

STAMP: Multi-pattern Attention-aware Multiple Instance Learning for STAS Diagnosis in Multi-center Histopathology Images

Liangrui Pan¹, Xiaoyu Li¹, Chenchen Nie², Yaning Yang³, Shaoliang Peng¹

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

²Department of pathology, Hunan Cancer Hospital, The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, 410013, Hunan, China

³School of Computer Science, Xiangtan University, Xiangtan, 411100, Hunan, China
{panlr, hnulixy, slpeng}@hnu.edu.cn, niechenchen@hnca.org.cn, ynyang@xtu.edu.cn

Abstract

Spread through air spaces (STAS) constitutes a novel invasive pattern in lung adenocarcinoma (LUAD), associated with tumor recurrence and diminished survival rates. However, large-scale STAS diagnosis in LUAD remains a labor-intensive endeavor, compounded by the propensity for oversight and misdiagnosis due to its distinctive pathological characteristics and morphological features. Consequently, there is a pressing clinical imperative to leverage deep learning models for STAS diagnosis. This study initially assembled histopathological images from STAS patients at the Second Xiangya Hospital and the Third Xiangya Hospital of Central South University, alongside the TCGA-LUAD cohort. Three senior pathologists conducted cross-verification annotations to construct the STAS-SXY, STAS-TXY, and STAS-TCGA datasets. We then propose a multi-pattern attention-aware multiple instance learning framework, named STAMP, to analyze and diagnose the presence of STAS across multi-center histopathology images. Specifically, the dual-branch architecture guides the model to learn STAS-associated pathological features from distinct semantic spaces. Transformer-based instance encoding and a multi-pattern attention aggregation modules dynamically selects regions closely associated with STAS pathology, suppressing irrelevant noise and enhancing the discriminative power of global representations. Moreover, a similarity regularization constraint prevents feature redundancy across branches, thereby improving overall diagnostic accuracy. Extensive experiments demonstrated that STAMP achieved competitive diagnostic results on STAS-SXY, STAS-TXY and STAS-TCGA, with AUCs of 0.8058, 0.8017, and 0.7928, respectively, surpassing the clinical level. The 10 open baseline results establish a benchmark for STAS diagnostic research and facilitate the future generalizability and clinical integration of computational pathology technologies. Dataset features and code are accessible at <https://github.com/panliangrui/IJCAI2026>.

[//github.com/panliangrui/IJCAI2026](https://github.com/panliangrui/IJCAI2026).

1 Introduction

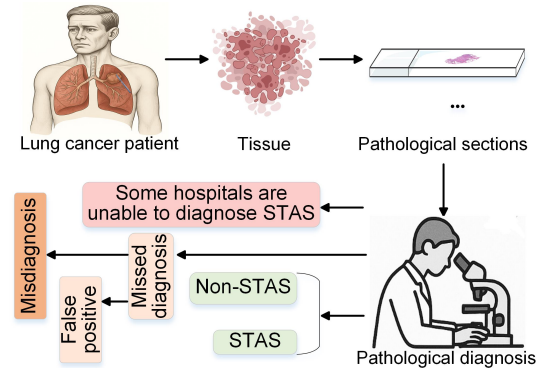


Figure 1: The challenges of STAS diagnosis include the unavailability of diagnostic results in some hospitals, a high risk of missed or incorrect diagnoses, and the labor-intensive nature of the diagnostic process.

Spread through air spaces (STAS) is a recently recognized pattern of invasion in lung cancer [Han *et al.*, 2021]. In 2013, Onozato *et al.* [Onozato *et al.*, 2013] first identified STAS using three-dimensional reconstruction, revealing that STAS consists of isolated tumor cells scattered in alveolar spaces, often separated from the primary tumor by several alveoli. In 2015, Kadota *et al.* formally described STAS in its current pathological context [Kadota *et al.*, 2015]. The World Health Organization later defined STAS as the presence of tumor cells such as micropapillary clusters, solid nests, or single cells within air spaces beyond the edge of the main tumor [Travis *et al.*, 2015], and officially incorporated it into the pathological classification of lung adenocarcinoma (LUAD) [Travis *et al.*, 2015]. STAS is frequently observed in LUAD, with reported incidences exceeding 30% in both domestic and international studies [Shi *et al.*, 2025]. It is significantly associated with aggressive clinicopathological features, such as pleural invasion, vascular invasion, larger tumor size, and advanced pathological stage [Han *et al.*, 2021; Toyokawa *et al.*, 2018]. Additionally, STAS is considered a

strong predictor of occult lymph node metastasis in stage IA LUAD [Vaghjani *et al.*, 2020]. Due to its association with poor prognosis and high recurrence risk after limited resection, STAS diagnosis has become a clinically essential task.

Histopathological images are considered the gold standard for diagnosing pulmonary STAS [Warth *et al.*, 2015]. As illustrated in Figure 1, the diagnosis of STAS heavily relies on pathologists performing detailed microscopic examinations of large quantities of histopathological slides. This process introduces substantial subjectivity into the diagnostic workflow. Moreover, hospitals with limited medical resources often lack the capacity to perform reliable STAS diagnosis [Villalba *et al.*, 2021]. In addition, STAS diagnosis is a labor-intensive task that is prone to both misdiagnosis and missed diagnoses. A major cause of false-positive STAS results stems from nonstandard tissue sampling and slide preparation procedures, which may lead to the artificial detachment of tumor cells from the primary lesion [Villalba *et al.*, 2021]. Intraoperative detection of STAS can guide surgical planning, including decisions between lobectomy and sublobar resection [Zhou *et al.*, 2022]. Postoperative STAS diagnosis also plays a crucial role in assisting oncologists in developing tailored treatment regimens. However, recent studies have reported that the AUC of manual STAS diagnosis ranges between 0.63 and 0.8, with an overall accuracy of approximately 0.7 [Villalba *et al.*, 2021; Zhou *et al.*, 2022]. Sensitivity tends to be relatively low, suggesting a high risk of missed diagnoses [Villalba *et al.*, 2021]. Therefore, there is an urgent clinical need for an objective, accurate, and efficient method to assist in the diagnosis of STAS.

With advances in deep learning, artificial intelligence (AI) has become increasingly integrated into pathological image analysis, yielding significant progress in cancer diagnosis and tissue grading [Coudray *et al.*, 2018; Li *et al.*, 2025]. In light of the limitations of conventional pathology, AI-based approaches for detecting STAS in histopathological images are emerging as a promising direction. As illustrated in Figure 2, STAS manifests in three histological forms: micropapillary clusters (lacking fibrovascular cores and occasionally forming ring-like structures within airways), solid nests (compact aggregates of tumor cells), and single tumor cells (individually scattered cells) [Kadota *et al.*, 2015]. The morphological heterogeneity and subtlety of STAS features pose substantial challenges for AI models, particularly in accurately identifying irregular or sparse lesions. Moreover, STAS diagnosis requires high-quality whole-slide images (WSIs) and expert annotations, which are costly, labor-intensive, and inherently subjective, leading to limited and imbalanced training data. Variations in staining protocols and institutional differences further hinder model generalization. Despite the strong performance of deep learning models in standard image classification tasks, their accuracy in STAS detection on WSIs remains insufficient for clinical deployment. These challenges highlight the urgent need for more robust diagnostic strategies to improve the consistency, reliability, and efficiency of STAS identification.

To address the above challenges, we first collected histopathological images of STAS patients from the Second

Xiangya Hospital and the Third Xiangya Hospital of Central South University, as well as the cancer genome atlas LUAD (TCGA.LUAD) dataset. Three senior pathologists were invited to perform cross-validation annotations, resulting in the construction of the STAS-SXY, STAS-TXY, and STAS-TCGA datasets. Each slide was annotated for the presence or absence of STAS and the type of dissemination focus. Given the complex pathological features of STAS in WSIs, we propose a multi-pattern attention-aware multiple instance learning (MIL) framework named STAMP for STAS diagnosis. The double-branch representation breaks the limitations of single-branch models by guiding the network to learn diverse lesion patterns from different semantic spaces, such as morphological features and contextual features, thereby enhancing feature diversity. Transformer-based instance encoding and a multi-pattern attention aggregation (MPAA) module dynamically selects lesion-relevant key regions, avoiding noise interference and improving the discriminative power of the global representation. The similarity regularization constraint learning strategy prevents pattern redundancy, ensuring that the learned lesion patterns are highly discriminative and complementary. Our contributions are as follows:

- We collaborated with three pathologists to construct and annotate three dedicated STAS datasets (STAS-SXY, STAS-TXY and STAS-TCGA), which comprising a total of 2,011 WSIs. These datasets offer valuable resources for advancing clinical research on STAS in LUAD.
- To address the pathological complexity of STAS, we introduce STAMP, a multi-pattern attention-aware MIL framework. Within the STAMP, the proposed MPAA module employs gated and content-aware dual-path attention at the feature level to dynamically highlight lesion-relevant regions, enabling the extraction of rich, heterogeneous features and enhancing WSI-level diagnostic accuracy.
- We conducted a extensive benchmark of 10 state-of-the-art MIL methods on the multi-center STAS datasets, providing comprehensive baseline results. Extensive ablation studies further validate the effectiveness and robustness of STAMP, highlighting its promise for advancing STAS-aware computational pathology.

2 Related Work

Histopathology Image Analysis Methods

Histopathological image classification has progressed through three major stages. Initially, traditional machine learning approaches relied on handcrafted features in conjunction with algorithms such as support vector machines, random forests, and k-nearest neighbors to classify WSIs [Cortes and Vapnik, 1995]. With the emergence of deep learning around 2015, convolutional neural networks became dominant, enabling accurate classification of tumor subtypes and invasion levels across various cancer types, including breast, prostate, and colorectal cancers, which often achieving area under the receiver operating characteristic curves (AUCs) exceeding 0.8 [Pan *et al.*, 2024b]. Since 2019,

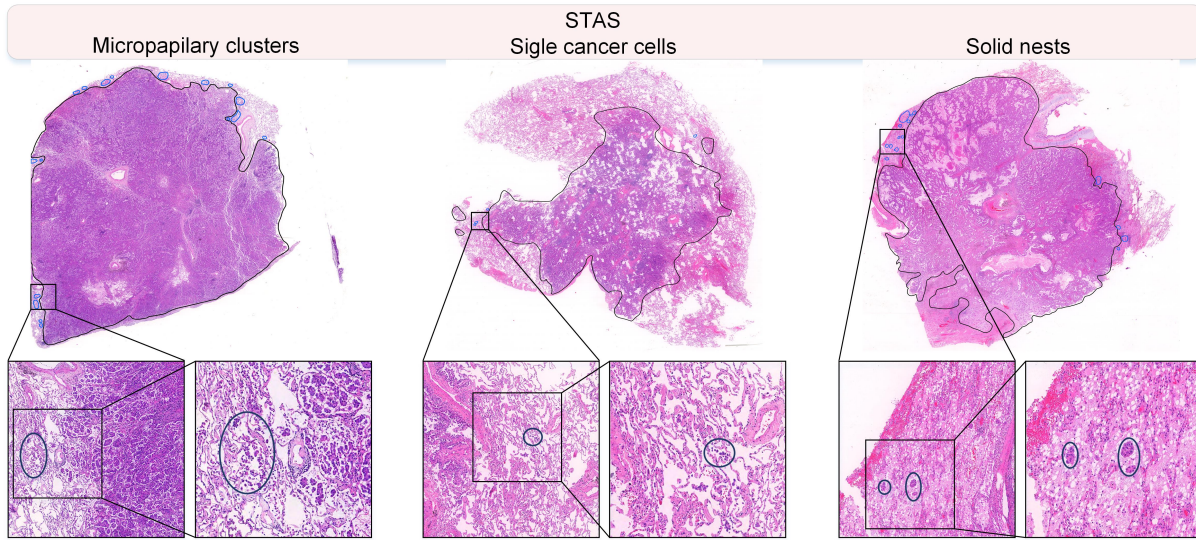


Figure 2: The three main pathological features of STAS in lung cancer histopathology images are as follows: STAS is mainly distributed outside the tumor body with the pathological features of micropapillary clusters, single cancer cells and solid nests.

MIL has become the prevailing paradigm for WSI analysis due to its compatibility with weakly supervised learning scenarios. MIL methods can be broadly categorized into instance-based and bag-based approaches. Instance-based methods assign predictions to individual instances and then infer the bag-level label, assuming that a few key instances are sufficient to determine the final classification. Representative methods include MaxPooling, ABMIL [Ilse *et al.*, 2018], DSMIL [Li *et al.*, 2021], ILRA [Xiang and Zhang, 2023], CLAM-SB, CLAM-MB [Lu *et al.*, 2021], and TransMIL [Shao *et al.*, 2021]. In contrast, bag-based methods operate directly at the bag level, using similarity metrics (e.g., Gaussian kernel distances) without requiring instance-level labels. Examples include MeanPooling and WIKG [Li *et al.*, 2024]. Recent advancements integrating MIL with other methods, such as DeeMILIP, which embeds domain probability priors [Al Hajj *et al.*, 2024], MixMIL, which combines generalized linear mixed models with MIL [Engelmann *et al.*, 2023], and cDP-MIL, which incorporates Bayesian nonparametrics for patch clustering [Chen *et al.*, 2024], have further advanced MIL’s capabilities, enhancing both its efficiency and robustness in addressing challenges like adversarial example scarcity and uncertainty estimation.

Diagnosis of STAS

In recent years, only a few studies have focused on technical exploration and model innovation for the diagnosis of STAS [Pan *et al.*, 2024a]. Lee *et al.* analyzed 115 LUAD WSIs, generating over 600,000 H&E patches to train a DL model via transfer learning, achieving an AUC of 0.738 for patch-level STAS detection [Lee *et al.*, 2023]. STAS-Net further improved semi-quantitative STAS identification across 489 WSIs, yielding a patch-level accuracy of 0.93 and a WSI-level AUC ranging from 0.72 to 0.78 [Feng *et al.*, 2024]. It also computed STAS density and spatial distribution parameters, enhancing detection in real-time diagnostic

workflows [Feng *et al.*, 2024]. More recently, the VERN model employed a Siamese graph encoder to capture spatial-topological features, achieved high AUC in internal validation and AUCs of 0.8275 and 0.8829 in frozen and paraffin-embedded test sections, respectively, demonstrating clinical-grade performance [Pan *et al.*, 2024a]. Additionally, Gong *et al.* proposed a multi-field channel transformer (MFCT), which attained an F1-Score of 0.734 in STAS identification [Gong *et al.*, 2025]. The SMILE framework introduced scale-aware MIL to adaptively select high-attention instances, diagnosing 251 and 319 STAS-positive WSIs on the CPTAC and TCGA datasets, respectively, which outperforming the average clinical AUC [Pan *et al.*, 2025]. Collectively, these studies demonstrate the promise of DL and MIL approaches in STAS diagnosis. However, existing methods still yield moderate AUCs, indicating a need for further advancement. To this end, we propose a multi-pattern attention-aware MIL framework, STAMP, to more effectively detect STAS from WSIs.

3 Method

3.1 Notations

In MIL, a WSI that has been partitioned into patches is represented as a bag $X = \{(x_1, y_1), \dots, (x_n, y_n)\}$, where x_i denotes the i -th patch (instance) in the bag, and y_i represents its instance-level label (which is typically unknown in weakly supervised settings). The number of instances obtained from the WSI is denoted by n . Each bag X is associated with a bag-level label $Y \in \{0, 1\}$, indicating whether the WSI contains a positive lesion region (e.g., STAS region). Under the weakly supervised MIL assumption, the relationship between the bag-level label and the instance-level labels is defined as follows:

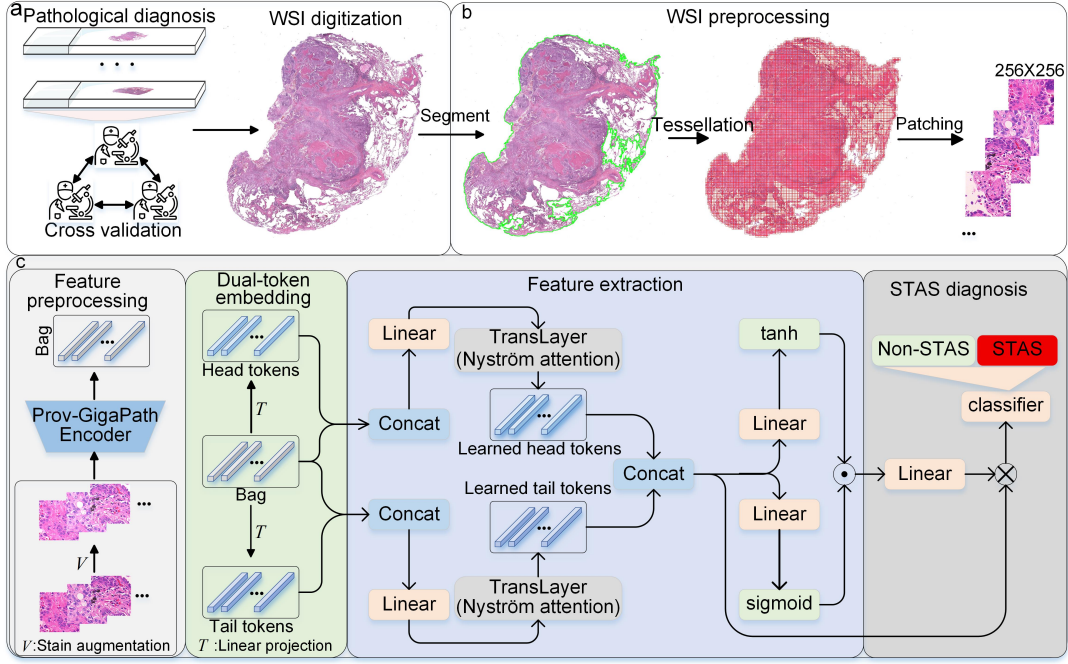


Figure 3: The complete workflow for diagnosing STAS from histopathological images. a): Annotation (cross-validation) of histopathological images from lung cancer patients and digitization of WSIs; b): Preprocessing of WSIs, including segmentation, tessellation, and patching; c): Preprocessing of WSI image features, dual-token embedding, feature extraction (including transformer-based instance encoding and MPAA modules), and STAS diagnosis (classification with regularized similarity loss).

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

In the STAS binary classification task, each instance x_i represents a patch from the histopathological image, and y_i denotes its corresponding label. The objective is to learn a classification model that can infer the label Y of a WSI based on its constituent patches, thereby identifying whether STAS-related pathological features are present in the WSI.

3.2 Multi-pattern Attention-aware Multiple Instance Learning

Overview

In this section, we introduce STAMP, a novel multi-pattern attention-aware MIL framework for weakly supervised WSI classification. Given a WSI represented as a bag of instances $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{n \times d}$, where each $x_i \in \mathbb{R}^d$ denotes a patch-level feature, STAMP aims to learn discriminative global lesion representations using a dual-token design and pattern-specific attention aggregation. The overall framework consists of four key components: (1) dual-token embedding, (2) Transformer-based instance encoding, (3) MPAA module, and (4) classification with regularized similarity loss.

3.3 Dual-token Embedding

To guide the learning of diverse lesion-aware representations, we introduce two sets of learnable tokens: a head token and a

tail token, each consisting of n_p vectors initialized as parameters:

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

These tokens are prepended to the original instance sequence and separately projected via two linear layers $\mathbf{W}_h, \mathbf{W}_t \in \mathbb{R}^{d \times L}$ to obtain the head and tail branches:

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

This double-branch design encourages the model to extract complementary semantic cues from distinct representation spaces.

3.4 Transformer-based Instance Encoding

Each token-augmented sequence $\mathbf{X}_h$ and $\mathbf{X}_t$ is independently fed into a lightweight Transformer encoder implemented via Nyström Attention [Xiong *et al.*, 2021] for efficiency and scalability. Specifically, each branch applies the following layer:

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

We retain the top n_p head and tail tokens from both branches and concatenate them as:

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

These aggregated tokens are expected to encode region-specific semantic patterns useful for final classification.

3.5 Multi-pattern Attention Aggregation

To compute a global bag-level representation, we design a pattern-level attention module that combines gated and content-based attention:

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

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

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

Here, $\odot$ denotes element-wise multiplication, and $\mathbf{W}_v, \mathbf{W}_u \in \mathbb{R}^{L \times D}$, $\mathbf{W}_a \in \mathbb{R}^{D \times 1}$ are trainable matrices. The attention weights $\mathbf{A} \in \mathbb{R}^{1 \times 2n_p}$ are used to aggregate token features into a global representation $\mathbf{M} \in \mathbb{R}^{1 \times L}$ via:

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

Where, $\cdot$ represents the product of vectors.

3.6 Classification and Similarity Regularization

The global feature $\mathbf{M}$ is passed to a linear classifier for bag-level prediction:

$$\hat{y} = \text{softmax}(\mathbf{W}_{\text{cls}} \mathbf{M}) \quad (10)$$

To enhance the diversity and orthogonality of the learned pattern tokens, we introduce a similarity regularization loss over the final token representations:

$$\mathcal{L}_{\text{sim}} = \mathbb{E}_{i \neq j} [\text{ReLU}(\cos(\mathbf{H}_i, \mathbf{H}_j))] \quad (11)$$

The total training loss combines standard cross-entropy with similarity regularization:

$$\mathcal{L}_{\text{total}} = \lambda \cdot \mathcal{L}_{\text{CE}} + (1 - \lambda) \cdot \mathcal{L}_{\text{sim}} \quad (12)$$

where λ is set to 0.9 by default.

4 Experiments and Results

4.1 Datasets

Given the propensity for misdiagnosis and underdiagnosis in STAS pathological assessment, three experienced pathologists employed a double-blind protocol with cross-validation to annotate each WSI for STAS status under microscopic examination, thereby ensuring accuracy and mitigating subjectivity, oversight and overdiagnosis. During the inclusion-exclusion process, WSIs containing mechanical artifacts were precluded by the aforementioned pathologists with extensive diagnostic expertise. Although false positives in STAS differentiation are encountered in clinical pathology practice, we utilized pertinent immunohistochemical markers to label suspicious lesions, facilitating further discrimination between genuine STAS-positive dissemination and histiocytic aggregates.

STAS-SXH: We included WSIs from the Xiangya Second Hospital of Central South University cohort as our internal training and validation dataset. Among the 12,169 patients who underwent lung nodule resection and were histologically confirmed with LUAD between April 2020 and December 2023, 206 individuals with a definitive diagnosis of STAS and 150 individuals without STAS were selected. From these selected cases, we collected a total of 1,290 WSIs

(approximately four slides per patient), along with associated clinical data and immunohistochemistry images (e.g., markers such as TTF-1, CK, CD68) [Yatabe *et al.*, 2002; Chistiakov *et al.*, 2017].

STAS-TXH: We included a cohort from the Third Xiangya Hospital of Central South University, consisting of 304 histopathology slides from 68 lung cancer patients diagnosed with STAS between 2022 and 2023, with each WSI annotated for the presence or absence of STAS.

STAS-TCGA: We collected 417 WSIs from 366 patients in the TCGA_LUAD cohort dataset based on the inclusion/exclusion method. All WSIs are accompanied by corresponding STAS labels.

4.2 Experimental Details

Data Preprocessing

The workflow for predicting STAS in WSIs using multi-center STAS datasets mainly consists of three steps (Figure 3): (1) pathological diagnosis of STAS, (2) WSI preprocessing, and (3) feature preprocessing of WSIs. Specifically, pathologists first screen eligible pathological slides for STAS diagnosis according to inclusion and exclusion criteria, followed by digitization of the slides. WSI preprocessing mainly involves selecting regions of interest (ROIs) of lung cancer tissue and dividing the WSIs into patches. We use Otsu’s automatic thresholding method to segment the tissue, background, and blurred areas in the WSIs [Liu and Yu, 2009]. Then, a sliding window strategy is applied to divide the $20\times$ magnification WSIs into 256×256 patches while recording the spatial coordinates of each patch. To reduce the impact of variations caused by different scanning devices and slide quality on model generalization, we apply a structure-preserving GAN trained on TCGA to perform stain augmentation on all patches of the WSIs [Wagner *et al.*, 2023]. Subsequently, all patches are processed using GigaPath for image feature extraction [Janowczyk *et al.*, 2019].

Training Details

We train our model using the Ranger optimizer, which combines RAdam and Lookahead for stable and efficient convergence. The initial learning rate is set to 1×10^{-4} , with a weight decay of 1×10^{-5} . A cosine annealing learning rate scheduler is adopted to gradually reduce the learning rate to a minimum value of 5×10^{-6} over the training process.

We train the model for 50 epochs with a batch size of 1. The input feature dimension d is set to 1536, and the hidden dimension L is set to 512. To ensure robust performance evaluation, we conduct experiments with random seeds ranging from 0 to 5 and report the averaged results across these experiments.

4.3 Results

To assess the performance of MIL approaches in diagnosing STAS, we benchmarked ten representative models across three curated STAS datasets: MaxPooling, MeanPooling, ABMIL [Ilse *et al.*, 2018], DSMIL [Li *et al.*, 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], and STAMP. For each MIL

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: Statistics of the results of diagnosing STAS in STAS-SXY, STAS-TXY, and STAS-TCGA based on 10 MIL methods.

method, we used accuracy (ACC), AUC, precision, recall, and F1-score to comprehensively evaluate the diagnostic performance of STAS. Below, we compare the performance of 10 MIL models on three STAS datasets.

In the STAS-SXY dataset, STAMP consistently outperformed all baselines, achieving the highest AUC (0.8058), F1-score (0.7376), and ACC (0.7449), demonstrating its superior capacity to detect STAS presence. Notably, ILRA, with an AUC of 0.7828, and CLAM-MB, attaining 0.7971, demonstrated strong performance, underscoring the effectiveness of multi-pattern attention aggregation and double branch framework. In contrast, simpler aggregation methods such as MaxPooling and MeanPooling showed limited effectiveness, underscoring their inability to capture critical histopathological cues.

On the STAS-TXY dataset, STAMP again achieved top performance, with the highest ACC (0.8124) and F1-score (0.7623). Although its AUC (0.8017) was slightly below that of CLAM-SB (0.8130), the results suggest comparable discriminative power. WIKG and CLAM-MB also performed competitively, indicating that bag-level modeling with clustering-aware attention is more robust for the diagnosis of STAS compared with instance-level methods such as DSMIL.

In the more challenging STAS-TCGA dataset, STAMP maintained strong performance, achieving the highest F1-score (0.6962) and AUC (0.7928) among all models. While CLAM-MB obtained the highest ACC (0.7464), its low F1-score (0.5337) reveals a trade-off between sensitivity and ACC, potentially caused by overfitting to dominant patterns. Traditional pooling strategies and early MIL models (e.g., ABMIL, DSMIL) suffered significant performance drops, with AUCs below 0.65 in most cases, reflecting their limitations in extracting complex pathological features of STAS.

4.4 Ablation Study

Influence of Pattern Number

To assess the effect of the number of latent patterns on model performance, we evaluated variants of STAMP with different numbers of patterns ($n_p \in \{1, 2, 3, 4, 5\}$) on the STAS-SXY dataset. As summarized in Table 2, the model achieves the highest overall performance when $n_p = 3$, indicating an optimal trade-off between representational capacity and generalization. Fewer patterns (e.g., $n_p = 2$) may lead to underfitting due to insufficient expressiveness, while a larger number (e.g., $n_p = 4, 5$) introduces redundancy without performance

gains. Based on these findings, we set $n_p = 3$ as the default configuration for all subsequent experiments.

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: Influence of Pattern Number on Model Performance.

Comparative Analysis of Single and Double-branches

To evaluate the effectiveness of the double-branch architecture, we conducted comparative experiments on STAS-SXY by varying the number of latent patterns in both single- and double-branch configurations. As shown in Figure 4, the double-branch (H,T) variant consistently outperforms its single-branch (H) counterpart across all metrics. For instance, with three latent patterns, the double-branch model achieves an accuracy of 0.7449, outperforming the single-branch version by 3.97%. This improvement can be attributed to the enhanced feature representation enabled by the double-branch design. While the single-branch model is limited to a fixed semantic level and receptive field, the double-branch architecture leverages parallel sub-networks to extract complementary information across multiple semantic granularities and receptive fields. This results in a significantly enriched feature space, thereby enhancing both representational capacity and discriminative power.

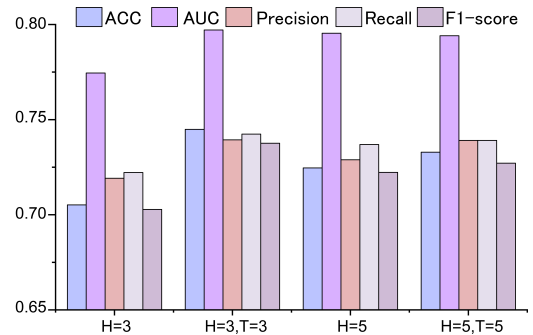


Figure 4: The impact of single-branch and double-branch structures on the performance of the STAMP.

The Impact of Token Embedding Strategies

We further investigate the impact of instance feature embedding strategies on model performance. Specifically, we design two embedding variants: pre-embedding and post-embedding. The corresponding formulations are defined as follows:

$$X_h = [W_h(X); T_h], \quad X_t = [W_t(X); T_t] \quad (13)$$

As shown in Figure 5, across different numbers of latent patterns, the post-attention embedding consistently outperforms the pre-attention approach. This suggests that allowing the region-aware learnable tokens to interact with instance features at a later stage is more beneficial for the model, possibly due to enhanced contextualization and better alignment with the learned semantics.

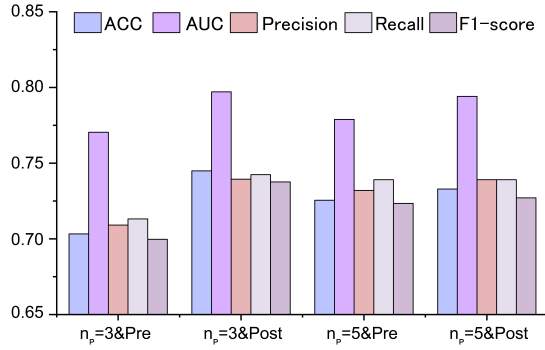


Figure 5: The impact of token embedding method on the STAMP performance. Pre and Post represent pre-embedding and post-embedding respectively.

The Impact of Multi-pattern Attention Aggregation Strategies

To further evaluate the impact of multi-pattern attention aggregation strategies on model performance, we compare two different fusion methods: feature-level aggregation (FA) and prediction-level aggregation (PA). In our proposed STAMP framework, FA is adopted as the default setting. The PA strategy can be formulated as follows:

$$M = \frac{1}{2} \left[\text{softmax}(W_a^h \cdot (\tanh(W_v^h H_h^{(1:n_p)}) \odot \sigma(W_u^h H_h^{(1:n_p)}))) \right. \\ \left. + \text{softmax}(W_a^t \cdot (\tanh(W_v^t H_t^{(1:n_p)}) \odot \sigma(W_u^t H_t^{(1:n_p)}))) \right] \quad (14)$$

As shown in Figure 6, FA consistently outperforms PA across different numbers of latent patterns. For instance, when $n_p = 3$, FA achieves an accuracy of 0.7449, surpassing PA's 0.7301. This result indicates that feature-level aggregation is more effective in enhancing the overall representational capacity and predictive performance of the model.

5 Discussion and Conclusion

In this study, we introduce three lung cancer STAS datasets to aid research on STAS diagnosis and propose STAMP, a multi-pattern attention-aware MIL framework, to address the

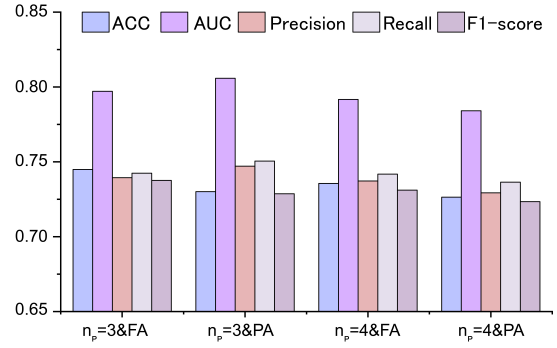


Figure 6: The impact of multi-pattern attention aggregation strategies on the STAMP performance.

complex pathological features of STAS. We benchmark 10 MIL models on the datasets for WSI-level STAS prediction, with STAMP achieving state-of-the-art AUC and F1-score on STAS-SXY and STAS-TCGA, and superior accuracy and F1-score on STAS-TXY. Unlike existing MIL methods, STAMP's MPAA module integrates gated and content-aware dual-path attention at the feature level, enabling adaptive focus on lesion-relevant regions for improved WSI diagnostics. Ablation studies confirm that the double-branch architecture enhances expressiveness and discriminability by expanding the feature space through parallel sub-networks for multi-scale extraction. Additionally, post-instance feature embedding facilitates early interaction between learnable region-aware tokens and instance-level features, promoting nuanced STAS pattern learning, while feature-level multimodal fusion leads to significant performance gains. These findings highlight STAMP's potential in clinical STAS detection. Despite these promising results, two main limitations remain. First, the STAS datasets were curated through expert judgment, a labor-intensive process, and future work will focus on collecting WSIs with common artifacts to train an automated quality control model for WSI screening. Second, while STAMP performs well on STAS-SXY and STAS-TCGA, its efficacy on STAS-TXY is reduced, indicating the need for better cross-center generalization. To address this, we plan to acquire more STAS-positive WSIs for training and bridge the gap between model-derived attention maps and expert annotations, a crucial direction for future research.

Acknowledgments

This work was supported by National Key R&D Program of China 2022YFC3400400, 2023YFC3503400; NSFC-FDCT Grants 62361166662; The Innovative Research Group Project of Hunan Province 2024JJ1002; Hunan Science and Technology Innovation Plan 2025ZYJ003; Key R&D Program of Hunan Province 2023GK2004, 2023SK2059, 2023SK2060; Top 10 Technical Key Project in Hunan Province 2023GK1010; Key Technologies R&D Program of Guangdong Province (2023B1111030004 to FFH).

Contribution Statement

Liangrui Pan and Xiaoyu Li contributed equally to this work. Chenchen Nie, Yaning Yang, and Shaoliang Peng are the corresponding authors.

References

- [Al Hajj *et al.*, 2024] Ghadi S Al Hajj, Aliaksandr Hubin, Chakravarthi Kanduri, Milena Pavlovic, Knut Dagestad Rand, Michael Widrich, Anne Schistad Solberg, Victor Greiff, Johan Pensar, Günter Klambauer, et al. Incorporating probabilistic domain knowledge into deep multiple instance learning. In *Forty-first International Conference on Machine Learning*, 2024.
- [Chen *et al.*, 2024] Yihang Chen, Tsai Hor Chan, Guosheng Yin, Yuming Jiang, and Lequan Yu. cdp-mil: Robust multiple instance learning via cascaded dirichlet process. In *European conference on computer vision*, pages 232–250. Springer, 2024.
- [Chistiakov *et al.*, 2017] Dmitry A Chistiakov, Murry C Killingsworth, Veronika A Myasoedova, Alexander N Orekhov, and Yuri V Bobryshev. Cd68/macrosialin: not just a histochemical marker. *Laboratory investigation*, 97(1):4–13, 2017.
- [Cortes and Vapnik, 1995] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [Coudray *et al.*, 2018] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L Moreira, Narges Razavian, and Aristotelis Tsirigos. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine*, 24(10):1559–1567, 2018.
- [Engelmann *et al.*, 2023] Jan P Engelmann, Alessandro Palma, Jakub M Tomczak, Fabian J Theis, and Francesco Paolo Casale. Mixed models with multiple instance learning. *arXiv preprint arXiv:2311.02455*, 2023.
- [Feng *et al.*, 2024] Yipeng Feng, Hanlin Ding, Xing Huang, Yijian Zhang, Mengyi Lu, Te Zhang, Hui Wang, Yuzhong Chen, Qixing Mao, Wenjie Xia, et al. Deep learning-based detection and semi-quantitative model for spread through air spaces (stas) in lung adenocarcinoma. *NPJ Precision Oncology*, 8(1):173, 2024.
- [Gong *et al.*, 2025] Haotian Gong, Jianing Xi, Cong Li, Haolong Zhang, Sisi Chen, Shuanlong Che, Ling Qi, and Guiying Zhang. Channel transformer based multi field-of-view model to detect tumor spread through air space in histopathological images. *Expert Systems with Applications*, 266:126125, 2025.
- [Han *et al.*, 2021] Yeon Bi Han, Hyojin Kim, Mari Mino-Kenudson, Sukki Cho, Hyun Jung Kwon, Ki Rim Lee, Soohyeon Kwon, Jeonghyo Lee, Kwhanmien Kim, Sanghoon Jheon, et al. Tumor spread through air spaces (stas): prognostic significance of grading in non-small cell lung cancer. *Modern Pathology*, 34(3):549–561, 2021.
- [Ilse *et al.*, 2018] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018.
- [Janowczyk *et al.*, 2019] Andrew Janowczyk, Ren Zuo, Hannah Gilmore, Michael Feldman, and Anant Madabhushi. Histoqc: an open-source quality control tool for digital pathology slides. *JCO clinical cancer informatics*, 3:1–7, 2019.
- [Kadota *et al.*, 2015] Kyuichi Kadota, Jun-ichi Nitadori, Camelia S Sima, Hideki Ujiie, Nabil P Rizk, David R Jones, Prasad S Adusumilli, and William D Travis. 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, 2015.
- [Lee *et al.*, 2023] S Lee, YB Han, CY Kim, J Lee, HJ Kwon, H Kim, and J-H Chung. Ma20. 04 predicting tumor spread through air space in the pulmonary adenocarcinoma slides using deep learning model. *Journal of Thoracic Oncology*, 18(11):S174, 2023.
- [Li *et al.*, 2021] Bin Li, Yin Li, and Kevin W Eliceiri. 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*, pages 14318–14328, 2021.
- [Li *et al.*, 2024] Jiawen Li, Yuxuan Chen, Hongbo Chu, Qiehe Sun, Tian Guan, Anjia Han, and Yonghong He. 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*, pages 11323–11332, 2024.
- [Li *et al.*, 2025] Yan Li, Liangrui Pan, Yijun Peng, Xiaoyu Li, Xiang Wang, Limeng Qu, Qiya Song, Qingchun Liang, and Shaoliang Peng. Application of deep learning-based multimodal fusion technology in cancer diagnosis: A survey. *Engineering Applications of Artificial Intelligence*, 143:109972, 2025.
- [Liu and Yu, 2009] Dongju Liu and Jian Yu. Otsu method and k-means. In *2009 Ninth International conference on hybrid intelligent systems*, volume 1, pages 344–349. IEEE, 2009.
- [Lu *et al.*, 2021] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.
- [Onozato *et al.*, 2013] Maristela L Onozato, Alexandra E Kovach, Beow Y Yeap, Vicente Morales-Oyarvide, Veronica E Klepeis, Swathi Tammireddy, Rebecca S Heist, Eugene J Mark, Dora Dias-Santagata, A John Iafrate, et al. 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, 2013.

- [Pan *et al.*, 2024a] Liangrui Pan, Qingchun Liang, Wenwu Zeng, Yijun Peng, Zhenyu Zhao, Yiyi Liang, Jiadi Luo, Xiang Wang, and Shaoliang Peng. Feature-interactive siamese graph encoder-based image analysis to predict stas from histopathology images in lung cancer. *NPJ Precision Oncology*, 8(1):285, 2024.
- [Pan *et al.*, 2024b] Liangrui Pan, Xiang Wang, Qingchun Liang, Jiandong Shang, Wenjuan Liu, Liwen Xu, and Shaoliang Peng. Deduce: Multi-head attention decoupled contrastive learning to discover cancer subtypes based on multi-omics data. *Computer Methods and Programs in Biomedicine*, 257:108478, 2024.
- [Pan *et al.*, 2025] Liangrui Pan, Xiaoyu Li, Yutao Dou, Qiya Song, Jiadi Luo, Qingchun Liang, and Shaoliang Peng. Smile: a scale-aware multiple instance learning method for multicenter stas lung cancer histopathology diagnosis. *arXiv preprint arXiv:2503.13799*, 2025.
- [Shao *et al.*, 2021] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147, 2021.
- [Shi *et al.*, 2025] Jinghan Shi, Kuan Xu, Xufeng Liu, Martin P Barr, Jeffrey B Velotta, Ninh M La-Beck, Chenxi Zhong, and Feng Yao. Clinical implication of tumor spread through air spaces in stage ia lung adenocarcinoma: prognostic impact and association with the international association for the study of lung cancer (iaslc) grade. *Translational Lung Cancer Research*, 14(4):1384, 2025.
- [Toyokawa *et al.*, 2018] Gouji Toyokawa, Yuichi Yamada, Tetsuzo Tagawa, Yuka Kozuma, Taichi Matsubara, Naoki Haratake, Shinkichi Takamori, Takaki Akamine, Yoshi-nao Oda, and Yoshihiko Maehara. Significance of spread through air spaces in resected pathological stage i lung adenocarcinoma. *The Annals of thoracic surgery*, 105(6):1655–1663, 2018.
- [Travis *et al.*, 2015] William D Travis, Elisabeth Brambilla, Allen P Burke, Alexander Marx, and Andrew G Nicholson. 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, 2015.
- [Vaghjiani *et al.*, 2020] Raj G Vaghjiani, Yusuke Takahashi, Takashi Eguchi, Shaohua Lu, Koji Kameda, Zachary Tano, Jordan Dozier, Kay See Tan, David R Jones, William D Travis, et al. 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, 2020.
- [Villalba *et al.*, 2021] Julian A Villalba, Angela R Shih, Treah May S Sayo, Keiko Kunitoki, Yin P Hung, Amy Ly, Marina Kem, Lida P Hariri, Ashok Muniappan, Henning A Gaissert, et al. 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, 2021.
- [Wagner *et al.*, 2023] Sophia J Wagner, Daniel Reisenbüchler, Nicholas P West, Jan Moritz Niehues, Jiefu Zhu, Sebastian Foersch, Gregory Patrick Veldhuizen, Philip Quirke, Heike I Grabsch, Piet A van den Brandt, et al. Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study. *Cancer cell*, 41(9):1650–1661, 2023.
- [Warth *et al.*, 2015] Arne Warth, Thomas Muley, Claudia A Kossakowski, Benjamin Goeppert, Peter Schirmacher, Hendrik Dienemann, and Wilko Weichert. Prognostic impact of intra-alveolar tumor spread in pulmonary adenocarcinoma. *The American journal of surgical pathology*, 39(6):793–801, 2015.
- [Xiang and Zhang, 2023] Jinxi Xiang and Jun Zhang. Exploring low-rank property in multiple instance learning for whole slide image classification. In *The Eleventh International Conference on Learning Representations*, 2023.
- [Xiong *et al.*, 2021] Yunyang Xiong, Zhanpeng Zeng, Rudrasis Chakraborty, Mingxing Tan, Glenn Fung, Yin Li, and Vikas Singh. Nyströmformer: A nyström-based algorithm for approximating self-attention. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 14138–14148, 2021.
- [Yatabe *et al.*, 2002] Yasushi Yatabe, Tetsuya Mitsudomi, and Takashi Takahashi. Ttf-1 expression in pulmonary adenocarcinomas. *The American journal of surgical pathology*, 26(6):767–773, 2002.
- [Zhou *et al.*, 2022] Fang Zhou, Julian A Villalba, Treah May S Sayo, Navneet Narula, Harvey Pass, Mari Mino-Kenudson, and Andre L Moreira. 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, 2022.