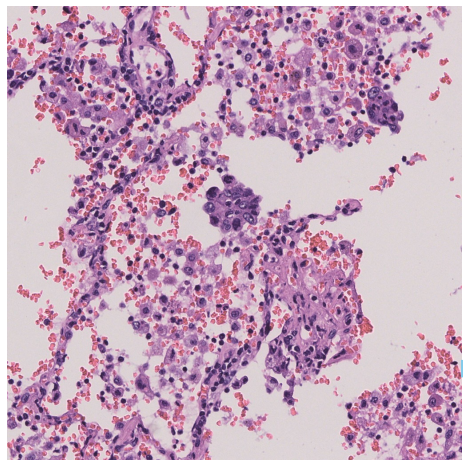


Visual Branch



Linear Projection

Visual Transformer

Visual Encoder

Expression Branch

Tissue heterogeneity with loosely arranged cells and localized cell infiltration.

Text Encoder

Expression Features

LLaVA-Med

Prompt: Could you explain the pathology-specific terms focusing on their visuals in slides?

Knowledge Features

Text Encoder

Knowledge Branch

- **Tissue heterogeneity:** Tissue heterogeneity appears in pathological images as significant differences in cell morphology, density, and arrangement across different regions.
- **Loosely arranged cells:** Loosely arranged cells typically show larger spaces between them, with irregular alignment and a lack of tight connections.
- **Localized cell infiltration:** Localized cell infiltration manifests as a dense aggregation of cells in specific areas, with irregular boundaries.

Cross-modal Fusion Module(CFM)

Prediction Head

Visual-Language Transformer

[REG] token

Linear Projection

$\times 2$

MLP

Norm

Multi-Head Attention

Norm

Knowledge Fusion Module(KFM)

Text Embedding

❄️ : Frozen

Ⓢ : Concatenation

⊕ : Element-wise Addition

