AI-ASSISTED WHOLE-SLIDE IMAGING ANALYSIS FOR HEPATOCELLULAR CARCINOMA PROGNOSIS: DEVELOPMENT OF A RISK SCORING SYSTEM WITH ENHANCED INTERPRETABILITY AND EFFICIENCY

LIU ANRAN

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Health Technology and Informatics

AI-Assisted Whole-Slide Imaging Analysis for Hepatocellular Carcinoma Prognosis: Development of a Risk Scoring System with Enhanced Interpretability and Efficiency

LIU Anran

A thesis submitted in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

June 2025

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

LIU Anran (Name of Student)

Abstract

Pathological images serve as the gold standard for cancer diagnosis and play a crucial role in clinical practice. With digital scanner development, Whole Slide Image (WSI) has gained attention due to its larger visual field and clearer imaging. Advanced technologies, especially deep learning models, have driven digital pathology development for downstream clinical tasks like automated diagnosis and prognostic analysis. However, existing methods are either limited by WSI's ultra-high resolution, leading to inefficient analysis, or rely on black box approaches lacking interpretability and clinical acceptability. This thesis addresses limited interpretability and excessive computational complexity in existing WSI automated analysis frameworks.

Following pathologists' diagnostic steps, we first propose an attention-driven mechanism named Attention Activator for identifying potential high-risk tissues. This system intuitively highlights potential high-risk areas in WSI using attention maps without prior guidance. Based on these localizations, we construct a multi-perspective risk scoring system from micro to macro levels. Unlike existing systems, we leverage deep learning for multi-perspective feature fusion. Validation results from hepatocellular carcinoma patient cohorts demonstrate our scoring system refines existing clinical staging, identifying potential high-risk patients within low-risk groups.

Given WSI's ultra-high resolution, we propose a novel encoding strategy termed FuzzyMIL to enhance analysis efficiency. This approach extends Fuzzy c-means to a learnable transformer framework, leveraging soft clustering characteristics to address

feature homogenization in global attention during encoding. This improves downstream task prediction accuracy while minimizing computational complexity. Experimental results showed FuzzyMIL outperforms existing state-of-the-art methods in diagnosis and subtyping across three public datasets while significantly reducing model parameters.

Furthermore, we applied the deep learning-based WSI analysis framework to explore automated localization, prognostic analysis, and treatment efficacy assessment based on Vessels that Encapsulate Tumor Clusters (VETC). VETC is a recently identified vascular pattern implicated in cancer metastasis progression and patient prognosis. However, current analyses rely on clinical experience, lacking standardized, quantitative methods. In this pioneering work integrating deep learning with VETC analysis, we developed VETC Net, which automatically distinguishes and locates VETC+ and VETC- regions within WSI. This network facilitates automated VETC+ distribution assessment and enables precise quantification of VETC-prognostic risk correlation. We also evaluated VETC response to various treatment modalities. Results demonstrate VETC Net enables precise VETC tissue localization and accurate prognostic assessment across multi-center, multi-treatment datasets.

In conclusion, this thesis presents solutions to limited interpretability and high computational complexity challenges in existing WSI analysis frameworks. We introduce Attention Activator for high-risk tissue identification and a multi-perspective risk scoring system improving risk stratification. Additionally, we propose FuzzyMIL, enhancing analysis efficiency while reducing model complexity. We explored novel

VETC-based clinical tasks for automated localization and prognostic analysis. Through extensive experimental validation, our methods show significant improvements in interpretability, efficiency, and clinical applicability for WSI analysis.

Acknowledgments

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Prof. CAI Jing. He has been a guiding light throughout my academic journey, not only offering invaluable support and encouragement in my research but also serving as a role model in my life. It is often said that finding an exceptional supervisor during one's doctoral studies is a rare gift, and I consider myself truly fortunate to have had the opportunity to study under his guidance.

I would also like to extend my sincere appreciation to my family. Their unwavering love and support have been invaluable. It is because of them that I have had the opportunity to experience the richness of the world and explore its many mysteries. In times of difficulty, my parents have been my anchors, and their encouragement has given me the strength to persevere.

I am grateful to the funding agencies for their generous support, which has made it possible for me to carry out my research. I am also thankful to all my project collaborators, especially the pathologists. Their expertise and cooperation have played a crucial role in the successful execution of my research.

I would also like to thank all my fellow lab members. They have been not only colleagues but also sources of inspiration, and their passion for research continues to motivate me. I am especially grateful to Dr. ZHANG Jiang for his patient guidance during times when my research encountered bottlenecks. I would also like to extend my sincere thanks to WANG Yinghui for her constant support and assistance in daily

life, and to LI Tong, who has been a steadfast companion and source of encouragement since the beginning of my graduate studies. I am also deeply appreciative of everyone who has walked alongside me throughout my entire student life. Although I cannot name each person individually, it is your presence that has made all this possible.

Finally, I am grateful for all the difficulties and challenges I have encountered throughout my time as a student. It is precisely because of these experiences that I have been able to continuously grow and gain the resilience and confidence for embracing success.

Publications

[1]. **Liu, A.**, Zhang, J., Li, T., Zheng, D., Ling, Y., Lu, L., ... & Cai, J. (2025). Explainable attention-enhanced heuristic paradigm for multi-view prognostic risk score development in hepatocellular carcinoma. *Hepatology International*, 1-11.

[2]. **Liu, A.**, Li, T., Cai, J., & Vajrala, S. S. V. (2025, April). FuzzyMIL: Decoupling Pathological Phenotypes through Deep Fuzzy Clustering for Efficient Whole Slide Image Analysis. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.

[3]. Dong, Y., Zhang, J., Lam, S., Zhang, X., **Liu, A.**, Teng, X., ... & Cai, J. (2023). Multimodal data integration to predict severe acute oral mucositis of nasopharyngeal carcinoma patients following radiation therapy. *Cancers*, *15*(7), 2032.

Table of Contents

List of Tables

List of Figures

List of Acronyms

WSI	Whole slide image
HER2	Human epidermal growth factor receptor 2
CD34	Cluster of Differentiation 34
LLM	Large language model
HCC	Hepatocellular carcinoma
VETC	Vessels that Encapsulate Tumor Clusters
DL	Deep learning
TNM	Tumor, node, metastasis
MVI	Microvascular invasion
CLAM	Clustering-constrained-attention multiple-instance learning
H&E	Hematoxylin and Eosin
ROI	Regions of Interest
CNN	Convolutional neural network
CONCH	Contrastive learning from Captions for Histopathology
GCN	Graph convolutional network
CLIP	Contrastive Language-Image Pre-training
HDS	Hybrid deep score
DFS	Disease free survival
OS	Overall survival
SYSUCC	Sun Yat-sen University Cancer Center

TCGA-LIHC	The cancer genome atlas liver hepatocellular carcinoma
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AFP	Alpha-fetoprotein
HBV	Hepatitis b virus
BCLC	Barcelona clinic liver cancer
ATAT	Attention activator
MLP	Multilayer perceptron
TIL	Tumor-infiltrating lymphocyte
AUC	Area Under the Curve
χ^2 **test**	Chi-squared test
PVP	Positive predictive value
NPV	Negative predictive value
ABMIL	Attention-based deep multiple instance learning
TransMIL	Transformer based correlated multiple instance learning
FCM	Learnable variant of Fuzzy C-means clustering
95%CI	95% confidence interval
FCA	Fuzzy clustering attention
GFE	Global feature encoder
BRCA	Breast invasive carcinoma
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma

LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
EMT	Epithelial-mesenchymal transition
TACE	Trans arterial chemoembolization
HAIC	Hepatic artery infusion chemotherapy
ORR	Overall response rate
DCR	Disease control rate

Chapter 1.
Introduction

1.1 Background and Significance

Pathological images are the gold standard for cancer diagnosis, providing rich clinical information that pathologists rely on for clinical tasks such as cancer diagnosis, cancer typing, and patient risk assessment. With the development of digital slide scanners, tissues can be converted into histopathological Whole-Slide Images (WSI). The visual field of WSI is more than four times larger than that of a traditional microscope. This enables the full preservation of original tissue structures and biopsy patterns. Concurrently, with advancements in computer vision for medical imaging, computational pathology has emerged as a critical field. The application of computer-assisted technologies in WSI analysis to aid pathologists in clinical diagnosis has become a key area of focus within computational pathology.

However, applying computer-assisted technologies in pathological image analysis and further utilizing these techniques in clinical tasks faces many challenges. Before the application of deep learning in WSI analysis, some studies focused on manually extracting features from WSIs to calculate and assess diagnostic or prognostic indicators. For example, immune staining scores (such as HER2, CD34, etc.), cell counts, cancer cell percentages. These approaches are often time-consuming and labor-intensive, which hinders the efficiency of clinical diagnoses, especially for rapidly

progressing severe diseases.

The application of deep learning techniques in WSI analysis has significantly improved efficiency, while the issue of trustworthiness remains a critical challenge. As many deep learning models are typically black-box in nature and often lack interpretability, they are difficult to directly apply in clinical settings. To address this, some attribution-based methods have made progress in enhancing interpretability, though these approaches are still largely dependent on the performance of the deep models. Therefore, developing a universal and easily deployable WSI analysis paradigm for real clinical tasks is of significant clinical importance and practical value.

Meanwhile, unlike natural images, WSIs possess extremely high resolutions, often exceeding hundreds of millions of pixels. Efficiently extracting clinically relevant information from such large-scale data remains a significant challenge in current research. Typically, WSIs are divided into tens of thousands of smaller patches to facilitate processing by deep learning models. To reduce redundancy, some approaches employ clustering techniques to aggregate features with similar representations in the feature space. However, such methods (especially those based on hard clustering like k-means) can easily conflate features from different tissue types due to their rigid attention mechanisms. Alternatively, some learnable methods tend to focus on visually distinctive patches rather than clinically significant ones, which may lead to diagnostic errors. Therefore, a strategy is needed that can extract as much information as possible from WSIs, while minimizing the impact of redundant information on computational cost and prediction accuracy.

Additionally, for many cancers that have limited WSIs, relying solely on WSI image data may not provide enough useful information. Therefore, supplementing information from WSIs is also an important task. Considering the advancements in multimodal information fusion in medical imaging, it offers a wealth of knowledge across different cancers. In recent years, large language models (LLMs) have developed rapidly and can also complement imaging information. The main challenge lies in how to prompt multimodal information fusion systems to provide accurate information and integrate it with the image features of WSIs.

Based on the above analysis, this paper presents a comprehensive framework for the interpretability of WSI analysis. First, using survival analysis as an example in the clinical context of hepatocellular carcinoma (HCC), we construct an interpretable WSI analysis paradigm based on the attention mechanism; Building upon this foundation, an efficient prototype-based WSI encoding method is proposed. The WSI analysis framework is subsequently applied to a novel clinical problem (detection of Vessels that Encapsulate Tumor Clusters (VETC), prognostic analysis, and evaluation of different treatments) to validate its clinical feasibility.

1.2 Deep Learning Techniques in WSI Analysis

The rapid development of deep learning (DL) techniques in recent years has significantly enhanced the efficiency of clinical diagnostics in medical imaging. As a representative form of medical imaging, the analysis, diagnosis, especially survival prediction of WSI have also benefited from the advancements in DL techniques. This section will analyze DL-based WSI feature encoding methods and their applications in

survival analysis.

1.2.1 Survival Prediction based on WSI

The goal of survival prediction based on WSI is to assess how a new patient will survive in the context of known patient image data. Utilizing survival prediction enables clinicians to make early treatment decisions, which is critical for improving patient healthcare outcomes[1].

The existing tumor staging systems, e.g. Tumor, Node, Metastasis (TNM) staging system[2], have been widely used in clinical practice. However, even within the same stage, there remains significant variability in the prognosis of different patients[3]. Therefore, there is a need to further explore the implicit information within WSI to refine the existing staging systems for more precise prognostic analysis.

The current work on applying DL to survival prediction primarily follows two approaches: the first is based on black-box DL models that construct a direct mapping from WSI to the probabilities of recurrence or death[4]. These models have achieved competitive performance in the survival prediction task[5][6]; however, due to their lack of interpretability, it is difficult to apply them directly to clinical practice. The other is to use DL as an aid to more efficiently compute clinical indicators, such as microvascular invasion (MVI)[7] or mitotic count[8], which are then used to assess the survival status of patients. However, these indicators often come from a single view. In fact, WSIs contain richer features spanning from the microscopic to the macroscopic, in this paper, we construct indicators from multiple perspectives to provide a more comprehensive evaluation and thus achieve a more accurate survival prediction.

1.2.2 WSI Feature Encoding Methods

As mentioned earlier, due to the extremely high resolution of WSI, it is impractical to directly input the entire image into deep learning models. Consequently, the common approach involves cropping the WSI into multiple patches, which are then sequentially processed by offline encoders[1][10] to extract features. However, even after this operation, tens of thousands of features are still generated. Therefore, the feature encoding of WSI primarily focuses on how to integrate the features from these numerous patches.

Sharma et al.[11] employed the k-means clustering approach to distinguish the feature information of different patches. They utilized the unsupervised method to aggregate similar features into several groups. The subsequent operations were performed at the cluster level, which significantly reduced the computational complexity. However, despite significantly reducing the computational complexity, distinguishing features of different tissues solely based on the distance in the feature space can easily lead to confusion. This is because, unlike normal images, the tissue features in WSIs do not exhibit such distinct differences[12].

Another approach is to integrate these patch-level features based on the attention mechanism. Some global attention mechanisms[13][14] are applied to assign higher weights to patches that are more relevant to downstream tasks. However, these mechanisms can often result in feature homogeneity, neglect important local correlations, and fail to capture variations between local regions. Some enhanced models introduce redundant local attention mechanisms[15] to emphasize localized areas. While this improves sensitivity to fine-grained details, it also increases

computational cost.

Therefore, we proposed a novel method named FuzzyMIL that integrates the Fuzzy C-means clustering approach into the cross-attention mechanism, re-encoding the patch-level features of WSI into disentangled features. This method not only reduces computational complexity but also preserves global relevance while emphasizing local features.

1.3 Challenges of Deep Learning-Based WSI Analysis in Clinical Practice

In the previous subsection, we summarized the supportive role of DL techniques In WSI analysis, particularly in survival prediction. Although these novel techniques have demonstrated competitive performance, directly applying them in clinical settings still presents challenges. These challenges primarily manifest in two aspects:

First, the interpretability of DL techniques. As mentioned, clinical practice demands precise diagnoses, while most black-box models lack interpretability, and as a result, they are not considered trustworthy. Some existing methods have explored the interpretability of DL in WSI analysis, aiming to compare the regions of focus during decision-making with those of human clinicians. For instance, attribution-based methods have been proposed to analyze to localize potential biomarkers in WSI, but their interpretability largely depends on the predictive capability of the model itself[16]. The other approach is based on attention mechanisms[17], although it more intuitively highlights the importance of different regions, attention-based approaches lack clinical priority, which reduces their credibility. Therefore, the interpretable framework for WSI analysis in this paper is primarily based on clinical priors, which first help to identify

several regions of interest. Attention mechanisms are then applied on this basis, providing physicians with references to the potential high-risk regions. In addition, to further validate the clinical applicability of DL-based techniques for WSI, we also applied the method to a novel clinical task involving the automated detection and prognostic analysis of VETC.

The second challenge is the efficiency of WSI analysis. Given that the resolution of a single WSI can reach billions of pixels, extracting meaningful information from such a vast number of pixels while minimizing model complexity has been the focus of current research. Some attention-based methods tend to focus on extracting visually specific information rather than clinically relevant features. Additionally, for WSI-level downstream tasks, such as automated diagnosis, considering the features of all tissues is not meaningful. This is because WSIs contain redundant information, and incorporating all these high-dimensional features into the model not only increases computational complexity but also inevitably introduces noise. Therefore, this paper expects to develop a more efficient WSI analysis framework that not only effectively extracts clinically relevant information from WSIs but also significantly reduces the computational complexity of downstream tasks.

1.4 Aim and Objectives

In an attempt to mitigate the remaining challenges introduced before, this thesis aims to **develop an efficient and transferable WSI analysis paradigm** for cancer diagnosis, subtyping, and survival prediction, while ensuring interpretability to address practical clinical challenges. Three objectives are achieved in the thesis:

1) To provide a new DL-assisted workflow from the identification of potentially high-risk tissues to the stratification of patients into high and low-risk groups. Taking HCC patients as an example, this approach supplements existing clinical risk staging systems to achieve more precise cancer patient risk assessment.
2) To more efficiently encode WSI and mitigate the drawbacks of current clustering or attention-based methods, a novel WSI encoding strategy is proposed, which is based on disentangled features to both reduce the impact of redundant information and fully exploit regional information.
3) To apply the proposed WSI analysis paradigm to a new clinical issue—evaluation of VETC patients. A DL-assisted system is constructed to automatically detect VETC+ patients while also conducting survival analysis and treatment efficacy evaluation.

1.5 Thesis Overview

This thesis will first review previous literature on WSI analysis, focusing on preprocessing, feature extraction, feature fusion, and learning frameworks. In the following chapter, which addresses the first objective, a WSI-based patient risk assessment paradigm is constructed in the context of HCC. This paradigm includes the identification of potential high-risk regions and survival analysis (this study is reproduced with permission from Springer Nature). The following chapter discusses how to improve the efficiency of WSI analysis. We validate the effectiveness of our proposed FuzzyMIL framework on three sub-tasks across three datasets. In Chapter five,

we apply the WSI analysis framework to a new clinical task: survival analysis of VETC patients and the evaluation of the efficacy of different treatment methods. Finally, we conclude the thesis by reviewing the key results and discussing the limitations and potential directions for future research.

Chapter 2.

Literature Review

Due to the gigapixels of WSI, the application of DL techniques requires unique processes in preprocessing, feature extraction, feature fusion, and learning strategies. This chapter reviews related work from the above aspects to provide a foundation for the discussions in the following chapters.

2.1 Preprocessing of WSI

Before downstream tasks, it is necessary to preprocess WSIs. Typically, WSIs with gigapixels are cropped into multiple patches to facilitate subsequent operations. However, not all patches are valid, for example, patches with excessive blank areas or artifacts may introduce redundant or noisy information. Clustering-constrained-attention multiple-instance learning (CLAM)[10] proposes a general procedure for dividing WSIs into patches of specified pixel size at a certain magnification level as shown in Figure 2-1. In the following sections, unless otherwise specified, we follow the CLAM procedure for cropping WSIs into patches.

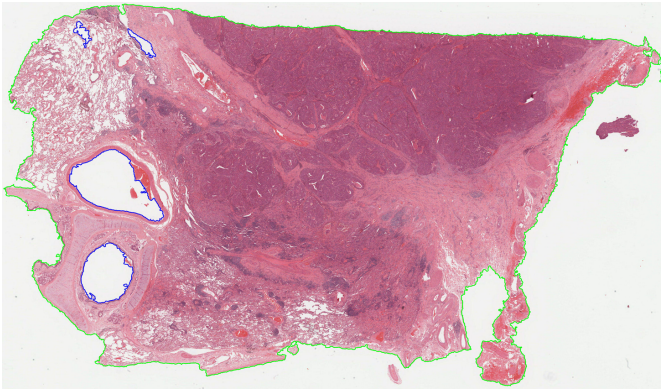


Figure 2-1 CLAM removes the void areas and discontinuous regions from the WSI.

2.1.1 Staining Normalization

As shown in Figure 2-2, using H&E (Hematoxylin and Eosin)-stained WSI as an example, WSIs may exhibit significant variations in staining space due to differences in scanning devices or staining conditions, which in turn affects the subsequent analysis. Therefore, it is essential to eliminate these color discrepancies.

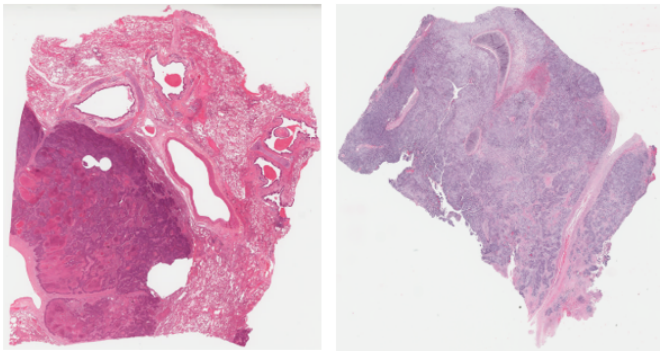


Figure 2-2 Staining variations exist within the WSIs of the same dataset.

Histogram matching is a widely used method for staining normalization. It works by adjusting the color distribution of the source image to align with that of the target image[20]. This non-parametric approach is convenient and effective, although it may result in some color distortion in some cases.

With the advancement of DL, DL-based methods for staining normalization have gained significant popularity. These approaches leverage annotated datasets to train models that autonomously learn to map images from diverse sources to a common color space[21][22]. DL techniques are capable of handling complex variations and provide more accurate normalization results.

2.2 Feature Extraction

Early research in feature extraction of WSI primarily focused on manually

identifying features, which is both time-consuming and labor-intensive. Moreover, this approach often fails to uncover deeper and more complex features hidden within WSIs. The quality of feature extraction directly influences the accuracy of downstream tasks.

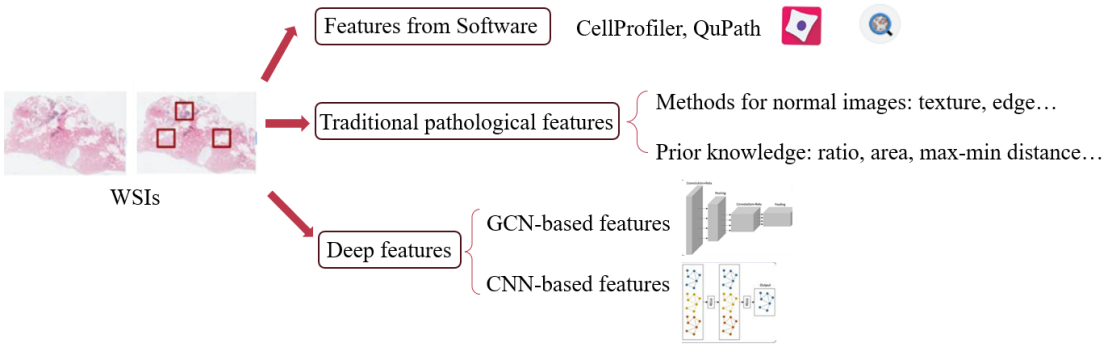


Figure 2-3 Methods for feature extraction from WSIs.

Therefore, it is necessary to determine how to extract useful features from high-resolution WSI. As shown in Figure 2-3, we categorize them into three main types based on their properties.

There exists freely available open-source software that enables researchers to extract features from Regions of Interest (ROIs) within WSIs. Notably, CellProfiler and QuPath are among the most widely used tools for extracting pathology image features. CellProfiler[23] is specifically designed to allow pathologists to automatically and quantitatively assess the phenotypes of thousands of pathological images, without requiring expertise in computer vision or programming.

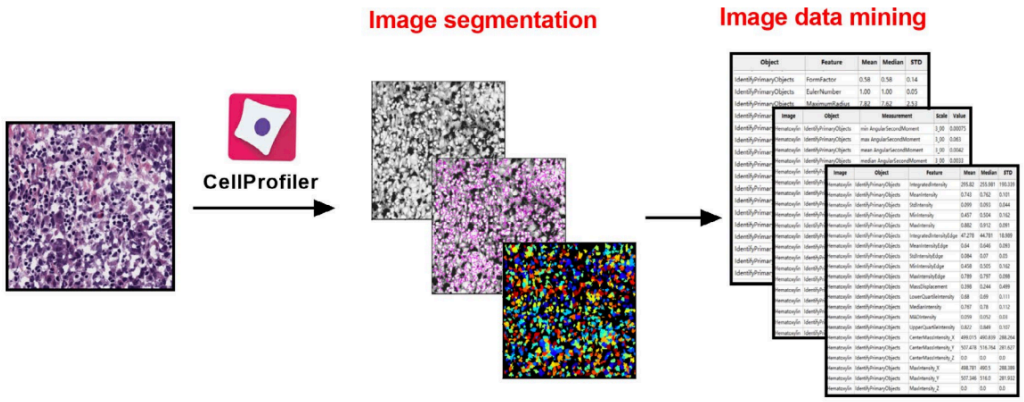


Figure 2-4 Workflow of CellProfiler.

As shown in Figure 2-4, initially, one or more ROIs are selected randomly or based on a priori knowledge. These ROIs are then processed through the 'UnmixColors' module to separate the color components. Following this separation, the images are automatically analyzed by two modules: 'IdentifyPrimaryObjects' for initial object identification and 'IdentifySecondaryObjects' for refinements based on the primary objects detected. The final step involves the extraction of quantitative image features, including the shape, size, texture of objects, the distribution of pixel intensity, etc., offering a comprehensive toolkit for the detailed analysis of pathological images. This streamlined workflow significantly improves the efficiency and accuracy of phenotype measurement in pathological research. QuPath[24] is another powerful software for reading, annotating, and analyzing pathology images, and it plays a significant role in the extraction of pathological features. The features extracted by QuPath are mainly at the cellular level, e.g., counting of cells in the ROIs.

Due to their user-friendly interfaces, CellProfiler and QuPath not only facilitate the extraction of a wide array of features but also their integration, thereby enriching the feature domain for analysis. Researchers often leverage features extracted by both

platforms for downstream tasks. Besides, signatures derived from these features are sometimes utilized as baseline indicators in various studies[25][26], which allows for more nuanced and detailed investigations.

While leveraging open-source software for feature extraction is both straightforward and efficient, the resulting features often lack specificity, which poses a challenge for researchers who must sift through thousands of potential features to identify the most relevant ones. Designing task-specific features for different downstream applications can significantly enhance the accuracy and effectiveness of the analysis. This customization ensures that the extracted features are directly aligned with the specific objectives of the study.

For example, in[27], MVI was automatically detected, and factors like area, count, and distance between MVI instances and HCC were calculated to predict prognosis after R0 hepatectomy. Another approach adapts feature extraction methods from normal images to pathological images, focusing on texture and edges. For example, Color Feature Discrimination with Markov models was used for prostate cancer detection[28], and Grey Level Co-occurrence Matrices with k-means clustering was used for tumor segmentation[29]. While features extracted using these methods are interpretable, they also lack a comprehensive understanding of WSIs, the same as extracting features using software. Furthermore, when dealing with high-resolution images, these methods remain time-consuming.

The integration of DL techniques into digital pathology represents a significant advancement in the field. DL techniques can significantly improve the efficiency of

feature extraction of WSIs, and they have been successfully applied to tasks such as cancer diagnosis, grading, and nucleus classification[30][31][32].

DL-based pathology usually employs CNNs, and the input data are patches with relatively small resolutions (for example, 256×256) obtained from splitting the WSI.

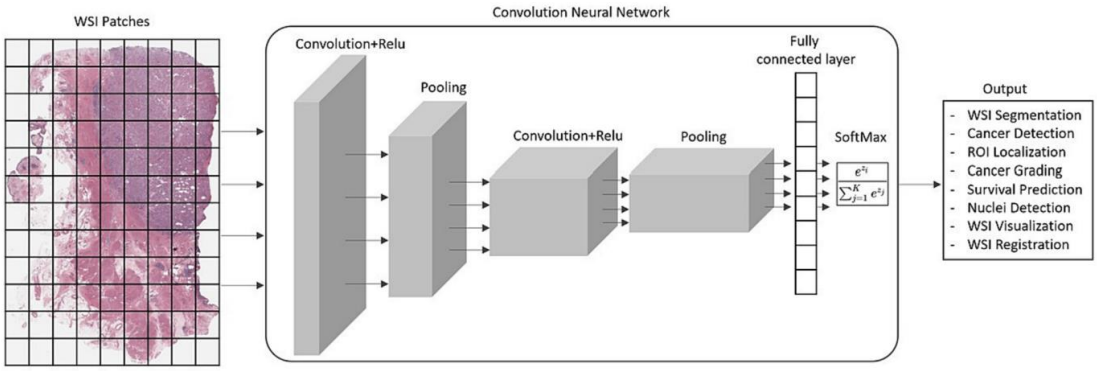


Figure 2-5 Workflow of deep feature extraction based on CNN.

As shown in Figure 2-5, these features can be used as input to other deep learning models or fully connected layers for classification, segmentation, and detection tasks. In addition, with the development of LLMs, some foundation models have been proposed to encode both visual and textual information, achieving alignment of text and visual information in the feature space. Benefiting from CLIP[33], the specialized visual-language model CONCH[34] for pathology images was pre-trained on over 1.17 million image-caption pairs, histopathology images, and biomedical text, demonstrating competitive performance across 14 downstream tasks. These visual-language model-based encoders play an increasingly important role in LLMs for supplementing prior knowledge.

Another approach is based on graph convolutional network (GCN), it was developed based on the concept of graph convolutions, where features of each node's relevant neighborhood are aggregated through convolutional layers. In GCN-based

methods, graphs are constructed using the spatial attributes and structure of the patches in WSIs[35].

The employment of DL-based methods for feature extraction has significantly improved the efficiency of WSI analysis. However, these approaches often lack interpretability due to the inherent complexities of the networks. Based on this, using a hybrid approach for extracting various types of features is more reasonable. This method not only enhances the analytical robustness of the model but also improves its transparency, thereby providing physicians with more accurate diagnostic guidance.

2.3 Feature Aggregation

Feature integration based on WSI analysis can be divided into two aspects: first, the integration of image features from the WSI. As discussed in Section 2.2, since the number of image features extracted from WSIs can reach tens of thousands, it is essential to aggregate or filter these features effectively; The second aspect is the integration of features from different categories. In addition to features of WSIs extracted by DL models (deep features), some studies may involve features from different categories or modalities, and it is also crucial to integrate these features in the feature space.

2.3.1 Fusion of Deep Features from WSI

As discussed above, the integration of deep features from WSIs is commonly based on clustering and attention mechanisms. Existing work typically combines these two approaches (see Figure 2-6), where clustering is used to sample tens of thousands of features in the WSI and select representative subset; The attention mechanism is then

applied to integrate the features from these subsets for slide-level prediction.

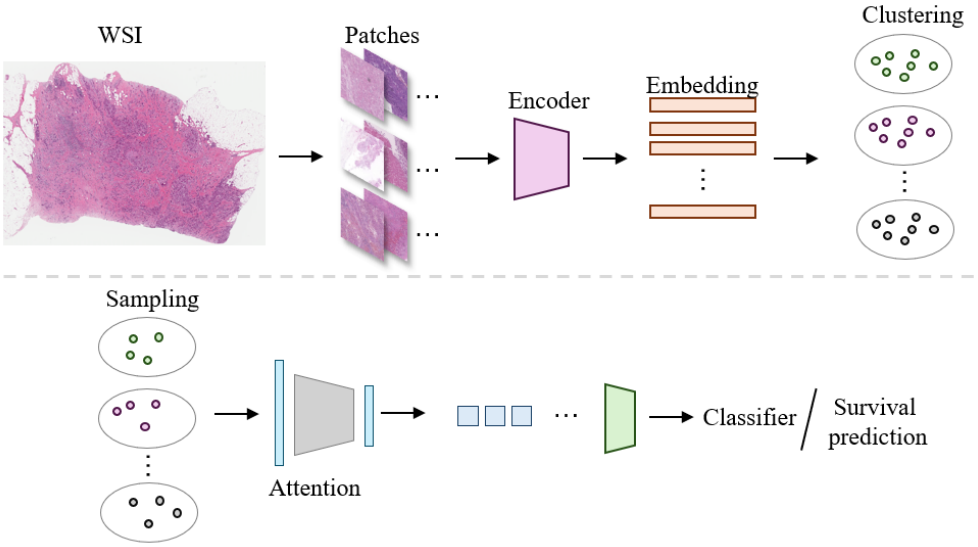


Figure 2-6 General procedure for WSI feature fusion based on clustering and attention mechanisms.

For example, Sharma et al.[11] demonstrated that partitioning WSI patch features through clustering can enhance the training of downstream tasks; Wu et al.[36] enhancing representation learning for histopathological images through cluster constraints and in [37], semantic relevance clustering with multi-granularity information was used to achieve cross-domain knowledge transfer.

For attention-based aggregation, Ilse et al. proposed the adaptive and flexible pooling method through a two-layer neural network to calculate weights for each patch feature. Let $\boldsymbol{H} = \boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_N$ be N patch features within the WSI or a cluster obtained by clustering the features of the WSI, then

$$\boldsymbol{z} = \sum_{n=1}^{N} \alpha_n \boldsymbol{h}_n \qquad (2-1)$$

where

$$\alpha_n = \frac{\exp\{\boldsymbol{\sigma}_2{}^{\top}\tanh(\boldsymbol{\sigma}_1\boldsymbol{h}_n{}^{\top})\}}{\sum_{j=1}^{N}\exp\{\boldsymbol{\sigma}_2{}^{\top}\tanh(\boldsymbol{\sigma}_1\boldsymbol{h}_j{}^{\top})\}} \quad (2-2)$$

and $\boldsymbol{z}$ denotes the aggregated representation, $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are learnable parameters from attention mechanism, then the attention weight α_n of $\boldsymbol{h}_n$ can be obtained through (2).

2.3.2 Fusion of Multimodal Features

The above primarily discussed the integration methods of pathological image features. In fact, for many tasks, multimodal information is also considered. Fremond et al.[39] extracted interpretable morphological features through deep features. Subsequent analyses linked these morphological features to molecular classes and evaluated their relative importance using a support vector machine. Furthermore, genetic information is often considered as another modality for WSI analysis and downstream tasks. Li et al.[40] proposed an interpretable progressive multimodal fusion network, which aligns pathological and genetic features with downstream clinical tasks. In this work, the attention mechanism (2) is also used to fuse genomic and pathological features. Another work integrating pathological and genetic features by constructing a pathology-genome heterogeneous graph[41]. They developed a feature fusion framework within each modality and concatenated the information from both modalities for downstream tasks.

Another common modality is textual information. In fact, text and corresponding visual information form the basis of visual-language foundation model. There are already well-established methods for aligning visual and textual information in normal

images. For instance, Contrastive Language-Image Pretraining (CLIP)[33] utilizes contrastive learning on 400 million image-text pairs to increase the similarity between corresponding image-text pairs and decrease the similarity between non-corresponding pairs. However, model training is often challenging due to the scarcity of labels in the medical domain. To address this challenge, Huang et al. developed PLIP[42] through a large dataset of pathology images paired with descriptions named OpenPath. Based on this, Lu et al. developed a visual-language foundation model for histopathology called CONCH[34] through task-agnostic pretraining using over 1.17 million image-caption pairs.

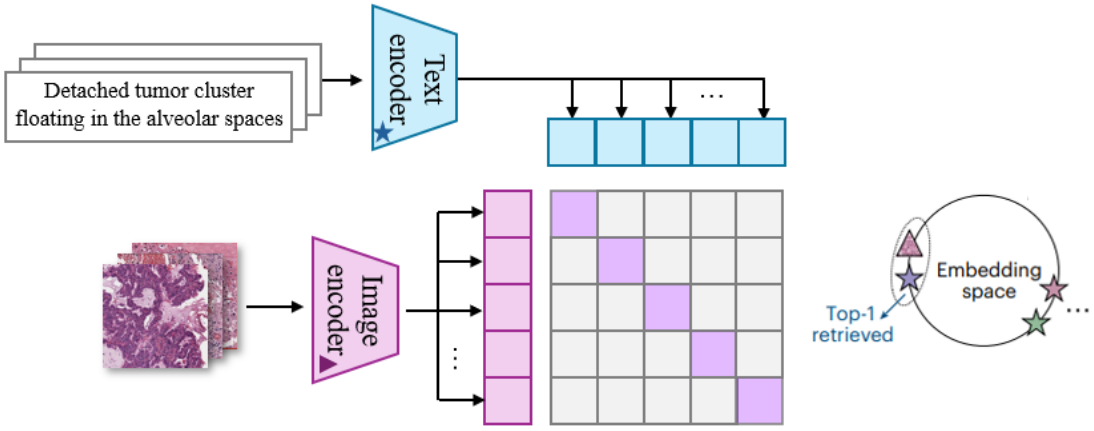


Figure 2-7 Pretraining and zero-shot image-to-text retrieval of CONCH.

The pretraining and zero-shot processes of CONCH are illustrated in Figure 2-7. The pretraining approach is similar to CLIP, for zero-shot, CONCH computes the similarity between the input image features and the latent category text embeddings and selects the most similar category as the output.

In addition to common textual and genetic features, WSI features can be fused with features from other modalities to improve the accuracy of downstream clinical tasks, e.g., Song et al.[43] combined WSI with CT features to predict outcomes in HPV-

associated oropharyngeal squamous cell carcinoma; Yeh et al.[44] integrated WSI with corresponding ECG data for histopathological classification of ischemic stroke clot origin and[45] combined diagnostic reports and other clinical information with pathological features.

In summary, integrating pathological information with diverse modalities such as radiological features, clinical data, and physiological signals enhances the performance of clinical models, providing a more comprehensive understanding of diseases.

2.4 Multiple Instance Learning

As discussed above, due to the large pixel size of WSI, directly inputting WSI into DL models is impractical. Therefore, for a WSI dataset $\{W_i\}_{i=1}^{L}$ with L WSI slides, each W_i is cropped into N_i patches $\{p_i^m\}_{m=1}^{N_i}$. For the supervised learning strategy, labels of a large number of patches are available for training, and the goal is obtaining a model f_θ, s.t. $p \rightarrow y$, where y is the corresponding label predicted by f_θ. This supervised learning approach requires pathologists to label tens of thousands of patches, which is a labor-intensive task. Therefore, the analysis of WSI generally adopts the Multiple Instance Learning (MIL) framework. MIL as a form of weakly supervised learning, plays a crucial role in WSI analysis.

For the weakly supervised learning framework, only the label for the WSI is known, while the label y for each patch is not accessible. Therefore, the supervised learning paradigm is not applicable in this case.

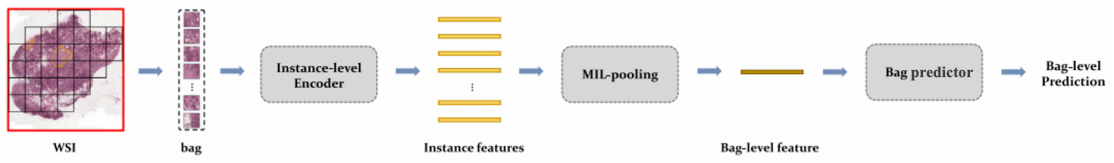


Figure 2-8 Workflow of bag-level WSI analysis based on MIL[46].

As illustrated in Figure 2-8, in MIL, each WSI is treated as a bag containing multiple patches (instances). For classification tasks, if a WSI is positive, at least one of the instances within the WSI is positive. Conversely, if the WSI is negative, all instances within the WSI are negative.

Specifically, for a WSI dataset $\{W_i\}_{i=1}^{L}$ with L WSI slides, and corresponding WSI-level labels $Y_i \in \{0, 1\}$, each W_i (bag) is cropped into N_i patches $\{p_i^m\}_{m=1}^{N_i}$ (instance) without patch-level labels. The relationship between patch-level labels and WSI-level labels is as follows:

$$Y_i = \begin{cases} 0, if \ \sum_j y_{i,j} = 0 \\ 1, else \end{cases} \quad (2-3)$$

The WSI analysis based on MIL aims to achieve two primary objectives: the first is WSI classification, and the second is the localization of positive patches.

2.4.1 WSI-based Classification

WSI-based classification is widely applied in various clinical tasks, such as diagnosis, subtyping and grading. Patch (instance)-level features are extracted and integrated followed by different downstream networks designed based on the specific task. In Section 2.3, we have discussed various methods of feature fusion. In this subsection, we focus on different downstream networks.

Due to the potential imbalance between positive and negative instances in WSI bags, to address this class imbalance issue, [47] introduced a self-supervised contrastive

learning approach to extract more meaningful representations in MIL, thereby further improving the accuracy of both classification and localization. Meanwhile, TransMIL[48] explored morphological and spatial information, and further addresses the issue of class imbalance. Recent work[49] has explored the use of structured state space models for modeling long sequences of patches. However, these approaches are limited to a single resolution, potentially overlooking contextual differences.

To address the limitation of existing MIL methods which overlook hierarchical label correlations in fine-grained classification, Jin et al.[50] introduced a hierarchical multi-instance learning framework that incorporates a class-wise attention mechanism and supervised contrastive learning. Additionally, a curriculum-based dynamic weighting module is introduced to balance hierarchical features during training.

Another approach to capturing spatial correlations between instances is based on GCN. H^2MIL[51] utilized GCN to aggregate the multi-resolution information through an iterative pooling layer based on patches' location, while this method failed to consider the hierarchical arrangement of patches and their diversity. In [52], Bontempo et al. proposed graph-based architecture that considered correlations within and across scales and introduced a distillation loss to compensate for the information gap between different scales and further enhance prediction efficiency.

2.5 Survival Prediction

The labels and loss function for survival prediction are different from those used in classification. Specifically, survival prediction data consists of three primary components: the initial patient data x, the time T to a failure event, and the event

occurrence indicator *E*. When an event (e.g. death) is recorded, *T* represents the period from when initial data was collected to the occurrence of the event, with the event indicator set to $E = 1$. Conversely, if the event is not recorded, *T* reflects the duration from data collection to the last known contact with the patient, for instance, at study completion, the event indicator *E* is set as 0. This scenario is referred to as right-censoring.

Two critical concepts in survival analysis are the survival and hazard functions. The survival function is expressed as $S(t) = \Pr(T > t)$, which indicates the likelihood of an individual surviving past a certain time *t*. The hazard function $\lambda(t)$ is defined as:

$$\lambda(t) = \lim_{\delta \to 0} \frac{\Pr(t \leq T < t + \delta \mid T \geq t)}{\delta} \qquad (2-4)$$

The hazard function is the probability an individual will not survive an extra infinitesimal amount of time δ, given they have already survived up to time *t*. Thus, a greater hazard signifies a greater risk of death[53].

The Cox proportional hazards model[54] is a common method for modeling an individual's survival given their baseline data *x*, in our task, *x* represents features extracted from WSIs. The model assumes that the hazard function is composed of two non-negative functions: $\lambda_0(t)$ is a baseline hazard function, and $r(x) = e^{h(x)}$ is a risk score, which is defined as the effect of an individual's observed covariates on the baseline hazard. $h(x)$ is denoted as the log-risk function. The hazard function is defined as:

$$\lambda(t|x) = \lambda_0(t) \cdot e^{h(x)} \qquad (2-5)$$

It is noted that most of these traditional survival prediction methods based on Cox

are linear[55][56], but in many applications, for example, modeling nonlinear gene interactions, we cannot assume the data satisfies the linear proportional hazards condition. In this case, a more complex nonlinear model is needed. The Faraggi-Simon method is a feed-forward neural network that provides the basis for a nonlinear proportional hazards model[57]. It experimented with a single hidden layer network with two or three nodes. Their model requires no prior assumption of the log-risk function $h(x)$ other than continuity. Instead, the neural networks compute nonlinear features from the training data and calculate their linear combination to estimate the log-risk function. Another popular machine learning approach to modeling patients' hazard function is the random survival forest (RSF)[58]. The random survival forest is a tree method that produces an ensemble estimate for the cumulative hazard function. The Nelson–Aalen estimator for the cumulative event-specific hazard function is given as:

$$\widehat{H}_j(t) = \int_0^t \frac{\mathrm{d}N_j(s)}{Y(s)} = \sum_{k=1}^{m(t)} \frac{d_j(t_k)}{Y(t_k)} \qquad (2-6)$$

where $d_j(t_k) = \sum_{i=1}^{n} I(T_i = t_k, \delta_i = j)$ is the number of type j events at t_k, $N_j(t) = \sum_{i=1}^{n} I(T_i \le t_k, \delta_i = j)$ is the number of type j events in $[0,t_k]$ and $Y(t) = \sum_{i=1}^{n} I(T_i \ge t)$ is the number of individuals at risk (event-free and uncensored) just prior to t.

These methods are typically unsupervised, so when the amount of the dataset is relatively small, traditional approaches can yield relatively stable results. However, these methods are very time-consuming when dealing with large volumes of data. Thus, it is necessary to explore more efficient techniques for constructing mappings from images to prognosis.

Due to the differences mentioned above compared to classification tasks, and the

fact that data containing survival information is typically more limited, survival analysis becomes more challenging. In[59], they assessed tumor proliferation in breast cancer and[60] quantified the stroma/tumor ratio in colorectal cancer. In these works, they utilized the image features that expert pathologists have already identified as being associated with survival.

In order to adaptively apply the above priori-based features to survival prediction, based on the traditional Cox prediction model, DeepSurv[53] was proposed.

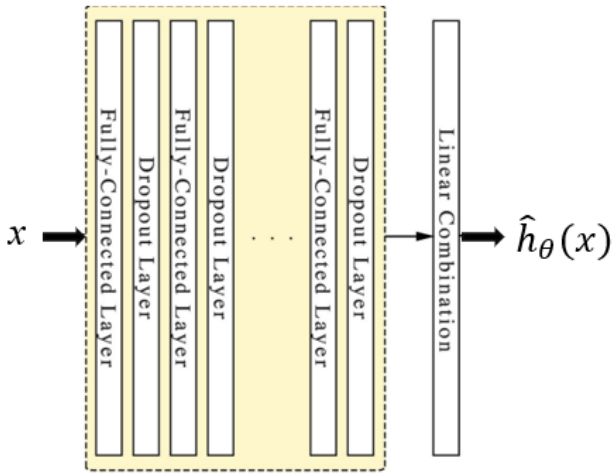


Figure 2-9 DeepSurv is a configurable feed-forward deep neural network.

As shown in Figure 2-9, DeepSurv is a deep feed-forward neural network which predicts the effects of a patient's covariates on their hazard rate. The network propagates the inputs through a number of hidden layers with weights θ. The hidden layers consist of fully connected nonlinear activation functions followed by dropout. The final layer is a single node which performs a linear combination of the hidden features. The output of the network $\hat{h}_\theta(x)$ is a single node with a linear activation which estimates the log-risk function in the Cox model (2-5). The network by setting the objective function as:

$$l(\theta) = -\frac{1}{N_{E=1}} \sum_{i:E_i=1} \left(\hat{h}_\theta(x_i) - \log \sum_{j \in \mathbb{R}(T_i)} e^{\hat{h}_\theta(x_j)} \right) + \lambda \cdot \|\theta\|_2^2 \qquad (2-7)$$

where $N_E = 1$ is the number of patients with an observable event and λ is the ℓ_2 regularization parameter.

In addition to the basic DeepSurv model, MIL is also widely applied to survival prediction tasks in WSI. Li et al.[61] introduced linear bias into attention to modify position embedding for handling shape-varying long-contextual WSIs. Shao et al.[62] considered spatial interactions tumors and their two primary components of the tumor microenvironment and introduced a tumor microenvironment interaction guided graph learning algorithm for predicting cancer prognosis.

In survival analysis, in addition to the traditional Cox proportional hazards model and the basic DeepSurv model, MIL has also been widely applied for survival prediction through the analysis of WSI.

2.6 Conclusion

In this section, we review previous work in several aspects, including the preprocessing of WSI, feature extraction, feature fusion, MIL learning strategies, and survival prediction. Due to the ultra-high resolution of WSIs, they are typically cropped into patches to facilitate subsequent analysis. For feature extraction, manual feature extraction is time-consuming and labor-intensive, while greatly enhancing efficiency, suffer from a lack of interpretability. Therefore, an ideal approach is to com bine these two methods to enhance their clinical utility. As for feature fusion, in addition to the fusion of features from different instances within the WSI, there is also the fusion of

features extracted through different strategies and the fusion of multi-modal features. A reasonable fusion approach can achieve information complementarity and reduce redundancy. Additionally, we review MIL strategies used in WSI analysis, which is the most commonly employed strategy in DL for WSIs. In subsequent chapters, we will also introduce improvements to this strategy. Finally, we provide a detailed introduction to survival analysis based on pathological images, laying the foundation for subsequent survival analysis tasks.

Chapter 3.

Paradigm for Multi-View Prognostic Risk Score Development

3.1 Introduction

Existing prognostic staging systems depend on expensive manual extraction by pathologists, potentially overlooking latent patterns critical for prognosis, or use black-box deep learning models, limiting clinical acceptance. This section introduces a novel deep learning-assisted paradigm that complements existing approaches by generating an interpretable, multi-view risk score named Hybrid Deep Score (HDS) to stratify prognostic risk in hepatocellular carcinoma (HCC) patients.

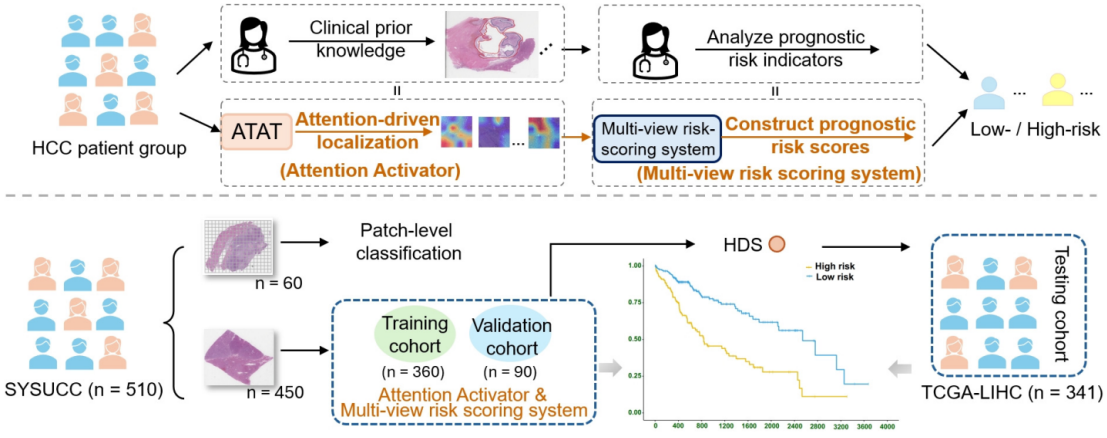


Figure 3-1 Illustration of the proposed paradigm for multi-view prognostic risk score development, which mainly includes two components: the identification of potential high-risk regions and the prediction of risk stratification.

As shown in Figure 3-1, this paradigm, from identifying high-risk tissues to constructing prognostic risk scores, offers fresh insights into HCC research. Additionally, the integration of HDS complements the existing clinical staging system

by facilitating more detailed stratification in Disease Free Survival (DFS) and Overall Survival (OS). The novel paradigm we propose provides diagnostic assistance in two aspects: First, for the localization of potential high-risk tissues, we introduce an attention-driven mechanism that explicitly highlights suspicious tissue areas through attention heatmaps and further obtains their spatial distribution across WSIs. Second, for the development of risk scores, differing from existing single-perspective indicators, we leverage DL methods to integrate features from three different perspectives to obtain a hybrid risk score, thereby refining the current clinical staging system.

3.2 Patient Cohort and Study Design

In this study, we collected WSIs from 510 patients diagnosed with HCC at the Sun Yat-sen University Cancer Center (SYSUCC) to form our in-house dataset. Among these, 365 patients received treatment between March 2013 and December 2014, while the remaining 145 were treated between August 2020 and June 2022. Each WSI corresponded to a unique patient. The criteria for inclusion were as follows: (1) pathologically confirmed HCC; (2) absence of concurrent malignancies; and (3) no extrahepatic metastases. Exclusion criteria included: (1) incomplete clinical information; (2) poor-quality WSIs; and (3) presence of other histological types of liver carcinoma. Following the same criteria, we also included 341 WSIs from The Cancer Genome Atlas Liver Hepatocellular Carcinoma dataset (TCGA-LIHC)[1] to serve as an external public dataset.

From the SYSUCC dataset, we randomly selected 60 WSIs, segmented them into

[1] https://portal.gdc.cancer.gov/projects/TCGA-LIHC

150 × 150-pixel patches at 20× magnification, and annotated six distinct tissue types: Tumor (1,041 patches), Necrosis (1,017), Fibrosis (1,007), Lymphocytes (1,029), Normal Liver Cells (1,041), and Others (e.g., blood vessels and steatosis, 994), these collectively form the patch-level dataset. These categories represent the most identifiable and representative tissue structures observed in HCC pathology.

HDS was developed using the remaining 450 WSIs from SYSUCC, with an 80:20 split for training and validation, respectively. The generalizability of the paradigm was then tested on the TCGA-LIHC dataset, which encompasses diverse treatment backgrounds.

Table 3-1 Demographic, clinical, and tumor characteristics of the training and validation cohorts.

Demographics	Training cohort (n = 360)	Validation cohort (n = 90)	p-value
Sex			0.82
Male	306 (85)	75 (83.33)	
Female	54 (15)	15 (16.67)	
Age			0.87
≤55	222 (61.67)	57 (63.33)	
>55	138 (38.33)	33 (36.67)	
Cirrhosis			0.87
No cirrhosis	154 (42.78)	40 (44.44)	
With cirrhosis	206 (57.22)	50 (55.56)	
Diameter			0.35
≤5cm	214 (59.44)	59 (65.56)	
>5cm	146 (40.56)	31 (34.44)	
Cancer Embolus			0.43
No cancer embolus	273 (75.83)	64 (71.11)	
With cancer embolus	87 (24.17)	26 (38.89)	
Lesion			1
Single	310 (86.11)	77 (85.56)	
Multiple	50 (13.89)	13 (14.44)	
AFP			0.44
<20 ng/ml	160 (44.44)	34 (37.78)	
20-400 ng/ml	94 (26.11)	24 (26.67)	
>400 ng/ml	106 (29.44)	32 (35.56)	

MVI			0.15
No MVI	181 (50.28)	37 (41.11)	
With MVI	179 (49.72)	53 (58.89)	
ALT			0.61
≤40U/l	149 (41.39)	34 (37.78)	
>40U/l	211 (58.61)	56 (62.22)	
AST			0.62
≤40U/l	256 (71.11)	61 (67.78)	
>40U/l	104 (28.89)	29 (32.22)	
HBV			0.94
Negative	100 (27.78)	24 (26.67)	
Positive	260 (72.22)	66 (73.33)	
BCLC			1
0-A	322 (89.44)	80 (88.89)	
B	38 (10.56)	10 (11.11)	

**ALT* alanine aminotransferase, *AST* Aspartate aminotransferase, *AFP* alpha-fetoprotein, *MVI* microvascular invasion, *HBV* Hepatitis B Virus, *BCLC* Barcelona Clinic Liver Cancer

The clinicopathological characteristics of the SYSUCC cohort used for the development of HDS are summarized in Table 3-1. This cohort comprises a training set (n = 360) and a validation set (n = 90). Comparative analysis reveals no significant differences in baseline features between the two groups.

Table 3-2 Demographic, clinical, and tumor characteristics of the patients for patch-level classification network.

Demographics	Patients for patch-level classification network (n=60)
Sex, male	51 (85)
Age, >55	22 (36.67)
Cirrhosis, positive	34 (56.67)
Diameter, >5cm	23 (38.33)
Cancer Embolus, positive	17 (28.33)
Lesion, multiple	9 (15)
AFP, >400 ng/ml	19 (31.67)
MVI, positive	31 (51.67)
ALT, >40 U/l	35 (58.33)
AST, >40 U/l	18 (30)
HBV, positive	46 (76.67)
BCLC, 0-A	53 (88.33)

**ALT* alanine aminotransferase, *AST* Aspartate aminotransferase, *AFP* alpha-fetoprotein, *MVI* microvascular invasion, *HBV* Hepatitis B Virus, *BCLC* Barcelona Clinic Liver Cancer

Additionally, Table 3-2 details the clinicopathological features of an additional 60 patients from the SYSUCC cohort, who were included for the development of a patch-level classification network. The distribution of clinical and pathological features in this subset is comparable to that of the cohort described in Table 3-1.

Table 3-3 Demographics of the patients of external public dataset TCGA-LIHC.

Demographics	Patients of external public dataset (n=341)
Sex, male	230 (67.54)
Age, >55	218 (63.74)
Depth of invasion, T1 & T2	253 (74.27)
Tumor grade, G1 & G2	207 (60.53)

Furthermore, Table 3-3 describes the external testing cohort consisting of 341 patients from the TCGA-LIHC dataset. As this cohort is not involved in the development of HDS, it serves to independently assess the generalizability of HDS to unseen clinical data.

3.3 Preprocess

All tissue samples were fixed in 4% neutral-buffered formaldehyde, embedded in paraffin, sectioned at a thickness of 4 μm, and stained with hematoxylin and eosin (H&E). All patches extracted from the WSIs followed the preprocessing steps illustrated in Clustering-constrained Attention Multiple Instance Learning (CLAM)[10] to filter out patches containing more than 75% blank areas.

Due to the gigapixel scale of WSIs, accurately classifying and localizing tissue patches is a critical preliminary step for developing interpretable risk scoring systems and conducting prognostic analyses. Considering PaSegNet[16] has demonstrated strong performance in classifying liver pathological tissue, we fine-tuned PaSegNet as our

patch-level classification model to predict both the class labels and spatial coordinates of each patch, enabling the identification and localization of regions corresponding to specific tissue categories.

3.4 Attention-Driven Mechanism for Localizing Potentially High-Risk Tissues

To explicitly highlight the contribution of various tissue categories in prognostic prediction and provide guidance for the development of potential risk scores, we designed the Attention Activator (ATAT).

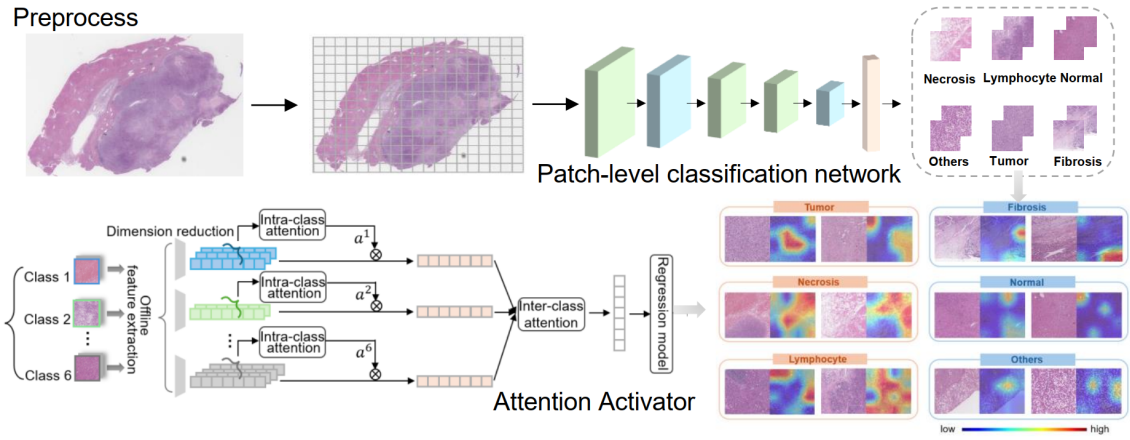


Figure 3-2 Workflow of ATAT and visualization results of attention maps.

As shown in Figure 3-2, ATAT incorporates the dual attention mechanism and is built upon the MIL strategy. Patches labeled by a pre-trained classifier (fine-tuned PaSegNet), are first processed through a pre-trained offline feature extractor (ResNet50[9], pre-trained on ImageNet[63] in this study). The extracted features are then fed into an intra-class attention mechanism, where the features within each tissue category are weighted by an attention score as follows:

$$\boldsymbol{B}^c = \sum_{n=1}^{N} a_n^c \boldsymbol{f}_n^c \in \mathbb{R}^D \qquad (3-1)$$

where N is the number of patches in each category, $\boldsymbol{f}_n^c$ is the patch feature extracted by the offline feature extractor, a_n^c is the learnable weight within each class for $\boldsymbol{f}_n^c$ and c={1,2,…,6} is the category index. The intra-class weight a_n^c is defined as:

$$a_n^c = \frac{\exp\{\omega^\intercal(\tanh(\boldsymbol{K}_1\boldsymbol{f}_n^c) \odot \mathrm{sigm}(\boldsymbol{K}_2\boldsymbol{f}_n^c))\}}{\sum_{j=1}^{N}\exp\{\omega^\intercal(\tanh(\boldsymbol{K}_1\boldsymbol{f}_n^c) \odot \mathrm{sigm}(\boldsymbol{K}_2\boldsymbol{f}_n^c))\}} \qquad (3-2)$$

where $\omega, \boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are learnable parameters. Thus, intra-class features are obtained through the weighted summation (if the pre-trained classification model does not detect a certain tissue type in a WSI, the feature for that class is set to $\boldsymbol{0}$). The WSI bag feature is then obtained through the sum of the weighted intra-class features using inter-class attention:

$$\boldsymbol{B} = \sum_{c=1}^{6} a^c \boldsymbol{B}^c \in \mathbb{R}^D \qquad (3-3)$$

where the inter-class weight a^c can be obtained by replacing the patch features in (2) with intra-class features. The obtained WSI bag feature is passed through a regression model composed of a three-layer Multilayer Perceptron (MLP). The output of ATAT predicts the risk score associated with the occurrence of a highly severe event. The loss function is defined in the form of (2-7). The prognostic capability of ATAT is evaluated using 5-fold cross-validation on SYSUCC and the validation results will be presented in the next section on experimental results.

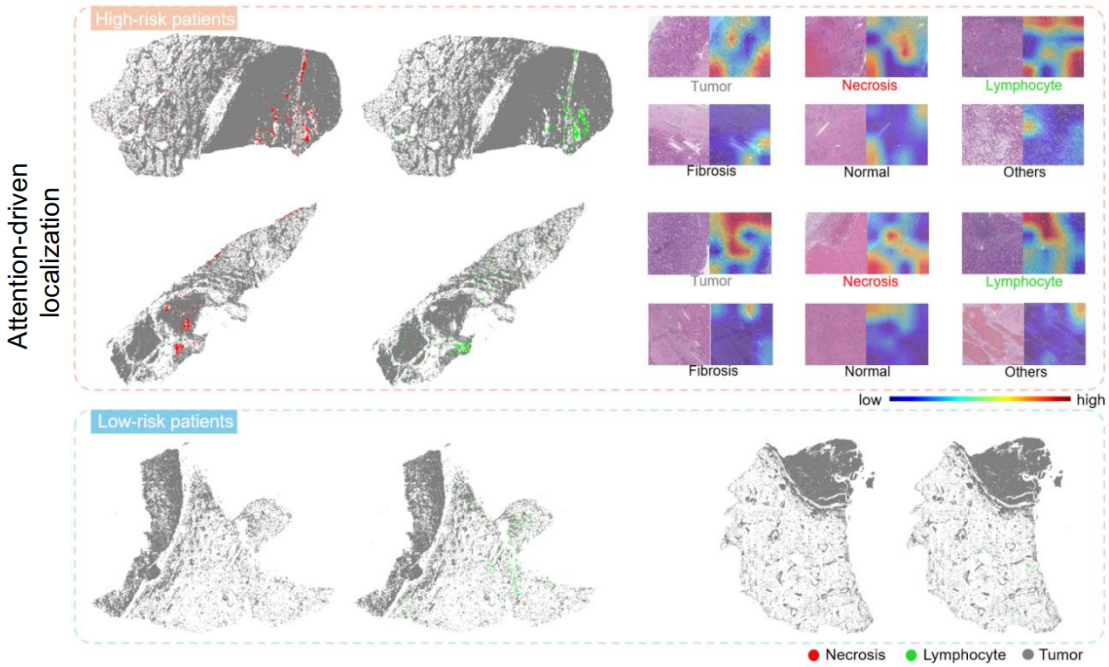


Figure 3-3 Distribution of necrotic, lymphocytic, and tumorous regions in the WSIs of high- and low-risk patients and local attention maps from ATAT highlighting detailed regions in high-risk cases.

To determine which tissue categories contribute most significantly to risk assessment, we leverage the explicit attention mechanism of ATAT to assign attention scores directly to individual patches. Representative local visualization results are presented in Figure 3-2 and Figure 3-3. Analysis of these visualizations reveals that necrotic regions and areas containing tumor-infiltrating lymphocytes (TILs) are consistently observed in nearly all high-risk patients. Notably, ATAT places greater emphasis on the interfaces between necrotic, lymphatic, and tumorous tissues, while exhibiting relatively lower attention to fibrotic and other non-tumorous tissue regions. This observation motivates the development of a risk-scoring system that integrates the contributions of necrotic, lymphocytic, and tumorous regions. Importantly, the hypothesis generated by ATAT without prior empirical input was validated by

experienced pathologists and shown to be consistent with conclusions from related studies[64][65][66].

3.5 Development of Hybrid Deep Score

To ensure clinical applicability, we use the hypotheses generated by ATAT as a reference for constructing an interpretable risk score and evaluated its effectiveness in real-world clinical settings.

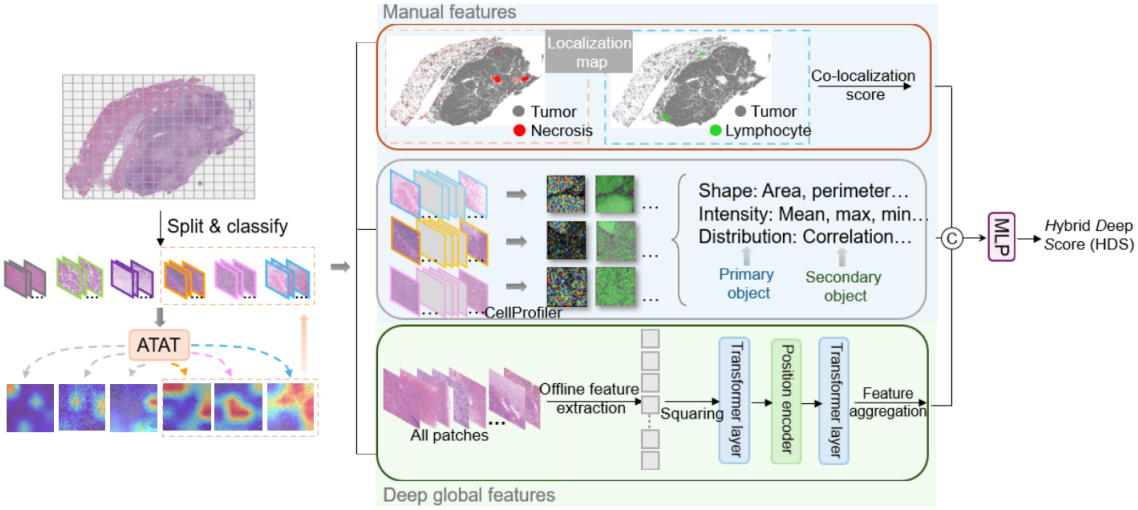


Figure 3-4 Inspired by ATAT, microscopic morphological features, co-localization score and deep global features are constructed and concatenated to establish HDS.

As shown in Figure 3-4, we develop HDS to model relationships between necrotic tissues, lymphocytes, and tumor regions. HDS integrates information across three hierarchical levels: 1) microscopic morphological characteristics of three distinct tissue types; 2) spatial interaction features, quantified as co-localization scores among lymphocytes, necrotic regions, and tumor tissues; 3) deep features that encapsulate global contextual information from WSIs.

3.5.1 Microscopic Morphological Features

Building upon ATAT, we initially focus on the microscopic morphological features of lymphatic, necrotic, and tumor regions. Specifically, patches classified into these three tissue types are processed using CellProfiler[23] to automatically extract primary objects (e.g. nuclei) and secondary objects (e.g. cell bodies). Subsequently, the corresponding shape, intensity, and distribution characteristics of both primary and secondary objects are calculated separately. Given that the number of patches varies across the three tissue categories in each WSI, we compute the median of each feature across all patches within a specific tissue category to represent the microscopic morphological feature set for that category in the WSI. Similarly, we normalized the extracted features, and if a particular category is predicted as absent in a WSI, its morphological features are set to $\boldsymbol{0}$. Differential analysis[67] is then employed to select 20 significant microscopic morphological features for each tissue type. Ultimately, for each WSI, we generated a 60-dimensional vector representing the microscopic morphological features.

3.5.2 Spatial Interaction Features

Given that ATAT emphasizes regions where necrosis, lymphocytes, and tumor tissues intersect, it is essential to construct features that capture the spatial distribution between necrotic, lymphatic, and tumor tissues. Notably, the spatial relationships between lymphatic regions and tumors, particularly Tumor-Infiltrating Lymphocytes (TIL), has been established as a critical prognostic indicator across various cancers[68]. Thus, to quantify this interaction, a co-localization score, termed "TILAb"[69] was

introduced to assess the spatial distribution and interaction of TILs as follows:

$$Co_{-L} = \begin{cases} \frac{M}{2} \times \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}^{l}\right)}{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}^{t}\right)}, \sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}^{t}\right) > 0 \\ 1, \sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}^{t}\right) \leq 0 \end{cases} \quad (3-4)$$

where

$$M = \frac{2\sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}^{l} \times p_{ij}^{t}\right)}{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}^{l}\right)^{2} + \sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}^{t}\right)^{2}} \quad (3-5)$$

In this part, we treat each patch as a pixel and reconstruct the patches with category labels obtained from a pre-trained classification network to generate the localization map (see Figure 3-3). This map is then divided into grids of pixel $m \times n$. The values p_{ij}^{l} and p_{ij}^{t} represent the percentage of lymphocytic and tumor regions, respectively, within the (i, j)-th grid cell. Similarly, by replacing p_{ij}^{l} with the proportion of necrotic regions in the (i, j)-th grid cell, denoted as p_{ij}^{n}, we can quantify the spatial relationships between necrosis and tumor as Co_{-N}. Thus, a two-dimensional feature vector $[Co_{-L}, Co_{-N}]$ can be computed for each WSI, which characterizes the spatial distribution between lymphatic/necrotic regions and the tumor.

3.5.3 Deep Global Features

To complement the manually extracted features associated with potential high-risk regions, we utilize a Transformer-based global feature extractor to derive deep global features from WSIs. The features of all patches in the WSI, extracted via the offline feature extractor in ATAT, are directly fed as input.

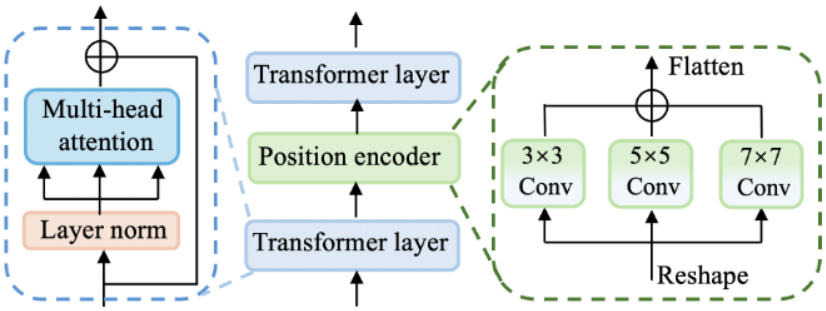


Figure 3-5 The Transformer-based deep global feature extractor.

Following the TransMIL framework[48], we first square the extracted features for subsequent processing. As shown in Figure 3-5, a two-layer Transformer with a position encoder is then applied to capture the global features of the WSI. This approach capitalizes on the Transformer encoder's proficiency in processing long sequential data, making it ideal for encoding the large number of patch features in a WSI. Additionally, this method adaptively integrates with the manual features described in 3.5.2 and 3.5.3, providing a multi-view representation for the final risk scoring system. The global features are subsequently passed through a fully connected layer for feature aggregation, resulting in a 64-dimensional deep global feature vector for each WSI.

Finally, we develop a multi-view risk-scoring system as Figure 3-4. This model integrates tissue regions identified as potentially high-risk by ATAT and incorporates features across three spatial scales: micro, local, and macro. These multi-scale features are concatenated and fed into an MLP layer for survival prediction. The model is trained using the survival loss defined in (2-7) and obtain the final Hybrid Deep Score (HDS).

3.6 Model Evaluation and Survival Analysis

Accurate tissue classification by the patch-level classification network, along with the clinically relevant identification of potential high-risk regions by ATAT, are essential

prerequisites for the effective development of HDS. Based on this, we first evaluate the performance of the fine-tuned patch-level classification network on the patch-level dataset.

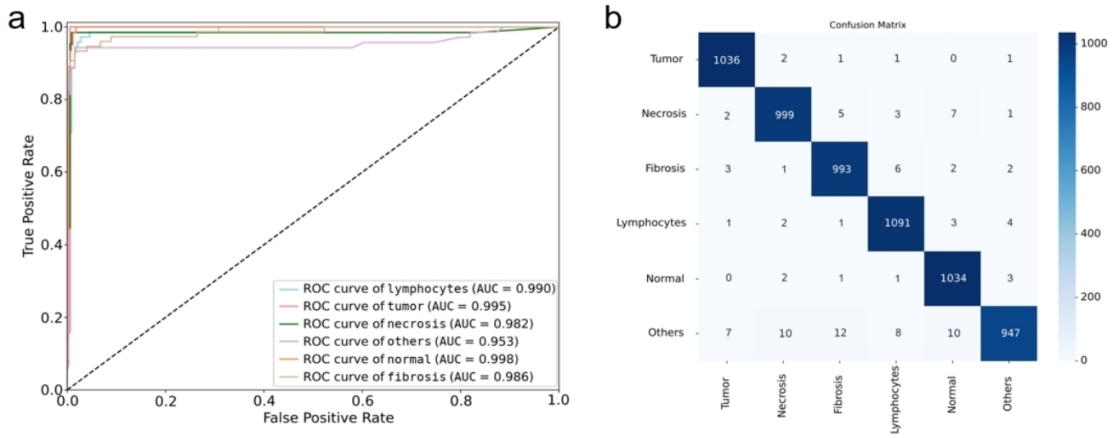


Figure 3-6 **a** ROC curves and **b** confusion matrix of patch-level classification network on patch-level dataset.

As shown in Figure 3-6, the average Area Under the Curve (AUC) across the six categories was 0.98 (see Figure 3-6 **a**). Moreover, the confusion matrix in Figure 3-6 **b** provides further evidence of accuracy of the classification model in differentiating between various tissue types.

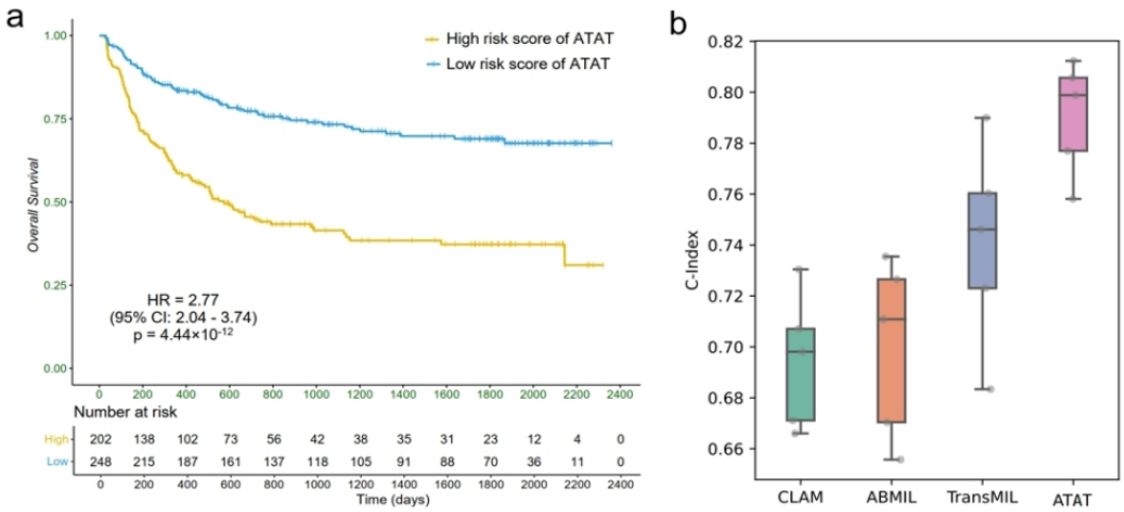


Figure 3-7 **a** KM curves of ATAT for DFS in SYSUCC. **b** c-index of ATAT and other

deep prediction models.

Furthermore, we assessed the performance of ATAT on the SYSUCC dataset. As shown in Figure 3-7 **a**, ATAT demonstrated strong capability in distinguishing between high- and low-risk patients, with HR of 2.77 (95% CI 2.04-3.74). Additionally, we compared its prognostic prediction accuracy with other state-of-the-art deep survival prediction models[10][48][70] using C-index in Figure 3-7 **b**. ATAT achieved the best performance in five-fold cross-validation, with the C-index of 0.79 ± 0.035.

It is important to note that the identification of high-risk patients presented above and showed in Figure 3-2 and Figure 3-3 was obtained using ATAT. However, as ATAT is a heuristic approach with limited clinical interpretability, we subsequently developed a multi-view risk scoring system (shown in Figure 3-4) and introduced the HDS to improve its clinical relevance and practical applicability.

As mentioned in Section 3.5, our HDS incorporates features from three perspectives: micro, local, and macro. To evaluate the effectiveness of each perspective and the necessity of combining them to construct HDS, we first analyze the performance of each perspective individually, as well as the performance of their partial combinations in prognostic prediction.

Table 3-4 **a**. Comparison of C-index and time-dependent AUC for different indicators on SYSUCC.

	C-index	1-year AUC	2-year AUC	5-year AUC
CL	0.571±0.048	0.604±0.101	0.612±0.110	0.627±0.123
MM	0.619±0.104	0.651±0.106	0.690±0.176	0.703±0.121
DG	0.683±0.099	0.659±0.137	0.701±0.121	0.731±0.119
CL&MM	0.624±0.105	0.661±0.145	0.696±0.091	0.716±0.167
CL&DG	0.703±0.188	0.666±0.134	0.711±0.169	0.742±0.173
MM&DG	0.744±0.176	0.673±0.122	0.718±0.202	0.759±0.089
HDS	**0.751±0.082**	**0.682±0.153**	**0.724±0.117**	**0.767±0.138**

Table 3-4 **b**. Comparison of C-index and time-dependent AUC for different indicators on TGGA-LIHC.

	C-index	1-year AUC	2-year AUC	5-year AUC
CL	0.564±0.041	0.619±0.080	0.643±0.120	0.655±0.131
MM	0.608±0.103	0.634±0.115	0.672±0.159	0.690±0.170
DG	0.674±0.141	0.663±0.111	**0.735±0.189**	0.699±0.158
CL&MM	0.620±0.038	0.656±0.133	0.691±0.132	0.697±0.155
CL&DG	0.689±0.173	0.672±0.162	0.719±0.098	0.701±0.183
MM&DG	0.707±0.160	**0.684±0.143**	0.728±0.119	0.714±0.212
HDS	**0.729±0.196**	0.678±0.168	**0.735±0.219**	**0.723±0.130**

*C-index focuses on overall predictive concordance, while time-dependent AUC focuses on the performance at specific time points. *MM* microscopic morphological features; *DG* deep global features; *CL* co-localization features. & concatenate two types of features. The best-performing indicators are highlighted in bold.

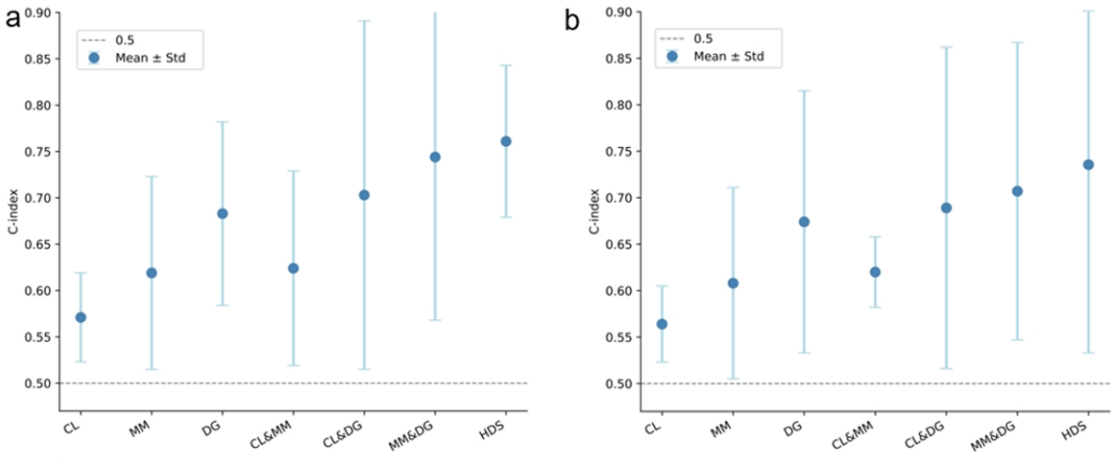


Figure 3-8 Independent risk factors and their combinations from the three distinct perspectives incorporated in HDS for their corresponding C-index values of DFS in **a** SYSUCC and **b** TCGA-LIHC.

As shown in Table 3-4 and Figure 3-8, we assessed the C-index values and the 1-, 2-, and 5-year AUCs by using single or combinations of two out of the three feature perspectives: CL co-localization features, MM microscopic morphological features, and DG deep global features, as well as their pairwise combinations (CL&MM, CL&DG, MM&DG). The evaluation was conducted on both the SYSUCC and TCGA-

LIHC datasets. The results show that combining any two feature types enhances prognostic performance compared to using a single feature alone. For example, the combination of MM and DG resulted in a 20% improvement in the C-index on SYSUCC compared to MM alone. Ultimately, integrating all three perspectives into HDS yielded the most competitive performance, with the 1-, 2-, and 5-year AUCs and the C-index for DFS being 0.682±0.183, 0.724±0.117, 0.767±0.138, and 0.751±0.082 in SYSUCC, and 0.678±0.238, 0.735±0.219, 0.723±0.130, and 0.729±0.196 in TCGA-LIHC, respectively.

Table 3-5 **a**. Univariate and multivariate survival analysis of three views of indicators of DFS on SYSUCC.

Disease-Free Survival	Univariate analysis			Multivariate analysis		
	HR	95%CI	*p*-value	HR	95%CI	*p*-value
CL (high/low)	1.132	1.044-1.567	**0.0020**	1.164	1.016-2.978	<0.001
MM (high/low)	1.451	1.255-1.681	**0.0015**	1.222	1.197-4.571	<0.001
DG (high/low)	2.467	2.253-3.680	**<0.001**	2.511	1.914-3.114	<0.001

Table 3-5 **b**. Univariate and multivariate survival analysis of three views of indicators of DFS on TCGA-LIHC.

Disease-Free Survival	Univariate analysis			Multivariate analysis		
	HR	95%CI	*p*-value	HR	95%CI	*p*-value
CL (high/low)	1.244	1.109-3.774	**<0.001**	1.945	1.720-3.170	<0.001
MM (high/low)	1.641	1.020-2.558	**<0.001**	2.015	1.219-4.926	<0.001
DG (high/low)	1.984	1.267-4.677	**<0.001**	1.921	1.173-5.111	<0.001

**MM* microscopic morphological features; *DG* deep global features; *CL* co-localization features. *HR* hazard ratio, *CI* confidence interval. The median value was taken as the cutoff value of high- and low-risk groups.

Additionally, the results of the multivariable analysis for these three indicators of DFS in the SYSUCC and TCGA-LIHC datasets are presented in Table 3-5. The analysis confirms that these three factors are independent prognostic risk factors (with multivariable $p < 0.001$).

Table 3-6 **a**. χ^2 test of three views of indicators on SYSUCC.

Pairwise comparison	χ^2 *p*-value
CL *v.s.* MM	0.08
CL *v.s.* DG	0.22
MM *v.s.* DG	0.45

Table 3-6 **b**. χ^2 test of three views of indicators on TCGA-LIHC.

Pairwise comparison	χ^2 *p*-value
CL *v.s.* MM	0.48
CL *v.s.* DG	0.59
MM *v.s.* DG	0.77

**MM* microscopic morphological features; *DG* deep global features; *CL* co-localization features. χ^2 test: chi-squared test.

Furthermore, Table 3-6 shows the pairwise Chi-square test results for these indicators across the two datasets. The Chi-square p-values exceeding 0.05 indicate that the features from these three perspectives are independent of each other.

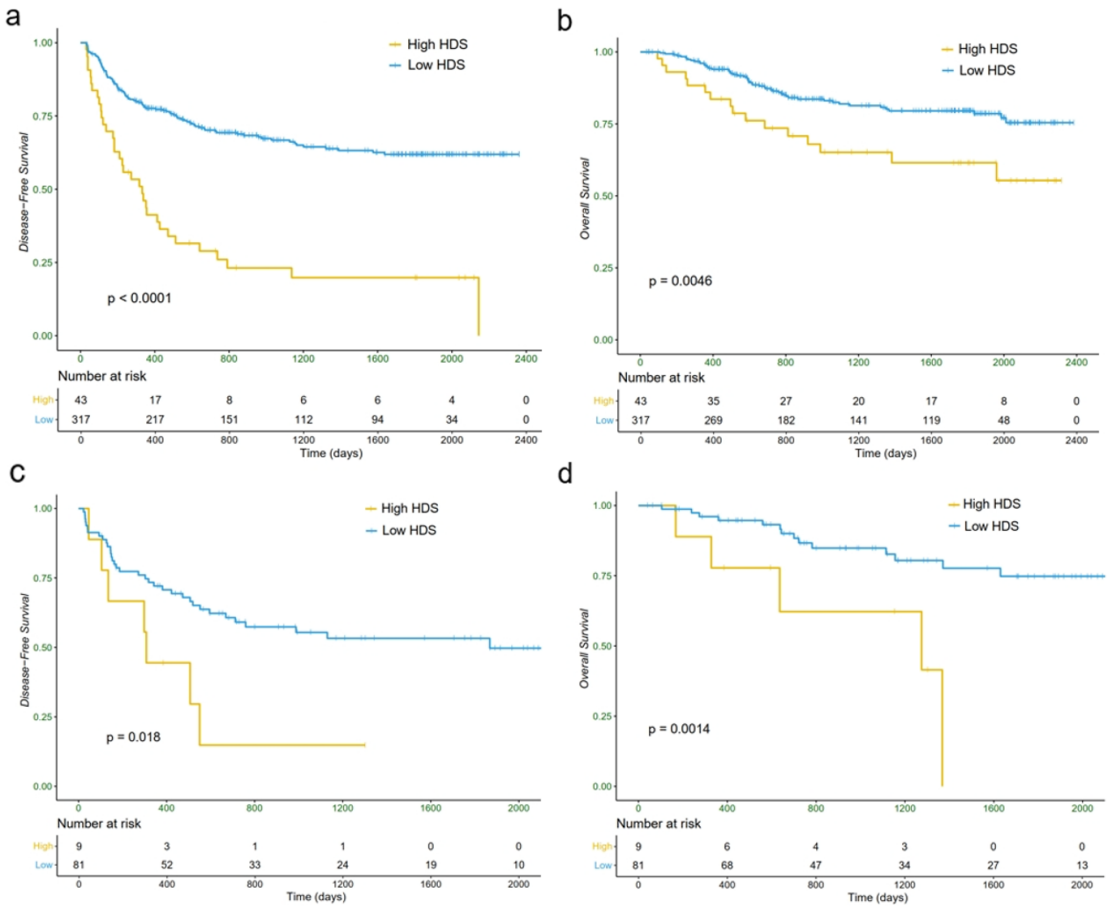


Figure 3-9 The KM survival curves of HDS for DFS and OS in training cohort and

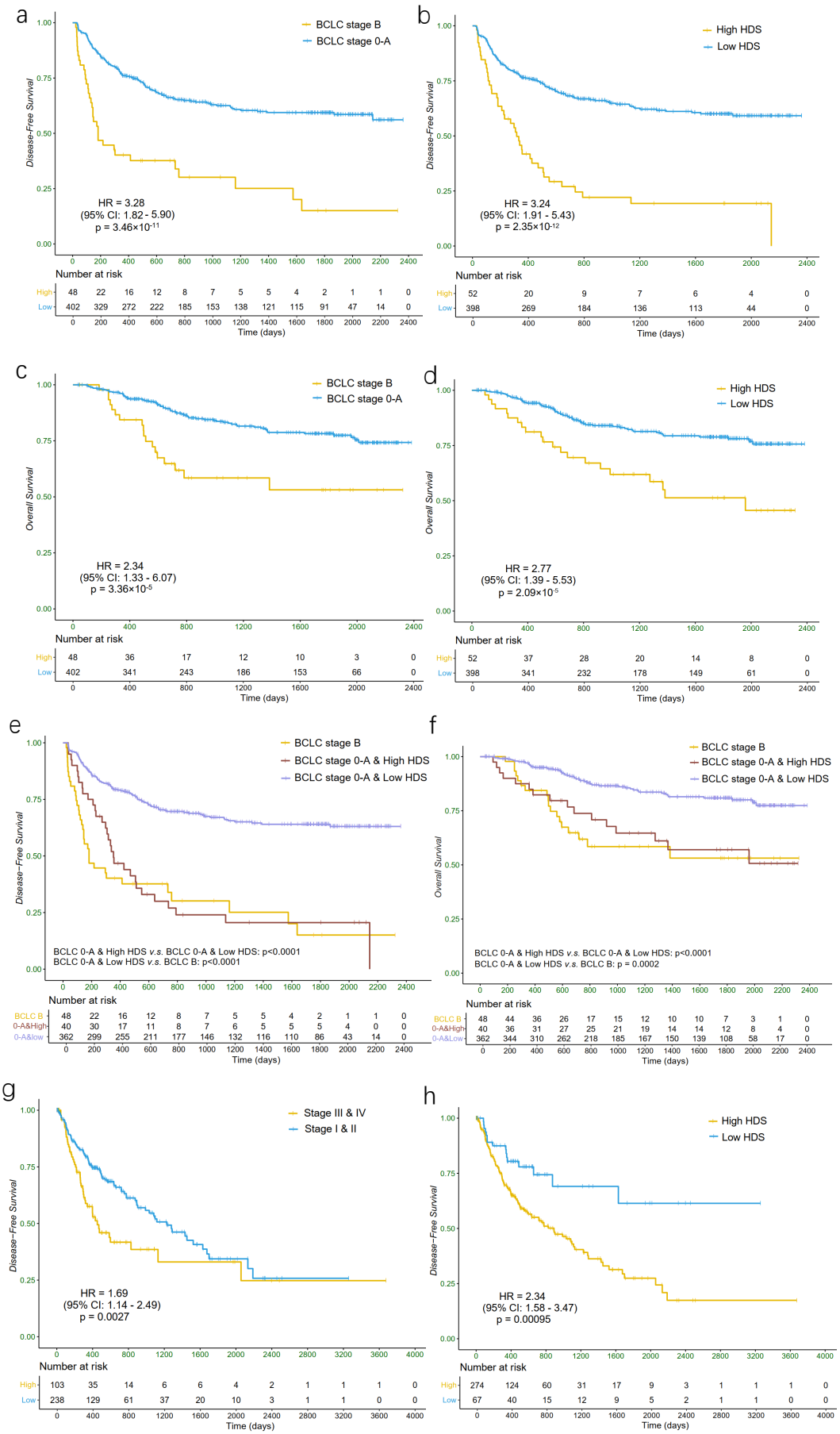
validation cohort of SYSUCC.

For HDS, the median value from the training cohort was used as the cutoff to categorize patients into high-risk and low-risk groups. Patients in the low-risk group have significantly longer DFS and OS in both the training and validation cohorts. As shown in Figure 3-9, the 2-year DFS rate for the high-risk group is 26.79% in the training cohort and 16.22% in the validation cohort, while the low-risk group exhibit DFS rates of 70.69% and 66.21% in the respective cohorts. Similarly, when OS is used as a secondary endpoint, HDS demonstrate strong differentiation between high- and low-risk groups: the 2-year OS rate for the high-risk group is 69.24% in the training cohort and 60.31% in the validation cohort, compared to 87.21% and 86.49% in the low-risk group. In the long-term (5-year) analysis, HDS also showed superior risk stratification, consistent with the results observed in the short-term (1-/2-year) DFS and OS assessments.

Table 3-7 Comparison of C-index and time-dependent AUC for clinical staging systems, HDS and clinical staging systems plus HDS.

dataset	Staging	C-index	1-year AUC	2-year AUC	5-year AUC
SYSUCC	BCLC	0.681±0.053	0.665±0.031	0.732±0.221	0.730±0.091
	HDS	0.751±0.082	0.682±0.153	0.724±0.117	0.767±0.138
	BCLC+HDS	**0.783±0.110**	**0.701±0.045**	**0.739±0.213**	**0.772±0.224**
TCGA-LIHC	TNM	0.670±0.012	0.669±0.031	0.712±0.097	0.735±0.126
	HDS	0.729±0.196	0.678±0.168	0.735±0.219	0.723±0.130
	TNM+HDS	**0.747±0.091**	**0.681±0.060**	**0.741±0.073**	**0.740±0.208**

*C-index focuses on overall predictive concordance, while time-dependent AUC focuses on the performance at specific time points.


a
BCLC stage B
BCLC stage 0-A
Disease-Free Survival
HR = 3.28
(95% CI: 1.82 - 5.90)
p = 3.46×10-11
Number at risk
High 48 22 16 12 8 7 5 5 4 2 1 1 0
Low 402 329 272 222 185 153 138 121 115 91 47 14 0
Time (days)
b
High HDS
Low HDS
Disease-Free Survival
HR = 3.24
(95% CI: 1.91 - 5.43)
p = 2.35×10-12
Number at risk
High 52 20 9 7 6 4 0
Low 398 269 184 136 113 44 0
Time (days)
c
BCLC stage B
BCLC stage 0-A
Overall Survival
HR = 2.34
(95% CI: 1.33 - 6.07)
p = 3.36×10-5
Number at risk
High 48 36 17 12 10 3 0
Low 402 341 243 186 153 66 0
Time (days)
d
High HDS
Low HDS
Overall Survival
HR = 2.77
(95% CI: 1.39 - 5.53)
p = 2.09×10-5
Number at risk
High 52 37 28 20 14 8 0
Low 398 341 232 178 149 61 0
Time (days)
e
BCLC stage B
BCLC stage 0-A & High HDS
BCLC stage 0-A & Low HDS
Disease-Free Survival
BCLC 0-A & High HDS v.s. BCLC 0-A & Low HDS: p<0.0001
BCLC 0-A & Low HDS v.s. BCLC B: p<0.0001
Number at risk
BCLC B 48 22 16 12 8 7 5 5 4 2 1 1 0
0-A&High 40 30 17 11 8 7 6 5 5 5 4 0 0
0-A&low 362 299 255 211 177 146 132 116 110 86 43 14 0
Time (days)
f
BCLC stage B
BCLC stage 0-A & High HDS
BCLC stage 0-A & Low HDS
Overall Survival
BCLC 0-A & High HDS v.s. BCLC 0-A & Low HDS: p<0.0001
BCLC 0-A & Low HDS v.s. BCLC B: p = 0.0002
Number at risk
BCLC B 48 44 36 26 17 15 12 10 10 7 3 1 0
0-A&High 40 36 31 27 25 21 19 14 14 12 8 4 0
0-A&Low 362 344 310 262 218 185 167 150 139 108 58 17 0
Time (days)
g
Stage III & IV
Stage I & II
Disease-Free Survival
HR = 1.69
(95% CI: 1.14 - 2.49)
p = 0.0027
Number at risk
High 103 35 14 6 6 4 2 1 1 1 0
Low 238 129 61 37 20 10 3 1 1 0 0
Time (days)
h
High HDS
Low HDS
Disease-Free Survival
HR = 2.34
(95% CI: 1.58 - 3.47)
p = 0.00095
Number at risk
High 274 124 60 31 17 9 3 1 1 1 0
Low 67 40 15 12 9 5 2 1 1 0 0
Time (days)

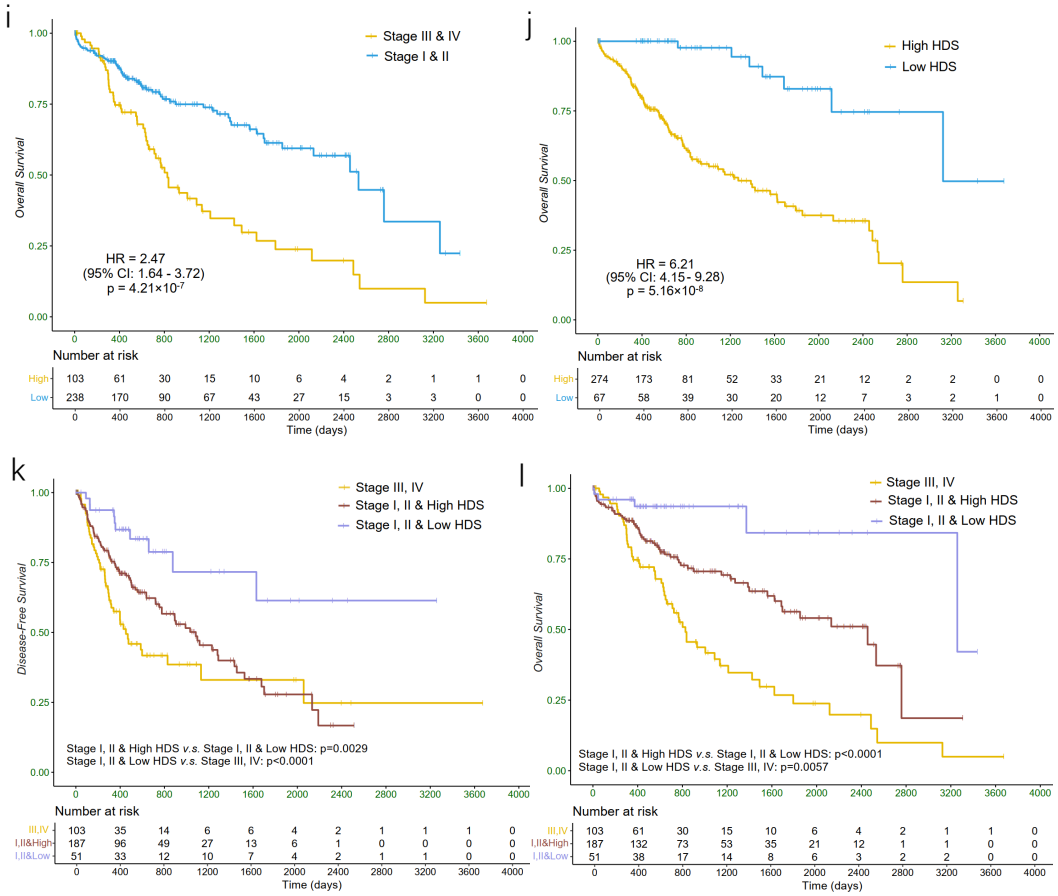


Figure 3-10 KM curves comparing the existing clinical staging systems with the refined stratification based on HDS for both DFS and OS. **a** BCLC staging system, **b** HDS and **e** HDS-based refined stratification for BCLC stage 0-A in DFS for SYSUCC. **c** BCLC staging system, **d** HDS and **f** HDS-based refined stratification for BCLC stage 0-A in OS for SYSUCC. **g** TNM staging system, **h** HDS and **k** HDS-based refined stratification of TNM stage I&II of DFS in TCGA-LIHC. **i** TNM staging system, **j** HDS and **l** HDS-based refined stratification of TNM stage I&II of OS in TCGA-LIHC.

To evaluate the prognostic accuracy of the HDS and to further verify its compatibility with existing clinical staging systems, we first compare the C-index and the 1-, 2-, and 5-year AUC values of the HDS, BCLC, TNM staging systems, as well as combinations of HDS with the latter two (incorporate a linear composite of the HDS

and clinical staging systems as covariates in the Cox proportional hazards model) on SYSUCC and TCGA-LIHC. The results are presented in Table 3-7. Integrating HDS with either the BCLC or TNM staging systems consistently yielded superior prognostic performance across both the SYSUCC and TCGA-LIHC. Specifically, the combined models achieved C-indices of 0.783 ± 0.110 (SYSUCC) and 0.747 ± 0.091 (TCGA-LIHC); 1-year AUCs of 0.701 ± 0.045 and 0.681 ± 0.060; 2-year AUCs of 0.739 ± 0.213 and 0.741 ± 0.073; and 5-year AUCs of 0.772 ± 0.224 and 0.740 ± 0.208, respectively. These results underscore the additive value of HDS in enhancing the prognostic precision of existing clinical staging systems.

To further assess the discriminative ability of the HDS in stratifying high- and low-risk patients, and more importantly, to examine its potential to refine existing clinical staging systems, we compare the KM curves of BCLC and HDS for DFS (Figure 3-10 a and b) and OS (Figure 3-10 d and e) in SYSUCC. Similarly, comparisons between TNM staging and HDS for both DFS and OS are conducted in TCGA-LIHC (see Figure 3-10 g, h, j and k). Expanding upon the existing BCLC and TNM staging systems, we further refine the stratification of low-risk subgroups using HDS. As illustrated in Figure 3-10 c, f, i and l, the p-values for comparisons between the low-risk group with high HDS and the low-risk group with low HDS are all below 0.001 for both DFS and OS. These results show that HDS effectively identifies potentially high-risk patients within the low-risk category, thereby allowing clinicians to provide more targeted monitoring for these individuals.

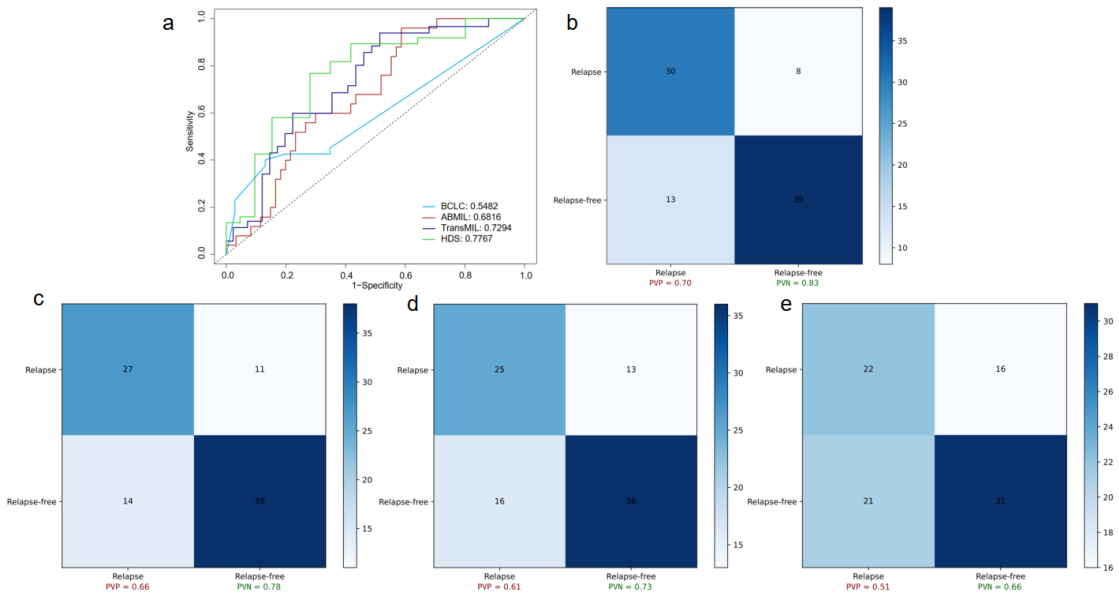


Figure 3-11 **a** The ROC curve for comparison between HDS and other predictive staging systems based on DFS. **b** Confusion matrix with PVP and PVN of HDS. **c** Confusion matrix with PVP and PVN of TransMIL[48]. **d** Confusion matrix with PVP and PVN of ABMIL[70]. **e** Confusion matrix with PVP and PVN of BCLC. *ROC* receiver operating characteristic curve, *HDS* hybrid deep score, *PVP* positive predictive value, *PVN* negative predictive value, *DFS* disease-free survival, *ABMIL* attention-based deep multiple instance learning, *TransMIL* transformer based correlated multiple instance learning, *BCLC* Barcelona clinic liver cancer.

In addition, we also compared HDS with some pure deep learning models[48][70], as shown in Figure 3-11, **a** illustrates ROC curves for HDS, along with widely used deep learning methods such as ABMIL and TransMIL and the established clinical staging system BCLC for DFS in the SYSUCC validation cohort. HDS demonstrates competitive performance, achieving the AUC of 0.7767. Figure 3-11 **b-e** show the confusion matrices for four methods, as well as their respective positive predictive

value (PPV) and negative predictive value (NPV). These results highlight the strong discriminatory ability of HDS in distinguishing high- and low-risk prognostic patients, with a PPV of 0.70 and an NPV of 0.83.

3.7 Prognostic risk factors of HDS

To identify independent predictors of DFS in the SYSUCC and TCGA-LIHC cohorts, we conduct Cox proportional hazards regression analysis in each cohort. Given the differences in clinicopathological factors between SYSUCC and TCGA-LIHC, each cohort is analyzed independently in the univariate Cox regression analysis..

Table 3-8 **a**. Univariate and multivariate survival analysis of variables with DFS.

Disease-Free Survival	Univariate analysis			Multivariate analysis		
	HR	95%CI	*p*-value	HR	95%CI	*p*-value
Sex (male/female)	0.912	0.714-1.165	0.453			
Age (>55/⩽55 years)	1.234	1.982-1.551	0.106			
Cirrhosis (with/without)	1.453	1.032-1.046	**0.032**			
Diameter (>5/⩽5cm)	2.542	1.786-3.613	**<0.001**	3.275	2.053-5.219	<0.001
Cancer embolus (with/without)	0.874	0.654-1.169	0.367			
Lesion (multiple/single)	2.025	1.425-3.457	**0.017**	1.541	1.326-4.358	0.021
AFP (>400/⩽400 ng/ml)	1.356	0.998-1.842	0.051			
MVI (with/without)	1.859	1.120-3.088	**<0.001**			
ALT (>40/⩽40 U/l)	1.231	0.822-1.846	0.224			
BCLC (B/0-A)	3.282	1.821-5.902	**<0.001**	2.958	1.774-5.193	<0.001
AST (>40/⩽40 U/l)	1.421	1.005-2.014	0.084			
HBV (positive/negative)	1.972	1.223-3.178	**0.005**			
HDS (high/low)	3.241	1.911-5.434	**<0.001**	3.852	2.333-5.319	<0.001

**ALT* alanine aminotransferase, *AST* Aspartate aminotransferase, *AFP* alpha-fetoprotein, *MVI* microvascular invasion, *HR* hazard ratio, *CI* confidence interval, *DFS* disease-free survival.

Table 3-8 **b**. Univariate and multivariate analysis of DFS in TCGA-LIHC.

Disease-Free Survival	Univariate analysis			Multivariate analysis		
	HR	95% CI	*p*-value	HR	95% CI	*p*-value
Sex (male/female)	0.912	0.714-1.165	0.45			
Age (>55/≤55 years)	1.125	0.830-1.540	0.2			
Depth of invasion (T3 & T4/T1 & T2)	2.134	1.502-3.034	0.001	2.214	1.6-3.064	0.002
Tumor Grade (G3 & G4/G1 & G2)	1.678	1.123-2.505	0.045			
TNM (stage III & IV/stage I & II)	1.694	1.143-2.491	0.0027	2.111	1.492-4.102	<0.001
HDS (high/low)	2.342	1.581-3.473	<0.001	2.449	1.797-5.312	<0.001

**DFS* disease-free survival.

As shown in Table 3-8, the multivariable Cox analysis identified tumor size (HR = 3.275, 95% CI: 2.053-5.219), tumor number (HR = 1.541, 95% CI: 1.326-4.358), and HDS (HR = 3.852, 95% CI: 2.333-5.319) as independent predictors of DFS in SYSUCC. In TCGA-LIHC, depth of invasion (HR = 2.214, 95% CI: 1.600-3.064) and HDS (HR = 2.449, 95% CI: 1.797-5.312) were similarly identified as independent risk factors.

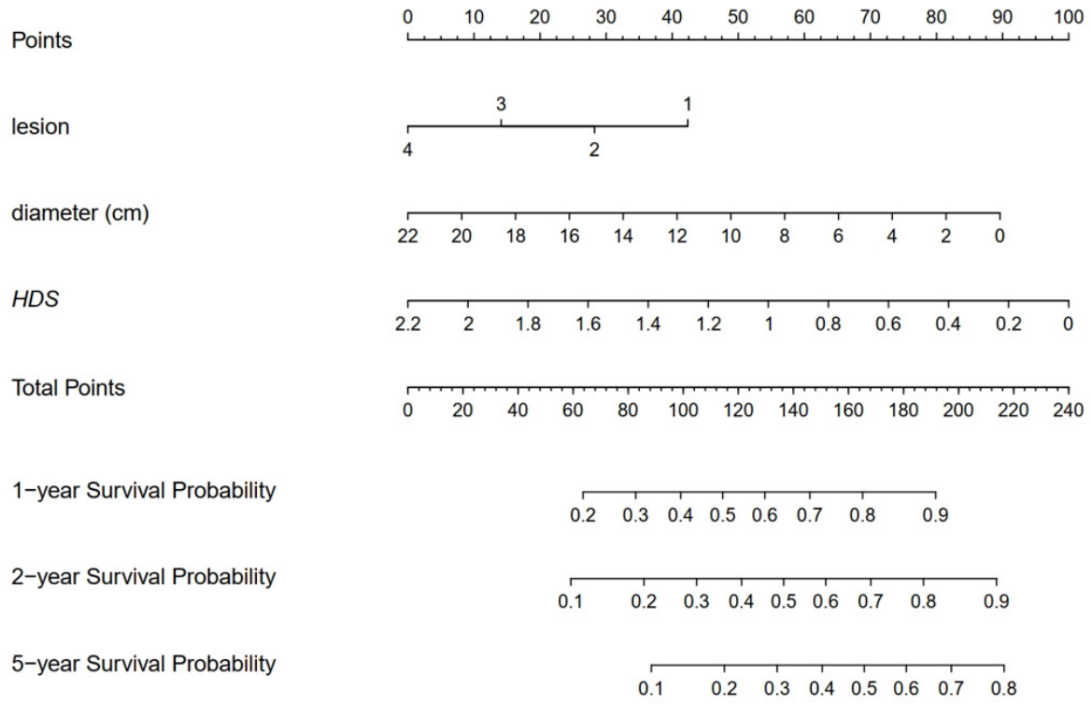


Figure 3-12 Nomogram of independent clinical risk factors and HDS for DFS in

SYSUCC.

In SYSUCC, we incorporate the independent clinical risk factors mentioned above ("Number of lesions" and "diameter of tumor") into our analysis and constructed a corresponding nomogram in Figure 3-12. The nomogram incorporates tumor diameter, lesion number, and the HDS score to estimate 1-, 2-, and 5-year survival probabilities. HDS contributes a substantial proportion of the total predictive score, with a wide value range and strong discriminative capacity. These results underscore the added value of HDS in enhancing prognostic precision beyond clinical variables.

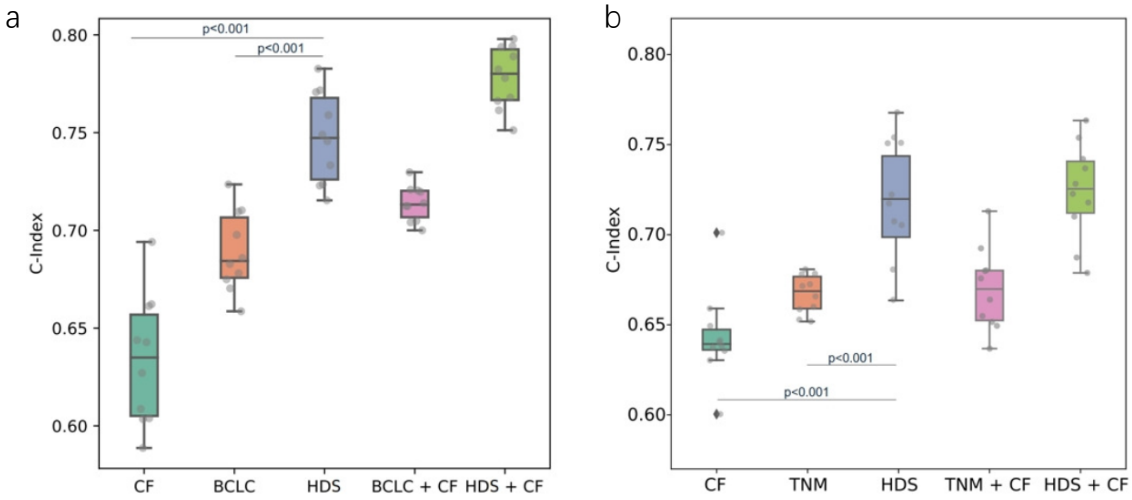


Figure 3-13 **a** C-indexes of independent CF, BCLC stage, HDS and BCLC/HDS combined CF in SYSUCC. **b** C-indexes of independent CF, TNM stage, HDS and TNM/HDS combined CF in TCGA-LIHC. *CF* clinical risk factors, *BCLC* Barcelona clinic liver cancer, *TNM* American joint committee on cancer tumor node metastasis, *HDS* hybrid deep score, *DFS* disease-free survival.

Furthermore, Figure 3-13 **a** displays the C-index values based solely on clinical risk factors, BCLC staging system, and the HDS on SYSUCC. Among these, HDS

demonstrates strong prognostic performance (C-index = 0.747, 95% CI: 0.711–0.783), better than that of clinical risk factors. Importantly, combining HDS with clinical risk factors resulted in a notable enhancement of predictive accuracy, yielding a 4.1% improvement in the C-index. Similarly, as shown in Figure 3-13 **b**, integrating HDS with clinical features also resulted in improved prognostic performance in TCGA-LIHC, with a 2.87% increase in the C-index. These results show the synergistic potential of HDS when combined with conventional clinical variables, offering a meaningful enhancement to survival prediction models and providing a promising direction for optimizing current prognostic frameworks.

a

Characteristics	High	Low	*p*-value	HR (95%CI)	*p*-value
All	52	398		3.24 (1.91−5.43)	<0.001
Sex			0.55		
Male	46	335		3.12 (1.82−5.35)	<0.001
Female	6	63		3.7 (2.95−7.82)	0.025
Age			0.30		
≤55	35	234		2.88 (1.57−5.29)	<0.001
>55	17	164		4.24 (1.46−12.29)	<0.001
Cirrhosis			0.57		
No cirrhosis	20	174		3.08 (1.31−7.26)	<0.001
With cirrhosis	32	224		3.25 (1.67−6.29)	<0.001
Diameter			0.75		
Max. diameter, ≤5cm	30	243		2.83 (1.2−6.71)	<0.001
Max. diameter, >5cm	22	155		2.81 (1.53−5.17)	<0.001
Cancer Embolus			0.62		
No cancer embolus	37	300		1.95 (1.06−3.97)	0.031
With cancer embolus	15	98		2.62 (1.42−4.87)	<0.001
Lesion			0.17		
Single lesion	41	346		3.32 (1.8−6.14)	<0.001
Multiple lession	11	52		2.14 (0.96−5.3)	0.033
AFP			0.012		
AFP, <20 ng/ml	15	179		3.36 (1.34−8.44)	<0.001
AFP, 20−400 ng/ml	12	106		2.52 (1.08−7.02)	0.008
AFP, >400 ng/ml	25	113		3.26 (1.48−7.18)	<0.001

b

Characteristics	High	Low	*p*-value	HR (95%CI)	*p*-value
All	274	67		2.34 (1.58−3.47)	<0.001
Sex			0.21		
Male	180	50		2.94 (1.78−4.86)	<0.001
Female	94	17		2.46 (1.34−4.50)	0.0011
Age			0.04		
≤55	91	32		3.96 (2.03−7.72)	<0.001
>55	183	35		2.51 (1.55−4.05)	<0.001
Depth of invasion			0.002		
T1 & T2	223	40		3.01 (1.84−4.93)	<0.001
T3 & T4	61	27		2.74 (1.45−5.2)	<0.001
Tumor grade			0.01		
G1 & G2	176	31		2.45 (1.44−4.16)	<0.001
G3 & G4	98	36		1.79 (1.45 − 2.15)	<0.001

Figure 3-14 Forest plot of HDS for DFS in **a** SYSUCC and **b** TCGA-LIHC. *AFP* alpha-fetoprotein, *DFS* disease-free survival.

As shown in Figure 3-13, the forest plots depict the prognostic risks across different subgroups with different clinical features. From the left half of the forest plots corresponding to the two datasets, in addition to the existing clinical staging system with the capability to significantly differentiate between high- and low-risk patients, the high-risk group of HCC patients also had higher AFP in SYSUCC and deeper invasion in TCGA-LIHC compared to the low-risk group ($p<0.001$). The right half of the forest plots show that HDS was proven to be an effective prognostic factor across all clinical subgroups in both the SYSUCC and TCGA-LIHC datasets.

3.8 Discussion

As a prevalent malignancy with a high mortality rate, HCC necessitates the development of effective prognostic risk scoring systems with significant clinical relevance. The systems can aid clinicians in formulating personalized treatment strategies aimed at reducing recurrence and mortality rates. Although existing risk

scoring models have substantially advanced the field, it remains challenging to ensure both transparency and automation in predictive systems. For example, manual features extracted by doctors can yield rich information but are often time-consuming, whereas fully black-box deep learning models may compromise clinical interpretability. To mitigate these limitations, we proposed a DL-assisted framework for risk scoring based on WSI analysis and introduced a novel prognostic risk score, termed HDS.

In this study, we developed an attention-enhanced, heuristic framework for generating prognostic risk scores with the assistance of DL, following the general workflow typically used by pathologists. The integration of DL in this framework is manifested in two core components: first, it employs attention mechanisms to highlight regions potentially associated with high risk, which provided interpretable visual cues for clinicians and simplified the downstream process. Second, it extracted and integrated interpretable risk factors from three distinct perspectives through DL models. Using this heuristic framework, we constructed HDS for HCC patients. Prognostic analyses demonstrated that HDS not only delivered competitive performance in predicting patients' prognosis, but more importantly, it supplemented and improved the existing clinical staging systems, such as BCLC and TNM. Incorporating HDS into current prognostic staging frameworks allows for improved identification of potential high-risk patients and offers valuable guidance for clinicians in tailoring more precise follow-up and treatment strategies.

Previous studies have often relied on pathologists' prior knowledge for the localization of tissues associated with potential high-risk populations. However, such

approaches face a major challenge in the absence of clinical expertise, particularly when dealing with newly emerging cancers or rare subtypes. Therefore, there is a need for a heuristic method capable of identifying high-risk regions even without specialized clinical knowledge. To address this, we introduced the ATAT, which was specifically designed to detect potential high-risk tissue regions in WSIs. The attention values from ATAT indicated which tissue types the model prioritizes during the prognostic prediction process. Unlike attribution map methods[10], which rely heavily on model performance, ATAT provided clear, interpretable references that assist pathologists in locating regions that may be associated with risk, thereby improving the identification of potential high-risk areas.

On the other hand, while identifying tissue regions linked to high-risk prognostic groups is possible, the challenge lies in the effective extraction of features from these regions and the development of a prognostic risk score that is both meaningful and clinically relevant. Some methodologies rely on features manually defined by clinicians. While these features often exhibit strong clinical relevance, they are heavily reliant on prior knowledge, which can lead to the omission of potentially valuable information within pathological images. In contrast, other approaches rely entirely on deep features extracted by black-box deep learning methods, which frequently lack interpretability. While these models may show promising performance within a specific cohort, their ability to generalize across different cohorts is limited, thus reducing their broader clinical applicability.

Building upon the hypotheses established by ATAT, we developed a risk scoring system that integrates information from three key dimensions: 1) Microscopic morphological features that capture the microstructural details of necrosis, lymphocytes, and tumors; 2) Co-localization features that characterize the spatial distribution between necrotic/lymphocytic and tumor regions; and 3) Deep global features that encapsulate the overall information from WSIs. The first two dimensions consist of manually computed and extracted features generated by ATAT, providing interpretability. To enhance these interpretable features, we also employed TransMIL to extract global features from WSI. To effectively combine these three perspectives, we developed a multi-view hybrid multiple instance learning framework, which generates the final risk score (HDS). Experimental results demonstrated that our HDS can significantly complement and enhance existing clinical staging systems for both DFS and OS. Notably, HDS was able to identify high-risk patients within the low-risk groups (stratified by BCLC or TNM) in both SYSUCC and TCGA-LIHC datasets, offering clinicians valuable insights for further refining treatment strategies for these patients.

Despite the promising results achieved in this study for HCC prognostic analysis, several limitations remain. One key limitation is the feature fusion approach: We are exploring the possibility of using manually extracted features to direct the extraction of deep global features in future studies. Additionally, in the case of some HCC patients, only needle aspiration biopsy WSIs are available, and these differ significantly in morphology from the WSIs used in our current cohort. Consequently, extending our

approach to these patients presents a challenge. The extension to needle aspiration biopsy WSIs will be discussed in the following section.

3.9 Conclusion

In conclusion, we developed a prognostic risk-scoring framework for HCC patients, inspired by the clinical diagnostic workflow employed by pathologists. This framework leverages heuristic methods to identify potential high-risk tissue regions and constructs a multi-view hybrid risk score, providing novel perspectives on how DL models can improve and refine existing clinical staging systems for HCC. Our experimental results show that the proposed HDS performed effectively in risk stratification and demonstrated robust generalizability to external cohorts, highlighting its potential to enhance and complement current clinical staging systems.

Chapter 4.

Decoupling Pathological Phenotypes for Efficient Whole Slide Image Analysis

4.1 Introduction

For tasks that require considering every detail within the WSI, such as the task in Chapter 3, we take into account as many patches in the WSI as possible to capture the distribution of each tissue type and facilitate further analysis for downstream tasks. However, for most tasks (e.g., WSI-level diagnosis and subtyping), it is unnecessary to include all patches in the WSI, as it substantially decreases analysis efficiency and introduces noises. Therefore, in this chapter, we primarily focus on strategies to enhance the efficiency of WSI analysis.

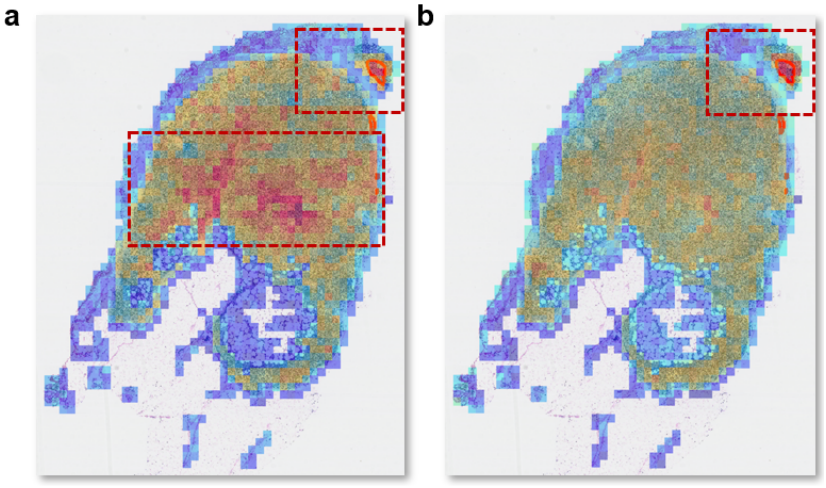


Figure 4-1 Methods based on the conventional attention mechanism in Transformers may result in sparse attention distributions and often focus on patches that lack clinical relevance, the red regions indicate tumor areas annotated by clinicians; a ABMIL *v.s.* b Ours (patches with high attention weights are highlighted).

In practice, MIL is a commonly used strategy in WSI analysis. To effectively weigh the significance of individual instances within WSI, attention mechanisms are extensively utilized. However, the use of global attention may result in feature homogenization, overlooking pathological tissue variations. The employment of local attention mechanisms can capture these differences, although it requires a higher number of parameters. On the other hand, the attention mechanism in Transformers may focus on tissue regions that are irrelevant to clinical diagnosis. As shown in Figure 4-1, attention is allocated to visually distinctive areas while lacking clinical significance, which greatly compromises the interpretability of the model. Moreover, since the attention mechanism considers all patches in the WSI simultaneously, the resulting attention distribution tends to be sparse overall.

In this chapter, we introduce FuzzyMIL, a deep fuzzy clustering framework designed for compact and efficient analysis of WSI. The core of FuzzyMIL is a learnable variant of Fuzzy C-means clustering named FCM, which iteratively updates fuzzy cluster centers to disentangle the complex morphological features present in WSIs. This process promotes the formation of more distinct and less correlated phenotypic patterns. By progressively refining the clustering centers, FCM effectively compresses the feature representation space, guiding features gradually toward the representation prototypes. Leveraging a soft assignment mechanism, these prototypes integrate information from all feature updates, enabling the model to preserve global structural context while maintaining sensitivity to local variations.

Since our approach is a deep network formulation of Fuzzy C-means clustering, we first discuss the relationship between Fuzzy C-means clustering and cross-attention to implement a learnable FCM within the attention architecture and dynamically update the cluster centers. Furthermore, we will deploy the FCM within the WSI analysis framework. Finally, we provide a complete workflow for efficient WSI analysis employing a local-to-holistic approach.

4.2 Relationship between Fuzzy Clustering Mechanism and Cross Attention

Inspired by the work of kMaX-DeepLab[71], which rethinks the relationship between pixels and object queries by framing cross-attention as a clustering process, we propose a method to replace the traditional attention layers in transformers with a multi-layer Fuzzy Clustering Attention (FCA) mechanism. By employing the learnable Fuzzy C-means algorithm, we aim to capture local attention centers, thereby refining the model's ability to focus on pertinent features and enhance its performance in capturing complex spatial dependencies.

The Fuzzy C-means algorithm assigns a sample to a cluster based on the degree of membership of the data point across all cluster centers. The update of the cluster centers can be formulated as:

$$c_j = \frac{\sum_{i=1}^{N} \left(u_{ij}\right)^m \cdot \boldsymbol{h}_i}{\sum_{i=1}^{N} \left(u_{ij}\right)^m} \qquad (4-1)$$

where

$$u_{ij} = \frac{1}{\sum_{k=1}^{Q} \left(\frac{\left\| \boldsymbol{h}_i - \boldsymbol{c}_j \right\|}{\left\| \boldsymbol{h}_j - \boldsymbol{c}_k \right\|} \right)^{\frac{2}{m-1}}} \quad (4-2)$$

here u_{ij} and $\boldsymbol{c}_i$ are components of the membership matrix $\boldsymbol{U} \in \mathbb{R}^{Q \times N}$ and clustering centers $\boldsymbol{C} \in \mathbb{R}^{Q \times d}$. In the context of WSI analysis tasks, we denote by $H \in \mathbb{R}^{N \times d}$ the offline features of the WSI extracted using the pretrained encoder (e.g. ResNet50[9] in this work), where $\{\boldsymbol{h}_i\}_{i=1}^{N}$ are component of $\boldsymbol{H}$, $\boldsymbol{h}_i$ corresponding to the offline feature of each individual patch. Since m is a hyperparameter, $\boldsymbol{U}$ can be represented by $\boldsymbol{C}$ and $\boldsymbol{H}$.

On the other hand, cross attention is employed to integrate related pixel features for updating object queries. The process can be formulated as $\widetilde{\boldsymbol{C}} = \mathrm{Softmax}_d(\boldsymbol{Q} \times \boldsymbol{K}^{\mathrm{T}}) \times \boldsymbol{V}$. In order to maintain consistency with Fuzzy C-means in the context of WSI analysis and facilitate the exploration of the relationship between the two, we express it as follows:

$$\widetilde{\boldsymbol{C}} = \boldsymbol{A} \times \sigma_v(\boldsymbol{H}) \quad (4-3)$$

where

$$\boldsymbol{A} = \mathrm{Softmax}_d(\sigma_c(\boldsymbol{C}) \times \sigma_h(\boldsymbol{H})^{\mathrm{T}}) \quad (4-4)$$

Among them, $\boldsymbol{A} \in \mathbb{R}^{Q \times N}$ and $\sigma_{c,h,v}$ are learnable mappings.

By comparing the two mechanisms above, it is evident that from the perspective of fuzzy clustering, the goal is to obtain a membership matrix that assigns patch-level features to different clusters, with each cluster representing a distinct morphological prototype within the WSI. Similarly, cross-attention also aims to derive attention weights indicating the degree to which different patch features belong to the clusters.

Additionally, when the feature queries are treated as the centers of the clusters, the mathematical formulations of the attention map $\boldsymbol{A}$ and the membership matrix $\boldsymbol{U}$ become consistent.

Based on the observations, if we replace $\boldsymbol{U}$ with the parameterized estimator derived from $\boldsymbol{C}$ and $\boldsymbol{H}$, and substitute this approximation for $\boldsymbol{A}$ in cross-attention mechanism, an alternative and learnable version of the Fuzzy C-mean algorithm can be implemented using deep network. Then the formulation of our FCA can be expressed as follows:

$$\boldsymbol{C}_{i+1} = f(\boldsymbol{C}_i, \boldsymbol{H}_{i+1}) \times \sigma_v(\boldsymbol{H}_{i+1}) \tag{4-5}$$

where

$$f(\boldsymbol{C}_i, \boldsymbol{H}_{i+1}) = \mathrm{Softmax}_d\Big(\exp(nol(\boldsymbol{C}_i) \times nol(\boldsymbol{H}_{i+1})^{\mathrm{T}})\Big) \tag{4-6}$$

is a learnable membership estimator. $nol(\cdot)$ maps $\boldsymbol{C}$ and $\boldsymbol{H}$ into a higher dimensional space. In this work, it represents a Convolution-based structure. The exponential function is utilized to ensure that u_{ij} is nonnegative. In this way, FCA (Equ. (4-6)) can learn to control the degree of fuzziness through training. This is accomplished by substituting the membership matrix of the original version (typically depending on the distance between features and clustering centers) with the designed membership estimator. This adjustment preserves the soft clustering nature of the Fuzzy C-means algorithm while allowing for dynamic adaptation during the whole process, which enhances its ability to capture the relationships between features and cluster centers (prototype centers) more effectively.

An ideal iteration can be updated with each transformer layer. Given the variable number of instances in each WSI, after each FCA, both $\boldsymbol{C}$ and $\boldsymbol{H}$ are passed through a standard self-attention module. Therefore, a complete iteration consists of three steps:

$$Step\ 1: \boldsymbol{C}' \leftarrow f(\boldsymbol{C}, \boldsymbol{H})$$

$$Step\ 2: \boldsymbol{C}' \leftarrow AT(\boldsymbol{C}'\boldsymbol{W}^q, \boldsymbol{C}'\boldsymbol{W}^k, \boldsymbol{C}'\boldsymbol{W}^v)$$

$$Step\ 3: \boldsymbol{C}_{out} \leftarrow FFN(\boldsymbol{C}') + \boldsymbol{C}'$$

where *AT* is the self-attention mechanism, $\boldsymbol{W}^q, \boldsymbol{W}^k$ and $\boldsymbol{W}^v$ are learnable matrices in *AT*, *FFN* is the feed forward network. As a result, this iterative process gradually separates the features, enabling the representation of the WSI prototype to show decreasing correlations.

4.3 Global Feature Encoder

Multiple clustering iterations imply multiple inputs of the global structural features of WSI, requiring that the Global Feature Encoder (GFE) be both efficient and capable of effectively capturing the global information embedded within the WSI. To address this, a local-to-global strategy is employed in this work.

First, WSI feature bag $\boldsymbol{H}_i$ is first squared and then divided into $K \times K$ regions with $P \times P$ resolution uniformly to encode the reginal information. Specifically, $\boldsymbol{H}_i = \{H_k^i\}_{k=1}^{K^2}$, where $H_k^i \in \mathbb{R}^{P^2 \times d}$ and $K \times P = \lceil \sqrt{N} \rceil$. Here, d represents the dimension of the features after dimensionality reduction.

For every region, we begin by utilizing the positional encoding method for patch embeddings. Following[72], we employ a lightweight 1-D convolution to more efficiently encode the features within each region individually. Subsequently, the

features are flattened and encoded individually using self-attention. The encoding process for positional information only considers the features within each region, which limits its ability to model context-based semantic features. This is crucial for downstream tasks that require more comprehensive global encoding. To effectively capture the contextual relationships between regions, we employ a cross-region attention mechanism. Then the global features can be updated as $\boldsymbol{H}_{i+1}$.

4.4 Pipeline

Consistent with the Fuzzy C-means algorithm, the proposed FuzzyMIL is also an iterative process. Within the Transformer framework, the fuzzy cluster centers (prototype features) are iteratively updated, gradually decoupling to obtain a representative prototype representation of the WSI.

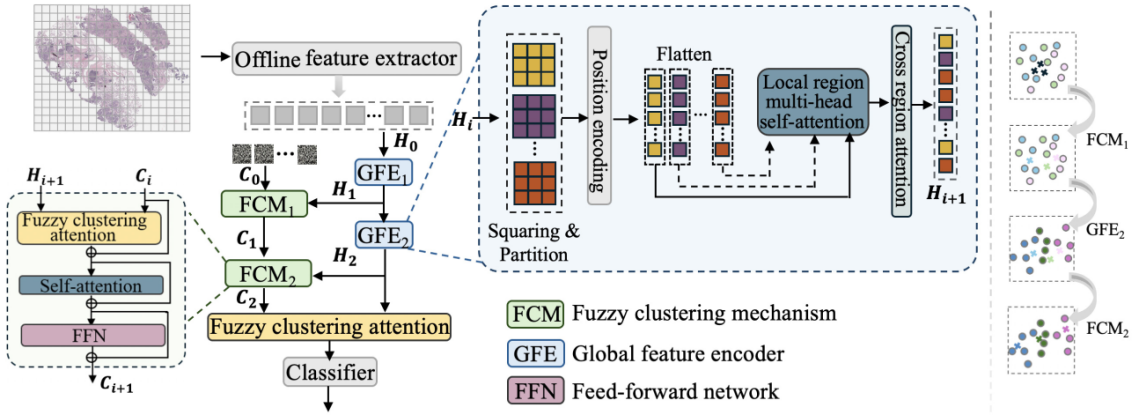


Figure 4-2 Overview of FuzzyMIL: The WSI is first divided into patches, which are then processed through an offline feature extractor. Patch embeddings are passed into the GFE-FCM iteration. The GFE captures features with a global perspective using a local-to-global strategy, while the FCM decouples pathological features through fuzzy clustering centers. The schematic diagram on the right illustrates how the dynamics of

the instance features (represented by dots) and fuzzy clustering centers (depicted as morphological phenotypes, represented by crosses) evolve across iterations of FCM and GFE. As the clustering centers are updated, their distribution shifts from dense to sparse, indicating a transition in the correlation (measured by the distance between cluster centers): from strong (proximity) to weak (separation).

As shown in Figure 4-2, we propose a compact WSI representation framework that leverages the Fuzzy C-means algorithm and applies it to subsequent tasks. Drawing on the potential similarities between the Fuzzy C-means algorithm and the cross attention, we introduce our proposed FCM to implicitly disentangle the features of the WSI. Building on the GFE to capture global information as much as possible, the fuzzy clustering centers (prototype features) are updated through the collaboration of GFE and FCM. The final compact representation, referred to as the pathological phenotypes of the WSI, is achieved through the iterative interaction between FCM and GFE.

Denoting Q initial fuzzy clustering centers as $\boldsymbol{C}_0 \in \mathbb{R}^{Q\times d}$ within a WSI, we start by cropping the WSI into non-overlapping patches and a pre-trained feature extractor to encode them into $\boldsymbol{H}_0 = \{\boldsymbol{h}_0^j\}_{j=1}^{N}$, where N is the total number of patches and $\boldsymbol{h}_0^j \in \mathbb{R}^{1\times d}$ among them. The GFE module is then applied to re-encode the WSI into $\boldsymbol{H}_1$, capturing both local and global features, thus enhancing the overall global perspective. Next, the FCM algorithm further clusters $\boldsymbol{H}_1$ into Q clusters to update centers as $\boldsymbol{C}_1$. After repeating the iterations of the "GFE-FCM" two times and the single application of FCA, an effective representation of the WSI is obtained. This can be understood as Q decoupled morphological prototypes in our framework. A pooling layer based on

self-attention is then applied to assign weights to these prototypes, followed by a prediction head to carry out the relevant downstream tasks.

4.5 Datasets and Implementation Details

In this work, we evaluate the performance of FuzzyMIL in diagnosis and subtyping tasks on CAMELYON-16 (C16), TCGA-BRCA and TCGA-NSCLC datasets.

- C16 - We evaluate the diagnostic performance of the model on C16, which includes WSIs from breast cancer patients and normal cases. This dataset is widely used for evaluating algorithms in the domain of computational pathology and cancer diagnosis.
- TCGA-BRCA - We validate the performance of subtyping on BRCA, specifically by training the proposed FuzzyMIL to differentiate between the two main subtypes present in this dataset: Invasive Ductal Carcinoma (IDC) and Invasive Lobular Carcinoma (ILC).
- TCGA-NSCLC - In NSCLC, Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC) are two major subtypes. FuzzyMIL was trained to subtype these two types of non-small cell lung cancer.

In these datasets, all WSIs are divided into 512 × 512-pixel patches at a 20× magnification. The processing of patches followed the method outlined in CLAM[10]: Features are extracted using the ResNet50[9] pre-trained on ImageNet, which embedded each patch into a 1024-dimensional vector. These feature vectors are then reduced to 256 dimensions via a fully connected layer. For the fuzzy clustering process, the number of fuzzy clustering centers utilized in this study is set to 64.

We evaluated our framework by comparing it with efficient deep learning models using several metrics, including AUC, accuracy (Acc.), F1-score (F1), precision (Pre.), and the number of parameters (#Params). A universal MIL classifier is utilized, optimized with the Adam optimizer (learning rate: 2e-4, cosine weight decay: 1e-5) over 100 epochs. For TCGA-BRCA and TCGA-NSCLC, 5-fold cross-validation is applied, whereas for C16, 3-fold cross-validation is used.

The results to be presented in the next section show the mean and standard deviation across the different folds. In the GFE module, instance features are squared, divided into K × K (8 × 8) regions, and encoded using an 8-head attention mechanism to extract global features ($\boldsymbol{H}_i$). The experiments are performed with a batch size of 1 on NVIDIA RTX 3090 GPUs.

4.6 Experimental Results

In order to perform a comprehensive comparison, we assessed our method with other current approaches, which we organized into three groups: attention-based multi-instance learning approaches, methods based on transformer architecture, and those utilizing feature re-embedding.

- ABMIL[70] - A basic MIL model utilized an attention mechanism to assign different weights to instances within a bag, allowing the model to focus on the most relevant features for improving pattern recognition in complex data;
- DSMIL[73] - A multi-instance learning method that incorporates deep supervision at multiple layers of the network, enabling better gradient flow and enhancing the model's ability to learn relevant features from complex data;

- CLAM[10] - It utilizes class activation maps to identify discriminative regions within the slide images, improving both the accuracy and interpretability of the model for computational pathology tasks with limited labeled data;
- DTFD-MIL[74] - A method that uses a two-tier feature distillation process with multi-instance learning to improve histopathology whole-slide image classification, enhancing model performance with weakly labeled data;
- TransMIL[48] - A method that employs transformer models in multi-instance learning to capture long-range dependencies and improve the classification of histopathology whole-slide images;
- R^2T[72] - A method that refines features extracted from whole-slide images by re-embedding them into a more suitable space, enhancing the model's ability to perform at foundation model-level performance in computational pathology tasks.

Table 4- 1 **a** Subtyping results on TCGA-BRCA. The highest performance is in bold.

Method	AUC	F1	Acc.	Pre.	#Params(M)
ABMIL	83.90±0.29	80.82±0.88	81.53±0.37	73.45 ±0.21	-
DSMIL	88.03±0.63	84.22±1.20	85.65±0.93	74.86 ±0.49	-
CLAM	87.37±0.46	85.04±1.01	83.21±0.53	74.25 ±0.67	-
DTFD-MIL	88.39±0.42	84.54±0.67	87.90±0.32	77.01± 0.89	-
TransMIL	87.98±1.02	85.93±2.09	82.21±3.59	76.43 ±1.28	-
R^2T	91.93±0.28	85.47±0.36	90.52±0.19	83.44 ±0.43	2.64
Ours	**94.94±0.37**	**87.16±0.22**	**90.81±0.19**	**84.50 ±0.32**	**1.52**

Table 4-1 **b** Subtyping results on TCGA-NSCLC. The highest performance is in bold.

Method	AUC	F1	Acc.	Pre.
ABMIL	93.53±1.21	87.96±1.53	87.22±1.01	75.20 ±0.69
DSMIL	93.80±0.95	88.03±1.78	87.38±1.99	76.24 ±0.85
CLAM	93.69±1.50	88.21±0.98	87.97±0.87	77.29 ±1.23
DTFD-MIL	94.21±1.38	89.95±0.79	89.09±2.10	79.46 ±1.09
TransMIL	92.03±1.73	86.34±1.86	86.90±1.24	77.46 ±1.17
R2T	95.50±1.25	90.19±1.93	90.44±2.03	84.25 ±1.30
Ours	**96.08±0.69**	**91.22±1.23**	**91.85±1.40**	**85.43 ±0.92**

Table 4-1 **c** Diagnosis results on C16. The highest performance is in bold.

Method	AUC	F1	Acc.
ABMIL	86.75±0.21	82.82±1.03	84.23±0.57
DSMIL	90.25±0.90	84.22±1.20	88.03±0.43
CLAM	87.72±0.76	81.03±0.95	83.10±0.25
DTFD-MIL	93.39±0.33	86.58±0.16	89.94±0.14
TransMIL	87.14±1.98	81.93±4.13	82.21±2.59
R2T	95.31±0.19	88.98±0.56	90.06±0.48
Ours	**97.41±0.20**	**91.31±0.38**	**91.73±0.12**

Table 4-1 **a** and **b** show the subtyping performance of various methods on the TCGA-BRCA and TCGA-NSCLC datasets. The results indicate that our model achieves the best performance across both datasets in terms of AUC, F1, Accuracy, and Precision, especially with AUC of 94.94±0.37 for TCGA-BRCA and 96.08±0.14 for TCGA-NSCLC, outperforming all other methods. Moreover, our method achieves competitive performance while maintaining a relatively low number of parameters (1.52M), which suggests efficiency in terms of both accuracy and model size.

The diagnosis results on C16 in Table 4-1 **c** show a similar trend. Given that only

small regions of tumor tissue are present within the overall tissue area of a WSI in C16, our model still demonstrates competitive performance. We attribute these significant improvements to the effective decoupling of WSI features, which allows for the capture of crucial phenotypic information while minimizing the influence of redundancy.

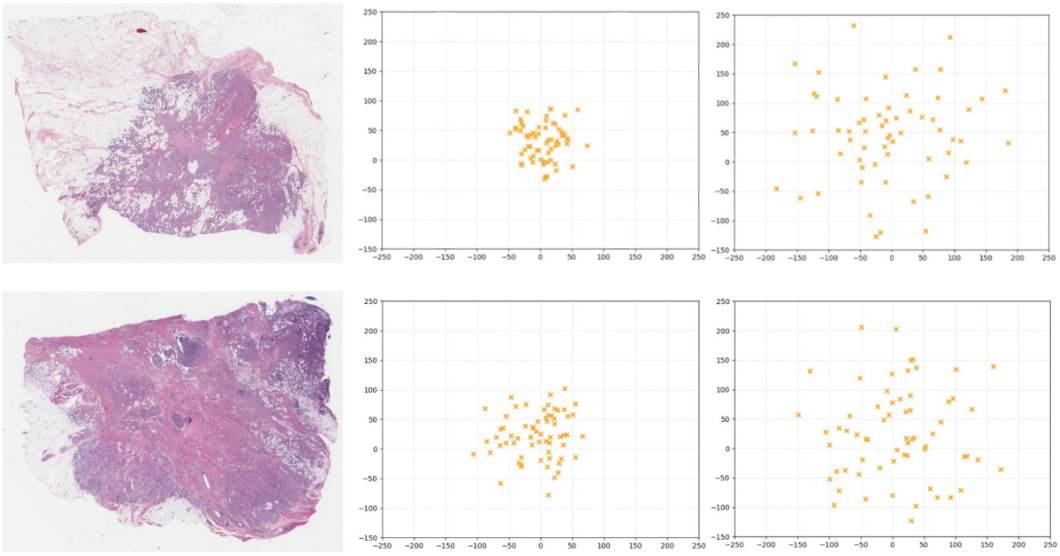


Figure 4-3 t-SNE visualization of cluster centers after the first and second iterations. As the iterations progress, the cluster centers (representing pathological phenotypes) shift from a dense to a sparse distribution, indicating a reduction in the correlation between the prototypes.

Figure 4-3 illustrates the evolution of cluster centers across “GFE-FCM” iterations. As the process progresses, the clustering centers (representing pathological phenotypes) transition from a dense to a sparse configuration, signaling a decrease in the correlation between the prototypes. These changes in the spatial distribution of the fuzzy clustering centers confirm that the clustering process effectively decouples interrelated features, resulting in more independent phenotypes with minimal correlation.

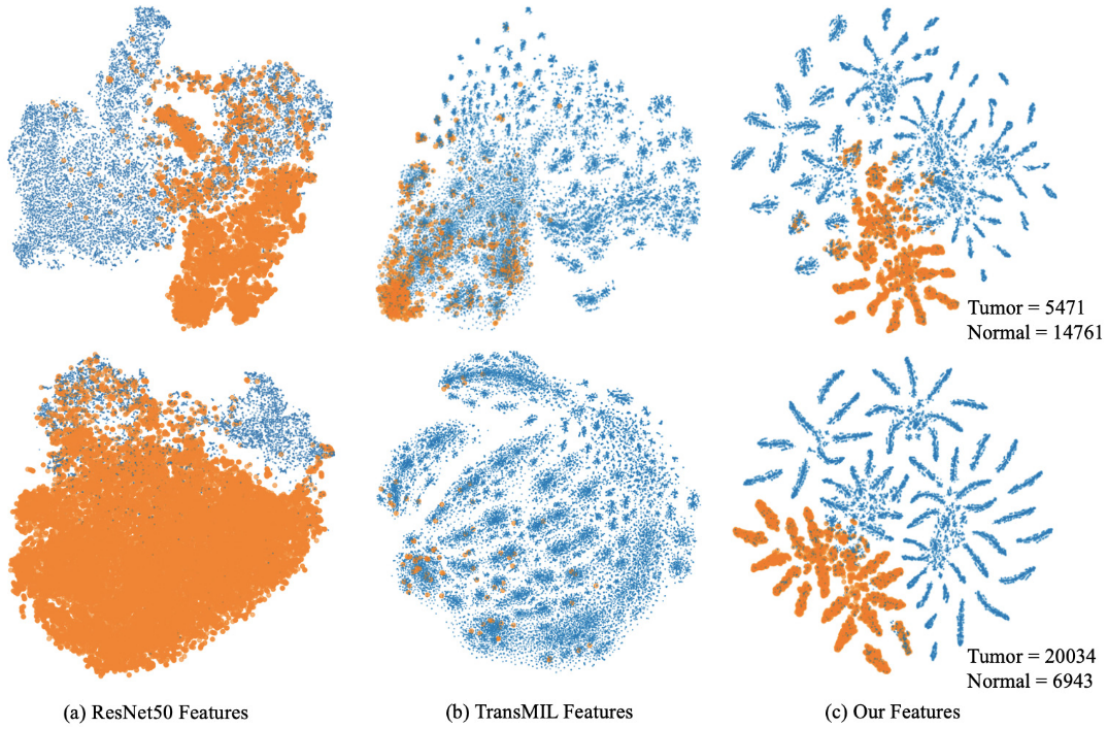


Figure 4-4 t-SNE visualization of instance features filtered by attention scores from CAMELYON-16. **a** Original offline features extracted using a pre-trained ResNet50, **b** features after applying TransMIL, and **c** features after applying our model. Orange dots represent tumor instances, while blue dots represent normal instances.

To further investigate the comparative effectiveness of the original global attention mechanism in Transformer and our approach, we employed t-SNE for a more detailed analysis. We employ TransMIL as the benchmark for comparison. The t-SNE visualizations include the following: raw offline embeddings, embeddings from the standard Transformer (TransMIL), and embeddings from our model. For C16, we extracted instance features with the associated attention scores learned by both the TransMIL and our model. Using instance-based annotation files, we label these features to visualize and compare how different attention mechanisms cluster positive and negative class features. To evaluate the models' attention to positive instances, we

filtered the features by applying a threshold to the attention scores. As a baseline, we compared these results to the instance features directly extracted from the original offline feature extraction.

As shown in Figure 4-4, our analysis reveals several important observations. In the t-SNE visualization **a**, positive and negative features are intermixed, indicating that the offline feature extractor fails to capture discriminative instance-level representations. The TransMIL model **b** improves feature separation, offering a more structured distribution. However, when focusing on instance features with attention scores greater than 0.1, only a small number of tumor-labeled instances are emphasized by TransMIL, indicating limited attention to clinically relevant features. On the other hand, our model **c** demonstrates stronger feature mining capabilities, which activates attention on a greater number of tumor instances and shows clearer separation between classes. This figure shows that FuzzyMIL more effectively addresses feature homogenization and enhances the ability of attention mechanism to focus on clinically relevant patterns.

4.7 Ablation Experiments

Exploring the trade-off between model complexity (measured by the number of parameters) and predictive accuracy is essential. Therefore, we first conduct ablation experiments varying the number of clustering centers.

Table 4-2 Ablation experiments for different numbers of clustering centers on C16.

#Centers	AUC	F1	Acc.	Pre.	#Params (M)
16	94.86	89.82	89.72	86.33	1.10
32	95.27	90.20	90.88	86.39	1.37
64	97.04	**91.12**	92.51	86.47	1.52
128	**97.28**	90.97	**93.42**	**87.56**	2.04

In Table 4-2, as the number of cluster centers increases, the model's performance across several metrics shows consistent improvement. At the same time, there is a trade-off in terms of model complexity, measured by the number of parameters. As the number of cluster centers grows, the model's parameter counts increases as well. Given this, 64 cluster centers offer an optimal balance between model performance and complexity, yielding an accuracy of 92.51%, AUC of 97.04%, and precision of 86.47%, with a modest increase in parameters to 1.52 million. Thus, we selected 64 cluster centers to balance model complexity and accuracy.

In this ablation study, we further investigate the effect of integrating FCA into different encoding strategies for MIL in the context of WSI analysis. Our focus is on understanding how FCA enhances the performance of various encoding approaches. Specifically, we examine two distinct methods: TransMIL, a transformer-based model with global attention mechanisms, and GFE (utilized in this work), which utilizes a combination of both global and local encoding to capture detailed information from WSIs. For each of these encoding strategies, we incorporate FCM to assess its impact on model performance, evaluating improvements across various performance metrics.

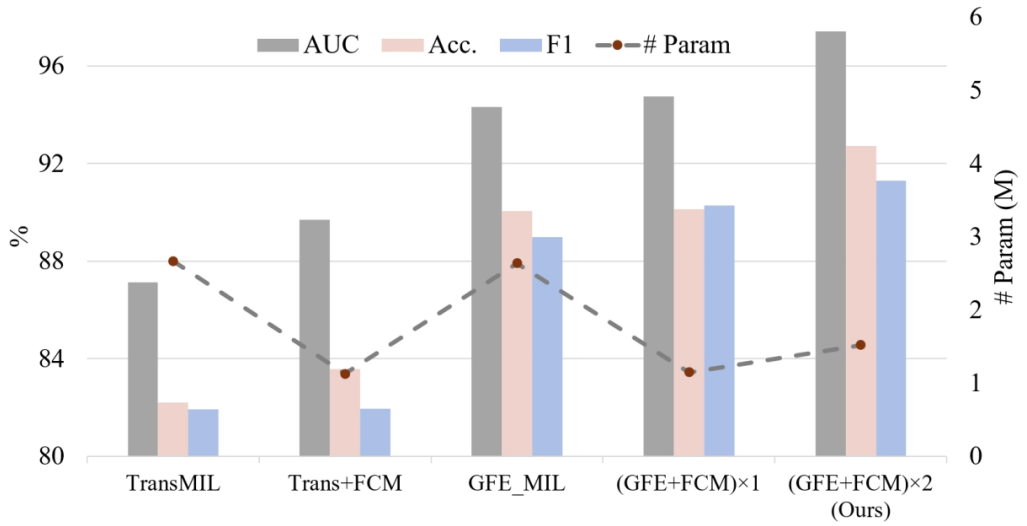


Figure 4-5 Ablation experiments on different encoding strategies. After plugging in FCM, MIL algorithms with different embedding strategies are efficiently improved while maintaining accuracy.

Figure 4-5 shows that FCM consistently leads to performance improvements across various models. Additionally, it significantly reduces the number of parameters, particularly in terms of enhancing the efficiency of other methods when integrated with alternative embedding techniques. This highlights the ability of FCM to deliver a more compact data representation while maintaining accuracy. In conclusion, fuzzy clustering effectively balances performance enhancement with reduced model complexity, making it a valuable approach for efficient and accurate WSI analysis.

4.8 Conclusion

In this work, we propose a novel approach for WSI analysis by integrating learnable fuzzy clustering into MIL, with a particular emphasis on decoupling local morphological features. Our approach involves the iterative updating of fuzzy clustering centers, which allows for continual refinement of the learned features within

the transformer architecture. This dynamic refinement ensures that the model retains the most clinically significant and clinically relevant centers of attention throughout the learning process.

One of the key advantages of our method is its efficiency, as it strikes a balance between reducing the number of parameters and maintaining high accuracy. Furthermore, our approach makes it easily integrable with other MIL embedding methods. This flexibility not only enhances the potential for more efficient analysis but also opens new avenues for deriving deeper insights from WSI data. In conclusion, our proposed method represents a significant advancement in WSI analysis by combining the power of fuzzy clustering and multi-instance learning to better capture clinically relevant patterns, all while maintaining computational efficiency and flexibility for future applications.

Benefiting from the plug-and-play nature of FCM, we may further incorporate multimodal information or guidance from LLM in the future to complement the information in WSI. It should be noted that FuzzyMIL demonstrated suboptimal performance in TCGA-KIRC subtyping, which may be attributed to the multifocal growth pattern of KIRC. Additionally, we will further investigate its computational costs in actual clinical deployment in future studies to validate its clinical deployability.

Chapter 5.

Clinical Application of AI-Based WSI Analysis Technologies for Prognostic Analysis and Treatment Evaluation in VETC Patients

5.1 Introduction

Based on the WSI analysis framework outlined above, this chapter will discuss a specific clinical application: the localization of Vessels that encapsulate tumor clusters (VETC) in WSI of HCC patients, as well as the prognostic analysis and treatment evaluation based on VETC.

According to global cancer statistics, primary liver cancer ranks as the third leading cause of cancer-related deaths worldwide. Hepatocellular carcinoma (HCC) constitutes approximately 85%-90% of all primary liver cancer cases. In China alone, there are 466,000 new cases of HCC annually, making up 55.4% of the global total. Liver resection remains the primary treatment for HCC. However, the recurrence rate of HCC post-resection is still as high as 50%[75]. This frequent recurrence and high mortality rate are primarily due to early metastasis of the cancer. The classical pathway for cancer cell metastasis involves the epithelial-mesenchymal transition (EMT). Through histological examination of sequential HCC sections and three-dimensional reconstruction, a novel and prevalent vascular pattern has been identified. This involves VETC[76] and form cobweb-like network.

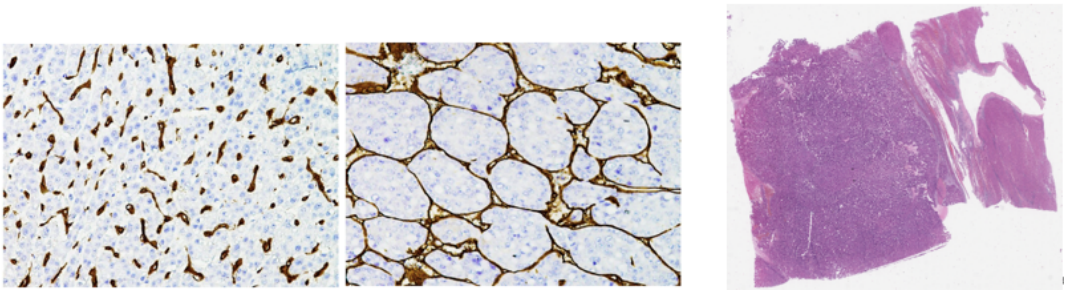


Figure 5- 1 VETC- and VETC+ patterns in CD34 (left), H&E staining image (right).

As shown in Figure 5- 1, The VETC+ patient is currently defined as ≥ 55% tumor area by CD34 immunostainin. But the preparation of CD34 is both time-consuming and expensive, which presents a significant challenge for the definitive diagnosis of VETC+. In contrast, H&E-stained WSIs are comparatively more affordable, and most existing studies on WSIs are based on H&E staining images (see Figure 5- 1 (right)). If we can identify VETC+ regions and assist in diagnoses solely based on H&E staining, it would not only substantially shorten the diagnostic process but also reduce costs and the financial burden on patients. Thus, developing a fully automated VETC diagnostic system based on H&E WSI would be highly valuable to streamline the diagnostic process while ensuring accuracy and efficiency.

Recently, VETC was founded to be a significant risk factor for the recurrence and survival of patients with HCC after hepatectomy[76][77]. Among that, Wang et al. have shown that VETC significantly influence both DFS and overall survival OS in patients undergoing hepatectomy, with varying impacts across different stages of HCC. Lu et al. have also highlighted that the combined patterns of VETC and EMT affect the prognosis of HCC patient's post-resection. Further, Wang et al. have found that integrating the VETC pattern with tumor characteristics can help stratify patient

outcomes and responses to adjuvant transarterial chemoembolization (TACE) therapy[78]. Despite the established importance of VETC in HCC prognosis, most current research remains limited to outcome-based analysis, without a quantitative measure of VETC's prognostic impact. Moreover, existing studies on VETC are often narrow in scope, typically focusing on patients at a specific stage of HCC or undergoing certain treatments. To our knowledge, there are currently less studies that have developed markers based on VETC to quantitatively evaluate the relationship between VETC and prognosis. Therefore, it is crucial to develop a new VETC-based marker with broader applicability to accurately quantify its effects on the prognosis of HCC and generalization to be utilized across various disease stages and treatment regimens.

In this work, we propose VETC Net, which can automatically differentiate VETC+ and VETC- in WSI and further calculate the distribution of VETC+. Based on this, we assess the impact of VETC+ distribution on the prognosis of patients with HCC. VETC Net is developed using an internal cohort from SYSUCC and validated in an external cohort composed of multiple centers. Additionally, we evaluated the effect of VETC on different treatment modalities, including adjuvant TACE, adjuvant HAIC, neo-adjuvant HAIC, and neo-adjuvant Triplet therapies.

5.2 Patient Cohort and Study Design

We collected data from 510 patients, along with their corresponding whole slide images (WSIs), at SYSUCC, which served as the internal cohort for the development of VETC Net. This internal cohort provided the foundation for training and optimizing the model. To ensure the robustness and generalizability of VETC Net, we further

assembled an external validation cohort consisting of 363 patients and their corresponding WSIs from multiple prominent institutions, including The First Affiliated Hospital of Sun Yat-sen University, Zhujiang Hospital, and Guangzhou Hospital of Traditional Chinese Medicine. This diverse external cohort enabled us to evaluate VETC Net's ability to perform consistently across different patient populations and clinical settings. The primary aim of this validation was to assess the model's capability to accurately stratify patients into high-risk and low-risk groups based on external data, confirming its broad applicability.

All patients in the above cohorts underwent hepatic resection as part of their treatment plan, ensuring uniformity in the surgical context. In addition to validating the model's performance in stratifying risk, we utilized VETC Net to assess the prognostic significance of VETC+ status in patients who received different treatment modalities. Specifically, we evaluated the relationship between VETC+ and patient prognosis in those who underwent adjuvant transarterial chemoembolization (TACE) (225 samples) and adjuvant hepatic artery infusion chemotherapy (HAIC) (165 samples). These analyses highlighted the potential of VETC Net as a tool for evaluating treatment outcomes and further demonstrated its utility in clinical decision-making.

Finally, we extended our evaluation by analyzing needle biopsy images from a cohort of patients treated with neo-adjuvant therapies. This included patients who received neo-adjuvant HAIC (92 samples) and neo-adjuvant Triplet therapies (118 samples). Using these biopsy images, we assessed the ability of VETC Net to predict and distinguish between high- and low-risk prognostic groups in the context of neo-

adjuvant treatment. This final analysis emphasized the flexibility of VETC Net in adapting to various clinical scenarios, showcasing its potential to be used across a broad range of therapeutic approaches.

5.3 Method

This study developed an AI model named VETC Net to automatically and precisely localize the VETC+ regions within WSIs and further conduct prognostic analysis and treatment evaluation for HCC patients based on the distribution of VETC+. VETC Net employs an instance-level supervised learning strategy to distinguish between VETC+ and VETC- regions within WSIs, thereby predicting the spatial distribution of VETC for downstream analysis.

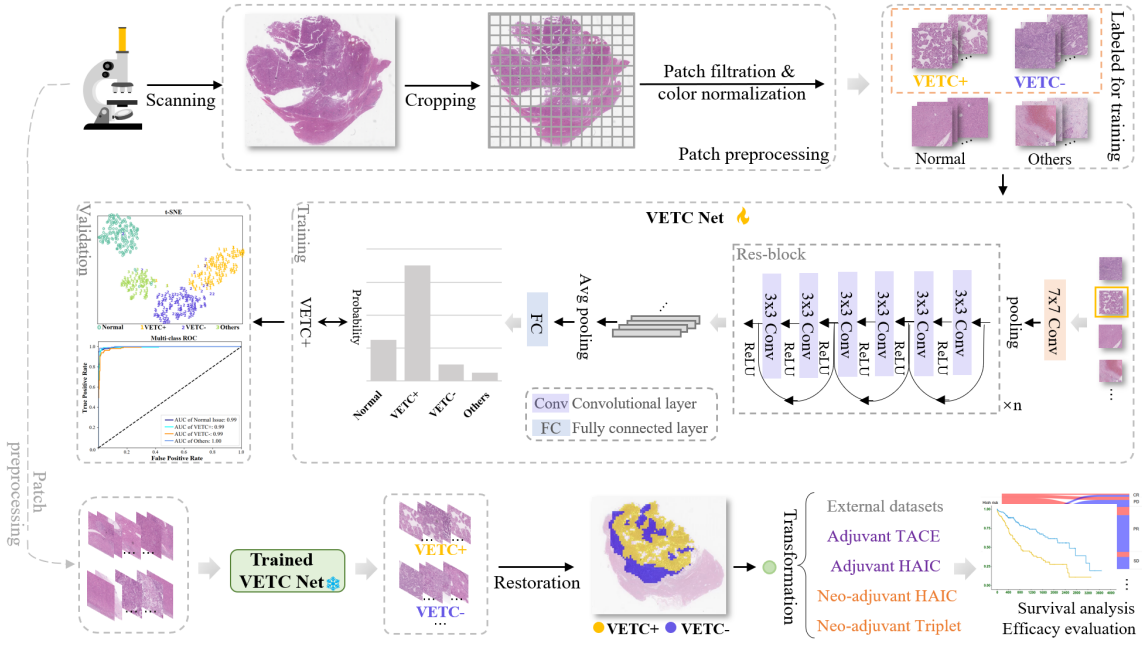


Figure 5-2 Workflow of establishing VETC Net: WSIs are first scanned and cropped into patches, followed by filtration and color normalization for preprocessing. The patches are annotated as VETC+, VETC-, normal tissue or others by pathologists.

VETC Net is trained and validated on the internal dataset to identify VETC+ and VETC- regions in the WSI. The trained model is then applied to external datasets for survival analysis and treatment efficacy evaluation.

As shown in Figure 5-2, the entire process of model development and evaluation is systematically divided into three key stages:

1) Preprocessing: In the initial phase, we begin by working with the internal dataset, which consists of WSIs from patients. Each of these WSIs is cropped into 512×512 patches at a magnification of 40x, following the methodology outlined in CLAM[10]. This magnification level allows for detailed cellular and tissue-level examination while maintaining a manageable image size for computational analysis. Once the patches are extracted, they undergo a series of preprocessing steps aimed at improving data quality. These patches are first filtered to remove any low-quality or irrelevant regions that may contain artifacts, such as background noise, smearing, or areas that lack sufficient tissue structure. Additionally, staining normalization is performed to correct any inconsistencies in staining across the slides. Then the patches are annotated by experienced pathologists, who identify and label them based on CD34, which is widely used to identify vascular endothelial cells. The pathologists classify the patches into four categories based on their morphological features: "VETC+", "VETC-", "Normal tissue", and "Others". The distribution of the patch-level dataset is shown as follows:

Table 5-1 The distribution of the patch-level dataset for the development of VETC Net.

	Normal	VETC+	VETC-	Others
Train	2341	2340	2317	1817
Validation	292	290	288	215
Test	292	291	288	215

2) Training and validation of VETC Net: The backbone of VETC Net is based on ResNet34, which is used to extract features from the different categories of patches and to train and validate the patch-level four-class classification task. All patch-level operations are carried out using the patches derived from the 120 WSIs mentioned earlier. In this way, the trained VETC Net can learn discriminative features related to the presence of VETC- and VETC+. Once trained, the model is used to predict the labels for each patch, which are then integrated back into the WSI format. This integration allows for the reconstruction of the entire slide, creating a spatial distribution of VETC+ and VETC- regions at the WSI level. We further divide WSIs of the remaining internal dataset into WSI-level training and validation dataset with the 8:2 ratio to determine the optimal cutoff value of the VETC+ proportion relative to the tumor area for survival analysis. We finally identify the optimal cutoff value as 0.452, allowing for the stratification of patients based on the distribution of VETC+ regions within their tumors.

3) External validation and efficacy evaluation of different treatment strategies: To assess the generalizability and clinical applicability of the VETC Net model, we conducted external validation using a cohort of patients from multiple institutions, including The First Affiliated Hospital of Sun Yat-sen University, Zhujiang Hospital, and Guangzhou Hospital of Traditional Chinese Medicine. This external cohort allowed

us to evaluate the model's ability to accurately distinguish between patients who are at high risk of recurrence and those with high mortality, based on the distribution of VETC+ and VETC- regions in their tissue samples. The external validation was critical for confirming the robustness of VETC Net across diverse patient populations and clinical settings. Following the initial validation, we further tested the model's efficacy in evaluating different treatment strategies. Specifically, we validated the model's capability to evaluate the efficacy of various treatment strategies in patients receiving adjuvant TACE and adjuvant HAIC. Finally, we further validated the model's generalizability through biopsy samples from patients undergoing neo-adjuvant HAIC and neo-adjuvant Triplet therapies. Based on this, we are able to assess the potential of VETC Net in predicting treatment responses and identifying prognostic factors that could guide clinical decision-making.

5.4 Experimental Results

For the patch-level dataset, we extract patches with 512x512 resolutions from 120 WSIs at 40x magnification in the internal dataset, obtaining a total of more than 1 6000 patches. We performed data selection on these patches and used 5-fold cross-validation to train and validate the VETC Net, enabling it to accurately distinguish between VETC+ and VETC- tissues. The patch-level evaluation results have been shown in Figure 5-2.

For the WSI-level internal dataset, we divide the WSIs corresponding to 450 patients into training and validation sets at 8:2, in order to further determine the optimal cutoff value for the VETC+ distribution ratio associated with high recurrence and high

mortality risks.

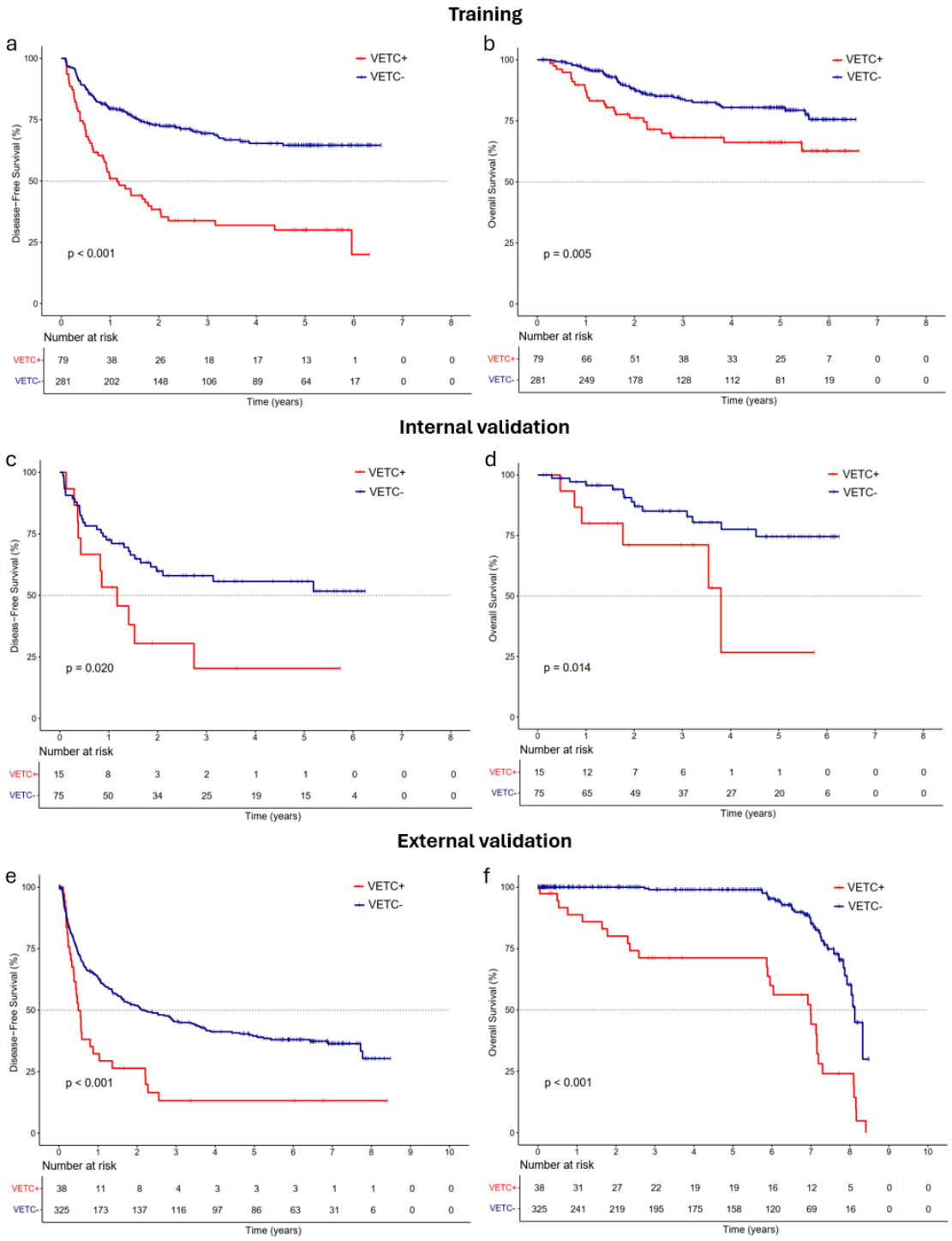


Figure 5-3 KM curves for internal training, internal validation, and external validation sets. KM curves of DFS **a**, **c**, **d** and OS **b**, **d**, **e** under the VETC+ ratio predicted by VETC Net at the optimal cutoff value.

To identify the prognostic significance of VETC+ proportion within the tumor area,

we first employed R to determine the optimal cutoff value based on DFS and OS outcomes in the training set. Through this analysis, we established 0.452 as the optimal threshold for the VETC+ ratio (Figure 5-3 **a** and **b**). Next, we validated the prognostic utility of this cutoff in the internal validation set. As illustrated in Figure 5-3 **c** and **d**, patients identified by VETC Net as VETC+ (defined as having a VETC+ proportion greater than 0.452 within the tumor region) exhibited significantly poorer prognosis. Specifically, these patients demonstrated a higher risk of recurrence and mortality compared to those classified as VETC-, with statistical significance observed in both DFS ($p=0.02$) and OS ($p=0.014$). To further evaluate the robustness and generalizability of VETC Net, we applied the model to an independent external validation cohort comprising 363 patients from three medical centers: The First Affiliated Hospital of Sun Yat-sen University, Zhujiang Hospital, and Guangzhou Hospital of Traditional Chinese Medicine. It can be seen that patients predicted to be VETC+ had significantly worse prognoses than VETC- patients (Figure 5-3 **e** and **f**). Both DFS and OS were markedly reduced in the VETC+ group, with p-values < 0.001 for both outcomes. These results collectively reinforce the predictive power and broad applicability of VETC Net in identifying high-risk HCC patients across multiple clinical settings.

To compare the ability of VETC Net and human pathologists to identify VETC+ patients, we adopted a predefined cutoff value for VETC Net. Specifically, a patient was classified as VETC+ if the VETC Net predicted that VETC+ regions accounted for more than 45.2% of the tumor area; otherwise, the patient was classified as VETC-. Similarly, two pathologists (Pathologist #1, a senior pathologist, and Pathologist #2, a

junior pathologist) classified patients based on a previously reported[76][77] cutoff value. A patient was considered VETC+ if the pathologist determined that more than 55% of the tumor area was VETC+, and VETC- otherwise.

Both the model and the pathologists made their decisions based solely on H&E-stained images. The ground truth for this evaluation was established based on CD34 staining to label each patient. Consistently, a cutoff value of 45.2% was used to define the ground truth for VETC Net evaluation, while a cutoff of 55% was applied for the ground truth used in the assessment by human pathologists.

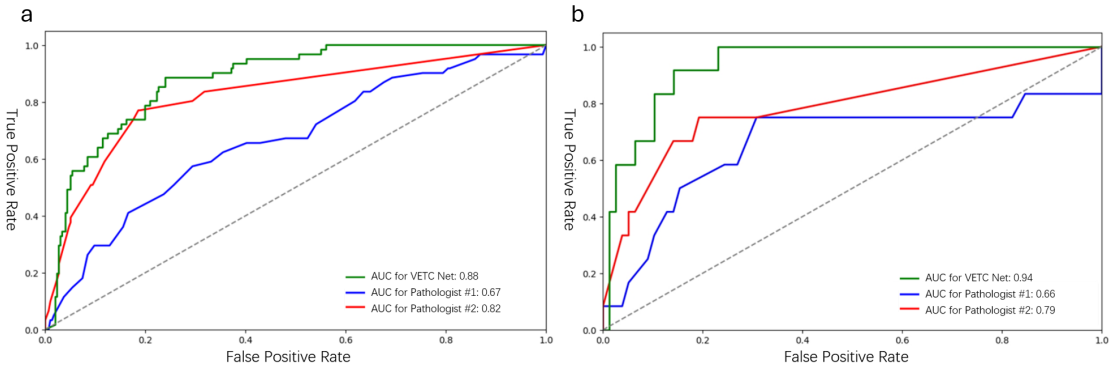


Figure 5-4 The ROC curves for the **a** training and **b** validation sets of the internal cohort, comparing the accuracy of VETC Net and two pathologists in classifying VETC+ patients. The threshold for VETC Net was set at 0.452, while the classification threshold for the pathologists was 0.55[76].

As shown in Figure 5-4, we compared the VETC Net with Senior Pathologist #1 and Junior Pathologist #2 on the WSI-level dataset. In the validation set, the accuracy of VETC Net significantly exceeded that of Pathologist #2, with an AUC of 0.94, which was also higher than that of Pathologist #1 (AUC = 0.79), achieving highly competitive results.

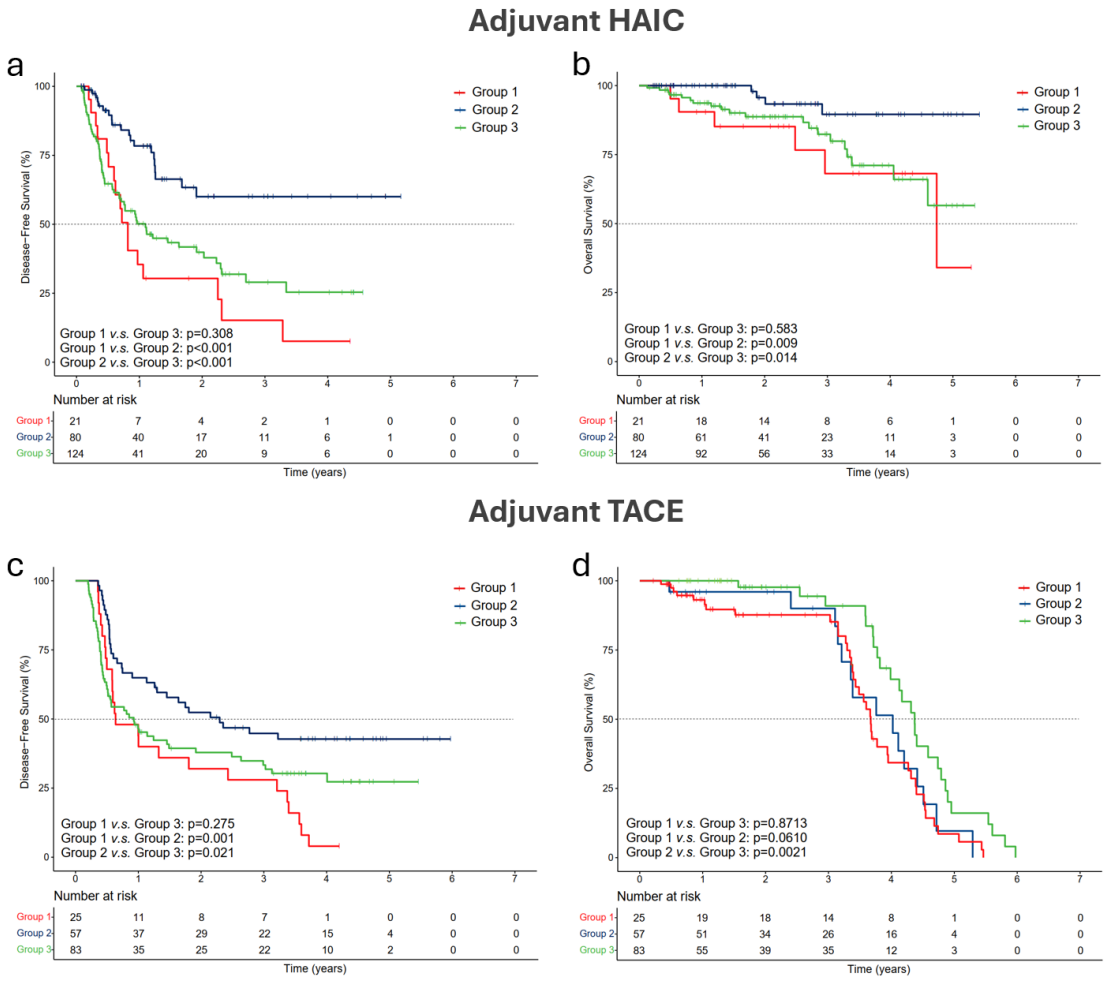
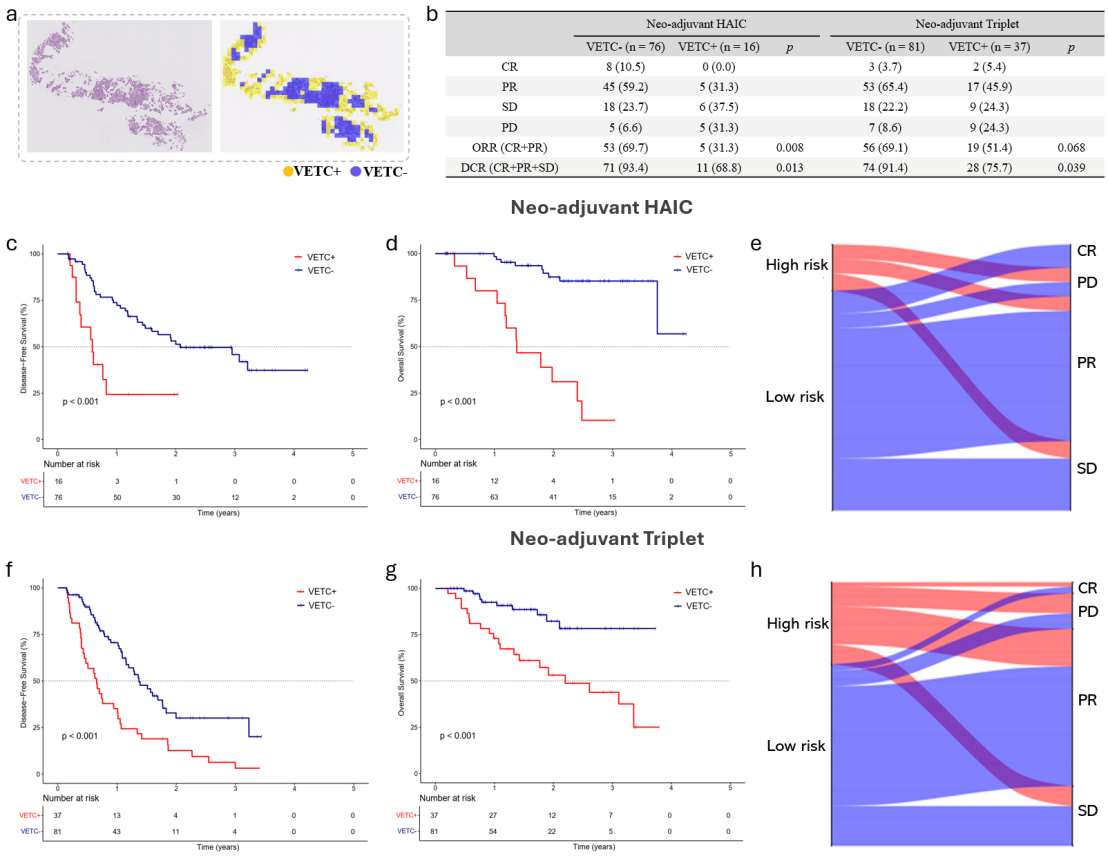


Figure 5-5 Evaluation of the efficacy of adjuvant HAIC **a** and **b** and adjuvant TACE **c** and **d** treatments based on the VETC+ distribution predicted by VETC Net. Group 1: VETC+ Patients receiving the corresponding treatment; Group 2: VETC- patients receiving the corresponding treatment; Group 3: Patients not receiving the corresponding treatment.

Next, we evaluate the relationship between adjuvant HAIC treatment and VETC+ in 225 patients with HCC, of which 101 received adjuvant HAIC treatment and 124 were in the control group. As shown in Figure 5-5 **a** and **b**, adjuvant HAIC demonstrates greater efficacy in improving DFS and overall survival OS for VETC- patients. Specifically, a statistically significant difference in both DFS ($p < 0.001$) and OS ($p =$

0.009) was observed between treated VETC- patients and those in the control group, with Group 2 (VETC-) patients having better outcomes. In contrast, the treatment's efficacy for VETC+ patients was less pronounced. No significant difference in DFS ($p = 0.308$) and OS ($p = 0.583$) was observed when comparing VETC+ patients who received HAIC to those in the control group. This suggests that while adjuvant HAIC may provide significant survival benefits for VETC- patients, its effect on VETC+ patients remains inconclusive, with p-values indicating no clear advantage in comparison to untreated patients.

Similarly, we assessed the efficacy of adjuvant TACE in HCC patients stratified by VETC status. A total of 102 patients received adjuvant TACE following curative resection, while 83 patients served as untreated controls. As illustrated in Figures 4c and 4d, adjuvant TACE significantly improved both DFS and OS in VETC− patients. Specifically, VETC- patients in the TACE group demonstrated notably lower recurrence and mortality rates compared to their untreated counterparts, with p-values < 0.05 indicating statistically significant differences. In contrast, the benefit of adjuvant TACE was not observed in VETC+ patients. There were no significant differences in either DFS or OS between VETC+ patients who received adjuvant TACE and those who did not, with p-values exceeding 0.2. These findings suggest that, similar to adjuvant HAIC, the efficacy of adjuvant TACE may be limited in the VETC+ subgroup. Collectively, these results underscore the importance of VETC stratification when selecting postoperative adjuvant therapies, with VETC- patients appearing to derive more substantial benefit from adjuvant TACE.



	Neo-adjuvant HAIC			Neo-adjuvant Triplet		
	VETC- (n = 76)	VETC+ (n = 16)	*p*	VETC- (n = 81)	VETC+ (n = 37)	*p*
CR	8 (10.5)	0 (0.0)		3 (3.7)	2 (5.4)	
PR	45 (59.2)	5 (31.3)		53 (65.4)	17 (45.9)	
SD	18 (23.7)	6 (37.5)		18 (22.2)	9 (24.3)	
PD	5 (6.6)	5 (31.3)		7 (8.6)	9 (24.3)	
ORR (CR+PR)	53 (69.7)	5 (31.3)	0.008	56 (69.1)	19 (51.4)	0.068
DCR (CR+PR+SD)	71 (93.4)	11 (68.8)	0.013	74 (91.4)	28 (75.7)	0.039

Figure 5-6 **a** Segmentation map obtained by VETC Net on biopsy images. **b** Contingency table for the treatment outcomes of Neo-adjuvant HAIC and Neo-adjuvant Triplet in VETC+ and VETC- patients. KM curves for DFS **c** and OS **d** of VETC+ and VETC- patients treated with Neo-adjuvant HAIC and the corresponding waterfall plot **e**. KM curves for DFS **f** and OS **g** of VETC+ and VETC- patients treated with Neo-adjuvant Triplet and the corresponding waterfall plot **h**.

Finally, we assessed the therapeutic effects of Neo-adjuvant HAIC and Neo-adjuvant Triplet on VETC+ and VETC- patients using needle biopsy samples from HCC patients. As shown in Figure 5-6 **a**, VETC Net can also effectively identify and localize VETC+ regions in needle biopsy images, confirming its reliability and generalizability in distinguishing between VETC+ and VETC- areas. Subsequently, we

evaluate the treatment differences between VETC+ and VETC- patients for both treatment modalities. As indicated in Figure 5-6 **b** and **e**, from the perspective of overall response rate (ORR) and disease control rate (DCR), VETC- patients treated with Neo-adjuvant HAIC exhibit significantly higher ORR and DCR than VETC+ patients ($p = 0.008$ and $p = 0.013$, respectively). Specifically, VETC- patients achieve an ORR of 96.7% and a DCR of 93.4%, suggesting a favorable response to the treatment. In contrast, VETC+ patients, although still showing some therapeutic effect, exhibit substantially lower response rates, with ORR at 31.3% and DCR at 68.8%, indicating that the treatment is less effective for this subgroup. Similarly, in the Neo-adjuvant Triplet group, as shown in Figure 5-6 **b** and **h**, the DCR for VETC- patients is significantly higher than that for VETC+ patients ($p = 0.039$). Although VETC+ patients also benefit from the Triplet treatment, the therapeutic effects remain limited compared to VETC− patients.

For survival analysis, Figure 5-6 **c** and **d** (Neo-adjuvant HAIC) and Figure 5-6 **f** and **g** (Neo-adjuvant Triplet) demonstrate significant differences in treatment efficacy between VETC+ and VETC- patients ($p < 0.001$). Both treatment modalities show better outcomes in VETC- patients, with longer DFS and OS compared to VETC+ patients, who experience relatively poorer treatment responses.

In conclusion, VETC- patients show superior treatment responses to both Neo-adjuvant HAIC and Neo-adjuvant Triplet, while VETC+ patients exhibit more limited therapeutic benefits, reinforcing the need for personalized treatment strategies based on VETC status.

5.5 Discussion

In this study, we proposed and validated VETC Net, a DL model designed to automatically identify and localize a novel vascular pattern named VETC in WSI of HCC patients. Our results underscore the utility of VETC Net not only as a diagnostic tool but also as an instrument for prognostic analysis and treatment evaluation. By leveraging VETC Net, we demonstrated its robust performance in stratifying patients based on the proportion of VETC+ regions in their tumors, offering new insights into the clinical significance of VETC in HCC.

The clinical importance of VETC in HCC prognosis has been established by several studies. Our work builds on this foundation by providing a quantitative measure of VETC+ within tumor regions, offering a more precise means of assessing patient risk. Using VETC Net, we identified a cutoff value of 45.2% for the proportion of VETC+ regions within the tumor area, which effectively predicted poor outcomes in terms of both DFS and OS across a diverse cohort of patients from multiple medical institutions. These results enhance and complement previous findings by Wang et al. and Lu et al.[76][77], highlighting the substantial prognostic impact of VETC on HCC outcomes. Specifically, the validation of VETC Net across an independent external cohort further solidifies its potential to generalize across various patient populations. Notably, patients with a higher proportion of VETC+ regions (defined as >45.2%) exhibited significantly worse prognosis in terms of both DFS and OS, suggesting that VETC+ serves as a reliable marker for high-risk patients. These results reinforce the notion that the vascular microenvironment of the tumor, particularly the presence of

VETC plays a pivotal role in influencing disease progression, prognosis analysis and therapeutic response.

One of the major strengths of this study lies in its evaluation of VETC+ in relation to various treatment modalities, including adjuvant therapies such as adjuvant TACE and adjuvant HAIC, as well as neo-adjuvant treatments like as neo-adjuvant Triplet and as neo-adjuvant HAIC. Our analysis revealed significant differences in treatment efficacy based on VETC status. For both adjuvant TACE and HAIC, VETC- patients demonstrated significantly better treatment outcomes compared to VETC+ patients, with improved DFS and OS. In contrast, the efficacy of these treatments in VETC+ patients was less pronounced, with no significant differences observed between treated and untreated groups. The results of the neo-adjuvant therapies further underscore this differential response. Both Neo-adjuvant HAIC and Neo-adjuvant Triplet therapies were found to represent superior responses in VETC- patients, with significantly higher ORR and DCR. VETC+ patients, on the other hand, exhibited considerably lower response rates, which is consistent with the hypothesis that the vascular characteristics of the tumor, as influenced by the presence of VETC, may hinder the efficacy of treatment. The observed lower efficacy in VETC+ patients may also reflect the complex interplay between tumor vasculature and therapeutic agents, which could impede drug delivery and limit the effectiveness of treatment.

Our results emphasize the importance of incorporating VETC status into personalized treatment strategies for HCC patients. Specifically, VETC Net provides a automatic and reproducible method for evaluating VETC, which can inform clinical

decision-making and guide the selection of the most appropriate treatment for individual patients. Moreover, the integration of VETC Net into clinical practice has the potential to streamline the diagnostic process. As we showed, the model's reliance on routine H&E-stained WSIs, which are more accessible and cost-effective compared to CD34 immunostaining. This approach could significantly reduce the time and cost associated with diagnosing VETC+ patients and by extension, improve patient prognosis through earlier and more precise risk stratification.

While this study demonstrates the promising potential of VETC Net, there are several limitations that warrant consideration. First, while we validated the model across multiple centers, the external cohort still comprises a relatively small sample size. Further validation across larger, more diverse cohorts will be essential to fully assess the generalizability of VETC Net.

On the other hand, future studies could also focus on integrating VETC Net with other biomarkers, such as those related to immune response or genetic alterations, to provide a more comprehensive approach to patient stratification. Moreover, investigating the molecular mechanisms underlying the differential responses of VETC+ and VETC- patients to treatment could yield valuable insights into how vascular patterns influence tumor biology and therapy resistance.

5.6 Conclusion

This study presented the VETC Net system, which accurately identifies and localizes VETC+ regions in WSIs of HCC patients, providing effective prognostic evaluation. Through validation across both internal and external datasets, VETC Net

demonstrated strong generalizability and can predict patients' recurrence risk and survival based on the distribution of VETC+. The results indicate that VETC- patients exhibit better outcomes following treatments: adjuvant TACE, adjuvant HAIC, neo-adjuvant HAIC and neo-adjuvant triplet, while VETC+ patients show limited therapeutic benefits. Therefore, VETC Net can aid in developing personalized treatment strategies and has significant potential for improving both the diagnosis and treatment outcomes for HCC patients.

Chapter 6. Summary

This thesis presents the development and clinical application of an efficient DL - based framework for the analysis of WSI. The study begins with an exploration of HCC prognosis analysis, providing a comprehensive overview of the WSI analysis pipeline from preprocessing and feature extraction to downstream tasks. Subsequently, we introduce a novel approach to enhance WSI encoding efficiency by generating prototype features with deep fuzzy clustering, thereby improving the predictive precise of subsequent tasks. Finally, we discuss a novel clinical application scenario: the automated localization and distribution prediction of VETC. Based on this, a new biomarker for HCC patient prognosis is developed. Additionally, the evaluation of the efficacy of different treatment methods is presented. The results underscore the importance of advanced DL methods in improving personalized treatment strategies and clinical decision-making.

In Chapter 2, we first provide a comprehensive review of the current DL-based frameworks for WSI analysis, encompassing key stages such as preprocessing, feature extraction, and multimodal feature fusion. We systematically review the fundamental methodologies employed in these processes. Furthermore, we specifically highlight the analysis strategies for WSIs based on MIL, illustrating the complete workflow of training and inference through a classification task as an example. Finally, we introduced the general methodologies of survival analysis and explored their applications within the context of WSI analysis.

The third chapter introduced a heuristic WSI analysis paradigm for survival

analysis in HCC patients. This framework assisted pathologists in clinical diagnosis in two key stages: Firstly, it helped the localization of regions with high potential for recurrence. By leveraging an attention-based mechanism, we identify tissue regions associated with high recurrence risk in HCC without any prior knowledge. The localization results were supported by clinical experts and existing literature. Secondly, in contrast to traditional single-perspective risk scoring, we developed a multi-perspective prognostic risk assessment system using DL methods. By integrating this system with existing clinical staging, our proposed risk assessment framework can identify potential high-risk patients within the low-risk group, thereby assisting clinicians in devising more precise treatment strategies.

Chapter 4 first analyzed the limitations of current attention-based mechanisms in WSI analysis: Methods relying on global attention mechanisms tend to result in feature homogeneity, which neglects clinically significant tissue characteristics and leads to a lack of model interpretability. On the other hand, for local attention-based methods, although they focus more on tissue regions, these approaches inevitably increase computational cost and introduce noise, which can be catastrophic for ultra-high-resolution WSIs. To address these challenges, we integrate Fuzzy C-means into the Transformer's cross-attention framework to develop a novel deep fuzzy clustering approach named FuzzyMIL. Through a soft clustering approach, our method considers tissue features as comprehensively as possible while gradually converging them to representative disentangled features. Validation results on three publicly available datasets demonstrate that our approach not only significantly reduces the computational

complexity of downstream tasks but also achieves competitive performance.

Chapter five applied DL- based WSI analysis methods to address a novel clinical question: the automated localization and distribution prediction of VETC, as well as its relationship with prognosis and the evaluation of its response to different treatment modalities. We first constructed a patch-level dataset to train an automated VETC+ region identification and localization network named VETC Net. Subsequently, we can identify an optimal cutoff value on the internal WSI-level dataset to distinguish between high and low-risk groups based on the distribution of VETC+ regions in WSI-level dataset. We then performed validation on a multi-center external dataset to assess the effectiveness of our approach. To evaluate the response of VETC status to different treatment modalities, we first compared the efficacy of adjuvant HAIC and adjuvant TACE in patient cohorts. Similarly, we also conducted efficacy assessments on the needle biopsy images of patients who underwent Neo-adjuvant HAIC and Neo-adjuvant TACE. The efficacy evaluations across different cohorts revealed that VETC+ patients exhibited a limited response to these treatment modalities, whereas VETC- patients showed more favorable outcomes. This contrast highlights the potential significance of VETC status as a predictive factor for treatment efficacy.

This thesis has some limitations that require further investigation and development. Firstly, from a technical perspective, considering the growing importance of multimodal information in clinical diagnosis, prognosis analysis, and other downstream tasks, although the thesis has demonstrated competitive results in WSI image analysis, incorporating multimodal information could further enhance the model's

interpretability and improve the accuracy. On the other hand, given that LLMs can provide valuable additional prior knowledge, particularly in situations involving sparsity or rare disease diagnoses, designing an LLM-guided approach to complement WSI information represents an efficient strategy. From the clinical perspective, this thesis primarily focuses on HCC as a clinical issue. However, the framework has the potential for transferability across multiple cancer types. Therefore, in future work, we plan to evaluate the model's generalizability across pan-cancer, thus extending its clinical applicability.

In conclusion, this thesis presented a comprehensive DL-based framework for WSI analysis, demonstrating its potential in improving clinical decision-making, prognosis prediction, and treatment evaluation, particularly in the context of HCC. In improving the efficiency of analysis, we incorporated Fuzzy C-means with DL framework, provided a WSI encoding method with lower computational complexity. The clinical applications including the localization of VETC regions and the evaluation of treatment responses highlight the importance of this framework in personalized medicine.

References

[1]. Zhu X, Yao J, Zhu F, et al. Wsisa: Making survival prediction from whole slide histopathological images[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7234-7242.

[2]. Sobin L H. TNM: evolution and relation to other prognostic factors[C]//Seminars in surgical oncology. Hoboken: Wiley Subscription Services, Inc., A Wiley Company, 2003, 21(1): 3-7.

[3]. Steck H, Krishnapuram B, Dehing-Oberije C, et al. On ranking in survival analysis: Bounds on the concordance index[J]. Advances in neural information processing systems, 2007, 20.

[4]. Yan X, Wang W, Xiao M, et al. Survival prediction across diverse cancer types using neural networks[C]//Proceedings of the 2024 7th International Conference on Machine Vision and Applications. 2024: 134-138.

[5]. Fan L, Sowmya A, Meijering E, et al. Cancer survival prediction from whole slide images with self-supervised learning and slide consistency[J]. IEEE Transactions on Medical Imaging, 2022, 42(5): 1401-1412.

[6]. Di D, Li S, Zhang J, et al. Ranking-based survival prediction on histopathological whole-slide images[C]//International conference on medical image computing and computer-assisted intervention. Cham: Springer International Publishing, 2020: 428-438.

[7]. Wang K, Xiang Y, Yan J, et al. A deep learning model with incorporation of microvascular invasion area as a factor in predicting prognosis of hepatocellular

carcinoma after R0 hepatectomy[J]. Hepatology International, 2022, 16(5): 1188-1198.

[8]. Nateghi R, Danyali H, Helfroush M S. A deep learning approach for mitosis detection: application in tumor proliferation prediction from whole slide images[J]. Artificial intelligence in medicine, 2021, 114: 102048.

[9]. Targ S, Almeida D, Lyman K. Resnet in resnet: Generalizing residual architectures[J]. arXiv preprint arXiv:1603.08029, 2016.

[10]. Lu M Y, Williamson D F K, Chen T Y, et al. Data-efficient and weakly supervised computational pathology on whole-slide images[J]. Nature biomedical engineering, 2021, 5(6): 555-570.

[11]. Sharma Y, Sxhrivastava A, Ehsan L, et al. Cluster-to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification[C]//Medical imaging with deep learning. PMLR, 2021: 682-698.

[12]. Li X, Li C, Rahaman M M, et al. A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches[J]. Artificial Intelligence Review, 2022, 55(6): 4809-4878.

[13]. Xiong Y, Zeng Z, Chakraborty R, et al. Nyströmformer: A nyström-based algorithm for approximating self-attention[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(16): 14138-14148.

[14]. Tang L, Diao S, Li C, et al. Global contextual representation via graph-transformer fusion for hepatocellular carcinoma prognosis in whole-slide

images[J]. Computerized Medical Imaging and Graphics, 2024, 115: 102378.

[15]. Chen C F, Panda R, Fan Q. Regionvit: Regional-to-local attention for vision transformers[J]. arXiv preprint arXiv:2106.02689, 2021.

[16]. Liang J, Zhang W, Yang J, et al. Deep learning supported discovery of biomarkers for clinical prognosis of liver cancer[J]. Nature Machine Intelligence, 2023, 5(4): 408-420.

[17]. Liang M, Chen Q, Li B, et al. Interpretable classification of pathology whole-slide images using attention based context-aware graph convolutional neural network[J]. Computer methods and programs in biomedicine, 2023, 229: 107268.

[18]. Shi J, Li C, Gong T, et al. CoD-MIL: Chain-of-Diagnosis Prompting Multiple Instance Learning for Whole Slide Image Classification[J]. IEEE Transactions on Medical Imaging, 2024.

[19]. Nguyen A T, Nguyen D M H, Diep N T, et al. MGPATH: Vision-Language Model with Multi-Granular Prompt Learning for Few-Shot WSI Classification[J]. arXiv preprint arXiv:2502.07409, 2025.

[20]. Bejnordi B E, Litjens G, Timofeeva N, et al. Stain specific standardization of whole-slide histopathological images[J]. IEEE transactions on medical imaging, 2015, 35(2): 404-415.

[21]. Zheng Y, Jiang Z, Zhang H, et al. Adaptive color deconvolution for histological WSI normalization[J]. Computer methods and programs in biomedicine, 2019, 170: 107-120.

[22]. Franchet C, Schwob R, Bataillon G, et al. Bias reduction using combined stain

normalization and augmentation for AI-based classification of histological images[J]. Computers in Biology and Medicine, 2024, 171: 108130.

[23]. Stirling D R, Swain-Bowden M J, Lucas A M, et al. CellProfiler 4: improvements in speed, utility and usability[J]. BMC bioinformatics, 2021, 22: 1-11.

[24]. Bankhead P, Loughrey M B, Fernández J A, et al. QuPath: Open source software for digital pathology image analysis[J]. Scientific reports, 2017, 7(1): 1-7.

[25]. Schüssele D S, Haller P K, Haas M L, et al. Autophagy profiling in single cells with open source CellProfiler-based image analysis[J]. Autophagy, 2023, 19(1): 338-351.

[26]. Courtney J M, Morris G P, Cleary E M, et al. An automated approach to improve the quantification of pericytes and microglia in whole mouse brain sections[J]. Eneuro, 2021, 8(6).

[27]. Wang K, Xiang Y, Yan J, et al. A deep learning model with incorporation of microvascular invasion area as a factor in predicting prognosis of hepatocellular carcinoma after R0 hepatectomy[J]. Hepatology International, 2022, 16(5): 1188-1198.

[28]. Yu E, Monaco J P, Tomaszewski J, et al. Detection of prostate cancer on histopathology using color fractals and probabilistic pairwise Markov models[C]//2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2011: 3427-3430.

[29]. Partio M, Cramariuc B, Gabbouj M, et al. Rock texture retrieval using gray level co-occurrence matrix[C]//Proc. of 5th Nordic Signal Processing Symposium. 2002, 75(1): 511-524.

[30]. Khan A, Han S, Ilyas N, et al. CervixFormer: A Multi-scale swin transformer-Based cervical pap-Smear WSI classification framework[J]. Computer Methods and Programs in Biomedicine, 2023, 240: 107718.

[31]. Li J, Chen Y, Chu H, et al. Dynamic graph representation with knowledge-aware attention for histopathology whole slide image analysis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 11323-11332.

[32]. Li J, Zhang Y, Shu W, et al. M4: Multi-proxy multi-gate mixture of experts network for Multiple instance learning in histopathology image analysis[J]. Medical Image Analysis, 2025: 103561.

[33]. Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//International conference on machine learning. PmLR, 2021: 8748-8763.

[34]. Lu M Y, Chen B, Williamson D F K, et al. A visual-language foundation model for computational pathology[J]. Nature Medicine, 2024, 30(3): 863-874.

[35]. Gao Z, Lu Z, Wang J, et al. A convolutional neural network and graph convolutional network based framework for classification of breast histopathological images[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(7): 3163-3173.

[36]. Wu W, Gao C, DiPalma J, et al. Improving representation learning for histopathologic images with cluster constraints[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 21404-21414.

[37]. Wang P, Zhang J, Li Y, et al. Histopathology image classification based on semantic correlation clustering domain adaptation[J]. Artificial Intelligence in Medicine, 2025: 103110.

[38]. Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning[C]//International conference on machine learning. PMLR, 2018: 2127-2136.

[39]. Fremond S, Andani S, Wolf J B, et al. Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: a combined analysis of the PORTEC randomised trials and clinical cohorts[J]. The Lancet Digital Health, 2023, 5(2): e71-e82.

[40]. Li L, Pan H, Liang Y, et al. PMFN-SSL: Self-supervised learning-based progressive multimodal fusion network for cancer diagnosis and prognosis[J]. Knowledge-Based Systems, 2024, 289: 111502.

[41]. Zhang Z, Zhao Y, Duan J, et al. Pathology-genomic fusion via biologically informed cross-modality graph learning for survival analysis[J]. arXiv preprint arXiv:2404.08023, 2024.

[42]. Huang Z, Bianchi F, Yuksekgonul M, et al. A visual–language foundation model for pathology image analysis using medical twitter[J]. Nature medicine,

2023, 29(9): 2307-2316.

[43]. Song B, Leroy A, Yang K, et al. Deep learning informed multimodal fusion of radiology and pathology to predict outcomes in HPV-associated oropharyngeal squamous cell carcinoma[J]. EBioMedicine, 2025, 114.

[44]. Yeh K, Jabal M S, Gupta V, et al. Transformer-Based Self-Supervised Learning for Histopathological Classification of Ischemic Stroke Clot Origin[J]. arXiv preprint arXiv:2405.00908, 2024.

[45]. Marini N, Marchesin S, Wodzinski M, et al. Multimodal representations of biomedical knowledge from limited training whole slide images and reports using deep learning[J]. Medical Image Analysis, 2024, 97: 103303.

[46]. Qu L, Liu S, Liu X, et al. Towards label-efficient automatic diagnosis and analysis: a comprehensive survey of advanced deep learning-based weakly-supervised, semi-supervised and self-supervised techniques in histopathological image analysis[J]. Physics in Medicine & Biology, 2022, 67(20): 20TR01.

[47]. Li B, Li Y, Eliceiri K W. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 14318-14328.

[48]. Shao Z, Bian H, Chen Y, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification[J]. Advances in neural information processing systems, 2021, 34: 2136-2147.

[49]. Fillioux L, Boyd J, Vakalopoulou M, et al. Structured state space models for

multiple instance learning in digital pathology[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 594-604.

[50]. Jin C, Luo L, Lin H, et al. HMIL: Hierarchical Multi-Instance Learning for Fine-Grained Whole Slide Image Classification[J]. IEEE Transactions on Medical Imaging, 2024.

[51]. Hou W, Yu L, Lin C, et al. H^ 2-MIL: exploring hierarchical representation with heterogeneous multiple instance learning for whole slide image analysis[C]//Proceedings of the AAAI conference on artificial intelligence. 2022, 36(1): 933-941.

[52]. Bontempo G, Bolelli F, Porrello A, et al. A graph-based multi-scale approach with knowledge distillation for wsi classification[J]. IEEE Transactions on Medical Imaging, 2023, 43(4): 1412-1421.

[53]. Katzman J L, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network[J]. BMC medical research methodology, 2018, 18: 1-12.

[54]. Therneau T M, Grambsch P M, Therneau T M, et al. The cox model[M]. Springer New York, 2000.

[55]. Tibshirani R. The lasso method for variable selection in the Cox model[J]. Statistics in medicine, 1997, 16(4): 385-395.

[56]. Yang Y, Zou H. A cocktail algorithm for solving the elastic net penalized Cox's regression in high dimensions[J]. Statistics and its Interface, 2013, 6(2): 167-173.

[57]. Faraggi D, Simon R. A neural network model for survival data[J]. Statistics in medicine, 1995, 14(1): 73-82.

[58]. Ishwaran H, Gerds T A, Kogalur U B, et al. Random survival forests for competing risks[J]. Biostatistics, 2014, 15(4): 757-773.

[59]. Veta M, Heng Y J, Stathonikos N, et al. Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge[J]. Medical image analysis, 2019, 54: 111-121.

[60]. Geessink O G F, Baidoshvili A, Klaase J M, et al. Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer[J]. Cellular oncology, 2019, 42: 331-341.

[61]. Li H, Zhang Y, Zhu C, et al. Long-MIL: scaling long contextual multiple instance learning for histopathology whole slide image analysis[J]. arXiv preprint arXiv:2311.12885, 2023.

[62]. Shao W, Shi Y Y, Zhang D, et al. Tumor micro-environment interactions guided graph learning for survival analysis of human cancers from whole-slide pathological images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 11694-11703.

[63]. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115: 211-252.

[64]. Yao W, He J, Yang Y, et al. The prognostic value of tumor-infiltrating lymphocytes in hepatocellular carcinoma: a systematic review and meta-analysis[J]. Scientific reports, 2017, 7(1): 7525.

[65]. Ha S Y, Choi S, Park S, et al. Prognostic effect of preoperative neutrophil-lymphocyte ratio is related with tumor necrosis and tumor-infiltrating lymphocytes in hepatocellular carcinoma[J]. Virchows Archiv, 2020, 477: 807-816.

[66]. Kuo F Y, Eng H L, Li W F, et al. Tumor necrosis is an indicator of poor

prognosis among hepatoma patients undergoing resection[J]. Journal of Surgical Research, 2023, 283: 1091-1099.

[67]. Zhu Q, Dai H, Qiu F, et al. Heterogeneity of computational pathomic signature predicts drug resistance and intra-tumor heterogeneity of ovarian cancer[J]. Translational Oncology, 2024, 40: 101855.

[68]. Chen R J, Lu M Y, Williamson D F K, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning[J]. Cancer cell, 2022, 40(8): 865-878. e6.

[69]. Shaban M, Khurram S A, Fraz M M, et al. A novel digital score for abundance of tumour infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma[J]. Scientific reports, 2019, 9(1): 13341.

[70]. Yao J, Zhu X, Jonnagaddala J, et al. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks[J]. Medical image analysis, 2020, 65: 101789.

[71]. Yu Q, Wang H, Qiao S, et al. k-means Mask Transformer[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 288-307.

[72]. Tang W, Zhou F, Huang S, et al. Feature re-embedding: Towards foundation model-level performance in computational pathology[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 11343-11352.

[73]. Li B, Li Y, Eliceiri K W. Dual-stream multiple instance learning network for

whole slide image classification with self-supervised contrastive learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 14318-14328.

[74]. Zhang H, Meng Y, Zhao Y, et al. Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 18802-18812.

[75]. Ferenci P, Fried M, Labrecque D, et al. Hepatocellular carcinoma (HCC): a global perspective[J]. Journal of clinical gastroenterology, 2010, 44(4): 239-245.

[76]. Fang J H, Zhou H C, Zhang C, et al. A novel vascular pattern promotes metastasis of hepatocellular carcinoma in an epithelial–mesenchymal transition–independent manner[J]. Hepatology, 2015, 62(2): 452-465.

[77]. Fang J H, Xu L, Shang L R, et al. Vessels that encapsulate tumor clusters (VETC) pattern is a predictor of sorafenib benefit in patients with hepatocellular carcinoma[J]. Hepatology, 2019, 70(3): 824-839.

[78]. Lu L, Wei W, Huang C, et al. A new horizon in risk stratification of hepatocellular carcinoma by integrating vessels that encapsulate tumor clusters and microvascular invasion[J]. Hepatology international, 2021, 15: 651-662.