

STAMP: 基于多中心组织病理图像的 STAS 诊断的多模式注意力感知多实例学习

Liangrui Pan^{*1}, Xiaoyu Li^{*1}, Guang Zhu¹, Guanting Li¹, Ruixin Wang¹, Jiadi Luo²,
Yaning Yang^{3†}, Liang Qingchun^{2‡}, Shaoliang Peng^{1§}

¹College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

²Department of Pathology, The Second Xiangya Hospital, Central South University, Changsha 410011, Hunan, China ³College of Computer Science and Electronic Engineering, Xiangtan University, Xiangtan 411000, China

Abstract

气腔扩散 (STAS) 构成了一种新型入侵模式的肺腺癌 (LUAD), 与肿瘤复发和存活率降低有关。然而, LUAD 中 STAS 的大规模诊断仍是一项劳动密集的工作, 并且由于其独特的病理特征和形态特征, 极易造成疏漏和误诊。因此, 利用深度学习模型进行 STAS 诊断成为迫切的临床需求。本研究首先在中南大学湘雅二医院和湘雅三医院, 以及 TCGA-LUAD 队列中收集了来自 STAS 患者的组织病理学图像。三位资深病理学家进行了交叉验证标注, 以构建 STAS-SXY、STAS-TXY 和 STAS-TCGA 数据集。我们提出了一种称为 STAMP 的多模式注意力感知多实例学习框架, 用于分析和诊断跨多中心组织病理图像中的 STAS 存在。具体来说, 双分支架构引导模型从不同的语义空间学习与 STAS 相关的病理特征。基于 Transformer 的实例编码和多模式注意力聚合模块可动态选择与 STAS 病理密切相关的区域, 抑制无关的噪声, 并增强全局表征的判别能力。此外, 相似性正则化约束可防止分支间的特征冗余, 从而提高整体诊断精度。大量实验证明, STAMP 在 STAS-SXY、STAS-TXY 和 STAS-TCGA 上的诊断结果具备竞争力, 其 AUC 分别达到 0.8058、0.8017 和 0.7928, 超越了临床水平。这 10 个开放基线结果为 STAS 诊断研究建立了基准, 并促进了计算病理技术的未来普遍化和临床整合。数据集特征和代码可在 <https://anonymous.4open.science/r/AAAI2026-3436> 获取。

介绍

空气腔内扩散 (STAS) 是近年来在肺癌中被认知的一种侵袭模式 (Han et al. 2021; Willner, Narula, and Moreira 2024; Ma et al. 2019; Shih and Mino-Kenudson 2020)。2013 年, Onozato 等人 (Onozato et al. 2013) 首次通过三维重建识别出 STAS, 揭示了 STAS 由散布在肺泡空间中的孤立肿瘤细胞组成, 通常与原发肿瘤隔着几个肺泡。2015 年, Kadota 等人正式描述了当前病理背景下的 STAS (Kadota et al. 2015)。随后, 世界卫生组织将 STAS 定义为肿瘤细胞如微乳头簇、实性巢或单个细胞存在于主要肿瘤边缘之外的空气腔内 (Travis

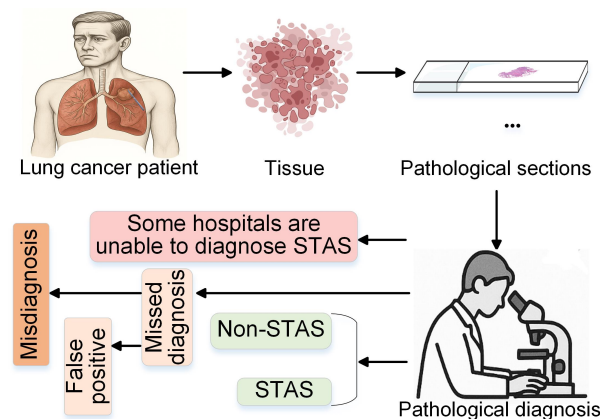


Figure 1: STAS 诊断的挑战包括某些医院无法提供诊断结果、错诊或漏诊风险高以及诊断过程劳动强度大。

et al. 2015), 并正式将其纳入肺腺癌 (LUAD) 的病理分类 (Travis et al. 2015)。在 LUAD 中经常观察到 STAS, 国内外研究报告的发生率超过 30%。它与侵袭性临床病理特征显著相关, 如胸膜侵袭、血管侵袭、肿瘤体积大和病理分期先进 (Han et al. 2021; Wang et al. 2019; Zhao et al. 2016; Toyokawa et al. 2018)。此外, STAS 被认为是 IA 期 LUAD 隐匿性淋巴结转移的强预测因子 (Vaghjani et al. 2020)。由于其与预后不良和局限性切除后高复发风险有关, STAS 的诊断已成为临床中的重要任务。

组织病理学图像被认为是诊断肺部 STAS 的金标准 (Warth et al. 2015; Zhou et al. 2022)。如图 1 所示, STAS 的诊断主要依赖于病理学家对大量组织病理学切片进行详细的显微检查。这一过程在诊断工作流程中引入了相当大的主观性。此外, 医疗资源有限的医院往往没有能力进行可靠的 STAS 诊断 (Villalba et al. 2021; Li et al. 2024b)。此外, STAS 诊断是一项劳动密集型任务, 容易出现误诊和漏诊。误报 STAS 结果的一个主要原因来源于不标准的组织采样和切片准备程序, 这可能导致肿瘤细胞从原发病灶的人为脱离 (Villalba et al. 2021; Ikeda et al. 2021)。STAS 的术中检测可以指导手术计划, 包括在全肺切除术和肺叶切除术之间进行决策 (Zhou et al. 2022; Villalba et al. 2021; Ding et al.

*These authors contributed equally.

†Corresponding author. Email: ynyang@xtu.edu.cn.

‡Corresponding author. Email: 503079@csu.edu.cn.

§Corresponding author. Email: slpeng@hnu.edu.cn.

2023)。术后 STAS 诊断也在帮助肿瘤学家制定个性化治疗方案方面发挥着重要作用。然而，最近的研究报告称人工 STAS 诊断的 AUC 介于 0.63 和 0.8 之间，总体准确率约为 0.7 (Villalba et al. 2021; Zhou et al. 2022)。敏感性往往较低，表明存在误诊的高风险 (Villalba et al. 2021; Cao et al. 2025)。因此，临床上迫切需要一种客观、准确和高效的方法来协助诊断 STAS。随着深度学习的进步，人工智能 (AI) 已越来越多地整合到病理图像分析中，在癌症诊断和组织分级方面取得了显著进展 (Coudray et al. 2018; Li et al. 2025)。鉴于传统病理学的局限性，在组织病理图像中检测 STAS 的基于 AI 的方法正在成为一种有前途的方向。如图 2 所示，STAS 表现为三种组织学形式：微乳头簇（缺乏纤维血管核心，有时在气道内形成环状结构）、固体巢（肿瘤细胞的紧密聚集）和单个肿瘤细胞（单独散布的细胞）(Kadota et al. 2015)。STAS 特征的形态异质性和细微性对 AI 模型构成了重大挑战，特别是在准确识别不规则或稀疏病变时。此外，STAS 诊断需要高质量的整张幻灯片图像 (WSIs) 和专家注释，这既成本高，又劳动密集且本质上是主观的，导致训练数据有限且不平衡。染色协议的变化和机构差异进一步阻碍了模型的泛化。尽管深度学习模型在标准图像分类任务中表现强劲，但它们在 WSIs 上的 STAS 检测准确性仍不足以进行临床部署。这些挑战突显了迫切需要更强大的诊断策略来提高 STAS 识别的一致性、可靠性和效率。

为了解决上述挑战，我们首先收集了来自中南大学湘雅二医院和湘雅三医院的 STAS 患者病理图像，以及癌症基因组图谱 LUAD (TCGA_LUAD) 数据集。我们邀请了三位高级病理学家进行交叉验证注释，构建了 STAS-SXY、STAS-TXY 和 STAS-TCGA 数据集。对每张切片进行了 STAS 存在与否及扩散病灶类型的注释。鉴于 WSIs 中 STAS 的复杂病理特征，我们提出了一种名为 STAMP 的多模式关注感知多实例学习 (MIL) 框架用于 STAS 诊断。双分支表示通过指导网络从不同语义空间（如形态特征和上下文特征）学习多样的病变模式，打破了单分支模型的限制，从而增强了特征的多样性。基于 Transformer 的实例编码和多模式注意力聚合 (MPAA) 模块动态选择与病变相关的关键区域，避免噪声干扰，提高全局表示的判别能力。相似性正则化约束学习策略防止模式冗余，确保所学病变模式具有高度鉴别性和互补性。我们的贡献如下：

- 我们与三位病理学家合作，构建并注释了三个专用的 STAS 数据集 (STAS-SXY、STAS-TXY 和 STAS-TCGA)，共包含 2,011 个 WSI。这些数据集为推进 LUAD 中 STAS 的临床研究提供了宝贵的资源。
- 为了解决 STAS 的病理复杂性，我们引入了 STAMP，这是一种多模式注意力感知的多实例学习框架。在 STAMP 中，提出的 MPAA 模块在特征层面采用门控和内容感知的双路径注意力，以动态突出病灶相关区域，从而提取丰富的异质性特征并提高全切片图像级别的诊断准确性。
- 我们对 10 种最先进的 MIL 方法在多中心 STAS 数据集上进行了广泛的基准测试，提供了全面的基线结果。广泛的消融研究进一步验证了 STAMP 的有效性和鲁棒性，突显了其在推动 STAS 感知计算病理学方面的潜力。

相关工作

组织病理学图像分析方法 组织病理学图像分类经历了三个主要阶段。最初，传统的机器学习方法依赖于手工设计的特征，结合支持向量机、随机森林和 k 近邻等算法来对全切片图像进行分类 (Cortes and Vapnik 1995; Breiman 2001; Zhang 2016)。随着 2015 年左右深度学习的出现，卷积神经网络成为主流，使得对包括乳腺、前列腺和结直肠癌在内的各种癌症类型的肿瘤亚型和侵袭水平的精确分类成为可能，通常实现了接收者操作特征曲线下面积 (AUCs) 超过 0.8 的结果 (Pan et al. 2024)。自 2019 年以来，由于其与弱监督学习场景的兼容性，多实例学习 (MIL) 已成为全切片图像分析的主要范式。MIL 方法可以大致分为基于实例的方法和基于袋 (bag) 的方法。基于实例的方法向单个实例赋予预测，然后推断袋级标签，假设几个关键实例足以决定最终分类。代表性的方法包括 MaxPooling、ABMIL (Ilse, Tomczak, and Welling 2018)、DSMIL (Li, Li, and Eliceiri 2021)、ILRA (Xiang and Zhang 2023)、CLAM-SB、CLAM-MB (Lu et al. 2021) 和 TransMIL (Shao et al. 2021)。相比之下，基于袋的方法直接在袋级操作，利用相似性度量（例如，高斯核距离）而不需要实例级标签。示例包括 MeanPooling 和 WIKG (Li et al. 2024a)。将 MIL 与其他方法结合的最新进展，如嵌入领域概率先验的 DeeMILIP (Al Hajj et al. 2024)、将广义线性混合模型与 MIL 结合的 MixMIL (Engelmann et al. 2023) 和结合贝叶斯非参数进行补丁聚类的 cDP-MIL (Chen et al. 2024)，进一步提升了 MIL 的能力，提高了在处理如对抗样本稀缺性和不确定性估计等挑战时的效率和稳健性。

近年来，只有少数研究专注于 STAS 诊断的技术探索和模型创新。Lee 等人分析了 115 个 LUAD 全切片图像，生成了超过 60 万个 HE 切片，通过迁移学习训练 DL 模型，实现了切片级 STAS 检测的 AUC 为 0.738。STASNet 在 489 个全切片图像上进一步改善了半定量 STAS 识别，获得了切片级精度为 0.93 和全切片图像级 AUC 范围在 0.72 到 0.78 之间。它还计算了 STAS 密度和空间分布参数，增强了实时诊断工作流程中的检测能力。最近，VERN 模型采用了 Siamese 图编码器来捕捉空间拓扑特征，在内部验证中获得了高 AUC，并在冷冻和石蜡包埋测试部分分别取得了 0.8275 和 0.8829 的 AUC，展示了临床级性能。此外，Gong 等人提出了一种多场景通道变压器 (MFCT)，在 STAS 识别中达到 0.734 的 F1 评分。SMILE 框架引入了尺度感知的 MIL 以自适应选择高注意力实例，在 CPTAC 和 TCGA 数据集上分别诊断出 251 和 319 个 STAS 阳性全切片图像，优于平均临床 AUC。这些研究总体上显示了 DL 和 MIL 方法在 STAS 诊断中的前景。然而，现有方法仍然产生适中的 AUC，表明需要进一步的进展。为此，我们提出了一种多模式注意感知 MIL 框架，STAMP，以更有效地从全切片图像中检测 STAS。

方法

在多实例学习 (MIL) 中，一个被划分成多个块的全图像切片 (WSI) 被表示为一个包 $X = \{(x_1, y_1), \dots, (x_n, y_n)\}$ ，其中 x_i 表示包中的第 i 个块（实例），并且 y_i 代表其实例级标签（在弱监督设置中通常是未知的）。从 WSI 获取的实例数量表示为 n 。每个包 X 与一个包级标签 $Y \in \{0, 1\}$ 相关联，表示

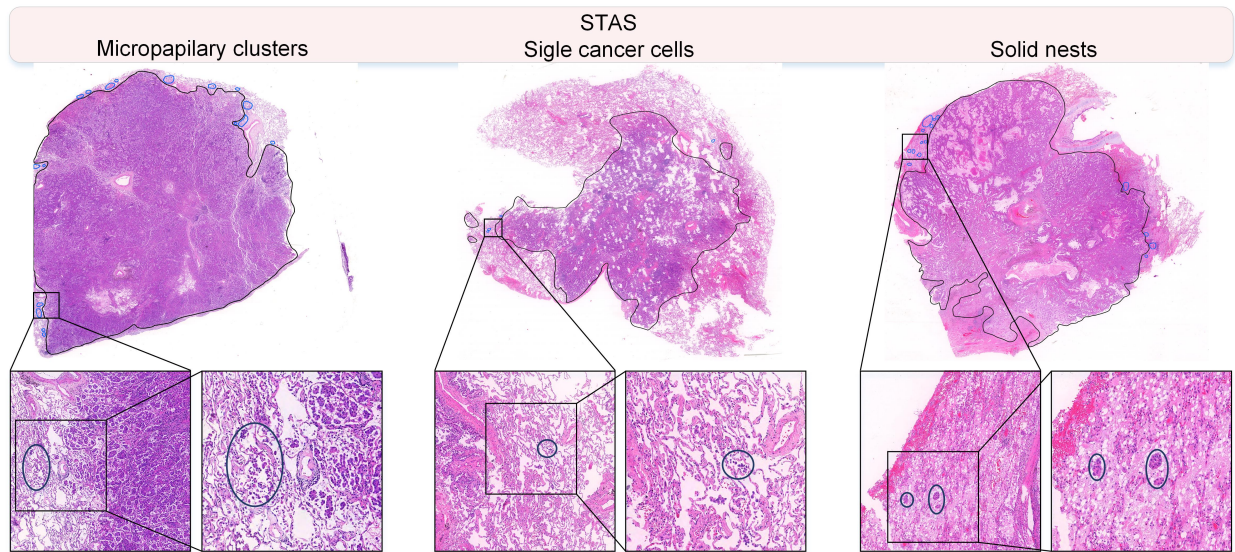


Figure 2: 肺癌组织病理学图像中 STAS 的三大主要病理特征如下：STAS 主要分布在肿瘤体外，具有微乳头簇、单个癌细胞和实性巢的病理特征。

WSI 是否包含阳性病变区域（例如，STAS 区域）。在弱监督 MIL 假设下，包级标签与实例级标签之间的关系定义如下：

$$Y = \begin{cases} 0, & \text{if } \sum y_i = 0 \\ 1, & \text{otherwise} \end{cases} \quad (1)$$

在 STAS 二分类任务中，每个实例 x_i 代表组织病理学图像中的一个区域， y_i 表示其对应的标签。目标是学习一个分类模型，可以根据其组成区域推断 WSI 的标签 Y ，从而识别 WSI 中是否存在与 STAS 相关的病理特征。

多模式注意力感知多实例学习

概述 在本节中，我们介绍了 STAMP，这是一种新颖的多模式注意力感知 MIL 框架，用于弱监督的 WSI 分类。给定一个由实例 $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{n \times d}$ 表示的 WSI，其中每个 $x_i \in \mathbb{R}^d$ 代表一个 patch 级别的特征，STAMP 旨在通过双令牌设计和特定模式的注意力聚合来学习辨别性的全局病变表示。整体框架由四个关键组件组成：(1) 双令牌嵌入，(2) 基于 Transformer 的实例编码，(3) MPAA 模块，和 (4) 带有正则化相似性损失的分类。

双重标记嵌入

为了引导学习多样的病变感知表示，我们引入了两组可学习的标记：头标记和尾标记，每个标记由初始化为参数的 n_p 个向量组成：

$$\mathbf{T}_h, \mathbf{T}_t \in \mathbb{R}^{n_p \times d} \quad (2)$$

这些标记被添加到原始实例序列的前面，并通过两个线性层 $\mathbf{W}_h, \mathbf{W}_t \in \mathbb{R}^{d \times L}$ 分别投射以获得头部和尾部分支：

$$\mathbf{X}_h = \mathbf{W}_h([\mathbf{T}_h; \mathbf{X}]), \quad \mathbf{X}_t = \mathbf{W}_t([\mathbf{T}_t; \mathbf{X}]) \quad (3)$$

这种双分支设计鼓励模型从不同的表示空间中提取互补的语义线索。

基于 Transformer 的实例编码

每个令牌增强序列 $\mathbf{X}_h$ 和 $\mathbf{X}_t$ 独立地输入到通过 Nyström Attention (Xiong et al. 2021) 实现的轻量级 Transformer 编码器中，以提高效率和可扩展性。具体来说，每个分支应用以下层次：

$$\mathbf{H}_h = \text{TransLayer}(\mathbf{X}_h), \quad \mathbf{H}_t = \text{TransLayer}(\mathbf{X}_t) \quad (4)$$

我们保留来自两个分支的顶部 n_p 头部和尾部令牌，并将它们连接为：

$$\mathbf{H} = [\mathbf{H}_h^{(1:n_p)}; \mathbf{H}_t^{(1:n_p)}] \in \mathbb{R}^{2n_p \times L} \quad (5)$$

这些聚合令牌预计可以编码对最终分类有用的区域特定语义模式。

多模式注意力聚合

为了计算全局包级表示，我们设计了一个结合门控和基于内容注意力的模式级注意力模块：

$$\mathbf{A}_V = \tanh(\mathbf{W}_v \mathbf{H}) \quad (6)$$

$$\mathbf{A}_U = \sigma(\mathbf{W}_u \mathbf{H}) \quad (7)$$

$$\mathbf{A} = \text{softmax}(\mathbf{W}_a (\mathbf{A}_V \odot \mathbf{A}_U)) \quad (8)$$

这里， $\odot$ 表示逐元素相乘， $\mathbf{W}_v, \mathbf{W}_u \in \mathbb{R}^{L \times D}$ ， $\mathbf{W}_a \in \mathbb{R}^{D \times 1}$ 是可训练的矩阵。注意力权重 $\mathbf{A} \in \mathbb{R}^{1 \times 2n_p}$ 用于通过

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{H} \quad (9)$$

将标记特征聚合成全局表示 $\mathbf{M} \in \mathbb{R}^{1 \times L}$ 。其中， $\cdot$ 代表向量的乘积。

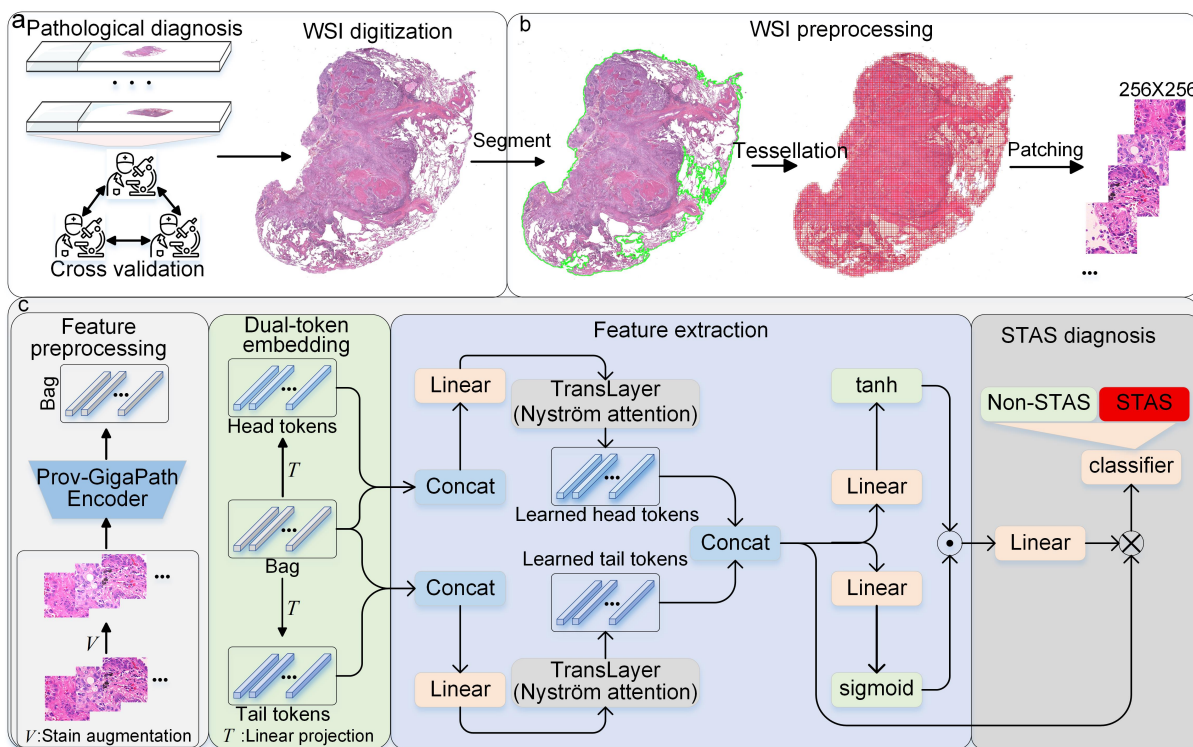


Figure 3: 诊断 STAS 的完整工作流程来自组织病理图像。a): 肺癌患者的组织病理图像的注释（交叉验证）和 WSI 的数字化的；b): WSI 的预处理，包括分割、镶嵌和拼接；c): WSI 图像特征的预处理，双令牌嵌入，特征提取（包括基于变压器的实例编码和 MPAA 模块），以及 STAS 的诊断（使用正则化相似性损失进行分类）。

分类和相似性正则化

全局特征 M 被传递到一个线性分类器用于包级预测：为了增强所学模式代币的多样性和正交性，我们在最终代币表示上引入了一个相似性正则化损失：总训练损失结合了标准的交叉熵和相似性正则化：其中 λ 默认设置为 0.9。

实验与结果

数据集

鉴于 STAS 病理评估中误诊和漏诊的倾向，三位经验丰富的病理学家采用双盲协议和交叉验证来注释每个 WSI 的 STAS 状态，以确保准确性并减轻主观性、疏忽和过度诊断。在纳入排除过程中，上述具有丰富诊断经验的病理学家排除了包含机械伪影的 WSI。虽然在临床病理实践中会遇到 STAS 分化的假阳性，但我们利用相关的免疫组织化学标记物标记可疑病灶，以进一步区分真正的 STAS 阳性扩散和组织细胞聚集体。

STAS-SXH: 我们将来自中南大学湘雅二医院队列的 WSI 作为我们的内部训练和验证数据集。在 2020 年 4 月至 2023 年 12 月间接接受肺结节切除并经组织学确认为 LUAD 的 12,169 名患者中，选择了 206 名明确诊断为 STAS 和 150 名未诊断为 STAS 的个体。在这些选择的病例中，我们总共收集了 1,290 张 WSI（每位患者约四张幻灯片），以及相关的临床数据和免疫组织化学图像（例如，标记，如 TTF-1、CK、CD68）(Yatabe, Mitsudomi, and Takahashi 2002; Chistiakov et al. 2017)

STAS-TXH: 我们纳入了来自中南大学第三湘雅医院的一组队列，包括 68 名在 2022 年至 2023 年期间诊断为 STAS 的肺癌患者的 304 张组织病理学切片，每个 WSI 都标注了 STAS 的存在或缺失。

STAS-TCGA: 我们根据纳入/排除方法从 TCGA_LUAD 队列数据集中收集了来自 366 名患者的 417 张 WSI。所有 WSI 均附有相应的 STAS 标签。

实验细节

使用多中心 STAS 数据集预测 WSI 中 STAS 的工作流程主要包括三个步骤（图 3）：(1) STAS 的病理诊断，(2) WSI 预处理，以及 (3) WSI 的特征预处理。具体来说，病理学家首先根据纳入和排除标准筛选合格的病理切片进行 STAS 诊断，随后对切片进行数字化。WSI 预处理主要涉及选择肺癌组织的感兴趣区域 (ROI) 并将 WSI 分割成若干小块。我们使用大津法的自动阈值分割方法将 WSI 中的组织、背景和模糊区域进行分割 (Liu and Yu 2009)。然后，应用滑动窗口策略将 $20 \times$ 倍放大率的 WSI 分割成 256×256 小块，同时记录每块的空间坐标。为了减少不同扫描设备和切片质量对模型泛化能力的影响，我们利用训练于 TCGA 的结构保留 GAN 对 WSI 的所有小块进行染色增强 (Wagner et al. 2023, 2021)。随后，所有小块都使用 GigaPath 进行图像特征提取 (Janowczyk et al. 2019)。

Method	STAS-SXY					STAS-TXY					STAS-TCGA				
	ACC	AUC	Precision	Recall	F1-Score	ACC	AUC	Precision	Recall	F1-Score	ACC	AUC	Precision	Recall	F1-Score
MaxPooling	0.6894	0.7337	0.6988	0.6981	0.6842	0.6693	0.6511	0.6803	0.6712	0.6274	0.6137	0.6159	0.5995	0.6311	0.5606
MeanPooling	0.7181	0.7733	0.7378	0.7253	0.7093	0.7895	0.6926	0.6919	0.7139	0.7010	0.5700	0.6069	0.6099	0.6343	0.5319
ABMIL	0.7061	0.7416	0.7142	0.7080	0.6965	0.7428	0.7430	0.7262	0.7322	0.6996	0.6236	0.6146	0.6283	0.6326	0.5602
DSMIL	0.7144	0.7666	0.7111	0.7163	0.7120	0.7725	0.6954	0.7285	0.6813	0.6760	0.5842	0.5630	0.5785	0.5955	0.5253
ILRA	0.7421	0.7828	0.7350	0.7359	0.7320	0.7047	0.7621	0.7022	0.7355	0.6801	0.5797	0.5964	0.6423	0.6316	0.5327
WKG	0.7227	0.7880	0.7376	0.7379	0.7196	0.7567	0.7801	0.7386	0.7771	0.7289	0.5654	0.6449	0.6411	0.6636	0.5312
Clam_SB	0.7227	0.8042	0.7227	0.7227	0.6885	0.7532	0.8130	0.7532	0.7532	0.6792	0.7392	0.6097	0.7392	0.7392	0.5764
Clam_MB	0.7328	0.7971	0.7328	0.7328	0.7214	0.7606	0.7925	0.7606	0.7606	0.6910	0.7464	0.6258	0.7464	0.7464	0.5337
TransMIL	0.7116	0.7382	0.7137	0.7072	0.7002	0.7218	0.7796	0.7485	0.7385	0.6867	0.5898	0.5872	0.6001	0.6181	0.5370
STAMP	0.7449	0.8058	0.7394	0.7424	0.7376	0.8124	0.8017	0.7619	0.7785	0.7623	0.7186	0.7928	0.7197	0.7546	0.6962

Table 1: 基于 10 种 MIL 方法在 STAS-SXY、STAS-TXY 和 STAS-TCGA 中诊断 STAS 结果的统计数据。

训练细节。 我们使用 Ranger 优化器训练我们的模型，Ranger 结合了 RAdam 和 Lookahead 以实现稳定和高效的收敛。初始学习率设为 1×10^{-4} ，权重衰减为 1×10^{-5} 。采用余弦退火学习率调度器，在训练过程中逐步将学习率降低到 5×10^{-6} 的最小值。

我们将模型训练 50 个 epoch，批量大小为 1。输入特征维度 d 设置为 1536，隐藏层维度 L 设置为 512。为了确保性能评估的稳健性，我们在随机种子为 0 到 5 之间进行实验，并汇报这些实验的平均结果。

结果

为了评估 MIL 方法在诊断 STAS 中的性能，我们在三个精心整理的 STAS 数据集上对十个代表性模型进行了基准测试：MaxPooling, MeanPooling, ABMIL (Ilse, Tomczak, and Welling 2018), DSMIL (Li, Li, and Eliceiri 2021), ILRA (Xiang and Zhang 2023), WIKG (Lu et al. 2021), CLAM-SB (Lu et al. 2021), CLAM-MB (Lu et al. 2021), TransMIL (Shao et al. 2021), 以及 STAMP。对于每种 MIL 方法，我们使用准确率 (ACC)、AUC、精确率、召回率和 F1-分数来全面评估 STAS 的诊断性能。下文中，我们比较了 10 个 MIL 模型在三个 STAS 数据集上的性能。在 STAS-SXY 数据集中，STAMP 始终优于所有基线模型，获得了最高的 AUC (0.8058)、F1-分数 (0.7376) 和 ACC (0.7449)，展示了其检测 STAS 存在的卓越能力。值得注意的是，ILRA 的 AUC 为 0.7828，CLAM-MB 达到了 0.7971，表现出强劲的性能，突显出多模式注意力聚合和双分支框架的有效性。相反，像 MaxPooling 和 MeanPooling 这样简单的聚合方法显示出有限的有效性，表明它们无法捕捉到关键的组织病理学线索。

在 STAS-TXY 数据集上，STAMP 再次取得了顶尖的表现，具有最高的 ACC (0.8124) 和 F1-score (0.7623)。尽管其 AUC (0.8017) 略低于 CLAM-SB (0.8130)，但结果表明其具有相当的辨别能力。WIKG 和 CLAM-MB 也表现出色，这表明与 DSMIL 等实例级方法相比，采用聚类感知注意力的袋级建模对于 STAS 的诊断更加稳健。

在更具挑战性的 STAS-TCGA 数据集上，STAMP 保持了强劲的表现，实现了所有模型中最高的 F1 得分 (0.6962) 和 AUC (0.7928)。虽然 CLAM-MB 获得了最高的 ACC (0.7464)，但其低 F1 得分 (0.5337) 揭示了灵敏度和 ACC 之间的权衡，可能是由于过度拟合到显著模式造成的。传统的池化策略和早期的 MIL 模型 (例如 ABMIL、DSMIL) 表现大幅下降，AUC 在大多数情况下低于 0.65，反映出这些模型在提取 STAS 复杂病理特征方面的局限性。

消融研究

图案数量的影响。 为了评估潜在模式数量对模型性能的影响，我们在 STAS-SXY 数据集上评估了不同模式数量 ($n_p \in \{1, 2, 3, 4, 5\}$) 的 STAMP 变体。如表 2 所总结，当 $n_p = 3$ 时，模型实现了最高的整体性能，表明了表示能力与泛化之间的最佳平衡。较少的模式 (例如， $n_p = 2$) 可能由于表达能力不足导致欠拟合，而更多的模式 (例如， $n_p = 4, 5$) 则在没有性能提升的情况下引入了冗余。基于这些发现，我们将 $n_p = 3$ 设定为所有后续实验的默认配置。

Pattern Number	ACC	AUC	Precision	Recall	F1-Score
1	0.7264	0.7804	0.7297	0.7332	0.7219
2	0.7227	0.7827	0.7186	0.7232	0.7163
3	0.7449	0.7971	0.7394	0.7424	0.7376
4	0.7356	0.7917	0.7372	0.7418	0.7311
5	0.7329	0.7941	0.7391	0.7391	0.7271

Table 2: 模式数量对模型性能的影响。

为了评估双分支架构的有效性，我们通过改变单分支和双分支配置中的潜在模式数量，在 STAS-SXY 上进行了对比实验。如图 4 所示，双分支 (H,T) 变体在所有指标上均始终优于其单分支 (H) 对应版本。例如，在有三个潜在模式的情况下，双分支模型达到了 0.7449 的准确率，比单分支版本高出 3.97%。这种改进可归因于双分支设计所增强的特征表示。尽管单分支模型仅限于固定的语义层次和感受野，双分支架构利用并行子网络在多个语义粒度和感受野上提取互补信息。这导致特征空间显著丰富，从而增强了表示能力和区分能力。

我们进一步研究实例特征嵌入策略对模型性能的影响。具体来说，我们设计了两种嵌入变体：预嵌入和后嵌入。相应的公式定义如下：

$$X_h = [W_h(X); T_h], \quad X_t = [W_t(X); T_t] \quad (10)$$

如图 5 所示，在不同数量的潜在模式中，后注意力嵌入始终优于前注意力方法。这表明允许区域感知可学习标记在后期与实例特征进行交互对模型更有利，可能是由于增强了上下文和与学习到的语义更好的对齐。

为了进一步评估多模式注意力聚合策略对模型性能的影响，我们比较了两种不同的融合方法：特征级聚合 (FA) 和预测级聚合 (PA)。在我们提出的 STAMP 框架中，FA 被采用为默认设置。PA 策略可以被公式化为如下所示：

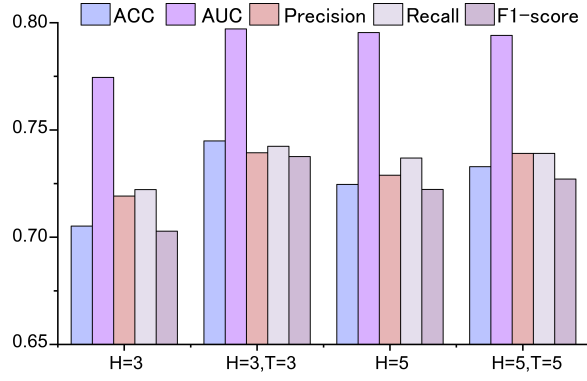


Figure 4: 单分支结构和双分支结构对 STAMP 性能的影响。

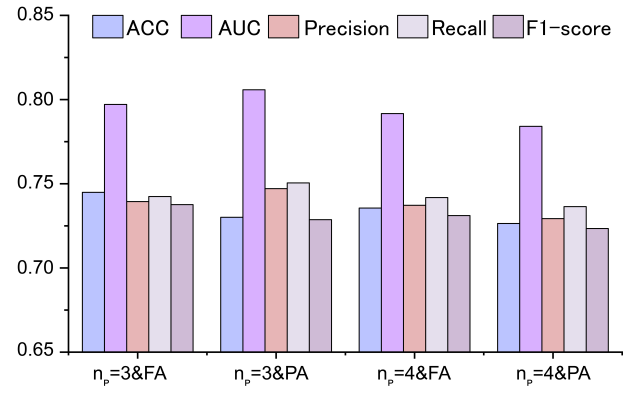


Figure 6: 多模式注意力聚合策略对 STAMP 性能的影响。

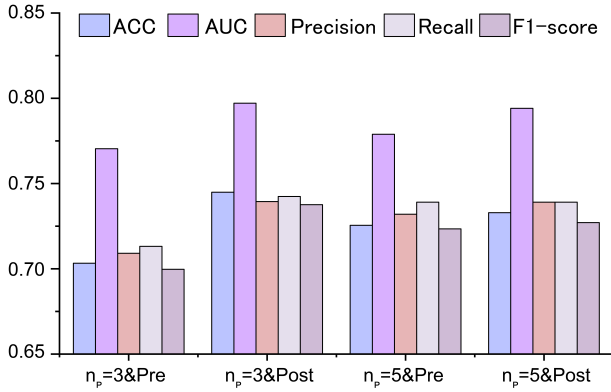


Figure 5: 令牌嵌入方法对 STAMP 性能的影响。Pre 和 Post 分别代表预嵌入和后嵌入。

$$M = \frac{1}{2} \left[\text{softmax}(W_a^h \cdot (\tanh(W_v^h H_h^{(1:n_p)}) \odot \sigma(W_u^h H_h^{(1:n_p)}))) \right. \\ \left. + \text{softmax}(W_a^t \cdot (\tanh(W_v^t H_t^{(1:n_p)}) \odot \sigma(W_u^t H_t^{(1:n_p)}))) \right] \quad (11)$$

如 6 所示，FA 在不同数量的潜在模式下始终优于 PA。例如，当 $n_p = 3$ 时，FA 达到了 0.7449 的准确率，超过了 PA 的 0.7301。这个结果表明特征级聚合在增强模型的整体表示能力和预测性能方面更有效。

讨论与结论

在这项研究中，我们引入了三个肺癌 STAS 数据集，以促进 STAS 诊断的研究。为了应对 STAS 的复杂病理特征，我们提出了 STAMP，一种多模式注意力感知的 MIL 框架。我们在三个数据集 (STAS-SXY、STAS-TXY 和 STAS-TCGA) 上对 10 个 MIL 模型进行了 WSI 级别的 STAS 预测基准测试。STAMP 在 STAS-SXY 和 STAS-TCGA 上达到最先进的 AUC 和 F1 得分，并在 STAS-TXY 上表现出优越的准确性和 F1 得分。与现有的 MIL 方法不同，STAMP 中的 MPAA 模块在特征级别整合了门控和内容感知的双路径注意力，能够自适应地聚焦于病变相关区域，并捕获多样的 STAS 特异性表示，从而改善 WSI 诊断。消融研究证实，双分支

结构通过并行子网络扩展了特征空间，用于多尺度和多粒度的提取，增强了表达性和可辨别性。此外，实例后特征嵌入允许学习的区域感知标记与实例级特征之间的早期交互，促进细致的 STAS 模式学习，而特征级多模态融合则带来了显著的性能提升。这些发现强调了 STAMP 在支持 WSI 中的临床 STAS 检测方面的潜力。

尽管取得了这些令人鼓舞的结果，但仍存在两个主要局限。首先，STAS 数据集是在与病理学家密切合作下策划的，质量控制完全依赖于专家判断，这是一项昂贵且劳动密集的过程。未来的工作将涉及收集带有常见伪影的 WSI (例如，机械变形、气泡和染色不一致) 来训练用于初步 WSI 筛选的自动质量控制模型。其次，虽然我们的模型在 STAS-SXY 和 STAS-TCGA 上表现出色，但在 STAS-TXY 上表现不佳，这强调了增强跨中心泛化的必要性。鉴于 STAS 阳性病例的稀缺性，我们计划获取更多阳性 WSI 来增强训练数据。此外，尽管 STAMP 在诊断准确性上优于最先进的 MIL 方法，弥合模型生成的注意力图与专家病理学注释之间的差距仍然是未来研究的一个重要方向。

References

- Al Hajj, G. S.; Hubin, A.; Kanduri, C.; Pavlovic, M.; Rand, K. D.; Widrich, M.; Solberg, A. S.; Greiff, V.; Pensar, J.; Klambauer, G.; et al. 2024. Incorporating probabilistic domain knowledge into deep multiple instance learning. In Forty-first International Conference on Machine Learning.
- Breiman, L. 2001. Random forests. Machine learning, 45(1): 5–32.
- Cao, H.; Zheng, Q.; Deng, C.; Fu, Z.; Shen, X.; Jin, Y.; Yang, Y.; Qian, B.; Yuan, C.; Wang, W.; et al. 2025. Prediction of Spread Through Air Spaces By Intra-operative Frozen Section for Patients With cT1N0M0 Invasive Lung Adenocarcinoma: A Multicenter Observational Study (ECTOP-1016). Annals of Surgery, 281(2): 187–192.
- Chen, Y.; Chan, T. H.; Yin, G.; Jiang, Y.; and Yu, L. 2024. cDP-MIL: Robust Multiple Instance Learning via Cascaded Dirichlet Process. In European conference on computer vision, 232–250. Springer.

- Chistiakov, D. A.; Killingsworth, M. C.; Myasoedova, V. A.; Orekhov, A. N.; and Bobryshev, Y. V. 2017. CD68/macrosialin: not just a histochemical marker. *Laboratory investigation*, 97(1): 4–13.
- Cortes, C.; and Vapnik, V. 1995. Support-vector networks. *Machine learning*, 20(3): 273–297.
- Coudray, N.; Ocampo, P. S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyö, D.; Moreira, A. L.; Razavian, N.; and Tsirigos, A. 2018. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine*, 24(10): 1559–1567.
- Ding, Y.; Zhao, S.; Liu, X.; Ren, J.; Li, J.; Zhang, W.; Xu, M.; and Sun, D. 2023. The value of frozen section diagnosis of tumor spread through air spaces in small-sized (≤ 2 cm) non-small cell lung cancer. *World Journal of Surgical Oncology*, 21(1): 195.
- Engelmann, J. P.; Palma, A.; Tomczak, J. M.; Theis, F. J.; and Casale, F. P. 2023. Mixed models with multiple instance learning. *arXiv preprint arXiv:2311.02455*.
- Han, Y. B.; Kim, H.; Mino-Kenudson, M.; Cho, S.; Kwon, H. J.; Lee, K. R.; Kwon, S.; Lee, J.; Kim, K.; Jheon, S.; et al. 2021. Tumor spread through air spaces (STAS): prognostic significance of grading in non-small cell lung cancer. *Modern Pathology*, 34(3): 549–561.
- Ikeda, T.; Kadota, K.; Go, T.; Haba, R.; and Yokomise, H. 2021. Current status and perspectives of spread through air spaces in lung cancer. *Thoracic Cancer*, 12(11): 1639–1646.
- Ilse, M.; Tomczak, J.; and Welling, M. 2018. Attention-based deep multiple instance learning. In *International conference on machine learning*, 2127–2136. PMLR.
- Janowczyk, A.; Zuo, R.; Gilmore, H.; Feldman, M.; and Madabhushi, A. 2019. HistoQC: an open-source quality control tool for digital pathology slides. *JCO clinical cancer informatics*, 3: 1–7.
- Kadota, K.; Nitadori, J.-i.; Sima, C. S.; Ujiie, H.; Rizk, N. P.; Jones, D. R.; Adusumilli, P. S.; and Travis, W. D. 2015. Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage I lung adenocarcinomas. *Journal of thoracic oncology*, 10(5): 806–814.
- Li, B.; Li, Y.; and Eliceiri, K. W. 2021. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14318–14328.
- Li, J.; Chen, Y.; Chu, H.; Sun, Q.; Guan, T.; Han, A.; and He, Y. 2024a. Dynamic graph representation with knowledge-aware attention for histopathology whole slide image analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11323–11332.
- Li, Y.; Adusumilli, P. S.; Chou, T.-Y.; Kadota, K.; Mino-Kenudson, M.; Papotti, M.; Rekhtman, N.; Yagi, Y.; Yatabe, Y.; and Travis, W. D. 2024b. Pro: “Is Spread Through Air Spaces an In Vivo Phenomenon or an Inducible Artifact?” .
- Li, Y.; Pan, L.; Peng, Y.; Li, X.; Wang, X.; Qu, L.; Song, Q.; Liang, Q.; and Peng, S. 2025. Application of deep learning-based multimodal fusion technology in cancer diagnosis: A survey. *Engineering Applications of Artificial Intelligence*, 143: 109972.
- Liu, D.; and Yu, J. 2009. Otsu method and K-means. In *2009 Ninth International conference on hybrid intelligent systems*, volume 1, 344–349. IEEE.
- Lu, M. Y.; Williamson, D. F.; Chen, T. Y.; Chen, R. J.; Barbieri, M.; and Mahmood, F. 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6): 555–570.
- Ma, K.; Zhan, C.; Wang, S.; Shi, Y.; Jiang, W.; and Wang, Q. 2019. Spread through air spaces (STAS): a new pathologic morphology in lung cancer. *Clinical Lung Cancer*, 20(2): e158–e162.
- Onozato, M. L.; Kovach, A. E.; Yeap, B. Y.; Morales-Oyarvide, V.; Klepeis, V. E.; Tammireddy, S.; Heist, R. S.; Mark, E. J.; Dias-Santagata, D.; Iafrate, A. J.; et al. 2013. Tumor islands in resected early-stage lung adenocarcinomas are associated with unique clinicopathologic and molecular characteristics and worse prognosis. *The American journal of surgical pathology*, 37(2): 287–294.
- Pan, L.; Wang, X.; Liang, Q.; Shang, J.; Liu, W.; Xu, L.; and Peng, S. 2024. DEDUCE: Multi-head attention decoupled contrastive learning to discover cancer subtypes based on multi-omics data. *Computer Methods and Programs in Biomedicine*, 257: 108478.
- Shao, Z.; Bian, H.; Chen, Y.; Wang, Y.; Zhang, J.; Ji, X.; et al. 2021. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34: 2136–2147.
- Shih, A. R.; and Mino-Kenudson, M. 2020. Updates on spread through air spaces (STAS) in lung cancer. *Histopathology*, 77(2): 173–180.
- Toyokawa, G.; Yamada, Y.; Tagawa, T.; Kozuma, Y.; Matsubara, T.; Haratake, N.; Takamori, S.; Akamine, T.; Oda, Y.; and Maehara, Y. 2018. Significance of spread through air spaces in resected pathological stage I lung adenocarcinoma. *The Annals of thoracic surgery*, 105(6): 1655–1663.
- Travis, W. D.; Brambilla, E.; Burke, A. P.; Marx, A.; and Nicholson, A. G. 2015. Introduction to the 2015 World Health Organization classification of tumors of the lung, pleura, thymus, and heart. *Journal of Thoracic Oncology*, 10(9): 1240–1242.
- Vaghjiani, R. G.; Takahashi, Y.; Eguchi, T.; Lu, S.; Kameda, K.; Tano, Z.; Dozier, J.; Tan, K. S.; Jones, D. R.; Travis, W. D.; et al. 2020. Tumor spread through

- air spaces is a predictor of occult lymph node metastasis in clinical stage IA lung adenocarcinoma. *Journal of Thoracic Oncology*, 15(5): 792–802.
- Villalba, J. A.; Shih, A. R.; Sayo, T. M. S.; Kunitoki, K.; Hung, Y. P.; Ly, A.; Kem, M.; Hariri, L. P.; Muniappan, A.; Gaissert, H. A.; et al. 2021. Accuracy and reproducibility of intraoperative assessment on tumor spread through air spaces in stage 1 lung adenocarcinomas. *Journal of Thoracic Oncology*, 16(4): 619–629.
- Wagner, S. J.; Khalili, N.; Sharma, R.; Boxberg, M.; Marr, C.; De Back, W.; and Peng, T. 2021. Structure-preserving multi-domain stain color augmentation using style-transfer with disentangled representations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 257–266. Springer.
- Wagner, S. J.; Reisenbüchler, D.; West, N. P.; Niehues, J. M.; Zhu, J.; Foersch, S.; Veldhuizen, G. P.; Quirke, P.; Grabsch, H. I.; van den Brandt, P. A.; et al. 2023. Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study. *Cancer cell*, 41(9): 1650–1661.
- Wang, S.; Hao, J.; Qian, C.; and Wang, H. 2019. Tumor Spread Through Air Spaces Is a Survival Predictor in Non-Small-Cell Lung Cancer. *Clinical lung cancer*, 20(5): e584–e591.
- Warth, A.; Muley, T.; Kossakowski, C. A.; Goeppert, B.; Schirmacher, P.; Dienemann, H.; and Weichert, W. 2015. Prognostic impact of intra-alveolar tumor spread in pulmonary adenocarcinoma. *The American journal of surgical pathology*, 39(6): 793–801.
- Willner, J.; Narula, N.; and Moreira, A. L. 2024. Updates on lung adenocarcinoma: invasive size, grading and STAS. *Histopathology*, 84(1): 6–17.
- Xiang, J.; and Zhang, J. 2023. Exploring low-rank property in multiple instance learning for whole slide image classification. In *The Eleventh International Conference on Learning Representations*.
- Xiong, Y.; Zeng, Z.; Chakraborty, R.; Tan, M.; Fung, G.; Li, Y.; and Singh, V. 2021. Nyströmformer: A nyström-based algorithm for approximating self-attention. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 14138–14148.
- Yatabe, Y.; Mitsudomi, T.; and Takahashi, T. 2002. TTF-1 expression in pulmonary adenocarcinomas. *The American journal of surgical pathology*, 26(6): 767–773.
- Zhang, Z. 2016. Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine*, 4(11): 218.
- Zhao, S.; Jiang, T.; Zhang, L.; Yang, H.; Liu, X.; Jia, Y.; and Zhou, C. 2016. Clinicopathological and prognostic significance of regulatory T cells in patients with non-small cell lung cancer: A systematic review with meta-analysis. *Oncotarget*, 7(24): 36065.
- Zhou, F.; Villalba, J. A.; Sayo, T. M. S.; Narula, N.; Pass, H.; Mino-Kenudson, M.; and Moreira, A. L. 2022. Assessment of the feasibility of frozen sections for the detection of spread through air spaces (STAS) in pulmonary adenocarcinoma. *Modern Pathology*, 35(2): 210–217.