Prompting Vision Foundation Models for Pathology Image Analysis

Chong Yin[1], Siqi Liu[2], Kaiyang Zhou[1], Vincent Wai-Sun Wong[3], Pong C. Yuen[1]
[1]Department of Computer Science, Hong Kong Baptist University, Hong Kong
[2]Shenzhen Research Institute of Big Data, Chinese University of Hong Kong, Shenzhen
[3]Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong
{chongyin, kyzhou, pcyuen}@comp.hkbu.edu.hk, siqiliu@sribd.cn, wongv@cuhk.edu.hk

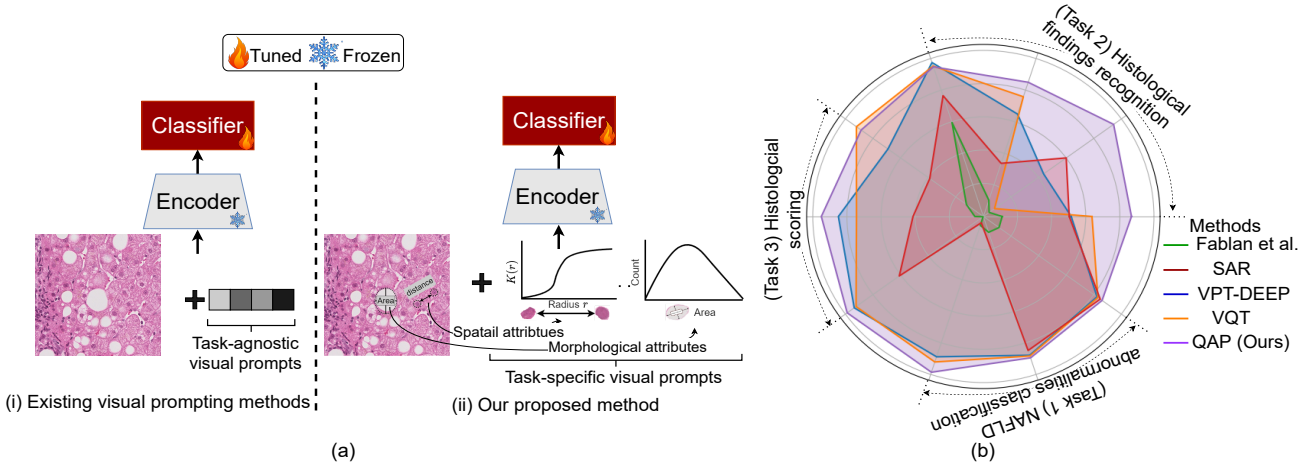


Figure 1. (a) The illustration of task-agnostic visual prompts and the proposed task-specific visual prompts. (b) Performance overview of liver pathology image analysis tasks with various methods (i.e., Fabian et al. [16], SAR [40], VPT-DEEP [18], VQT [37]). Our method (QAP) excels in overall performance. Detailed results are provided in Table 1 2 3.

Abstract

The rapid increase in cases of non-alcoholic fatty liver disease (NAFLD) in recent years has raised significant public concern. Accurately identifying tissue alteration regions is crucial for the diagnosis of NAFLD, but this task presents challenges in pathology image analysis, particularly with small-scale datasets. Recently, the paradigm shift from full fine-tuning to prompting in adapting vision foundation models has offered a new perspective for small-scale data analysis. However, existing prompting methods based on task-agnostic prompts are mainly developed for generic image recognition, which fall short in providing instructive cues for complex pathology images. In this paper, we propose Quantitative Attribute-based Prompting (QAP), a novel prompting method specifically for liver pathology image analysis. QAP is based on two quantitative attributes, namely K-function-based spatial attributes and histogram-based morphological attributes, which are aimed for quantitative assessment of tissue states. Moreover, a conditional prompt generator is designed to turn these instance-specific attributes into visual prompts. Extensive experiments on three diverse tasks demonstrate that our task-specific prompting method achieves better diagnostic performance as well as better interpretability. Code is available at https://github.com/7LFB/QAP.

1. Introduction

Non-alcoholic fatty liver disease (NAFLD) poses a pervasive global health challenge, impacting approximately 30% of adults [36, 39]. Traditional diagnostic process for NAFLD is labor-intensive as it relies on manual examination of specific histological findings, including steatosis, lobular inflammation, and ballooning degeneration, as well as histological scoring [20, 35]. With the increasing global prevalence of NAFLD, there is a concern that medical systems may face an excessive burden in terms of resources required for diagnosis. Furthermore, manual pathological scoring may suffer from inadequate reproducibility, even

when performed by expert pathologists [19, 25].

In recent years, deep learning-based methods have shown promising potential in automatic pathology image analysis [2, 41], offering fast, consistent, and more cost-effective results. However, deploying these methods in real-world scenarios is highly challenging, which is largely due to the limited availability of large-scale datasets.

To overcome the aforementioned issue, this paper resorts to a new data-efficient paradigm called prompting [4], which has recently emerged in computer vision [18, 42, 43]. The main idea of prompting is to tune a small number of parameters in a model's input space while keeping the majority of the pre-trained part untouched, which has been proven to be effective in handling low-data regimes [18, 42, 43]. However, existing prompting methods based on task-agnostic prompts are primarily developed for natural images, which differ significantly from pathology images. Natural images typically represent scenes and objects, whereas pathology images capture microscopic views of tissue samples or cells, each with their unique structures. The use of task-agnostic prompts alone may not be sufficient in providing clear and instructive cues that are necessary for fully comprehending the complex microenvironment in pathology images.

Our key insight is that visual prompts should include the elements that pathologists examine when analyzing pathological images. We first start with two fundamental questions. **What tissue structures to study?** Inspired by [1], nuclei and white regions are two crucial tissue structures for diagnosing NAFLD. These white regions may specifically represent fatty cells or vessels. Pathologists carefully inspect these structures and measure their related attributes to assist in making precise diagnoses. **What attributes of tissue structures to study?** Pathologists primarily investigate two categories of attributes: spatial and morphological. In terms of spatial attributes, the arrangement of nuclei around fatty cells may indicate the degree of fat accumulation and damage in liver cells [39]. The assessment of morphological damage also holds significant importance in clinical practice. For example, steatosis is characterized by the presence of multiple lipid droplets of varying sizes that occupy the cytoplasm and displace the nucleus [5]. Visual prompts that contain pathology-related information play a crucial role in helping models recognize and interpret important features and patterns. Instead of relying on task-agnostic visual prompts, designing task-specific visual prompts with quantitative attributes, such as spatial and morphological attributes, can be more effective, as demonstrated in Figure 1.

This paper proposes a quantitative attribute-based prompting (QAP), a novel method for liver pathology image analysis. Specifically, QAP explores two types of attributes including K-function-based spatial attributes and histogram-based morphological attributes. These attributes offer a comprehensive and precise quantitative evaluation of the images based on their clustering and shape characteristics. Furthermore, we introduces an attribute-conditioned prompt generator. By conditioning the generation of prompts on these attributes, we can guide the model to generate more relevant and precise prompts that are tailored to the specific characteristics of each pathology image.

In summary, we make the following contributions in this paper:

- We conduct a comprehensive investigation into prompting methods for pathology image analysis, and propose a new visual prompting method based on pathology-related features.
- We provide an efficient implementation of the new method, called Quantitative Attribute-based Prompting (QAP), which turns two instance-specific quantitative measures (i.e., K-function-based spatial attributes and histogram-based morphological attributes) into visual prompts.
- Experiments conducted on three tasks demonstrate that the incorporation of quantitative attributes enhance both the diagnosis efficiency and interpretability.

2. Related Work

NAFLD Diagnosis. Liver pathology images capture cellular tissue structures and play a central role in diagnosing NAFLD [22]. [15] utilized morphological features with a machine-learning algorithm for steatosis quantification. [16] fine-tuned a pre-trained model for histological findings recognition. [34] applied UNet network [33] for fat droplet segmentation. The performance is limited due to the relatively small-scale datasets.

Pathology Image Analysis. Deep learning methods demonstrate their effectiveness in tissue segmentation [9, 11, 32] and classification [13, 23]. HoverNet [31] has shown promising results in segmenting nuclei by utilizing the distance between them. Omni-Seg [8] achieved superior segmentation accuracy in tissue structures. Before the widespread adoption of deep learning, conventional machine learning methods were extensively used for pathology image analysis [21, 30]. [3] studied the morphological features for the diagnosis of breast cancer. [38] explored the shapes and boundary features for lung cancer image analysis. [24] applied spatial statistics to capture cell-cell interactions to predict treatment response in lung cancer patients. The effectiveness and robustness of handcrafted features have been demonstrated in various pathological image analysis tasks, making them a reliable tool for such applications. Although the process of hand-crafting features can be time-consuming, it offers a valuable quantitative assessment and generalization that is crucial in clinical practice [10, 29].

Visual Prompt Tuning. Prompting [27] was initially proposed in natural language process tasks by incorporating language instructions into input text to enhance its ability to comprehend unfamiliar tasks. VPT [18] firstly explored visual prompting with a small number of trainable parameters for vision tasks. VQT [37] and EXPRES [7] further studied how to inject prompts into foundation models and achieved higher efficiency of the parameters. However, these methods mainly focus on task-agnostic prompts, with little research focusing on task-specific visual prompts. In analyzing pathological images, these methods may not provide sufficient guidance for understanding the images.

3. Proposed Method

Overview. Figure 2 presents an overview of our proposed quantitative attribute-based prompting method for liver tissue recognition. Given a liver biopsy image $x \in R^{3\times h\times w}$ and its label y, the image is first embedded into the d-dimentional feature space with positional encoding through embedding operation $\mathcal{E}mbed$:

$$E_0 = \mathcal{E}mbed(x) \tag{1}$$

Additionally, the image x is also fed into quantitative attribute-based prompting (QAP) module $\mathcal{Q}$ to generate visual prompts P_0:

$$P_0 = \mathcal{Q}(x) \tag{2}$$

Combining the image embeddings E_i and P_i, and an extra learnable classification token $[cls]$, they are fed into the i^{th} transformer layer $\mathcal{L}_i$ of a vision transformer (ViT) for feature learning:

$$[cls_i, E_i, P_i] = \mathcal{L}_i([cls_{i-1}, E_{i-1}, P_{i-1}])i = 1, 2, ..., N \tag{3}$$

where N refers to the number of transformer layers. Lastly, the classifier $\mathcal{C}_{\theta_{cls}}$ is applied to the class token $[cls_N]$ to estimate the label $\hat{y}$

$$\hat{y} = \mathcal{C}_{\theta_{cls}}(cls_N) \tag{4}$$

The key aspect lies in how to generate visual prompts effectively. Our focus is on studying how to incorporate quantitative attributes into visual prompts.

3.1. Quantitative Attribute-based Prompting

In this section, we present the proposed quantitative attribute-based prompting. Our key insight is to learn explicit prompts from quantitative attributes. Imitating the diagnostic process of a pathologist, we consider two types of quantitative attributes, including spatial attributes and morphological attributes.

Given the image x, it is firstly fed into a segmenter $\mathcal{S}eg$ to parse the image and get tissue structure segmentation O:

$$O = \mathcal{S}eg(x) \tag{5}$$

where $O = \{o_i^c\}_{i=1,..,n_c}^{c=1,..,C}$. o_i^c refers to the i-th object belonging to class type c. n_c represents the number of corresponding objects. The total number of class types is C.

Given the objects detected from the liver biopsy images, we would perform a comprehensive quantitative assessment, helping with the pathological image analysis.

K-function-based Spatial Attributes. The spatial arrangement of objects in liver biopsy images plays a crucial role in the diagnosis of NAFLD. It includes the clustering of nuclei and their displacement around fatty cells. These spatial attributes align perfectly with the K-function [12], making it a valuable tool for analysis.

To quantify the spatial attributes, we adopt the K-function-based spatial attribute analysis module $\mathcal{D}^s$ on tissue structure segments O, yielding out spatial attributes A^s:

$$A^s = \mathcal{D}^s(O) \tag{6}$$

where the K-function-based spatial attribute analysis module $\mathcal{D}^s$ is implemented with a collection of K-function [12]:

$$\mathcal{D}^s(O) \colon = \{K_{src}^{tag}(r) | src = 1, ..., C, tag = 1, ..., C\} \tag{7}$$

Specifically, the K-function counts the number of source objects $O^{src} = \{o_i^{src}\}_{i=1,..,n_s}$ which are within a certain distance r from target objects $O^{tag} = \{o_i^{tag}\}_{i=1,..,n_{tag}}$:

$$K_{src}^{tag}(r) = \frac{1}{\lambda}\pi r^2 (\sum_{i=1,..,n_{src}} [\![d^{tag}(o_i^{src}) \leq r]\!])^2 \tag{8}$$

where $[\![\cdot]\!]$ is a boolean function that returns 1 if the input is true. λ denotes a constant. The distance function d^T defines the distance of the source objects to neighborhood target objects:

$$d^{tag}(o_i^{src}) = \min_{j=1,..,n_T} ||o_i^{src} - o_j^{tag}||_2 \tag{9}$$

where $||\cdot||_2$ refers to the Euclidean distance between two objects. The K-function-based spatial attributes $A^s = [a_1^s, .., a_{n_s}^s]$ offer valuable insights into the scale of clustering or dispersion. This quantitative assessment aligns with the needs of pathologists, enabling them to gain a deeper understanding of spatial patterns.

Histogram-based Morphological Attributes. Morphological attributes are used to quantify the shape, size, and structure of objects observed in liver biopsy images. Without losing any generality, let us consider the object sets $O^c = \{o_i^c\}_{i=1,..,n_c}$

We develop a histogram-based morphological attributes analysis module $\mathcal{D}^m$ to get its morphological attributes A^m:

$$A^m = \mathcal{D}^m(O^c) \tag{10}$$

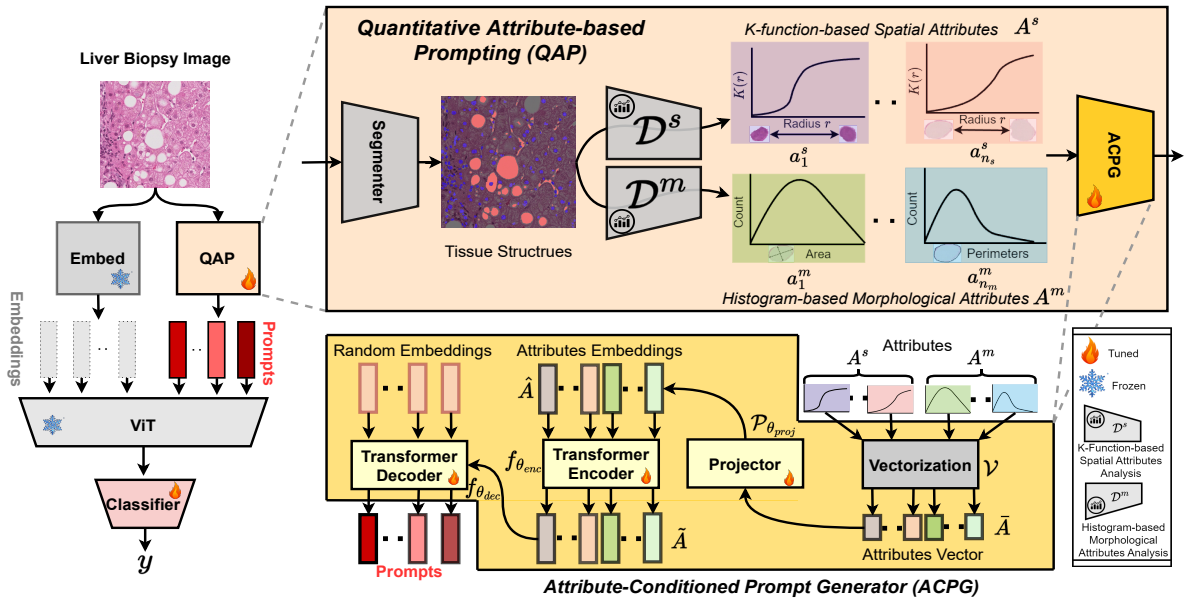


Figure 2. Overview of our method. QAP explores spatial and morphological attributes to generate visual prompts for aiding pathology image analysis. Furthermore, QAP introduces an attribute-conditioned prompt generator to generate visual prompts characterizing pathology images.

where the histogram-based morphological attribute analysis module $\mathcal{D}^m$ is a collection of histogram operators H:

$$\mathcal{D}^m := \{H(\phi^k(O^c))|k = 1, ..., n_\phi\} \tag{11}$$

where the histogram operator H counts the number of objects o_i^c with attribute $\phi^k(o_i^c)$ in the range $[a_j, a_{j+1}]$:

$$H(\phi^k(O^c)) = \{\sum_{i=1}^{n_c} \mathbf{1}_{[a_j, a_{j+1}]}(\phi(o_i^c)) \mid a_j < a_{j+1}, j \in N\} \tag{12}$$

The function ϕ^k serves as a morphological attribute extractor that captures diverse morphological attributes such as area and eccentricity. n_ϕ denotes the total number of morphological attributes considered. The histogram-based morphological attributes $a^m = [a_1^m, .., a_{n_m}^m]$ provide a comprehensive description of the shape and structure of the objects.

After spatial and morphological attributes analysis specific to the objects observed in liver pathology images, we get a comprehensive quantitative assessment of liver pathological image with attribute sets $A = [A^s, A^m] = [a_1^s, .., a_{n_s}^s, a_1^m, .., a_{n_m}^m]$.

Attribute-Conditioned Prompt Generator. Quantitative attributes are essential in identifying the significant features of pathology images. They provide a structured summary that helps our model concentrate on specific aspects of the data. With the availability of these attributes, we introduce a attribute-conditioned prompt generator, which utilizes these attributes to create more precise and context-aware prompts.

Given the attributes $A = [A^s, A^m]$, we vectorize the attributes by uniform sampling $\mathcal{V}$, yielding $d_{\bar{A}}$-dimensional feature representations $\bar{A}$:

$$\bar{A} = \mathcal{V}(A) \tag{13}$$

where $\bar{A} \in R^{(n_s+n_m)\times d_{\bar{A}}}$.

To remove the noise and irrelevant information from these attributes, we project the initial attributes vector $\bar{A}$ into a $d_{\hat{A}}$-dimensional feature space through the projector $\mathcal{P}_{\theta_{proj}}$:

$$\hat{A} = \mathcal{P}_{\theta_{proj}}(\bar{A}) \tag{14}$$

where $\hat{A} \in R^{(n_s+n_m)\times d_{\hat{A}}}$. Quantitative attributes in pathology images can have diverse meanings and relationships. To effectively model the interactions between input tokens, a transformer encoder with a self-attention mechanism can be advantageous. Therefore, we employ a transformer encoder $f_{\theta_{enc}}$ to learn quantitative attribute feature representation $\tilde{A}$:

$$\tilde{A} = f_{\theta_{enc}}(\hat{A}) \tag{15}$$

These attributes provide a rich source of structured information that can guide the model to focus on specific aspects of the data. Next, we will discuss generating prompts conditioned on these quantitative attributes.

Our prompt generator draws inspiration from the transformer decoder and functions in a similar manner. It defines the conditional probability distribution of a target sequence utilizing the contextualized encoding sequence. Using previous prompts and quantitative attributes, our generator generates each prompt $P_0 = [p_1^0, ..., p_{n_p}^0]$. Mathematically, the transformer decoder $f_{\theta_{dec}}$ defines the probability

Table 1. Classification performance of NAFLD abnormalities.

Methods	10% Training	50% Training	100% Training
Fabian et al. [16]	81.02±1.38	94.60±0.67	97.34±0.46
SAR [40]	80.71±2.99	99.09±0.35	99.54±0.19
VPT-DEEP [18]	96.58±0.60	99.30±0.23	99.49±0.26
VQT [37]	95.99±0.72	99.27±0.29	99.47±0.27
QAP (Ours)	**97.75**±0.77	**99.37**±0.23	**99.58**±0.17

distribution $p_{\theta_{dec}}$ of output prompts P_0 given the attributes $\hat{A}$:

$$p_{\theta_{dec}}(P_0|\hat{A}) = \prod_{i=1}^{n_p} p_{\theta_{dec}}(p_i^0|P_{0,1:i-1}, \hat{A}) \quad (16)$$

The final prompt learning is to maximize the probability distribution $p_{\theta_{dec}}(P_0|\hat{A})$ and the learning procedure is formalized using a transformer decoder $f_{\theta_{dec}}$:

$$P_0 = f_{\theta_{dec}}(\bar{P}_0, \hat{A}) \quad (17)$$

where $\bar{P}_0 \in R^{n_p \times d_p}$ is an initial prompt embeddings

This conditional decoder allows for the creation of prompts that are cohesive with the quantitative attributes and previous prompts. By conditioning the prompt generation on these attributes, we can guide the model to generate more relevant and precise prompts that are tailored to the specific characteristics of each image.

4. Experiments

Tasks and datasets. We evaluate our method with three types of tasks: classifying NAFLD abnormalities, recognizing histological findings, and histological scoring on WSI. We conduct experiments on **NAFLD-Anomaly** [45] for the classification of NAFLD abnormalities. It contains 256×256 image tiles extracted from WSI of H&E stained tissue of the mouse. It consists of 2170 normal samples and 2150 abnormal samples with NAFLD. For the recognition of histological findings, we evaluated the performance in the dataset **Liver-NAS** [44]. It provides an image tile dataset extracted from 9 patients with associated labels indicating the type of histological findings (i.e., steatosis, inflammation, ballooning). There are 5875 image tiles with steatosis ($N = 3838$), ballooning ($N = 298$), inflammation ($N = 69$), and others ($N = 1659$). Liver-NAS [44] also offers 256 WSIs of the liver, without any overlap with previously included patients. These WSIs are accompanied by patient-level labels that enable histological scoring. These labels quantify the presence of specific histological findings observed in the WSIs. It includes steatosis (0-3), inflammation (0-3), and ballooning (0-2). All liver tissues are stained with H&E without normalization processing. The average image resolution is 61,000×20,000 pixels.

We conduct a 5-fold cross-validation to verify the effectiveness of our proposed method. For the tasks of identifying NAFLD abnormalities and histological findings, the data was randomly split to ensure reliable results. Each fold was divided into a training set (70%), a validation set (10%), and a test set (20%). During training, the model's performance was monitored on the validation set and used for model selection. To evaluate the model's generalization ability, we performed histological scoring on larger, unseen image samples. We utilized the well-trained model that was developed for recognizing histological findings to score all 256 WSIs.

Implementation Details. The proposed method is implemented with Pytorch. We use ViT-S/16 as the backbone, which is pre-trained TCGA-BRCA [26] pathology images with self-supervised learning method [6]. We employed the AdamW [28] optimizer with an initial learning rate of $5e^{-4}$, applied cosine decay, and trained the model for 50 epochs. The batch size is 64. There are two types of tissue structure to consider: nuclei and white regions. More implementation details are provided in supplementary materials.

Evaluation Metrics. For classification experiments, the F1 score is used as the performance metric. For histological scoring, **Spearman correlation coefficient** ρ [17] is used to calculate the correlation between disease area percentage and the histological score assigned by pathologists at a significant level p-value. During 5-fold cross-validation experiments, the mean and standard deviation are reported. The overall significance is calculated by combining the p-values of each fold using the Fisher method [14]. More evaluation results are provided in the supplementary materials.

4.1. Comparison with State-of-the-Art Methods

We compare our methods against four different approaches, including Fabian et al. [16], SAR [40], VPT [18], and VQT [37], covering various aspects of liver pathology image analysis and visual prompting techniques. Fabian et al. [16] utilized deep convolutional neural networks (CNN) to recognize histological findings. SAR [40] is an attention regularization method driving the model focusing on clinical features. VPT [18] first explored visual prompt tuning for natural image recognition. VQT [37] introduced a highly efficient method of prompt tuning that optimizes memory usage. Additionally, it aggregates intermediate features to further enhance the tuning process.

Classification of NAFLD abnormalities. We compare our method with other SOTA methods for the classification of NAFLD abnormalities in Table 1. The table presents the F1 score for each method under different training data ratios. Our method has demonstrated superior performance on varying data utilization ratios. The superior performance in varying data utilization ratios suggests that our method is robust to the amount of labeled data available especially in low-data regimes, resulting in a 1% increase.

Histological findings recognition. Table 2 shows the F1

Table 2. Comparison of histological findings recognition with different methods.

Methods	Others	Inflammation	Ballooning	Seatosis	Macro-Average
Fabian et al. [16]	89.66±4.09	45.46±23.49	80.62±6.61	97.38±1.42	78.28±6.81
SAR [40]	92.02±1.89	51.60±21.93	82.02±6.35	98.06±0.46	80.92±6.31
VPT-DEEP [18]	92.82±3.96	45.89±12.54	84.55±9.85	98.81±0.21	80.52±6.35
VQT [37]	92.09±7.35	49.80±12.33	83.89±13.22	**98.90**±0.41	81.17±7.99
QAP (Ours)	**94.21**±3.27	**55.38**±17.17	**85.10**±7.92	98.79±0.25	**83.37**±6.99

Table 3. Generealization ability comparison of histological scoring performance with different methods. ρ represents the Spearman's Correlation Coefficients.

Methods	Inflammation		Ballooning		Steatosis	
	$\rho \uparrow$	p-value $\downarrow$	$\rho \uparrow$	p-value $\downarrow$	$\rho \uparrow$	p-value $\downarrow$
Fabian et al. [16]	0.143±0.142	≤ 0.001	0.188±0.067	≤ 0.001	0.271±0.150	≤ 0.001
SAR [40]	0.236±0.123	≤ 0.001	0.246±0.024	≤ 0.001	0.593±0.127	≤ 0.001
VPT-DEEP [18]	**0.422**±0.022	≤ 0.001	0.299±0.026	≤ 0.001	0.759±0.007	≤ 0.001
VQT [37]	0.342±0.062	≤ 0.001	0.316±0.012	≤ 0.001	0.766±0.006	≤ 0.001
QAP (Ours)	0.410 ±0.008	≤ 0.001	**0.332**±0.018	≤ 0.001	**0.793**±0.008	≤ 0.001

score of different methods in the classification of histological findings. Our method achieves a better overall performance, with an average F1 score of 83.37. Steatosis can be accurately detected using any of the available methods due to the abundance of training samples. We observe a considerable increase in both inflammation and ballooning, with values of 55.38 and 85.10, respectively. The increase can be attributed to the incorporation of quantitative attributes. Our method allows for a thorough and precise analysis of the histological characteristics relevant to specific findings.

Generalization for histological scoring. We conduct histological scoring on unseen patients to evaluate our approach's generalizability. In Table 3, we provide the Spearman correlation coefficient [17] and its associated p-value. This statistical measure assesses the relationship between disease area percentage and histological score. A higher correlation coefficient indicates a stronger association, while a lower p-value indicates statistical significance.

As shown in the table, prompting-based methods demonstrate better generalizability by accurately identifying histological findings in unseen patients. The model benefits from the general knowledge of large-foundation models, resulting in better generalization. Our model achieves a comparable correlation coefficient of 0.410 for scoring inflammation and effectively minimizes variance, making it a highly effective solution. Our method demonstrates an improvement in histological scoring on ballooning and steatosis, achieving correlation coefficients of 0.332 and 0.793, respectively. Quantitative attributes offer a stronger and more comprehensive representation of pathology images.

4.2. Quantification of Histological Findings

Our model offers a thorough analysis of the diagnosis procedure. It not only highlights the areas of focus, but also identifies the specific attributes that are emphasized, as illustrated in Figure 3. Consider an image of liver pathology with lobular inflammation. With the model's attention, we observe that the diagnosis is related to the nuclei region. Furthermore, the significance histogram associated with the attributes enables us to attribute the diagnostic results to specific attributes, such as the spatial arrangement of nuclei quantified by attribute a_1^s. The attribute a_1^s has a significance score of 0.27, which is higher than the scores

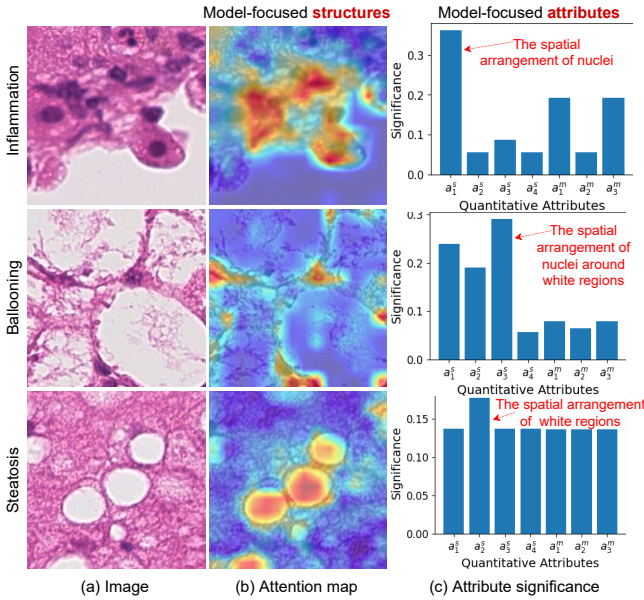
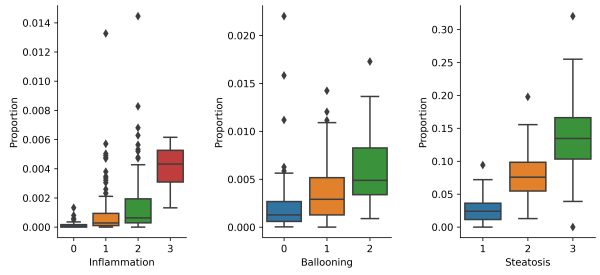


Figure 3. Image samples with its attention map and attribute significance histogram when identifying specific histological findings. Our method enhances interpretability by visually representing the decision-making process through attention maps and attribute significance histograms. a. The structures the model focuses on; b. The attributes of structures the model focus on.



(a) (b) (c)

Figure 4. Box plot of the disease area proportion and the histological score assigned by pathologists.

of other attributes. In the diagnosis of ballooning, the model primarily focuses on the spatial arrangement of nuclei surrounding white regions, denoted by attribute a_3^s. This is supported by its attention map and higher attribute significance scores. By accessing this information, we can gain a better understanding of the underlying factors involved and interpret the image effectively. More visualization results are provided in supplementary materials.

Furthermore, during histological scoring, our method also quantifies the proportion of tissue area assigned to each histological finding. As shown in Figure 4, It offers the distribution of the proportions of the disease area and the histological score assigned by pathologists. The distribution of these two variables can provide valuable information on the severity and progression of NAFLD within the population studied. This meets the clinical requirement where NAFLD is considered a continuous developmental process.

4.3. Justification for Attributes Choices

In this section, we will dive into the reasoning behind the attributes we have chosen for the analysis of the liver biopsy image. Our approach employs the K-function-based spatial attributes and histogram-based morphological attributes, chosen for their clinical relevance and interpretability in the context of analyzing liver biopsy images.

K-function-based Spatial Attributes. The spatial distribution captured by the K-function reveals patterns characteristic of NAFLD. As shown in Figure 5, the K-function distribution of the nuclei and the white regions provides significant information for the recognition of histological findings. Figure 5 (a) illustrates the K-function plot, which displays the spatial characteristics. The K-function value (y-axis) increases as the distance between objects (x-axis) increases. When the plotted line grows rapidly and remains above the reference line, it indicates clustering spatial characteristics. On the contrary, if the line grows under the reference line, it suggests sparsely distributed characteristics. The distribution of nuclei and white regions is depicted in Figure 5 (b)-(e), revealing the diverse developmental paths of each class. This provides an intriguing glimpse into the distinctive properties of these cells and their gro wth patterns.

As seen in Figure 5 (b), the K-function curves of each class characterize the spatial distribution of nuclei. In the 'Others' category, which includes normal tissues, the distribution of nuclei is typically uniform, with consistent spacing. As a result, the K-function curve exhibits a mid-speed growth rate, which can serve as a reference line for comparison with other classes. On the other hand, inflammation is characterized by a higher degree of clustering among nuclei, leading to a more rapid initial growth of the K-function curve. This observation is in line with the clinical definition of inflammation and provides valuable insights into the clustering behavior of nuclei. In contrast, steatosis and ballooning exhibit a more sparsely distributed nucleus pattern, which is reflected in the slower growth rate of their K-function curve. The K-function illustrated in Figure 5 (c) provides insight into the spatial distribution of white regions. The k-function curves of steatosis and ballooning exhibit a high growth rate, indicating clustering due to a large accumulation of fatty cells. The spatial arrangement of the nuclei and the white regions also demonstrates distinct patterns around different neighborhoods, as shown in Figure 5 (d)-(e). The K-function accurately identifies clustering or dispersion patterns, enabling informed decision-making.

Histogram-based Morphological Attributes. Similarly, the histogram-based morphological attributes quantify the shape, size, and structure of objects observed in the liver biopsy images. Changes in cell morphology are often indicative of disease states. The histogram in Figure 6 shows the distribution of the area, eccentricity, and perimeter of the white region. It provides valuable insights into the characteristics of the white region.

Area quantify the size of the white region. As shown in Figure 6 (a), it represents the distribution of the area of white regions. NAFLD starts with the accumulation of fatty cells. The area of cell structures is a critical attribute directly correlated with their size. A higher frequency of larger areas may indicate the presence of an accumulation of fatty cells. On the other hand, a higher frequency of smaller areas could suggest a lower degree of accumulation of fatty cells. The clinical pattern is evident in the area-based histogram. The area-based histogram shows that image diagnoses with steatosis and ballooning trend have a larger mean area of fatty cells. The histogram of eccentricity (Figure 6 (b)) and perimeters (Figure 6 (c)) also provide cues in distinguishing white regions observed in different classes.

These attributes not only provide a quantitative measure of the tissue state but also align well with the visual and qualitative assessments that pathologists perform during microscopic examination. Therefore, they bridge the gap between automated image analysis and traditional histopathological assessment, making the diagnosis process more efficient and accurate.

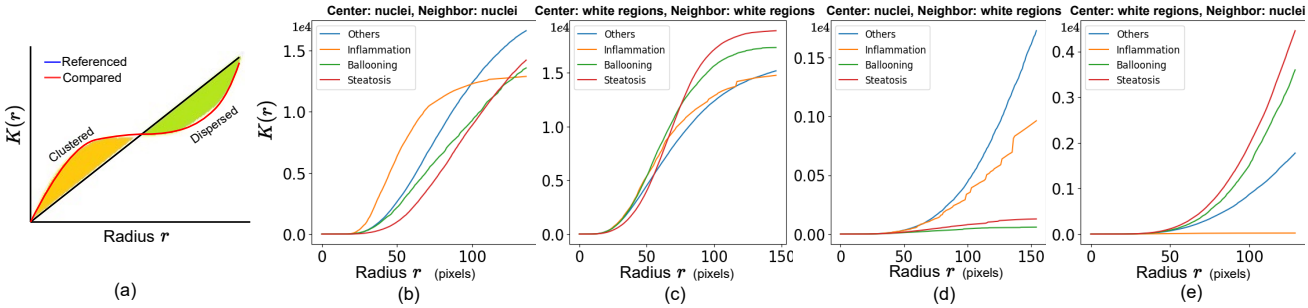


Figure 5. K-function distribution of nuclei and white regions in liver biopsy images. (a) The K-function plot indicates spatial clustering or dispersion. The K-function characterizes the (b) spatial arrangement of nuclei, (c) the spatial arrangement of white regions, (d) the spatial arrangement of nuclei around white regions, and (e) the spatial arrangement of white regions around nuclei.

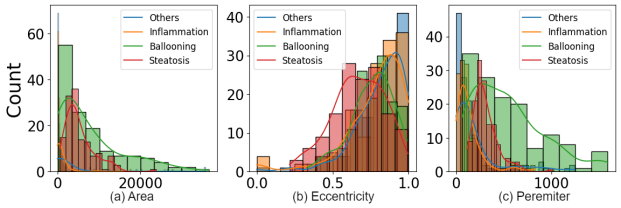


Figure 6. Histograms of morphological attributes of white regions in liver biopsy images. (a) The area of white regions in each image tile may indicate the degree of fat accumulation or vessels. (b) The eccentricity of white regions measures how elongated they are. (c) The perimeter of white regions indicates the size and shape of cells. Histogram-based morphological attributes are valuable for distinguishing between different histological findings.

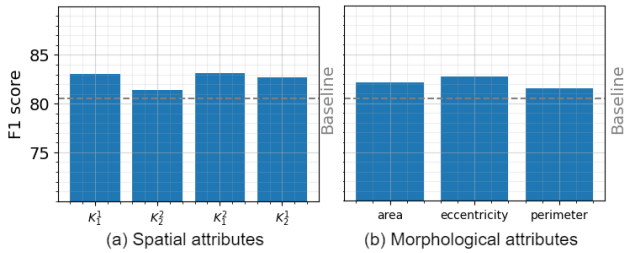


Figure 7. Comparative results with different (a) spatial and (b) morphological attributes. Our method uses quantitative attributes to better identify histological findings compared to the baseline approach with task-agnostic visual prompts.

4.4. Exploration of Quantitative Attributes

In this section, we conduct ablation studies to analyze the impact of various spatial and morphological attributes on our method's performance.

Attributes Evaluation. As shown in Figure 7, it illustrates the enhanced performance achieved by incorporating different quantitative attributes in our proposed method. The baseline performance is achieved using learning prompts from random initial embeddings without incorporating any quantitative attributes. The x-axis represents the varying quantitative attributes. Each attribute serves a distinct purpose in providing valuable information that aids in identifying histological findings.

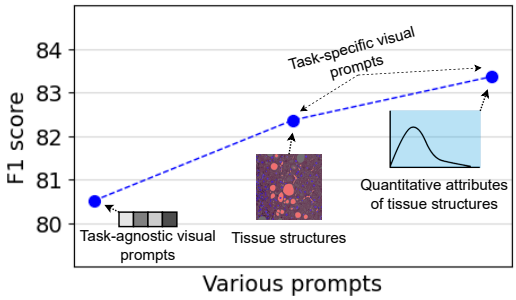


Figure 8. Exploration of various prompts. By incorporating explicit cues into the learning prompts, performance can be significantly enhanced.

Attributes Exploration. In this section, we explore learning various prompts. The tissue structure segments provide more informative cues compared to task-agnostic visual prompts learned from randomly initialized vectors. Additionally, the quantitative attributes obtained from summarizing the statistical information about tissue structures are more explicit. As shown in Figure 8, using prompts learned conditioned on explicit cues can enhance the learning process and improve performance.

5. Conclusion

We present a novel quantitative attribute-based prompting method for liver pathology image analysis. Our method leverages quantitative attributes that capture the spatial and morphological characteristics of key tissue alteration regions. An attribute-conditioned prompt generator is specifically designed to generate unique and accurate prompts that highlight the characteristics of pathology images. Our experiments across three tasks demonstrates the significant potential of combining traditional histological assessment with advanced deep-learning methods, resulting in notable improvements in interpretability and overall reliability of diagnostic performance.

Acknowledgement. This work was supported by the Hong Kong Research Grants Council General Research Fund under Grant RGC/HKBU12200122.

References

[1] Veeral Ajmera and Rohit Loomba. Imaging biomarkers of nafld, nash, and fibrosis. *Molecular metabolism*, 50:101167, 2021. 2

[2] Anusua Basu, Pradip Senapati, Mainak Deb, Rebika Rai, and Krishna Gopal Dhal. A survey on recent trends in deep learning for nucleus segmentation from histopathology images. *Evolving Systems*, pages 1–46, 2023. 2

[3] Andrew H Beck, Ankur R Sangoi, Samuel Leung, Robert J Marinelli, Torsten O Nielsen, Marc J Van De Vijver, Robert B West, Matt Van De Rijn, and Daphne Koller. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Science translational medicine*, 3(108):108ra113–108ra113, 2011. 2

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. 2

[5] Alastair D Burt, Carolin Lackner, and Dina G Tiniakos. Diagnosis and assessment of nafld: definitions and histopathological classification. In *Seminars in liver disease*, volume 35, pages 207–220. Thieme Medical Publishers, 2015. 2

[6] Richard J Chen and Rahul G Krishnan. Self-supervised vision transformers learn visual concepts in histopathology. *Learning Meaningful Representations of Life, NeurIPS 2021*, 2021. 5

[7] Rajshekhar Das, Yonatan Dukler, Avinash Ravichandran, and Ashwin Swaminathan. Learning expressive prompting with residuals for vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3366–3377, 2023. 3

[8] Ruining Deng, Quan Liu, Can Cui, Zuhayr Asad, Haichun Yang, and Yuankai Huo. Omni-seg: A single dynamic network for multi-label renal pathology image segmentation using partially labeled data. *arXiv preprint arXiv:2112.12665*, 2021. 2

[9] Ruining Deng, Quan Liu, Can Cui, Tianyuan Yao, Jun Long, Zuhayr Asad, R Michael Womick, Zheyu Zhu, Agnes B Fogo, Shilin Zhao, et al. Omni-seg: A scale-aware dynamic network for renal pathological image segmentation. *IEEE Transactions on Biomedical Engineering*, 2023. 2

[10] James A Diao, Jason K Wang, Wan Fung Chui, Victoria Mountain, Sai Chowdary Gullapally, Ramprakash Srinivasan, Richard N Mitchell, Benjamin Glass, Sara Hoffman, Sudha K Rao, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nature communications*, 12(1):1613, 2021. 2

[11] Kexin Ding, Mu Zhou, He Wang, Olivier Gevaert, Dimitris Metaxas, and Shaoting Zhang. A large-scale synthetic pathological dataset for deep learning-enabled segmentation of breast cancer. *Scientific Data*, 10(1):231, 2023. 2

[12] Philip Dixon. Ripley's k function. 2001. 3

[13] Okyaz Eminaga, Mahmoud Abbas, Jeanne Shen, Mark Laurie, James D Brooks, Joseph C Liao, and Daniel L Rubin. Plexusnet: A neural network architectural concept for medical image classification. *Computers in Biology and Medicine*, 154:106594, 2023. 2

[14] RA Fisher. Statistical methods for research workers (oliver and boyd, edinburgh, uk), 1925. 5

[15] Nikolaos Giannakeas, Markos G Tsipouras, Alexandros T Tzallas, Maria G Vavva, Maria Tsimplakidou, Evaggelos C Karvounis, Roberta Forlano, and Pinelopi Manousou. Measuring steatosis in liver biopsies using machine learning and morphological imaging. In *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 40–44. IEEE, 2017. 2

[16] Fabian Heinemann, Gerald Birk, and Birgit Stierstorfer. Deep learning enables pathologist-like scoring of nash models. *Scientific reports*, 9(1):1–10, 2019. 1, 2, 5, 6

[17] Jongyun Hwang, Sunghun Na, Hyangah Lee, and Dongheon Lee. Correlation between preoperative serum levels of five biomarkers and relationships between these biomarkers and cancer stage in epithelial overian cancer. *Journal of Gynecologic Oncology*, 20(3):169–175, 2009. 5, 6

[18] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision (ECCV)*, 2022. 1, 2, 3, 5, 6

[19] Ravi Juluri, Raj Vuppalanchi, John Olson, Aynur Ünalp, Mark L Van Natta, Oscar W Cummings, James Tonascia, and Naga Chalasani. Generalizability of the nonalcoholic steatohepatitis clinical research network histologic scoring system for nonalcoholic fatty liver disease. *Journal of clinical gastroenterology*, 45(1):55–58, 2011. 2

[20] David E Kleiner, Elizabeth M Brunt, Mark Van Natta, Cynthia Behling, Melissa J Contos, Oscar W Cummings, Linda D Ferrell, Yao-Chang Liu, Michael S Torbenson, Aynur Unalp-Arida, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*, 41(6):1313–1321, 2005. 1

[21] Daisuke Komura and Shumpei Ishikawa. Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, 16:34–42, 2018. 2

[22] Wei-Qiang Leow, Anthony Wing-Hung Chan, Paulo Giovanni L Mendoza, Regina Lo, Kihan Yap, and Haeryoung Kim. Non-alcoholic fatty liver disease: the pathologist's perspective. *Clinical and Molecular Hepatology*, 29(Suppl):S302, 2023. 2

[23] Honglin Li, Chenglu Zhu, Yunlong Zhang, Yuxuan Sun, Zhongyi Shui, Wenwei Kuang, Sunyi Zheng, and Lin Yang. Task-specific fine-tuning via variational information bottleneck for weakly-supervised pathology whole slide image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7454–7463, 2023. 2

[24] X Li, F Gaire, G Jansen, R Copping, T Bengtsson, and J Dai. 1381p spatial-statistics-based modeling for predicting treatment response in non-small cell lung cancer (nsclc) patients using h&e pathology images. *Annals of Oncology*, 31:S879, 2020. 2

[25] Yang-Yang Li, Tian-Lei Zheng, Shu-Yuan Xiao, Peng Wang, Wen-Jun Yang, Li-Lin Jiang, Li-Li Chen, Jun-Cheng Sha, Yi Jin, Sui-Dan Chen, et al. Hepatocytic ballooning in non-alcoholic steatohepatitis: Dilemmas and future directions. *Liver International*, 2023. 2

[26] Jianfang Liu, Tara Lichtenberg, Katherine A Hoadley, Laila M Poisson, Alexander J Lazar, Andrew D Cherniack, Albert J Kovatich, Christopher C Benz, Douglas A Levine, Adrian V Lee, et al. An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*, 173(2):400–416, 2018. 5

[27] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35, 2023. 3

[28] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5

[29] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis*, 33:170–175, 2016. 2

[30] Sidra Nawaz and Yinyin Yuan. Computational pathology: Exploring the spatial dimension of tumor ecology. *Cancer letters*, 380(1):296–303, 2016. 2

[31] Pushpak Pati, Guillaume Jaume, Antonio Foncubierta, Florinda Feroce, Anna Maria Anniciello, G. Scognamiglio, Nadia Brancati, Maryse Fiche, Estelle Dubruc, Daniele Riccio, Maurizio Di Bonito, Giuseppe De Pietro, Gerardo Botti, Jean-Philippe Thiran, Maria Frucci, Orcun Goksel, and Maria Gabrani. Hierarchical graph representations in digital pathology. *Medical image analysis*, 75:102264, 2021. 2

[32] Ruichen Rong, Hudanyun Sheng, Kevin W Jin, Fangjiang Wu, Danni Luo, Zhuoyu Wen, Chen Tang, Donghan M Yang, Liwei Jia, Mohamed Amgad, et al. A deep learning approach for histology-based nucleus segmentation and tumor microenvironment characterization. *Modern Pathology*, 36(8):100196, 2023. 2

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 2

[34] Mousumi Roy, Fusheng Wang, Hoang Vo, Dejun Teng, George Teodoro, Alton B Farris, Eduardo Castillo-Leon, Miriam B Vos, and Jun Kong. Deep-learning-based accurate hepatic steatosis quantification for histological assessment of liver biopsies. *Laboratory investigation*, 100(10):1367–1383, 2020. 2

[35] Arun J Sanyal, Stephen A Williams, Joel E Lavine, Brent A Neuschwander-Tetri, Leigh Alexander, Rachel Ostroff, Hannah Biegel, Kris V Kowdley, Naga Chalasani, Srinivasan Dasarathy, et al. Defining the serum proteomic signature of hepatic steatosis, inflammation, ballooning and fibrosis in non-alcoholic fatty liver disease. *Journal of hepatology*, 78(4):693–703, 2023. 1

[36] Pir Ahmad Shah, Rashmee Patil, and Stephen A Harrison. Nafld-related hepatocellular carcinoma: The growing challenge. *Hepatology*, 77(1):323–338, 2023. 1

[37] Cheng-Hao Tu, Zheda Mai, and Wei-Lun Chao. Visual query tuning: Towards effective usage of intermediate representations for parameter and memory efficient transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7725–7735, 2023. 1, 3, 5, 6

[38] Shidan Wang, Alyssa Chen, Lin Yang, Ling Cai, Yang Xie, Junya Fujimoto, Adi Gazdar, and Guanghua Xiao. Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. *Scientific reports*, 8(1):10393, 2018. 2

[39] Kara Wegermann, Anna Mae Diehl, and Cynthia A Moylan. Challenges and barriers in hepatocellular carcinoma (hcc) surveillance for patients with non-alcoholic fatty liver disease (nafld). 2023. 1, 2

[40] Chong Yin, Siqi Liu, Rui Shao, and Pong C Yuen. Focusing on clinically interpretable features: selective attention regularization for liver biopsy image classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pages 153–162. Springer, 2021. 1, 5, 6

[41] Xinyi Zhang, Frederico O Gleber-Netto, Shidan Wang, Roberta Rayra Martins-Chaves, Ricardo Santiago Gomez, Nadarajah Vigneswaran, Arunangshu Sarkar, William N William Jr, Vassiliki Papadimitrakopoulou, Michelle Williams, et al. Deep learning-based pathology image analysis predicts cancer progression risk in patients with oral leukoplakia. *Cancer medicine*, 2023. 2

[42] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16816–16825, 2022. 2

[43] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9):2337–2348, 2022. 2

[44] Yu-Jie Zhou, Feng Gao, Wen-Yue Liu, Grace Lai-Hung Wong, Sanjiv Mahadeva, Nik Raihan Nik Mustapha, Xiao-Dong Wang, Wah-Kheong Chan, Vincent Wai-Sun Wong, and Ming-Hua Zheng. Screening for compensated advanced chronic liver disease using refined baveno vi elastography cutoffs in asian patients with nonalcoholic fatty liver disease. *Alimentary pharmacology & therapeutics*, 54(4):470–480, 2021. 5

[45] Igor Zingman, Birgit Stierstorfer, Charlotte Lempp, and Fabian Heinemann. Learning image representations for anomaly detection: application to discovery of histological alterations in drug development. *arXiv preprint arXiv:2210.07675*, 2022. 5