摘要

组织病理学全切片图像（WSIs）为计算病理学（CPATH）中的癌症预后评估提供了重要工具。尽管现有生存分析（SA）方法取得了显著进展，但它们通常局限于采用高表达能力网络架构和粗粒度的患者级标签，从千兆像素级 WSIs 中学习视觉预后表征。在当前训练数据稀缺且 CPATH 标准框架为多示例学习（MIL）的情况下，这种学习范式面临严重的性能瓶颈。为此，本文首次提出了一种基于视觉-语言的生存分析（VLSA）新范式。具体而言：（1）VLSA 由病理学视觉-语言基础模型驱动，不再依赖高容量网络，展现出**数据高效性**优势；（2）在视觉端，VLSA 通过编码文本预后先验知识，将其作为**辅助信号**指导实例级视觉预后特征的聚合，从而弥补 MIL 中弱监督的不足。此外，针对 SA 特性，我们提出：i）**序数生存提示学习**，将连续生存标签转化为文本提示；ii）**序数风险函数**作为预测目标，使 SA 兼容基于视觉-语言的预测。值得注意的是，VLSA 的预测结果可通过我们基于 Shapley 值的方法直观解释。在五个数据集上的大量实验验证了该方案的有效性。VLSA 为 CPATH 中的 SA 开辟了新路径，为弱监督 MIL 提供了一种从千兆像素 WSIs 中学习有价值预后线索的有效手段。源代码发布于 https://github.com/liupeil01/VLSA。

什么是沙普利值？

沙普利值是合作博弈论中的核心概念，由诺贝尔经济学奖得主 Lloyd Shapley 于 1953 年提出，用于公平分配团队合作中每个成员的贡献。在机器学习中，它被广泛用于解释模型预测（如黑盒模型的特征重要性分析）。

1 引言

组织病理学全切片图像（Whole-Slide Image, WSI）在癌症诊断与治疗中具有关键作用（Zarella et al., 2018）。它通常涵盖从细胞形态、肿瘤微环境到组织表型的丰富且全面的微观信息（Pati et al., 2022; Chen et al., 2022）。由于这些信息能直接反映肿瘤进展，数字 WSI 常被用于计算病理学（Computational Pathology, CPATH）中，以评估癌症患者的生存预后（Kather et al., 2019; Song et al., 2023; Jaume et al., 2024）。精准的预后评估对改善患者管理和疾病结局至关重要（Skrede et al., 2020）。

然而，基于 WSI 数据的生存分析（Survival Analysis, SA）始终面临两大关键挑战。尽管现有方法在应对这些挑战方面取得了显著进展，但由于当前 SA 范式的固有局限性，

它们仍存在性能瓶颈问题：

（1）训练数据稀缺。受现实因素制约（如长期患者随访困难、隐私保护要求等），可用于 SA 的 WSI 数据规模通常有限，样本量普遍仅为千例级别。现有 SA 模型大多忽视这一现实，转而寻求网络层面的解决方案（例如采用图神经网络 GNN 等高

表达能力模型）以提升性能。然而，当面对当前有限规模的 WSI 数据时，这种方法更容易导致深度学习模型出现过拟合现象（Srivastava 等，2014），从而影响预测性能的最优化。

（2）弱监督下的十亿像素级图像学习。数字 WSI 具有极高的分辨率（例如 40,000×40,000 像素），因此每张图像通常会被处理成多个实例组成的"包"进行训练。在仅有 WSI 级别标签的情况下，包级别表征是通过事实上的弱监督多实例学习（MIL）框架（Ilee 等，2018；Liu 等）获得的。该框架首先：i）从单个实例中学习任务特定的嵌入表示；ii）然后将大量实例（通常达 10,000 个）聚合为单一向量。当前 SA 方案虽遵循此方式，但整个学习过程完全依赖于患者级别的标签。我们认为这种范式可能导致表征学习效率低下，因为 SA 模型不仅需要：i）在细粒度实例层面学习预后嵌入；ii）还需从海量候选实例中筛选关键实例（Li 等，2023），却仅能获得整体性的患者级别标签作为监督信号。

为突破现有方法的局限性，本文提出了一种面向计算病理学（CPATH）的新型生存分析范式——视觉语言生存分析（Vision-Language Survival Analysis, VLSA）。具体而言，我们发现近期 CPATH 领域的视觉语言基础模型（VLMs），例如 CONCH（Lu 等，2024），为解决上述挑战提供了潜在方案：

首先，这些 VLM 通过任务无关的目标函数在大规模图像-文本对上进行了预训练。如 Lu 等（2024）和 Javed 等（2024）所强调的，它们展现出卓越的数据高效性，尤其在零样本迁移任务中表现突出。这一特性有望缓解训练数据稀缺的挑战。

其次，视觉语言对比预训练将图像与文本在潜在嵌入空间中对齐（Radford 等，2021），使得语言能够作为视觉任务的"提示"。这表明，借助先验知识，语言很可能为提升学习效率提供辅助信号。这种额外信号对于改善多实例学习（MIL）中的弱监督问题尤为重要。尽管具备这些显著优势，基于 VLM 的生存分析研究仍属空白。我们认为主要原因有二：其一，适用于 CPATH 的强效 VLM 近期才得以开发；其二，与分类任务不同，生存分析如何适配 VLM 仍存在技术鸿沟。

基于上述洞见，本文首次提出了面向计算病理学（CPATH）的视觉语言生存分析（VLSA）框架。与现有视觉语言方案不同，VLSA 包含以下四项核心设计：

（1）视觉端：通过语言编码的预后先验知识指导多实例聚合，生成多层次视觉表征。

（2）语言端：针对生存风险的内在序数特性，提出序数生存提示学习，将连续生存时间标签编码为文本提示。

（3）预测与优化：为使生存分析兼容视觉语言预测范式，采用发病率函数作为预测目标，并引入序数归纳偏置进行优化正则化。

（4）预测解释：基于博弈论经典沙普利值（Shapley values），可从直观的语言描述视

角解释个体预后风险。

在五个数据集上的大量实验验证了本方案的有效性。对比实验与分析表明，VLSA 为 CPATH 中的生存分析开辟了新路径，为弱监督多实例学习（MIL）提供了从十亿像素级 WSI 中挖掘关键预后线索的有效方法。本文主要贡献可概括为：

- **提出首个可解释的视觉语言生存分析框架**：据我们所知，这是 CPATH 领域首个基于视觉语言的生存分析研究。
- **创新性引入语言增强的预后先验**：针对生存分析特性，设计两个序数归纳偏置项——序数生存提示与序数发病率函数，显著提升模型性能。
- **建立严谨的评估体系**：通过多指标判别力与校准度评估，实证表明 VLSA 能以更低计算成本达到当前最优性能。

3 方法

本节介绍我们提出的 VLSA 框架——一种基于视觉语言的计算病理学（CPATH）生存分析方法（图 1）。首先阐述其三个核心组成部分：i) 融合语言编码预后先验的 WSI 表征学习（图 1(a)，第 3.1 节）；ii) 序数生存提示学习（图 1(b)，第 3.2 节）；iii) 序数发病率函数预测（图 1(c)，第 3.3 节）。随后在第 3.4 节给出整体训练目标，并在第 3.5 节说明 VLSA 生存预测结果的解释方法。相关基础知识详见附录 A。

视觉端设计：

采用 WSI 多实例学习架构，通过语言编码的预后先验

$$f_m = \sum_{k=1}^{k} x_k \cdot \frac{exp(\alpha cos(p_m, x_k))}{\sum_i exp(\alpha cos(p_m, x_i))}$$

其中

x_k由图像编码器编码

p_m是m个预后先验的文本特征**（借助 GPT-4）**

语言端设计依据

生存序列分析

上下文提示：遵循 CoOp 方法

类别提示：*维护B个可学习的基础类别提示*

(1) 序数归纳偏置：根据常识，随着生存时间缩短，对应的死亡风险将逐步升高。这表明生存类别间存在序数关系，在设计生存分析提示时应考虑此类偏置。

(2) **类别提示插值**：一方面，多数文本编码器对数字序数性不敏感，直接采用数值类别作为文本提示可能不适用；另一方面，当类别数量较大时（如 C=10），人工设计 C 个不同级别的细粒度预后风险描述将极为困难。针对这些障碍，**采用少量基础提示（λ≤4）**，并基于插值策略来保持生存提示间的序数关系。

在生存分析任务中，时间标签具有天然序数性（如生存时间 1 年 ＜2 年 ＜3 年）。传统视觉语言模型（如 CLIP）的文本编码器对数值序数不敏感，直接使用数字作为类别提示（如"1 年"、"2 年"）会导致语义模糊。为此，本方法提出：

- 基础提示初始化：用自然语言描述预后风险等级（如"极差"、"中等"、"极好"）
- 序数插值生成：通过数学插值在基础提示间生成连续时间区间的文本特征

$$V_{cls}^{c} = \sum_{b=1}^{B} V_{cls}^{\lambda b} \cdot \frac{W(D_c, b)}{\sum_{i=1}^{b} W(D_c, b_i)}$$

$\mathbf{V_{cls}^{c}}$表示c类别的最终标记嵌入（类似加权后的极差，中等，良好的之和）

$W(D_c, b)$代表距离

损失函数

采用生存分析专用的联合损失：

$$L = L_{MIE} + \beta \cdot L_{EMD}$$

分别为最大似然估计损失和 Earth Mover Distance 损失

预测函数

$$\widehat{y_c} = \sum_{k=1}^{k} x_k \cdot \frac{exp(\tau \cdot cos(f_{image}, f_{text}^{c}))}{\sum_i exp(\tau \cdot cos(f_{image}, f_{text}^{i}))}$$

数据集介绍（TCGA）

TCGA（The cancer genome atlas，癌症基因组图谱）由 National Cancer Institute(NCI，美国国家癌症研究所)和 National Human Genome Research Institute（NHGRI，美国国家人类基因组研究所）于 2006 年联合启动的项目，收录了各种人类癌症（包括亚型在内的肿瘤）的临床数据，基因组变异，mRNA 表达，miRNA 表达，甲基化等数据，是癌症研究者很重要的数据来源。