Had only 31 matched cases

[✓] CHIMERA

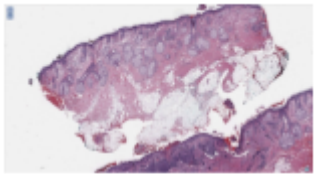
Had 95 cases with radiology, histology and clinical data

Contained 27 positive and 68 negative cases

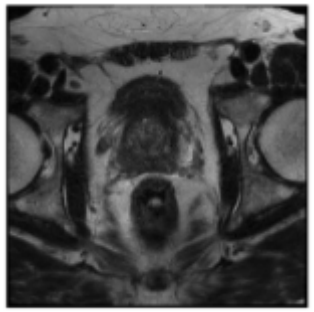
Clinical Task:

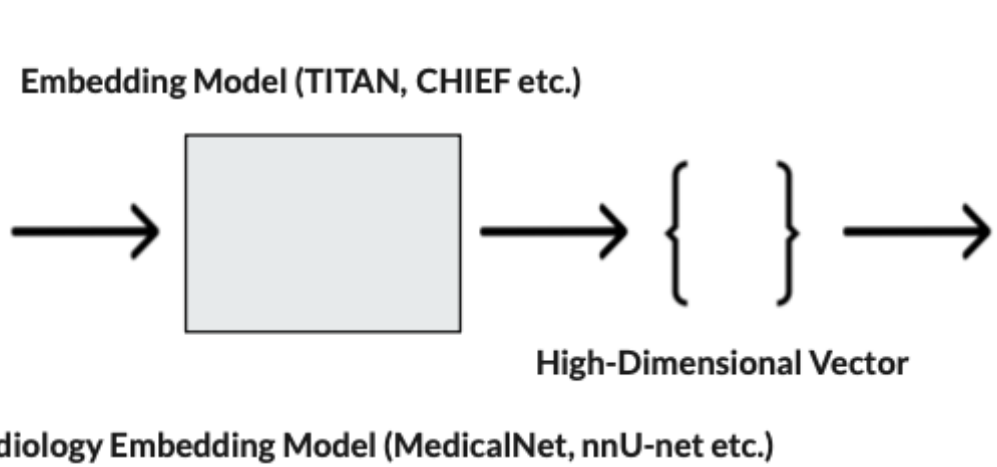
Predict when a patient will undergo biochemical recurrence (BCR) following prostatectomy

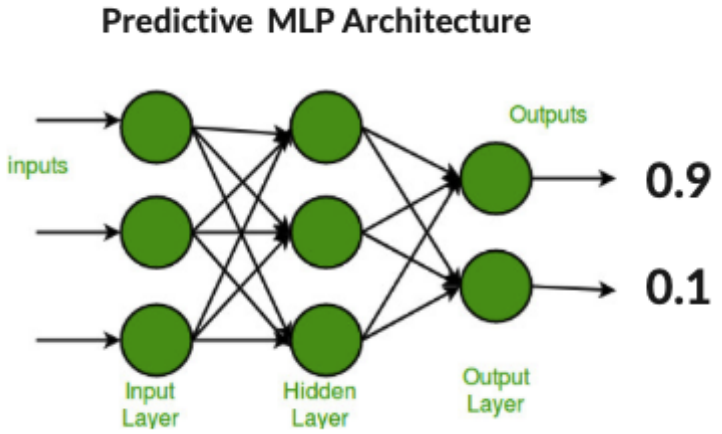
Unimodal Model



WSI Image
or
MRI Image







- Utilised C-Index as our base metric for predicting time to BCR

Considerations with Encoders

- Wanted a general purpose architecture

- MRI scans are 3D in nature
  - Existing models focused on 2D images
  - Adaptation would have required more time

Radiology Encoders

Models used (3D)

- nnU-net
- MedicalNet
- Radiomics

Other models tested (2D)

- ViT Base (pre-trained on ImageNet)
- Qwen FM
- PMC-CLIP

Histology Encoders

Models used (slide level)

- TITAN
- CHIEF
- MADELEINE
- PRISM

Other models tested (patch level)

- UNI
- CONCH

Unimodal Results (Time to BCR)

WSI Configuration	Model	Average C-Index (10 repeats)
10x 256	**PRISM**	**0.7759 ± 0.1088**
	CHIEF	0.7720 ± 0.0944
	TITAN	0.6916 ± 0.1049
	MADELEINE	0.7307 ± 0.1038
10x 512	**PRISM**	**0.7392 ± 0.1095**
	CHIEF	0.7380 ± 0.1121
	TITAN	0.6906 ± 0.1125
	MADELEINE	0.6498 ± 0.1275
20x 256	**PRISM**	**0.8181 ± 0.1018**
	CHIEF	0.6717 ± 0.1240
	TITAN	0.7418 ± 0.1101
	MADELEINE	0.7065 ± 0.1035
20x 512	**PRISM**	**0.7594 ± 0.1100**
	CHIEF	0.6741 ± 0.1140
	TITAN	0.7172 ± 0.1056
	MADELEINE	0.7159 ± 0.1007

Model	Average C-Index
Medical Net	**0.5584 ± 0.1250**
Radiomics	0.5508 ± 0.1318

The Classification Problem

- Initially focused on **classification** of BCR
- Data was **not amenable to classification** due to right censoring
- Led to us focusing on predicting time to BCR instead

Unimodal Results (Classification)

WSI Configuration	Model	Average F1 Score	AUC
10x 256	PRISM	0.5318 ± 0.1544	0.7700 ± 0.0920
	CHIEF	0.5186 ± 0.1798	0.7693 ± 0.1053
	TITAN	**0.5751 ± 0.1742**	**0.7939 ± 0.1106**
	MADELEINE	0.4182 ± 0.1620	0.6892 ± 0.1201
10x 512	PRISM	0.4278 ± 0.1688	0.6754 ± 0.1234
	CHIEF	0.5264 ± 0.1467	0.7792 ± 0.0886
	TITAN	0.5468 ± 0.1783	0.7618 ± 0.1360
	MADELEINE	**0.5563 ± 0.1582**	**0.8098 ± 0.1117**
20x 256	PRISM	0.5334 ± 0.1652	0.7509 ± 0.1230
	CHIEF	**0.5713 ± 0.1397**	**0.8307 ± 0.0802**
	TITAN	0.5607 ± 0.1704	0.7828 ± 0.1193
	MADELEINE	0.5547 ± 0.1638	0.7744 ± 0.1127
20x 512	PRISM	0.5306 ± 0.1773	0.7586 ± 0.1129
	CHIEF	0.4900 ± 0.1533	0.7243 ± 0.0900
	TITAN	**0.5350 ± 0.1981**	**0.8044 ± 0.1068**
	MADELEINE	0.4571 ± 0.1740	0.6828 ± 0.1332

Fusion Methodologies

1. Early Fusion

2. **Intermediate Fusion**
   - Marginal Intermediate Fusion
   - Joint Intermediate Fusion

3. Late Fusion

[1]

[1] Yang H, Yang M, Chen J, Yao G, Zou Q, Jia L. Multimodal deep learning approaches for precision oncology: a comprehensive review. Briefings in Bioinformatics. 2024 Nov 22;26(1).

Fusion Methodologies

Early Fusion

- Combining images before using encoders to get vector representations

Late Fusion

- Having two separate prediction pipelines for each modality and using a weighted average for a final prediction

Intermediate Fusion

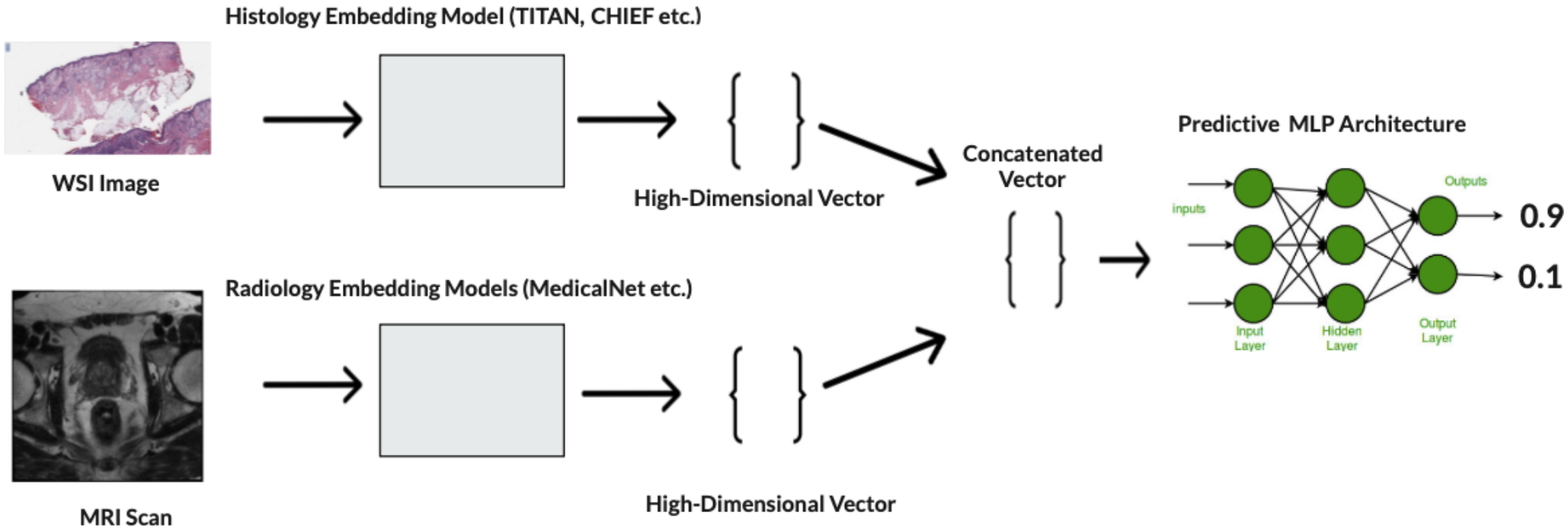
Marginal Intermediate Fusion

- Getting vector representations of each modality and combining them with **no additional learning**

Joint Intermediate Fusion

- Getting vector representations of each modality and combining them with **additional learning**

Marginal Intermediate Fusion



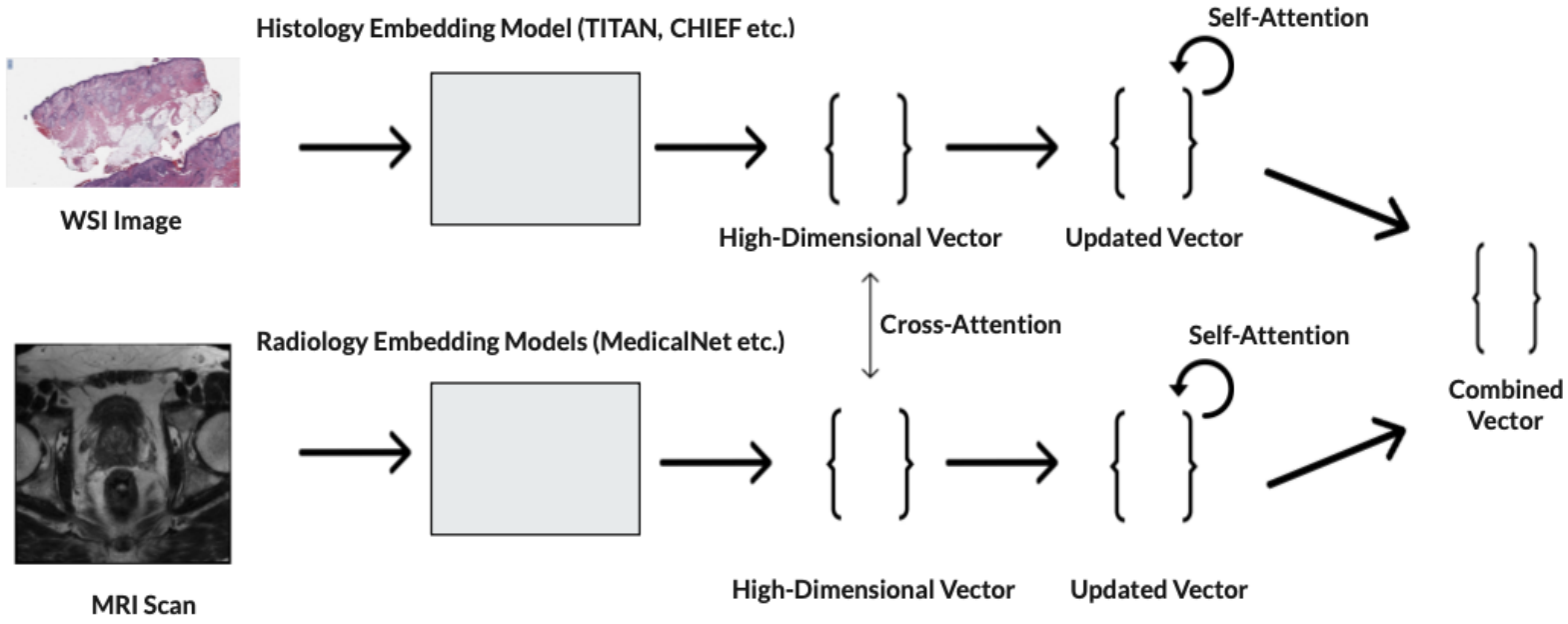
- Option to linearly project vectors such that dimensions of both vectors are equal before concatenation

Marginal Intermediate Fusion Results

Linear Projection	Average C-Index (10 repeats)
Yes	0.7520 ± 0.1069
No	0.7501 ± 0.1166

Unimodal Model	Average C-Index (10 repeats)
Histology (PRISM)	0.8181 ± 0.1018
Radiology (MedicalNet)	0.5584 ± 0.1250

Joint Intermediate Fusion



- Run cross-attention to allow histology and radiology vectors to give each other context
- Run self-attention to allow individual vectors to learn most important components for predictions

Joint Intermediate Fusion

- Instead of concatenation, we can calculate a **bilinear product** of both vectors to capture multiplicative relationships
- Use **n learnt weight matrices**, where n is the dimension of the combined vector

$$\{\ \}_{\text{Histology Vector, H}} \times \mathbf{A} \times \{\ \}_{\text{Radiology Vector, R}} = \text{Component 1 of combined vector}$$

$$\{\ \}_{\text{Histology Vector, H}} \times \mathbf{B} \times \{\ \}_{\text{Radiology Vector, R}} = \text{Component 2 of combined vector}$$

[1] Efficient Bilinear Attention-based Fusion for Medical Visual Question Answering [Internet]. Arxiv.org. 2016 [cited 2025 Sep 23]. Available from: https://arxiv.org/html/2410.21000v2

Joint Intermediate Fusion Results

Vector Combination Method	Average C-Index (10 repeats)
Concatenation	0.7420 ± 0.1114
Bilinear Product	-

Unimodal Model	Average C-Index (10 repeats)
Histology (PRISM)	0.8181 ± 0.1018
Radiology (MedicalNet)	0.5584 ± 0.1250

Discussion

- **Radiology features alone** were **not a good predictor** for time to BCR, with an average C-Index score of 0.55 across 2 encoders

- Both Marginal and Joint Intermediate Fusion **did not yield better results** compared to unimodal approaches, with histology alone being a better predictor

- Radiology **introduced noise** to histology features, causing multimodal approaches to **perform worse**

Conclusion

- Despite radiology's clinical relevance, its weak predictive power (C-index = 0.55) **degraded the performance** of histology-based models rather than providing complementary information

- Incorporating modalities with poor individual predictive performance in multimodal approaches can be **counterproductive**

Challenges and Lessons Learned

- Data availability and access
- Scaling within training and validation sets
- Flawed classification task
- Ensuring minimal data leakage
  - Patient-level vs slide-level set splitting

Future Research Directions

- Utilisation of 2D models for MRIs
- Experiment with patch-level WSI foundation models
- Using larger datasets
- Experiment with other fusion strategies
  - Late, Early
- Adapt the framework to other clinical tasks
  - Diagnosis, survival etc.
- Revisit the classification task with different data

Thank You

TIA Centre

TISSUE IMAGE ANALYTICS