$\mathcal{FOCUS}$: Knowledge-enhanced Adaptive Visual Compression for Few-shot Whole Slide Image Classification

Supplementary Material

A. Pathology Prior Knowledge Prompt

This section lists the language prompt generated by LLMs for each dataset, including TCGA-NSCLC, CAMELYON, and UBC-OCEAN (see Section A.1, A.2, and A.3). For TCGA-NSCLC and CAMELYON, Claude-3.5-Sonnet is used for prior knowledge generation. While for UBC-OCEAN, Claude-3.5-Sonnet, ChatGPT3.5-Turbo, ChatGPT-4o, OpenAI-o1-mini, and Llama3.1-405B are used, as mentioned in the main paper. Moreover, in Section A.4, we include a detailed analysis of different LLMs' prompts for potential reasons that lead to the performance ranking shown in the ablation studies of the main paper (Section 4.3.3).

A.1. TCGA-NSCLC

Lung adenocarcinoma (LUAD): A whole slide image of lung adenocarcinoma at high resolution with visually descriptive characteristics of clear cytoplasm, round or oval nuclei, prominent nucleoli, rich vascularity, irregular blood vessels, intratumoral septa, and heterogeneity.
Lung squamous cell carcinoma (LUSC): A whole slide image of lung squamous cell carcinoma at high resolution with visually descriptive characteristics of squamous cell differentiation, round structures with eosinophilic cytoplasm, distinct cell borders and abundant cytoplasm, enlarged nuclei, irregular nuclear shape, increased chromatin density.

A.2. CAMELYON

Normal: Normal lymph node, high resolution: Uniform small lymphocytes densely packed. Well-formed follicles in cortex, lymphocyte cords in medulla. Thin-walled blood vessels throughout. No atypical cells or architectural distortions.
Breast cancer metastases: Metastatic lymph node, high resolution: Large, pleomorphic cancer cells interspersed in lymphoid tissue. Cells show irregular nuclei, prominent nucleoli, abundant cytoplasm. Atypical cell arrangements. Desmoplasia, increased mitoses, potential necrosis. Abnormal blood vessels present.

A.3. UBC-OCEAN

A.3.1 Claude-3.5-Sonnet

Clear-cell ovarian carcinoma: A whole slide image of clear-cell ovarian carcinoma at high resolution reveals large polygonal cells with clear cytoplasm and distinct cell boundaries. Nuclei are hyperchromatic and pleomorphic. Characteristic "hobnail" cells with nuclei bulging into cystic spaces are visible.
High-grade serous carcinoma: A whole slide image of high-grade serous carcinoma at high resolution displays marked nuclear atypia with high nuclear-to-cytoplasmic ratio. Nuclei are highly pleomorphic with prominent nucleoli. Numerous mitotic figures and areas of necrosis are present.
Low-grade serous carcinoma: A whole slide image of low-grade serous carcinoma at high resolution exhibits uniform population of cells with mild to moderate nuclear atypia. Psammoma bodies (calcified concretions) are often visible.
Endometrioid carcinoma: A whole slide image of endometrioid carcinoma at high resolution shows back-to-back glands lined by columnar cells with stratified nuclei. Squamous differentiation may be present. Nuclear atypia varies with tumor grade.
Mucinous carcinoma: A whole slide image of mucinous carcinoma at high resolution exhibits tall columnar cells with abundant intracellular mucin, often with goblet cell features. Nuclei are usually basally located. Degree of nuclear atypia varies with tumor grade.

A.3.2 ChatGPT-3.5Turbo

Clear-cell ovarian carcinoma: Clear-cell ovarian carcinoma typically presents with hobnail cells containing clear cytoplasm and distinct cell borders.
High-grade serous carcinoma: High-grade serous carcinoma displays complex papillary structures lined by atypical cells with high nuclear to cytoplasmic ratio and prominent nucleoli.
Low-grade serous carcinoma: Low-grade serous carcinoma demonstrates small glands or papillae lined by cells with low-grade nuclear features and scant cytoplasm.
Endometrioid carcinoma: Endometrioid carcinoma exhibits glandular differentiation with stratification, nuclear atypia, and architectural complexity resembling endometrial tissue.
Mucinous carcinoma: Mucinous carcinoma is characterized by pools of extracellular mucin and tall columnar cells with basally located nuclei and apical mucin.

A.3.3 ChatGPT-4o

Clear-cell ovarian carcinoma: Clear-cell ovarian carcinoma: Characterized by large, clear cells with abundant cytoplasm and distinct cell borders, often forming tubulocystic, papillary, or solid patterns.

High-grade serous carcinoma: High-grade serous carcinoma: Noted for its high-grade nuclear atypia, frequent mitotic figures, and complex papillary structures with slit-like spaces.
Low-grade serous carcinoma: Low-grade serous carcinoma: Features small, uniform cells with mild nuclear atypia, forming papillary and micropapillary structures with less mitotic activity.
Endometrioid carcinoma: Endometrioid carcinoma: Resembles endometrial tissue with glandular patterns, often with squamous differentiation and varying degrees of nuclear atypia.
Mucinous carcinoma: Mucinous carcinoma: Composed of tall columnar cells with abundant intracellular mucin, forming glandular or cystic structures with varying degrees of atypia.

A.3.4 OpenAI-o1-mini

Clear-cell ovarian carcinoma: Clear-cell ovarian carcinoma: Characterized by cells with clear cytoplasm and distinct cell borders, often exhibiting hobnail nuclei and sharply defined cell outlines under high-resolution imaging.
High-grade serous carcinoma: High-grade serous carcinoma: Features densely packed, irregularly shaped cells with prominent nucleoli, high nuclear-to-cytoplasmic ratios, frequent mitotic figures, and abundant psammoma bodies on high-resolution analysis.
Low-grade serous carcinoma: Low-grade serous carcinoma: Presents with uniform, mildly atypical cells arranged in papillae or glands, minimal mitotic activity, and a more cohesive architecture when examined at high resolution.
Endometrioid carcinoma: Endometrioid carcinoma: Displays glandular structures resembling endometrial tissue, with tubular formations, moderate cytologic atypia, and occasional squamous differentiation under high-magnification views.
Mucinous carcinoma: Mucinous carcinoma: Identified by large mucin-filled vacuoles within the cells, abundant extracellular mucin, and expansive, often multiloculated glandular structures observable in high-resolution microscopy.

A.3.5 Llama3.1-405B

Clear-cell ovarian carcinoma: Clear-cell ovarian carcinoma: Characterized by hobnail or clear cells with abundant clear cytoplasm, often with a tubulocystic or solid growth pattern and a network of delicate, branching vessels.
High-grade serous carcinoma: High-grade serous carcinoma: Exhibits a complex, branching architecture with slit-like spaces and a high nuclear-to-cytoplasmic ratio, featuring large, pleomorphic nuclei with prominent nucleoli and frequent mitoses.
Low-grade serous carcinoma: Low-grade serous carcinoma: Displays a predominantly micropapillary or cribriform growth pattern with small, uniform nuclei and a low mitotic rate, often with a characteristic microinvasive pattern of invasion.
Endometrioid carcinoma: Endometrioid carcinoma: Comprises glandular structures resembling endometrial glands, with columnar or cuboidal cells having oval nuclei and a variable amount of eosinophilic cytoplasm, often with squamous differentiation.
Mucinous carcinoma: Mucinous carcinoma: Characterized by large, mucin-filled glands or cysts lined by tall, columnar cells with basally located nuclei and a prominent apical mucin droplet, often with a characteristic intestinal-type epithelial differentiation.

A.4. Prompt Analysis for Different LLMs

As shown in Fig. 3, the superior performance of Claude-3.5-Sonnet (86.4% Balanced ACC) can be attributed to its comprehensive and clinically precise descriptions. Its prompts consistently balance cellular, architectural, and nuclear features while maintaining a clear focus on primary diagnostic criteria. For example, in describing clear-cell ovarian carcinoma, it precisely details both cellular characteristics ("large polygonal cells with clear cytoplasm") and distinctive features ("hobnail cells with nuclei bulging into cystic spaces"), providing rich, diagnostically relevant information that helps the model identify key discriminative features even with limited training examples.

ChatGPT3.5-Turbo achieved the second-best performance (84.8%) with its more concise yet accurate descriptions. While less detailed than Claude's prompts, it successfully captures the essential diagnostic features of each cancer subtype. Its descriptions maintain clinical accuracy while focusing on primary diagnostic features, though they show slightly less consistency in description depth across different subtypes. This balance between conciseness and accuracy appears to provide sufficient guidance for the few-shot learning task.

OpenAI-o1-mini's prompts (84.6%) strike a reasonable balance between detail and clarity, explicitly contextualizing features within high-resolution imaging. However, its descriptions occasionally include less clinically relevant details and tend to be more verbose without adding substantive diagnostic information. This might explain why its performance, while strong, falls slightly below that of more focused prompts.

The lower performance of ChatGPT-4o can be attributed to its overly brief descriptions that sometimes miss critical diagnostic features. Its prompts, while accurate, lack the depth and specificity needed for optimal few-shot learning. The brevity of its descriptions might not provide sufficient guidance for the model to distinguish between similar can-

Algorithm 1 LAViC-MIL: Language-guided Adaptive Visual Compression Multi-instance Learning

Require: Training set of whole slide images $\{X^l\}_{l=1}^L$ with labels $\{Y_l\}_{l=1}^L$ ($Y_l \in \{1, 2, \ldots, S\}$), pathology description prompts $\mathbf{T^P}$, window size w, base similarity threshold θ_{base}, compression ratio γ, maximum sequence length M_{max}, maximum epochs E_{max}

Ensure: Optimal model parameters Θ^* of LAViC-MIL that minimize $\mathcal{L} = -\log P(Y_l|\mathbf{B}_c^l, \mathbf{T})$

Initialize model parameters Θ
for epoch $e = 1$ to E_{max} **do**
  for each WSI X^l in training set **do**
    Segment patches $X^l = \{x_1^l, ..., x_{N_l}^l\}$ from tissue regions
    Extract patch features via pathology foundation models $\mathbf{B}^l = \{f(x_i^l) : x_i^l \in X^l\} = \{\mathbf{b}_i^l\}_{i=1}^{N_l}$
    $\mathbf{T} = [\mathbf{T^L}; \mathbf{T^P}]$ ▷ Concatenate learnable and pathology prompts
    // Stage 1: Global Redundancy Removal via Pathology FMs
    for each window of size w in $\mathbf{B}^l$ **do**
      $\hat{\mathbf{b}}_i^l = \mathbf{b}_i^l / \|\mathbf{b}_i^l\|_2$ ▷ Normalize features
      $S_{ij} = \hat{\mathbf{b}}_i^l \cdot \hat{\mathbf{b}}_j^l$ ▷ Pairwise similarities
      $\tau_g = \mu(S) + \sigma(S)$ ▷ Dynamic threshold
      $R_i = \frac{1}{w} \sum_{j=1}^{w} S_{ij}$ ▷ Mean similarity
    end for
    Remove patches where $R_i > \tau_g$
    // Stage 2: Language-guided Visual Token Prioritization
    $\mathbf{A} = \text{softmax}(\frac{(\mathbf{T}W_q)(\mathbf{B}^l W_k)^\top}{\sqrt{d}}), \quad r_i = \frac{1}{t_1+t_2} \sum_{j=1}^{t_1+t_2} \mathbf{A}_{ji}$ ▷ Token relevance
    $k = \min(M_{max}, \gamma N_l')$ ▷ Number of tokens to select
    $\mathbf{B}_s^l = \{b_i^l | \text{rank}(r_i) \leq k\}$ ▷ Select top-k tokens
    // Stage 3: Sequential Visual Token Compression
    for stage i with threshold $\theta_i = \theta_{base} + i\Delta\theta$ **do**
      $s_{j,j+1} = \frac{\mathbf{b}_j^{l\top} \mathbf{b}_{j+1}^l}{\|\mathbf{b}_j^l\|_2 \cdot \|\mathbf{b}_{j+1}^l\|_2}$ ▷ Cosine similarity
      $\text{mask}_j = \begin{cases} 1, & \text{if } \min(s_{j-1,j}, s_{j,j+1}) < \theta_i \\ 0, & \text{otherwise} \end{cases}$
      $\mathbf{B}_{i+1}^l = \{\mathbf{b}_j^l \in \mathbf{B}_i^l : \text{mask}_j = 1\}$
    end for
    // After Compression: Cross-modal Aggregation and Loss Computation
    $\text{Head}_i = \text{softmax}(\frac{\mathbf{Q}W_q^i(\mathbf{K}W_k^i)^\top}{\sqrt{d_k}})\mathbf{V}W_v^i$
    $\mathbf{O} = \text{LayerNorm}(\text{Concat}(\text{Head}_1, ..., \text{Head}_h)W_o)$
    $P(Y_l|\mathbf{B}_c^l, \mathbf{T}) = \text{softmax}(W_c\mathbf{O} + \beta_c)$
    $\mathcal{L}_{\text{CE}} = -\log P(Y_l|\mathbf{B}_c^l, \mathbf{T})$
    Update parameters Θ using gradient descent on $\mathcal{L}_{\text{CE}}$ via AdamW optimizer
  end for
end for
return Optimal model parameters Θ^*

cer subtypes, particularly in cases where subtle differences are diagnostically important.

Llama3.1-405B's prompts showed the lowest performance, likely due to their tendency to mix critical and non-critical features while using overly complex language. Its descriptions often include rare or less reliable diagnostic features and lack consistent structure across different cancer subtypes. For example, its inclusion of features like "characteristic microinvasive pattern of invasion" might distract from more reliable primary diagnostic criteria, potentially confusing the model in the few-shot learning context.

Intuitions and guidelines: This analysis suggests that optimal prompts for few-shot pathology image classification should be comprehensive yet focused, maintaining a consistent structure while emphasizing established diagnostic criteria. The balance achieved by Claude-3.5-Sonnet between detail and clinical relevance appears to provide the most effective guidance for the model to learn discriminative features from limited training examples.