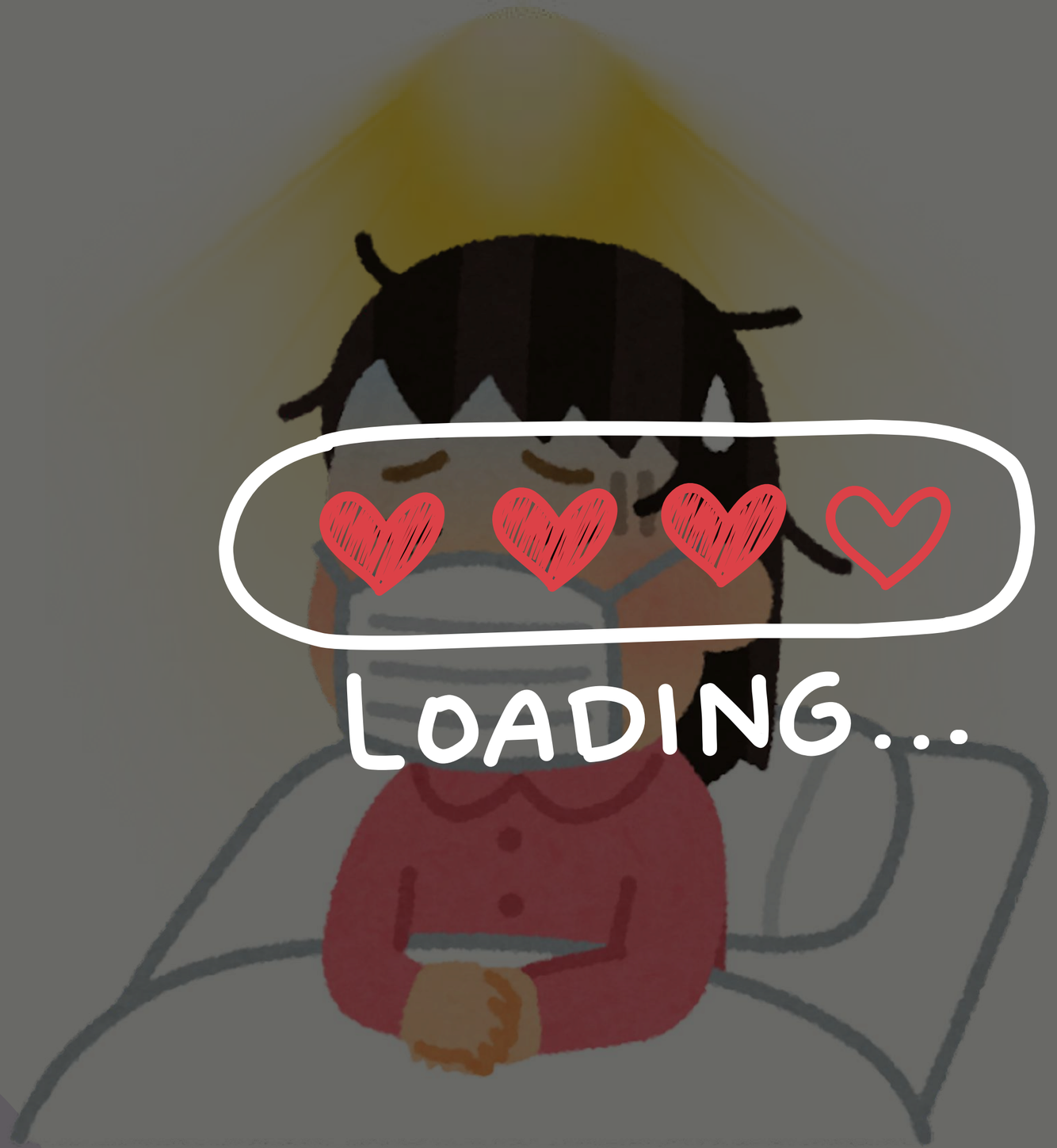
CANCER

Cancer Survival Prediction Using Foundation Models and MIL

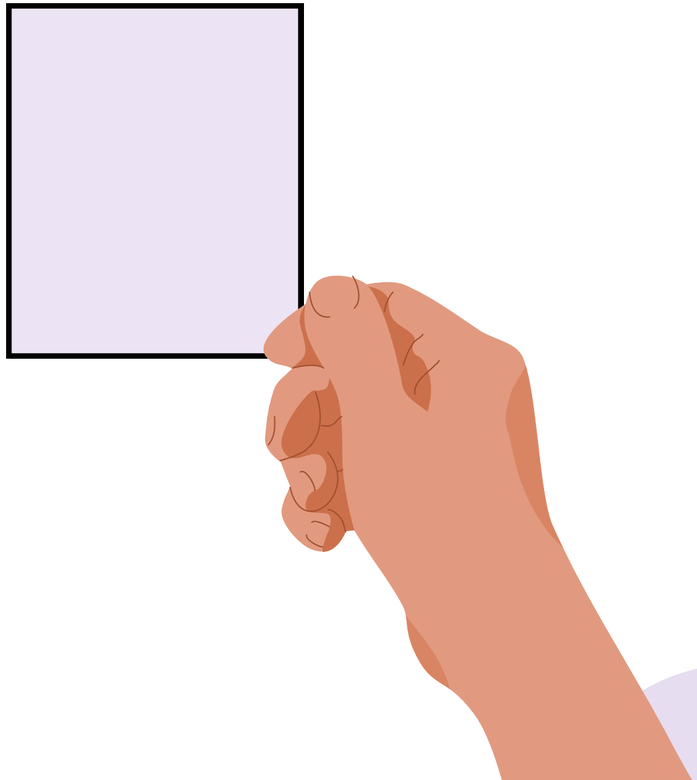
Project 8

SEAS 2026

31/7/2026

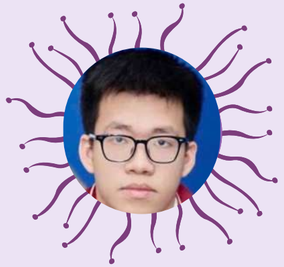


LOADING...



CANCER DETECTION!!!!

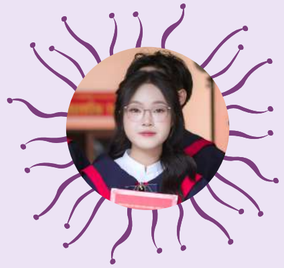
Ngô Đức Quý



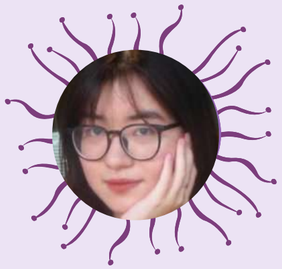
Nguyễn Thanh Minh

Trần Xuân Phú

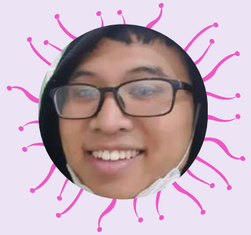
Nguyễn Thức Công Thành



Trần Thanh Hảo



Nguyễn Vũ Ánh Ngọc (mentor)



Nguyễn Cảnh Hoàng (Mentor)

Nguyễn Phương Trang (Mentor)

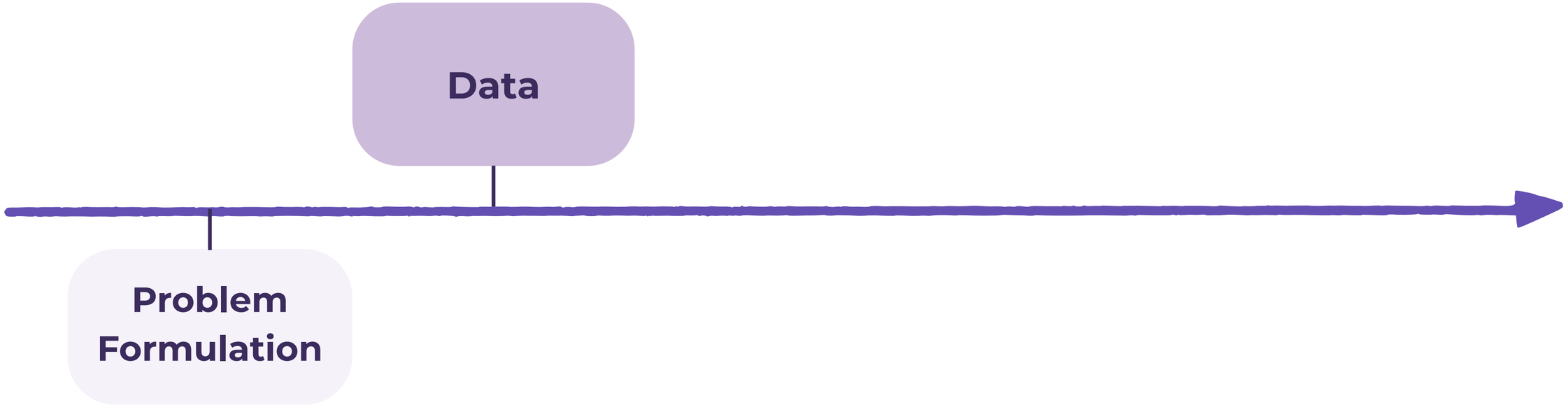
OUTLINE

OUTLINE

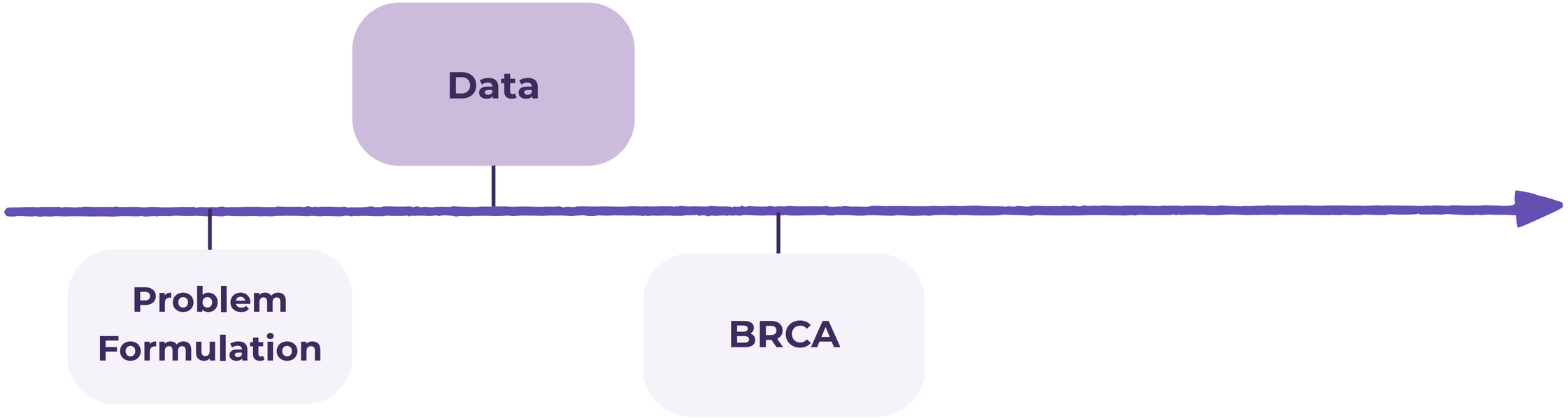
OUTLINE

Problem Formulation

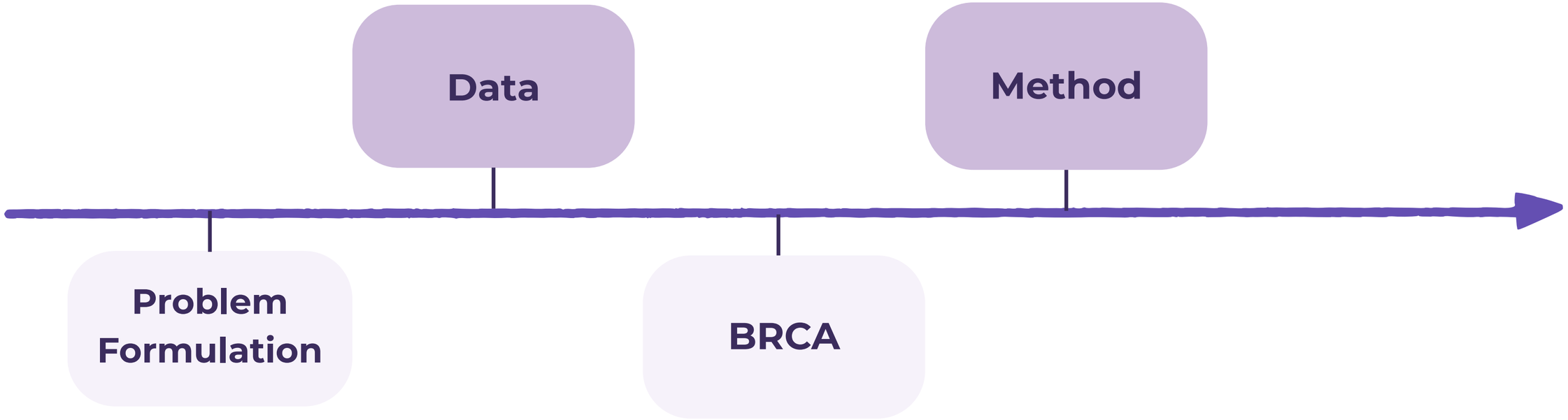
OUTLINE



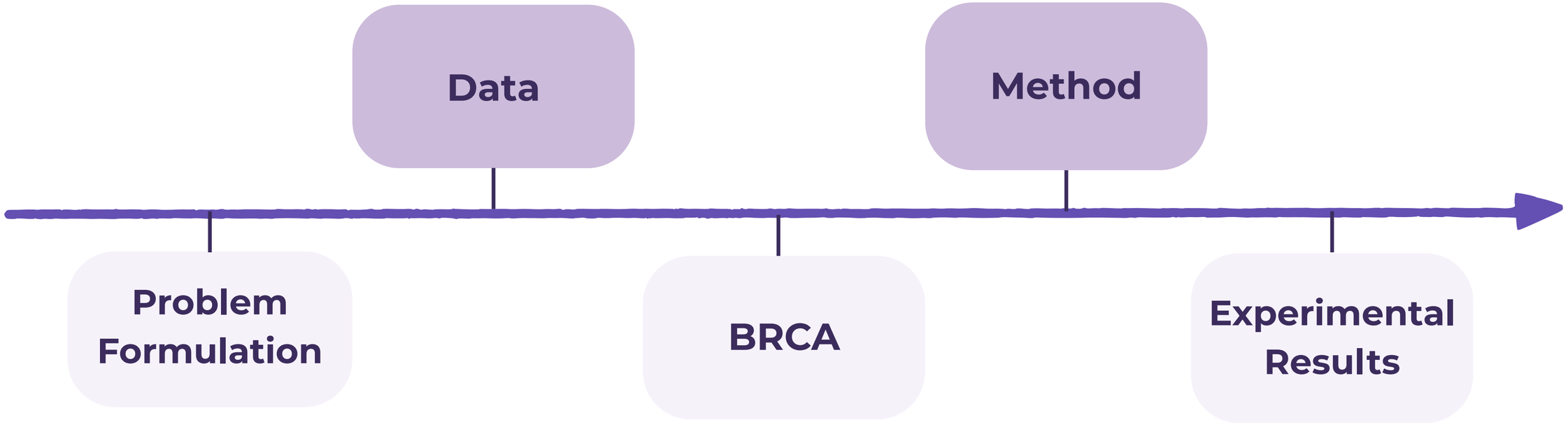
OUTLINE



OUTLINE



OUTLINE


Data
Method
Problem Formulation
BRCA
Experimental Results

PROBLEM FORMULATION

CHALLENGES

CHALLENGES

1. Gigapixel Whole-Slide Images

- Extremely large image size (gigapixel resolution).
- Direct processing requires substantial computational resources.

CHALLENGES

1. Gigapixel Whole-Slide Images

- Extremely large image size (gigapixel resolution).
- Direct processing requires substantial computational resources.

2. Limited and Difficult-to-Obtain Data

- Survival datasets are relatively small.
- Collecting high-quality WSI data with long-term follow-up is expensive and time-consuming.

CHALLENGES

1. Gigapixel Whole-Slide Images

- Extremely large image size (gigapixel resolution).
- Direct processing requires substantial computational resources.

2. Limited and Difficult-to-Obtain Data

- Survival datasets are relatively small.
- Collecting high-quality WSI data with long-term follow-up is expensive and time-consuming.

3. High Computational Cost

- Training on gigapixel images demands significant GPU memory and computational power.
- Efficient feature extraction and aggregation methods are essential.

MOTIVATION

MOTIVATION

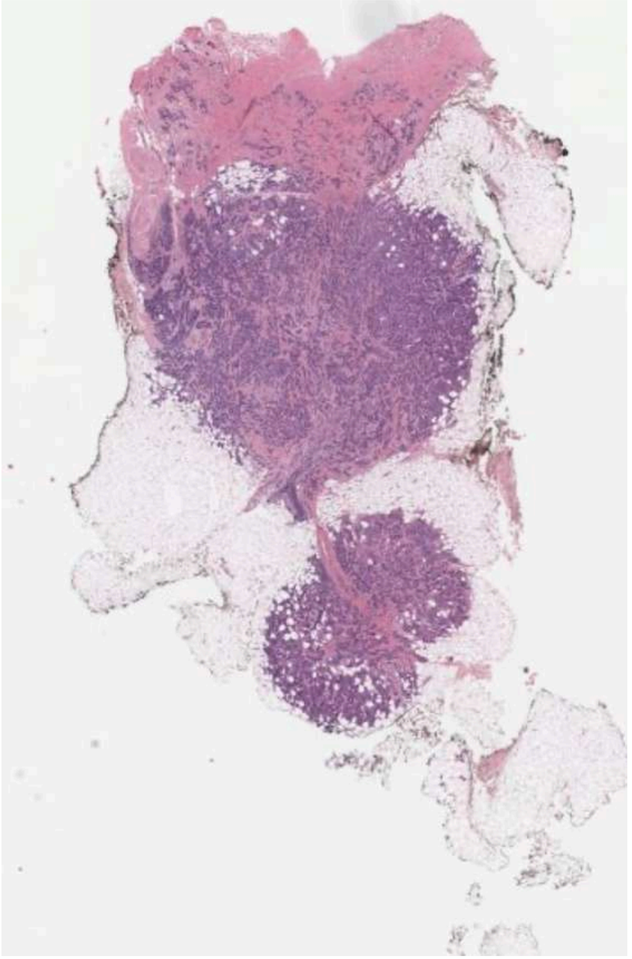
Clinical

1. Objective and Standardised Clinical Assessment
2. Accurate Patient Risk Stratification
3. Personalised Treatment Strategies
4. Improved Healthcare Efficiency

DATA

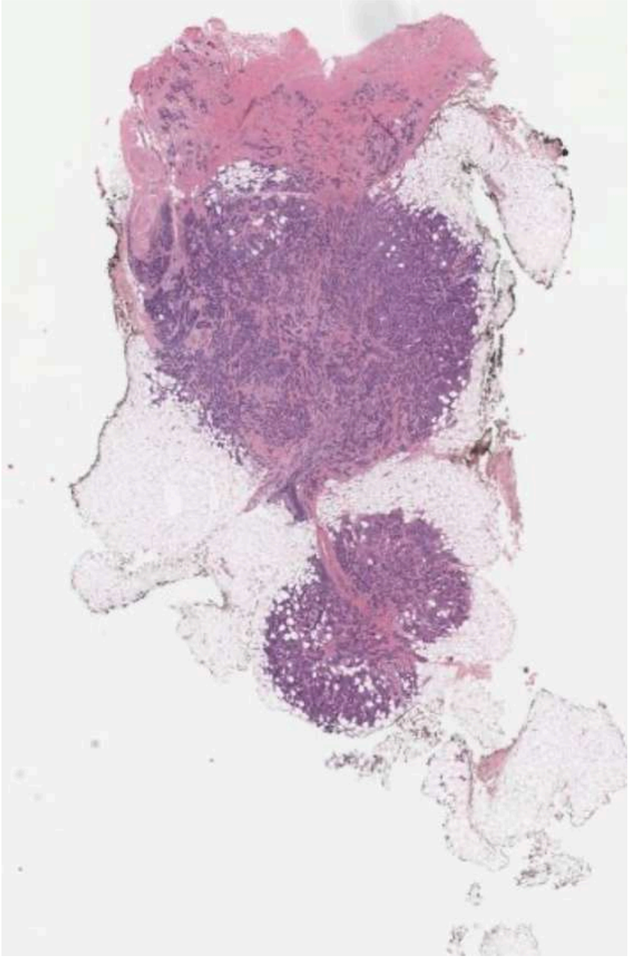
WSI-H&E

WSI-H&E



WSI of a breast cancer patient - BRCA

WSI-H&E

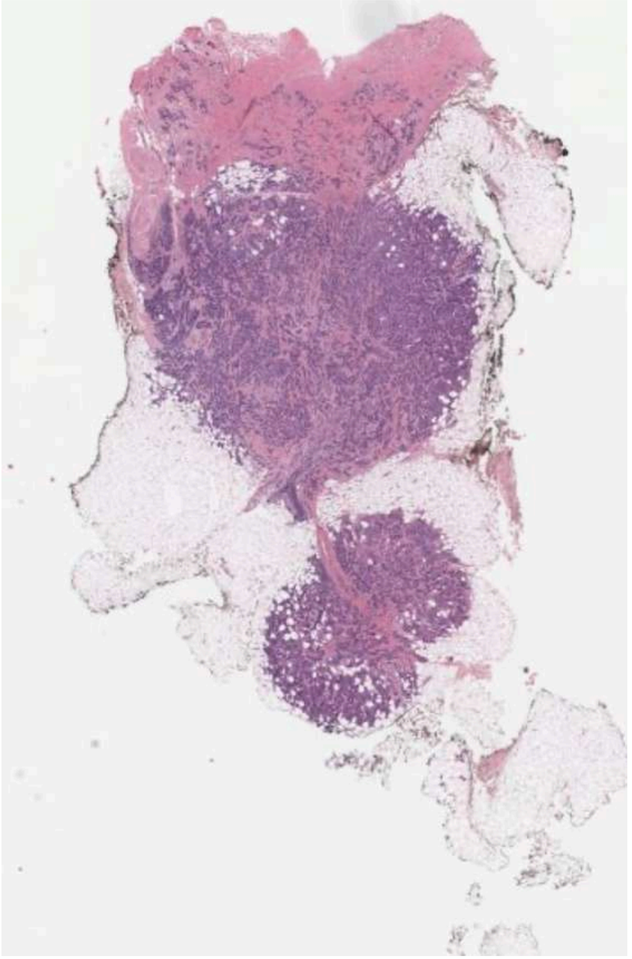


WSI of a breast cancer patient - BRCA

H&E

A conventional staining technique using Hematoxylin and Eosin to highlight tissue structures.

WSI-H&E



WSI of a breast cancer patient - BRCA

H&E

A conventional staining technique using Hematoxylin and Eosin to highlight tissue structures.

WSI

A high-resolution digital image generated by scanning an H&E-stained histopathology slide.

BREAST INVASIVE CARCINOMA (BRCA)

BREAST INVASIVE CARCINOMA (BRCA)

BREAST INVASIVE CARCINOMA (BRCA)

Cre: vietnamnet.vn

METHOD

Attention-based Multiple Instance Learning (ABMIL)

Why we use MIL instead of Conventional Fully Supervised Training?

Attention-based Multiple Instance Learning (ABMIL)

Why we use MIL instead of Conventional Fully Supervised Training?

1. A whole-slide image consists of thousands of patches, but survival labels are only available at the patient level → weakly supervised problem.

Attention-based Multiple Instance Learning (ABMIL)

Why we use MIL instead of Conventional Fully Supervised Training?

1. A whole-slide image consists of thousands of patches, but survival labels are only available at the patient level → weakly supervised problem.
2. When we use MIL, we can overcomes the lack of fine-grained labels.

$$Y_i = \begin{cases} 0, & \text{if } \sum_j y_{i,j} = 0, \\ 1, & \text{else }, \end{cases}$$

Attention-based Multiple Instance Learning (ABMIL)

Attention-based Multiple Instance Learning (ABMIL)

Why we use ABMIL instead Basic MIL?

Attention-based Multiple Instance Learning (ABMIL)

Why we use ABMIL instead Basic MIL?

1. Basic MIL uses rigid pooling methods (such as mean or max pooling), which assume all patches contribute equally or focus on a single isolated point.

Attention-based Multiple Instance Learning (ABMIL)

Why we use ABMIL instead Basic MIL?

1. Basic MIL uses rigid pooling methods (such as mean or max pooling), which assume all patches contribute equally or focus on a single isolated point.

2. ABMIL assigns weights to the patches to focus maximally on critical malignant tissue regions and suppress noise from normal tissues.

Attention-based Multiple Instance Learning (ABMIL)

Why we use ABMIL instead Basic MIL?

1. Basic MIL uses rigid pooling methods (such as mean or max pooling), which assume all patches contribute equally or focus on a single isolated point.

2. ABMIL assigns weights to the patches to focus maximally on critical malignant tissue regions and suppress noise from normal tissues.

Attention weights formula

$$a_i \propto \exp\left(w^T\left(\tanh(Vh_i) \odot \sigma(Uh_i)\right)\right)$$

Patient Embedding formula

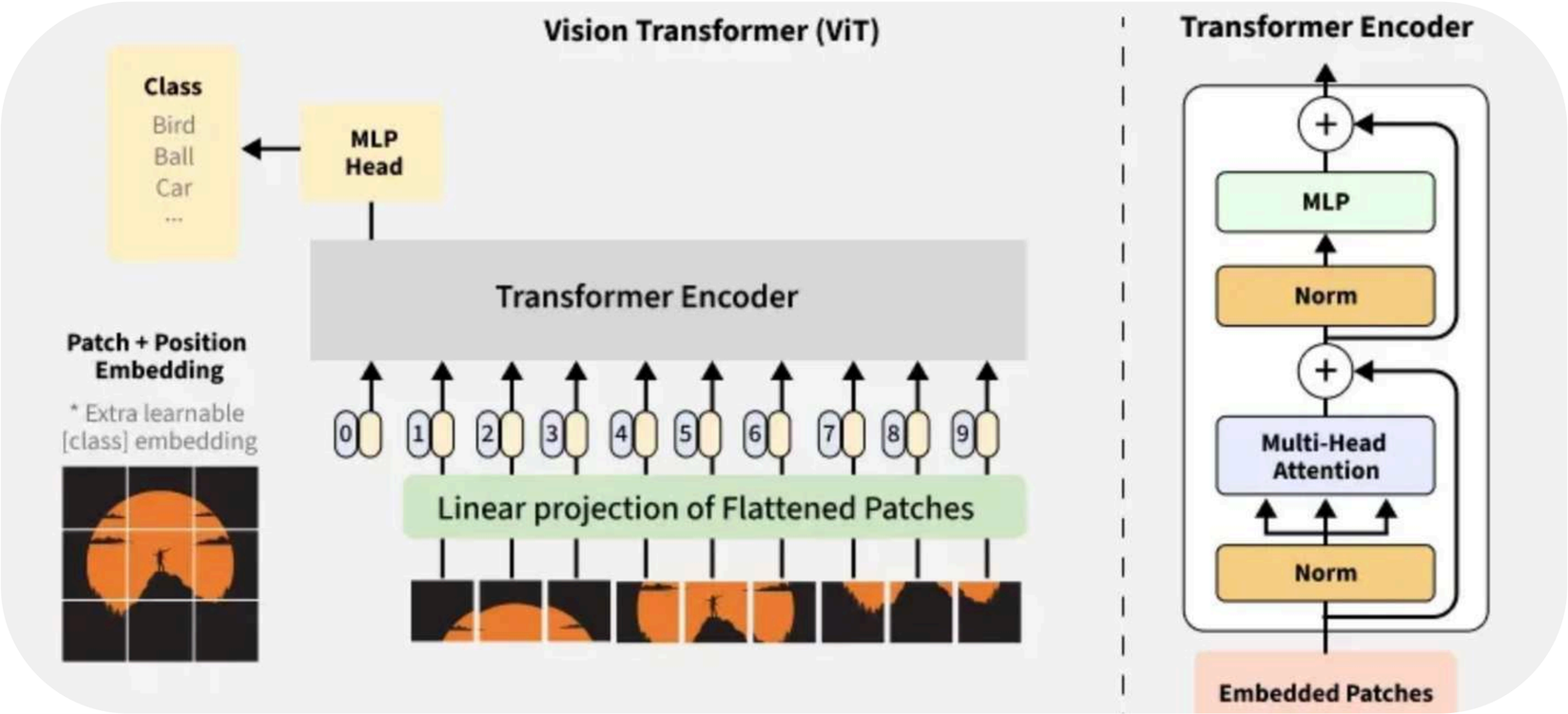
$$M = \sum_i a_i h_i$$

UNI2-H

UNI2-H

What is UNI2-H?

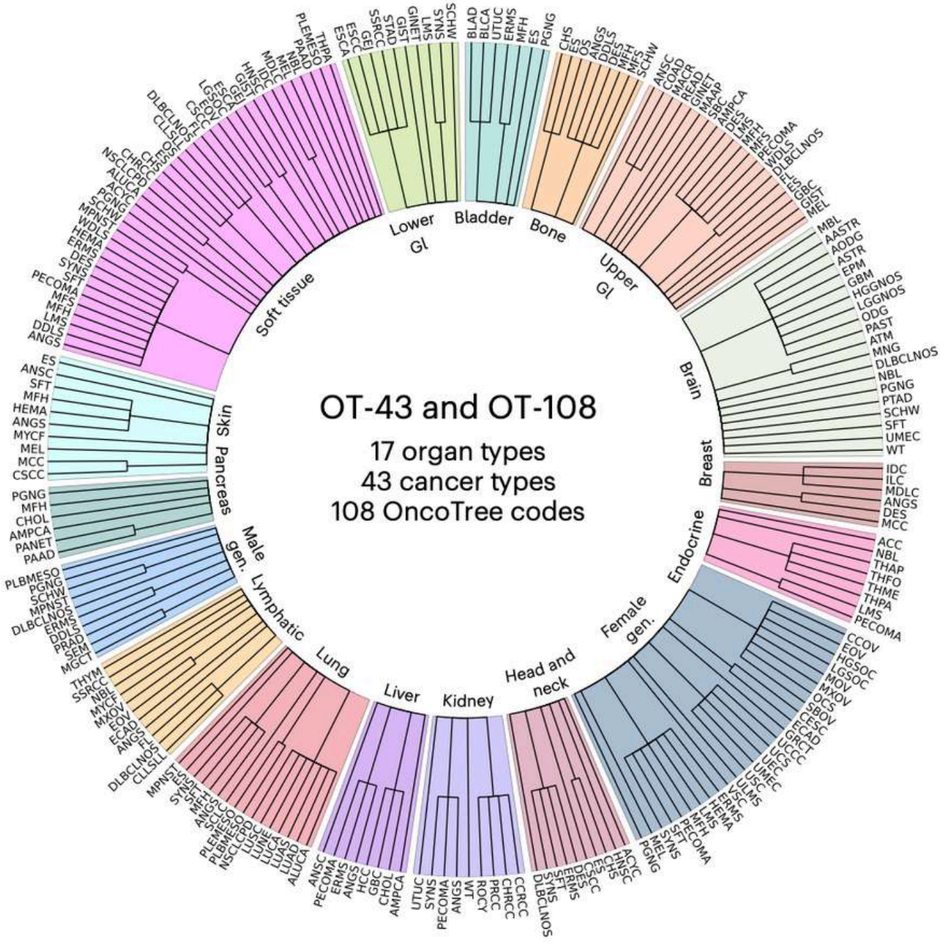
- General-purpose foundation model for computational pathology.
- Trained using self-supervised learning.
- Built upon the Vision Transformer Large (ViT-Large) architecture.
- Transforms histopathology image patches into high-quality feature embeddings.



UNI2-H

Why UNI2-H?

- **Large-scale pretraining**: Trained on the Mass-100K dataset, containing 100M+ histopathology image patches extracted from 100,000+ H&E whole-slide images (77+ TB) across 20 major organ types.
- **Advanced self-supervised learning**: Built on DINOv2, combining self-distillation and masked image modeling
- **Outstanding performance**: Evaluated on 34 clinical tasks, consistently outperforming previous state-of-the-art
- **Strong generalization**: Accurately recognizes 108 cancer types in the OncoTree taxonomy
- **Efficient fine-tuning:** Transfers pretrained pathology knowledge to downstream survival prediction, enabling strong performance with limited labeled datasets.

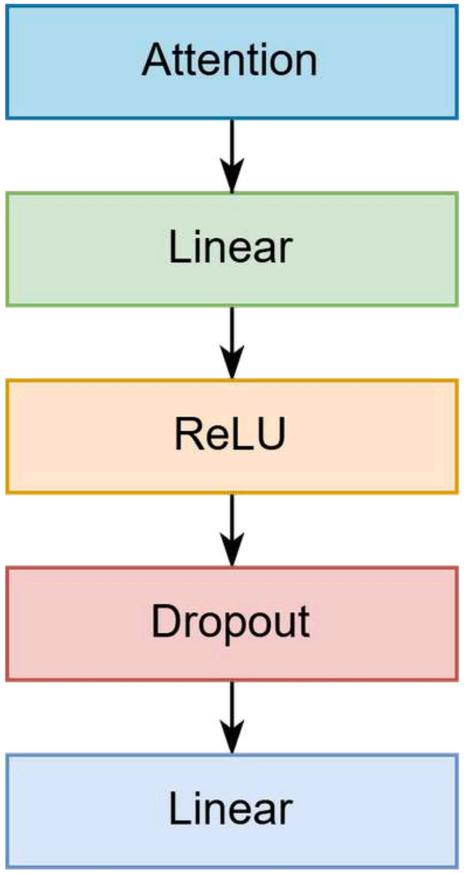


SURVIVAL PREDICTION HEAD

(Risk Score)

SURVIVAL PREDICTION HEAD
(Risk Score)

The Survival Prediction Head is the final prediction model of the network



SURVIVAL PREDICTION HEAD
(Risk Score)

Function

Transforms the slide-level representation generated by ABMIL into a continuous patient-specific risk score, which reflects the patient's relative risk of experiencing the clinical event (e.g., death).

TRAINING MODEL USING TIME & EVENT

TRAINING MODEL USING TIME & EVENT

Cox Partial Log-Likelihood Loss function

$$L = -\frac{1}{N_{\text{events}}} \sum_{i:E_i=1} \left(r_i - \log \left(\sum_{j \in R(T_i)} \exp(r_j) \right) \right)$$

Concordance Index (C-index) formula

$$\text{C-index} = \frac{\sum_{i,j} I(T_i < T_j) \cdot I(r_i > r_j) \cdot E_i}{\sum_{i,j} I(T_i < T_j) \cdot E_i}$$

Inputs

- WSI-H&E
- Survival Time
- Event Status

Optimization

- Learns relative patient risk
- Handles censored observations
- Updates model parameters through backpropagation

Output

- Relative risk score

DATASET

TCGA

TCGA is a global cancer genome mapping project conducted by the U.S. National Institutes of Health (NIH) to classify and collect detailed data on dozens of different types of cancer.

BRCA

BRCA is a subset of the TCGA dataset that focuses specifically on invasive breast cancer, one of the most common types of cancer.

EXPERIMENT SETUP

Dataset

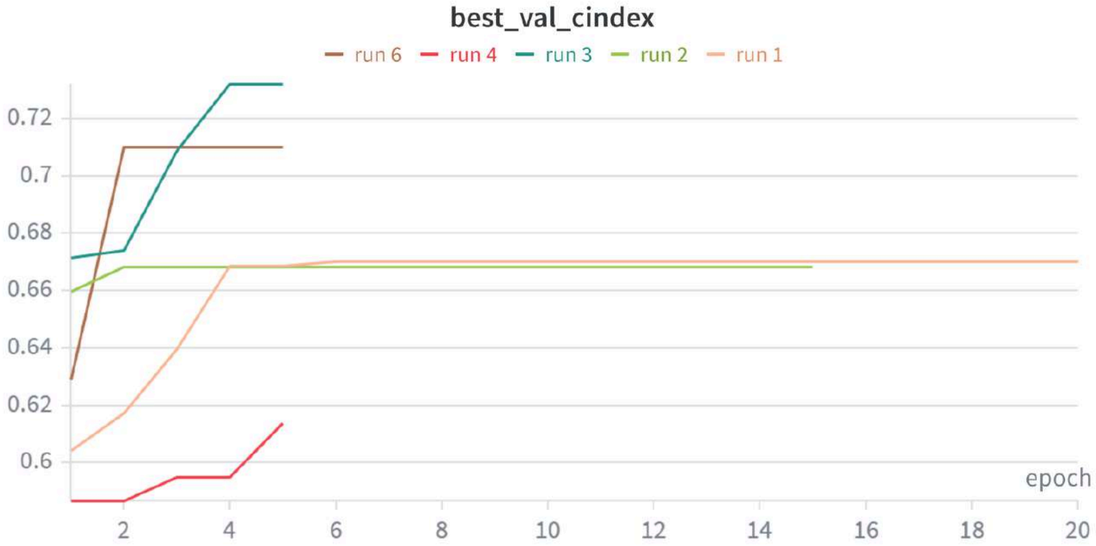
Train set	Validation set
150	38

Hyperparameters

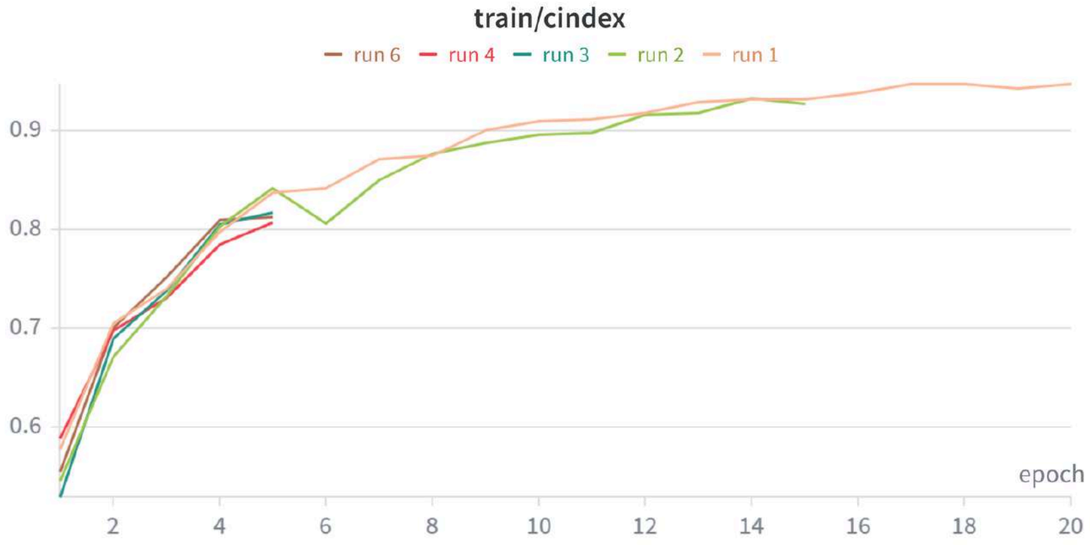
Batch size	Epochs	Drop out rate
4	10-20	0.25

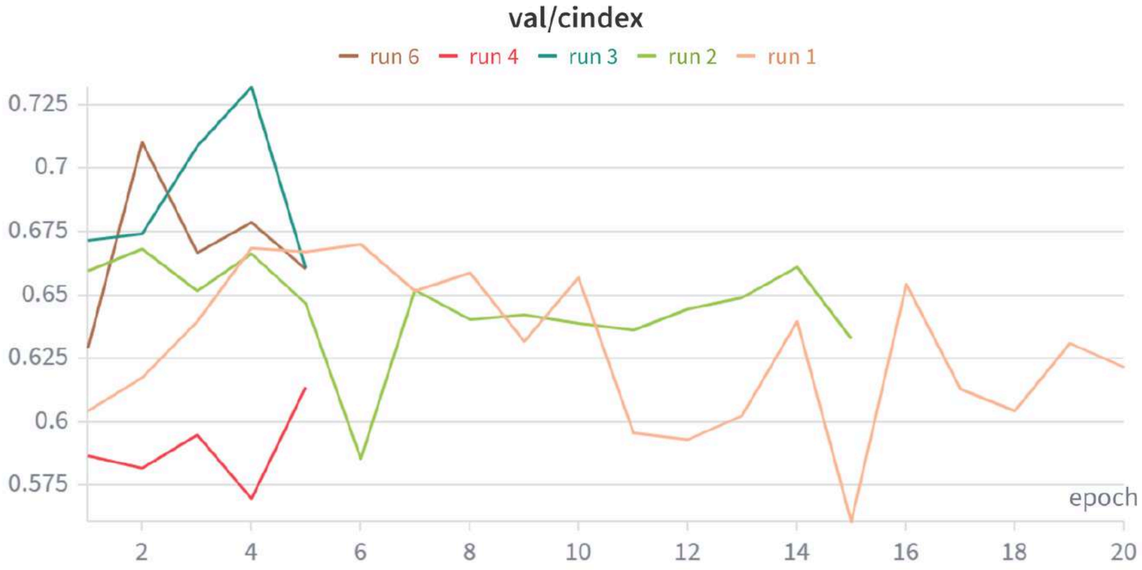
EXPERIMENTAL RESULTS

QUANTITATIVE RESULTS

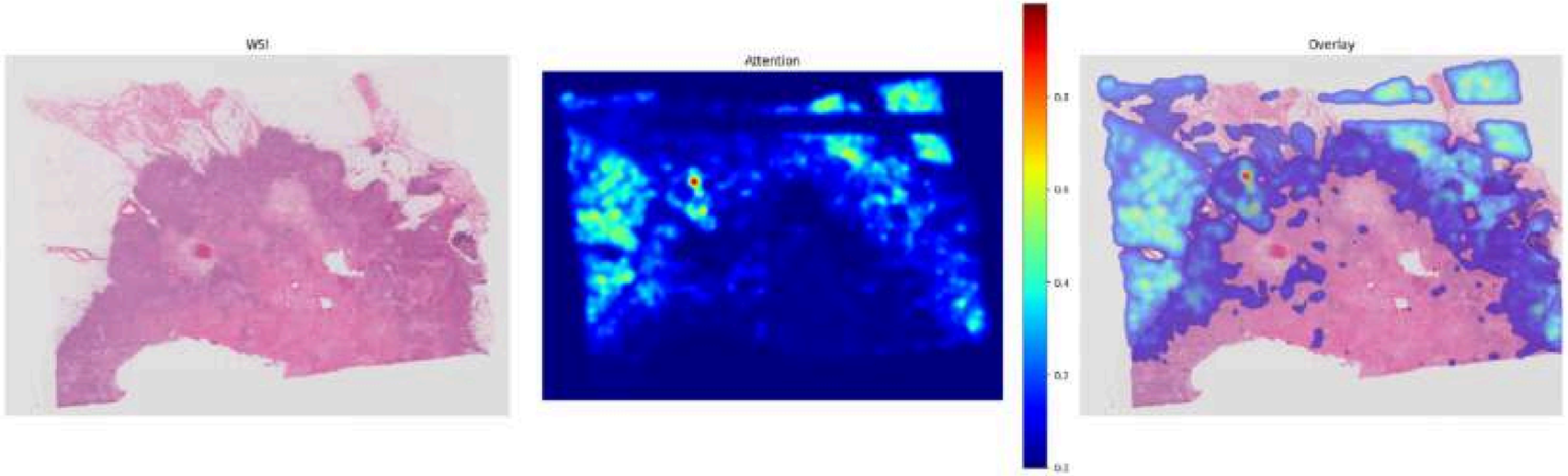


QUANTITATIVE RESULTS

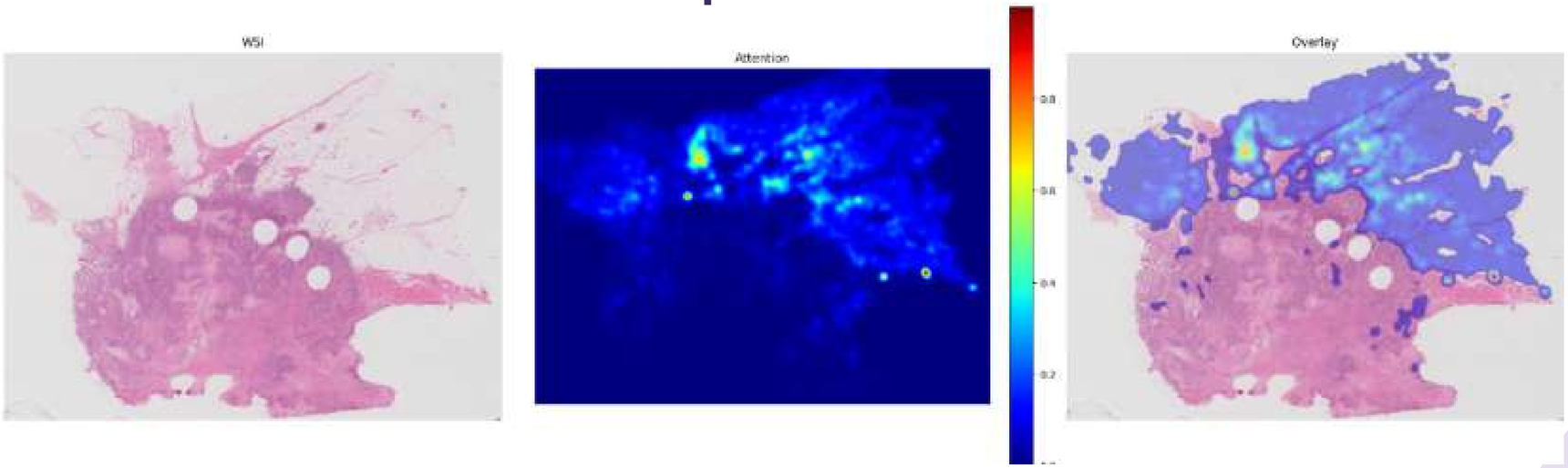




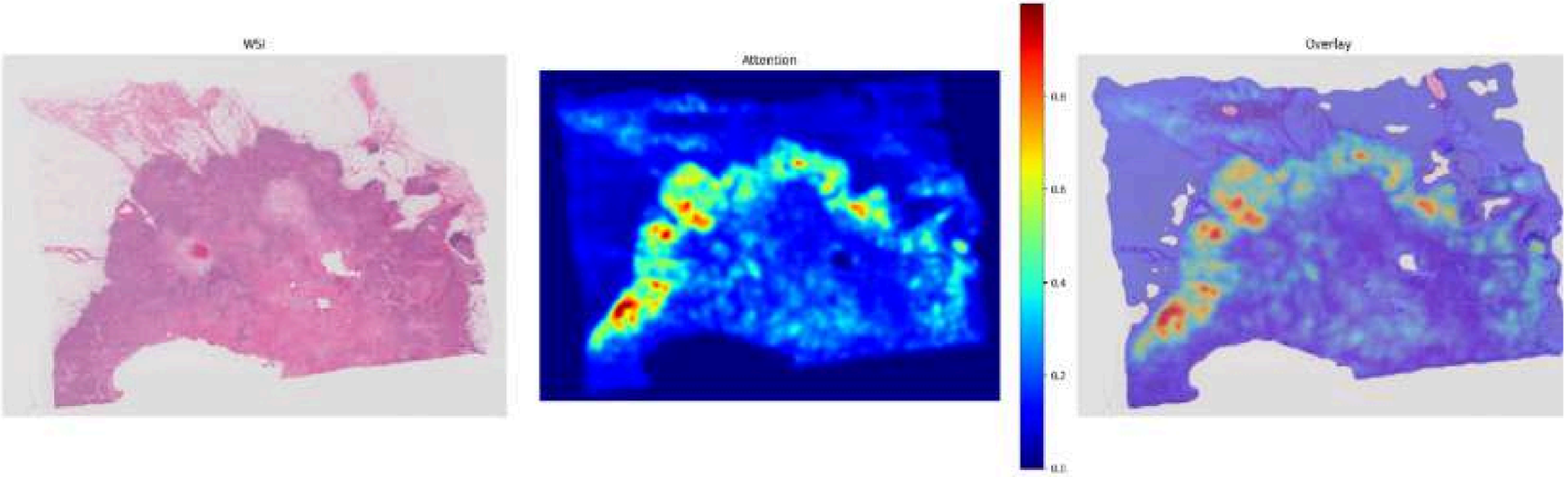
QUALITATIVE RESULTS



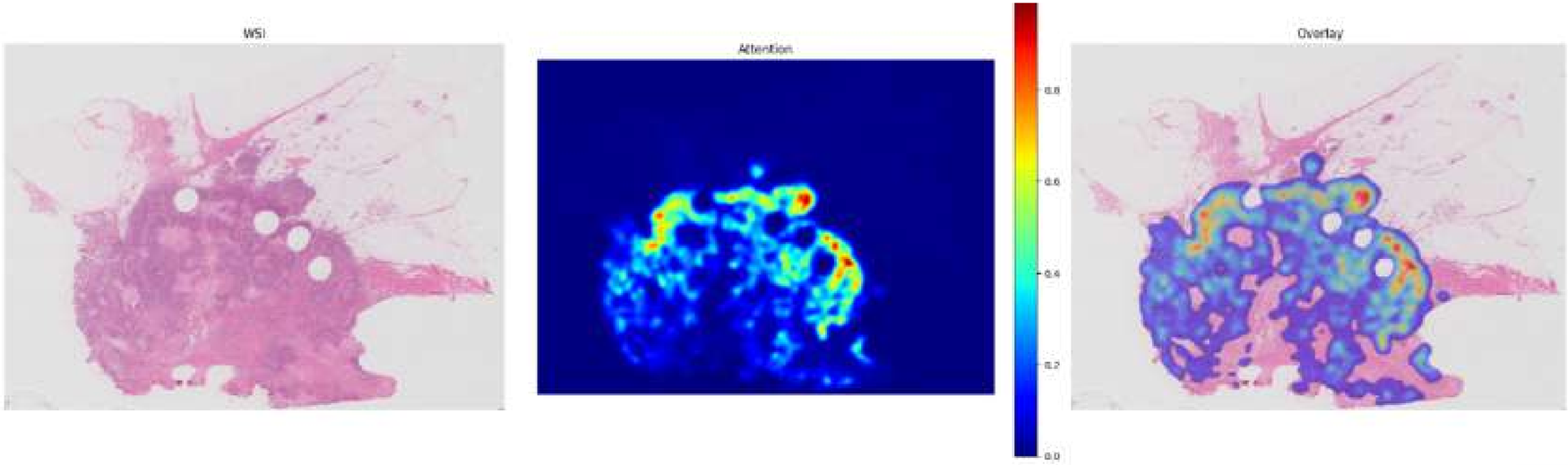
Epoch 20



QUALITATIVE RESULTS



Epoch 10



CONCLUSION

Whole-slide histopathology images provide rich morphological information for cancer prognosis

UNI2-H extracts robust and transferable histopathological features from H&E image patches.

Attention-Based Multiple Instance Learning aggregates patch-level features into an informative slide-level representation under weak supervision.

Survival analysis with Cox Partial Likelihood Loss enables accurate patient risk stratification while handling censored observations

.

T**he proposed pipeline** supports AI-assisted prognosis prediction and has the potential to improve personalized treatment planning

THANK YOU

REFERENCES

1. Mỗi năm Việt Nam ghi nhận hơn 25.000 ca mắc mới ung thư vú
2. Data
3. DINOv2: Learning Robust Visual Features without Supervision
4. Attention-based Deep Multiple Instance Learning
5. Towards a general-purpose foundation model for computational pathology