Deep Learning-Based Detection, Annotation and Staging of Breast Cancer Metastasis on Whole-Slide Histopathology Images of Lymph Nodes

Abdullah Tauqeer

A Thesis Submitted to
The Faculty of Graduate Studies
in Partial Fulfillment of The Requirements for The Degree of
Master of Applied Sciences

Graduate Program in
Electrical and Computer Engineering
York University
Toronto, Ontario

April 2025

Abstract

Breast cancer is one of the most frequently diagnosed malignancies and a leading cause of cancer-related mortality among women worldwide. Accurate detection and characterization of breast cancer lymph node metastasis are crucial for effective patient management. In this thesis, two novel deep learning frameworks have been developed and investigated to automate the analysis of hematoxylin and eosin (H&E) stained digital pathology whole-slide images (WSIs) of lymph nodes to streamline the breast cancer pathology workflow in the clinic.

First, a nuclei segmentation and classification model, *TexSegNet*, is introduced to extract detailed cellular-level information from the digital pathology images of biopsied lymph nodes. This model has been trained and evaluated using 7,904 histology images from the PanNuke dataset, which includes 205,343 annotated nuclei across 19 tissue types (2,351 images from breast tissue). TexSegNet leverages a hybrid encoder-decoder architecture, integrating multi-scale convolutions, nuclear texture extraction blocks, advanced attention mechanisms, and a feedback-driven classification branch. Trained on all tissue types of the PanNuke dataset and subsequently fine-tuned on its breast subset, TexSegNet has achieved an overall accuracy of 81.4±0.4% in detecting and classifying nuclei on the breast test set, substantially outperforming established benchmarks such as HoVer-Net, CellViT, and StarDist. Notably, TexSegNet has maintained consistently good performance across various cell types, including underrepresented ones, with F1-scores of 89.3±0.4%, 91.1±0.5%, 88.9±0.8%, and 84.3±0.3% in detecting and classifying neoplastic, epithelial, inflammatory, and connective cell nuclei, respectively.

Second, a selective neighborhood attention-based multiple instance learning framework (MIL) is proposed for automated detection, localization, and staging of lymph node metastases in breast cancer. The proposed framework leverages a dual-path feature extractor, incorporating both nuclei segmentation/classification outputs and transformer-based tissue features, alongside a dynamic attention mechanism that selects and emphasizes neighboring patches based on

similarity to the target patch. Trained and optimized on 269 WSIs (over 14 million image patches) from the CAMELYON16 dataset and evaluated on an independent test set of 129 WSIs, the model achieves 96.2±1.5% sensitivity, 95.3±2.4% precision, and 95.7±3.1% F1-score in patch-level classification, along with the area under the receiver operating characteristic curve (AUC) of 0.96±0.01 for slide-level classification. External validation on an out-of-distribution (OOD) subset of 30 annotated WSIs from the CAMELYON17 dataset (representing multiple institutions) demonstrates robust generalizability, with a patch-level F1-score of 87.0±1.8% and a slide-level AUC of 0.88±0.03. Furthermore, for 500 WSIs from 100 patients in the CAMELYON17 dataset, the framework demonstrates an excellent performance in patient-level pN-staging, with a quadratic-weighted Cohen's kappa of 0.94±0.02, indicating near-expert concordance in detecting and classifying the extent of nodal metastasis. Ablation analyses underscore the importance of incorporating nuclei-based features and selective neighborhood attention, with noticeable performance degradation observed when either element is removed. By integrating insights into histologic type, morphology and spatial heterogeneity at the cellular level with contextual information at the tissue level, the proposed framework effectively replicates key aspects of human pathological assessment and shows promise as a decision-support tool for metastatic breast cancer detection, annotation, and staging in the era of digital pathology.

Together, these two frameworks, TexSegNet for detailed nuclei segmentation and classification, and the selective neighborhood attention-based MIL model for targeting metastatic tumor detection and staging, form an integrated pipeline for breast cancer pathology assessment on standard H&E-stained slides. Together, they demonstrate strong potential to reduce pathologist workload, reduce subjectivity in diagnosis, and serve as processing tools in digital pathology to derive quantitative biomarkers for further analysis or prognostication. Future work will focus on expanding dataset diversity, enhancing model explainability, and integrating these algorithms into clinical workflows to support precision oncology care.

Originality Statement

This thesis has been written by Abdullah Tauqeer under the supervision of Dr. Ali Sadeghi-Naini and Dr. Amir Asif. Parts of the material presented in this thesis have been published or submitted for publication in peer-reviewed journals and conference proceedings, as listed below. The research presented in each publication has been performed by the principal author, and conceived, supervised, and/or supported by the senior authors.

- The material in **Chapter 3** has been presented in:

  A. Tauqeer, A. Asif, and A. Sadeghi-Naini. *TexSegNet: An Attention-Guided Feedback-Driven Texture-Aware Deep Learning Model for Nuclei Segmentation and Classification in Digital Pathology Images of Breast Tissues.* Accepted for presentation in 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Copenhagen, Denmark, 2025.

- The material in **Chapter 4** has been presented in:

  A. Tauqeer, A. Asif, and A. Sadeghi-Naini. *Detection, Localization, and Staging of Breast Cancer Lymph Node Metastasis in Digital Pathology Whole Slide Images Using Selective Neighborhood Attention-Based Deep Learning*. Submitted to Scientific Reports, Nature, 2025.

Acknowledgements

My supervisors, Dr. Ali Sadeghi-Naini and Dr. Amir Asif, gave me the freedom to explore various research methodologies and allowed me to direct this thesis project in a way that I was happy with. Dr. Ali Sadeghi-Naini generously allocated his time for revising and correcting my writings, refining ideas, and introducing me to medical terminologies that were new to me. Dr. Amir Asif brought a unique perspective with his expertise, and as my earliest acquaintance at York, he provided extensive and continuous support from the time he hired me as an international undergraduate research intern till the completion of my master's thesis. I would also like to thank Dr. Laleh Seyyed-Kalantari for serving as my committee member.

Moreover, I would like to thank my lab members Nasrin Sheibani, Nauman Bashir, and Maryam Parsian for being good friends and providing support beyond research.

Finally, I would like to thank my parents, especially my mom, for always being supportive of my decisions.

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

ABMIL / AMIL ..... Attention-based Multiple Instance Learning.
AJCC ..... American Joint Committee on Cancer.
AJI ..... Aggregated Jaccard Index.
AUC ..... Area Under the Curve.
BCE ..... Binary Cross-Entropy.
BN ..... Batch Normalization.
CAS-CNN ..... Context-Aware Stacked Convolutional Neural Network.
CD ..... Correctly Detected (nuclei).
CE ..... Cross-Entropy.
CLAM ..... Clustering-constrained Attention Multiple Instance Learning.
CNN ..... Convolutional Neural Network.
CT ..... Computed Tomography.
DCIS ..... Ductal Carcinoma in Situ.
DNA ..... Deoxyribonucleic Acid.
DTFD-MIL ..... A MIL-based model variant for tumor detection.
DSMIL ..... A variant of Multiple Instance Learning.
ED ..... Extra Detected (nuclei).
FN ..... False Negative.
FNA ..... Fine Needle Aspiration.
FP ..... False Positive.
FROC ..... Free-Response Receiver Operating Characteristic.
GAP ..... Global Average Pooling.
GNN ..... Graph Neural Network.
GPU ..... Graphics Processing Unit.
GGAT ..... Gated Graph Attention.
H&E ..... Hematoxylin and Eosin.
IHC ..... Immunohistochemistry.
ILC ..... Invasive Lobular Carcinoma.
IDC ..... Invasive Ductal Carcinoma.

ITCs..... Isolated Tumor Cells.
IoU ..... Intersection over Union.
LCIS..... Lobular Carcinoma in Situ.
MBConv..... Mobile Convolution module.
MaxViT.....Multi-Axis Vision Transformer
MLP ..... Multi-Layer Perceptron.
MIL ..... Multiple Instance Learning.
MHIM-MIL.....Masked Hard Instance Mining Multiple Instance Learning.
MSA..... Multi-Head Self-Attention.
MSE ..... Mean Squared Error.
MRI..... Magnetic Resonance Imaging.
NIC.....Neural Image Compression.
OOD.....Out-of-Distribution.
OTSU .....OTSU thresholding method.
R2T ..... Re-embedded Regional Transformer.
RNN ..... Recurrent Neural Network.
ReLU.....Rectified Linear Unit.
RGB ..... Red, Green, Blue.
SA .....Spatial Attention.
SAM..... Segment Anything Model.
SAM-MIL .....Spatially Aware Multiple Instance Learning.
SLIC..... Simple Linear Iterative Clustering.
SimCLR ..... a simple framework for contrastive learning
TIFF ..... Tag Image File Format
TN ..... True Negative.
TNM..... Tumor, Nodes, Metastasis.
TP.....True Positive.
TransMIL .....Transformer based Correlated Multiple Instance Learning
t-SNE..... t-distributed Stochastic Neighbor Embedding.
ViT ..... Vision Transformer.
WSI ..... Whole Slide Image.
WHO ..... World Health Organization.

Chapter 1

Introduction

1.1 Breast Cancer

Breast cancer is the most frequent cancer diagnosed in women worldwide. According to the WHO report released in 2020 [1], approximately 2.3 million new cases are identified annually worldwide (**Figure 1.1**), and the disease is responsible for an estimated 685,000 deaths each year. Despite significant advances in diagnostics and treatment, breast cancer remains the second leading cause of cancer-related mortality in women. [2].

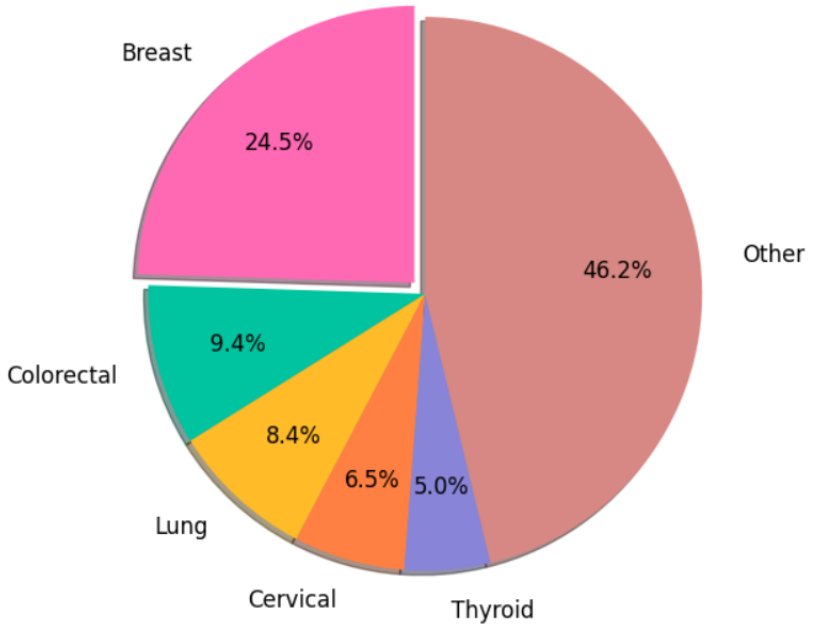

Figure 1.1: Distribution of cancer incidence among women globally. Breast cancer (24.5%) represents nearly one-quarter of all cancer diagnoses in females worldwide [1].

Breast tissue is composed of various components, each with its distinct functionality. These include lobules, which are glands responsible for secreting milk; ducts, which are tube-like structures that carry the secreted milk to the nipples; and stroma, which provides structural support in the tissue. [3]. DNA dictates the growth and division processes of cells in all the

aforementioned components. Mutations in genes responsible for cell growth regulation can lead to unprogrammed continuous cell division, thus forming tumor cells, which can be malignant (cancerous) or benign (non-cancerous abnormal cells) [4]. Breast cancer is broadly classified into invasive and non-invasive (in-situ) cancers. [5]. Invasive cancer can spread to other parts of the body. Based on its origin, it is further divided into two main types, i.e., invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) [5]. IDC originates from the ducts, spreads to neighboring tissues, and can metastasize to other body parts. It is also the most frequent cancer type, making up 70-80% of all breast cancer cases [6]. ILC originates from lobules and makes up 10-15% of invasive cancers [7]. Non-invasive breast cancers are confined to their place of origin and do not spread to surrounding tissues. The two main types of non-invasive breast cancer are ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) [8,9]. As the name suggests, DCIS and LCIS occur in the ducts and lobules, respectively. LCIS is not considered a true cancer; contrarily, DCIS is considered early-stage cancer. However, both types possess an added risk of becoming invasive cancer [9].

Early detection and accurate diagnosis are essential for improving survival rates in breast cancer [10]. The Canadian Task Force on Preventive Health Care recommends breast cancer screening every two to three years for women between the ages of 50 and 74 years [11]. Screening aims to detect cancer at an early stage when the tumors are small, less likely to have spread to lymph nodes and other organs, and more likely to be cured successfully [12].

X-ray mammography is the most common screening modality for early detection of breast cancer, although its efficacy varies based on factors such as age and breast density [13]. When an anomaly is discovered, further diagnostic investigation typically includes a biopsy to obtain tissue samples for histopathological evaluation. Additionally, imaging modalities such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) are utilized for lymph node localization and assessment before proceeding with biopsy or surgical intervention [13].

1.2 Lymph Nodes Metastasis

An important factor in assessing the severity of breast cancer is lymph node metastasis, which plays a key role in staging the disease [14,15]. Regional lymph node metastasis occurs when cancer cells from a primary tumor spread to the nearby lymph nodes, which are integral components of the body's lymphatic system. This process is crucial in breast cancer advancement, as lymph nodes frequently serve as initial sites for cancer metastatic spread. When cancer cells migrate to nearby lymph nodes, it typically indicates an advanced stage of the disease, often necessitating more aggressive treatment strategies [16]. The size and number of affected lymph nodes are key factors in determining treatment options and assessing the potential for distant metastases. Lymph node metastasis is one of the two most significant prognostic factors for breast cancer, along with tumor size [15]. Research shows a strong correlation between the number of positive lymph nodes and survival outcomes [17]. For example, the 5-year survival rate of patients with no positive lymph node is about 91% [18], compared to about 84% for those with 1-3 positive lymph nodes and around 73% for those with more than 3 positive lymph nodes [19].

1.3 Staging and pN-Stage

Breast cancer staging guidelines are used in the clinic to describe the characteristics of breast cancer in terms of the size and extent of the tumor, lymph node status, and whether the cancer has spread to other organs or not. The most common standard for breast cancer staging is the TNM system (**Figure 1.2**), which was introduced by the American Joint Committee on Cancer (AJCC) [20]. The acronym TNM stands for:

- Tumor (T): Describes the size and extent of the primary tumor. It is categorized from T0 to T4, where T0 indicates no evidence of a primary tumor and T4 signifies a large tumor that has invaded nearby tissues.

- Nodes (N): Refers to the extent of regional lymph node involvement. This is classified from N0 to N3, with N0 indicating no regional lymph node involvement and N3 denoting extensive cancer spread to regional lymph nodes.
- Metastasis (M): Indicates whether cancer has spread to distant parts of the body. M0 means no distant metastasis is present, while M1 indicates that distant metastasis is present.

The TNM system provides a standardized way to describe the severity and spread of cancer, which helps to guide treatment decisions and prognose the outcomes. Each type of cancer has its specific criteria for these categories, making the TNM system adaptable for different cancers.

In the scope of the current project, we will focus on the involvement of lymph nodes (i.e., "N"). The pathological staging of lymph nodes is detailed through "pN" classification that describes the degree of regional lymph node involvement determined through histopathology [20]. The following table describes the pN-staging classification criteria [20]. This staging requires evaluating multiple lymph nodes from a patient. In this context, isolated tumor cells (ITCs), micro-metastases, and macro-metastases are defined as tumor clusters smaller than 0.2 mm, larger than 0.2 mm but smaller than 2.0 mm, and larger than 2.0 mm, respectively.

- pN0: No micro-metastases, macro-metastases, or isolated tumor cells (ITCs) found.
- pN0(i+): Only isolated tumor cells (ITCs) were found.
- pN1mi: Micro-metastases are found, but no macro-metastases present.
- pN1: Metastases found in 1–3 lymph nodes, with at least one being a macro-metastasis.
- pN2: Metastases found in 4–9 lymph nodes, with at least one being a macro-metastasis.
- pN3: Metastases found in 10 or more lymph nodes, with at least one being a macro-metastasis.

Accurate pathological staging of lymph nodes is essential for determining an appropriate treatment plan and predicting patient prognosis.

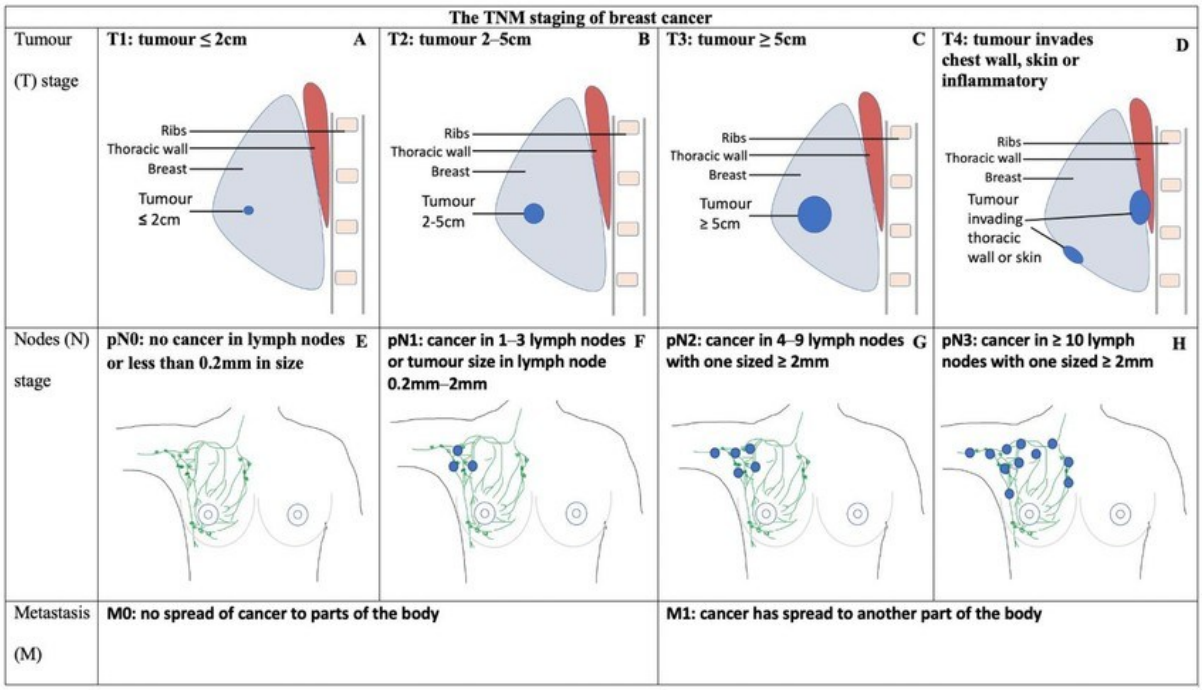

Figure 1.2: The TNM staging system for breast cancer categorizes tumor size (T1–T4), lymph node involvement (pN0–pN3), and metastasis (M0–M1) [21].

1.4 Tumor and Lymph Node Biopsy for Diagnosis and Staging

Biopsy procedures, including fine needle aspiration (FNA), core needle biopsy, and excisional biopsy, are used to acquire tissue samples from the suspicious breast lesions and regional lymph nodes that have been localized by imaging [22]. A schematic of these procedures is shown in **Figure 1.3**. Histopathological evaluation of biopsied tissue sections is essential for diagnosing a malignancy and determining the tumor histological type, grade, and lymph node involvement [23,24]. The choice of biopsy method depends on the size, location, and characteristics of the lesion. In needle biopsies, a hollow needle is used to extract tissue samples, but excisional biopsy involves a more invasive surgical procedure to remove an entire lymph node [24]. Biopsy samples may also be acquired from the breast tissue and lymph nodes resected during oncological surgeries to assess the extent of malignancy [24].

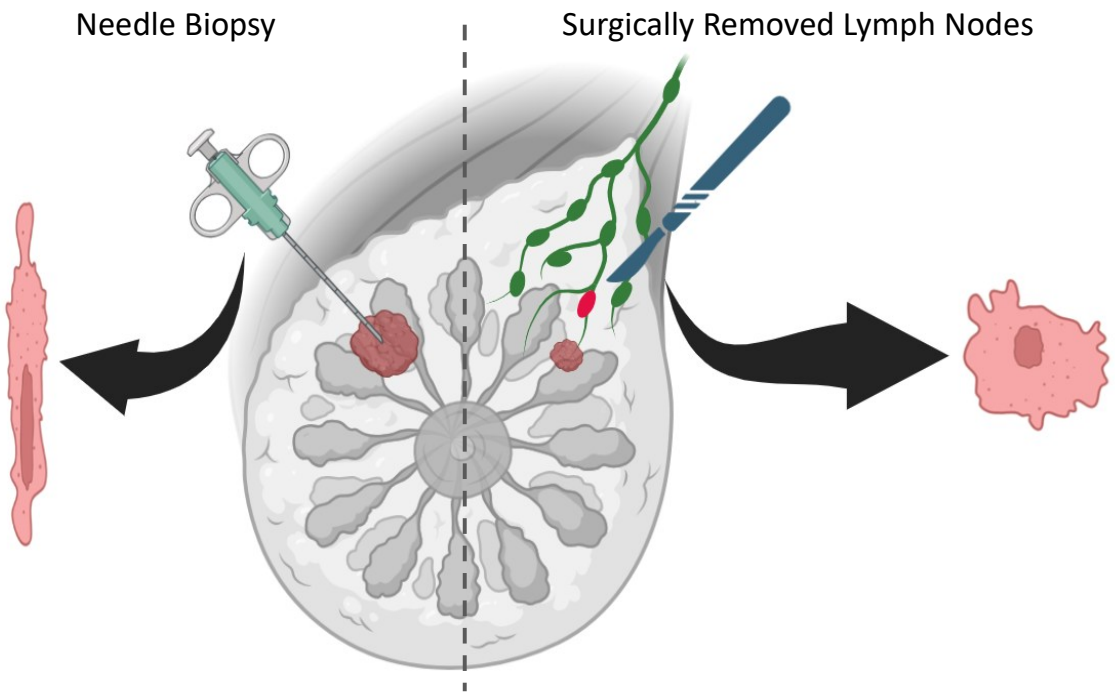

Figure 1.3: Schematic illustration of breast tumor biopsy and lymph node excision procedures for diagnosis and staging. A core needle biopsy is used to sample the primary tumor to diagnose and characterize the malignancy (left), while surgical removal of regional lymph nodes (right) is performed for pN-staging.

1.5 Histopathology Assessment

After a biopsy, the resected tissue samples are prepared for histopathology evaluation. The preparation procedure involves fixing the tissue samples in paraffin and slicing the fixed samples into sections with a thickness of 4-5 microns using a microtome. These tissue sections are subsequently mounted on glass slides and stained to enhance the visualization of cellular components under a microscope [20]. The most widely used staining method is hematoxylin and eosin (H&E) staining [22], which colors the cell nuclei blue and the cytoplasm and extracellular matrix pink. This contrast enables pathologists to identify anomalies in tissue and cellular structure and detect and characterize cancerous regions. Furthermore, immunohistochemistry (IHC) labeling can be employed to identify specific proteins and other markers, revealing additional insight into the presence of malignant cells [16]. After staining, the histopathology slide is examined under a microscope or digitized using a scanner to generate a pathology whole-slide image (WSI) which is a magnified representation of the original slide. The resolution and size of these digital images depend on the specifications and settings of the digital scanner being

used. The WSIs are usually stored in a pyramid structure at different magnifications, e.g., 1X, 10X, 20X, and 40X, for multiscale image analysis. The progress of slide digitization systems potentially helps pathologists to use digitized slides for a more accurate assessment. It also permits the application of computational image processing methods on digital pathology images for automating and streamlining histopathology analysis. **Figure 1.4** illustrates the digital histopathology workflow.

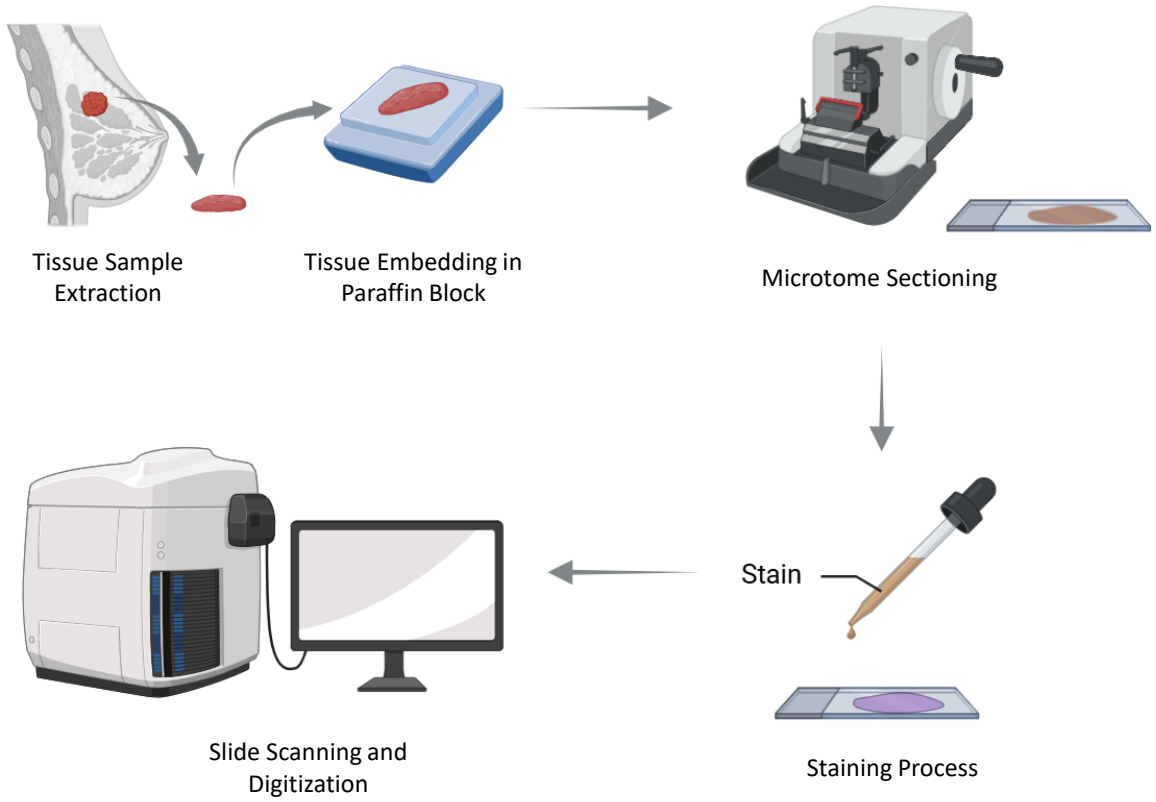

Figure 1.4: Histopathology workflow for breast cancer diagnosis: A tissue sample is extracted, embedded in a paraffin block, and sectioned using a microtome. The thin tissue sections are then stained (e.g., H&E staining) to enhance cellular structures. Finally, the stained slide is digitized using a scanner for detailed pathological analysis.

1.6 Problem Statement and Motivation

Accurate detection and annotation of malignant regions on pathology slides of biopsy samples are essential for breast cancer diagnosis, characterization, and staging [20], and subsequently for effective management of the disease. However, manual assessment of these slides by pathologists under the microscope or on WSIs is laborious, time-consuming, and resource-intensive [20]. Developing an automated platform for detecting and annotating malignant regions on WSIs can

streamline the pathology workflow and optimize clinical resources considerably. The platform can also be utilized to detect and annotate lymph node metastasis, quantify its extent, and identify the number of positive lymph nodes to automate pN-staging in the clinic [24]. Further, automated annotation of tumor regions on WSIs can facilitate the development of high-throughput data-driven algorithms for various clinical applications, including tumor grading, subtyping, and therapy response prediction using digital pathology WSIs [25]. An essential initial step in all these algorithms is to identify and annotate the regions of interest (ROIs), i.e., tumor regions, on a large set of WSIs for model training and evaluation. ROI annotation is also required on all input WSIs after the deployment of the developed models for subsequent analysis. These models can potentially be utilized as clinical decision support tools to guide treatment planning and tailor therapeutic strategies to the patient's specific needs. A personalized therapeutics paradigm facilitated by these data-driven models for breast cancer patients is expected to improve their overall survival and quality of life [10,26]. However, the successful clinical integration of such models depends on overcoming technical, regulatory, and interpretability challenges to ensure reliability and widespread adoption in medical practice.

1.7 Research Goal and Objectives

The overarching goal of this dissertation is to develop an automated, multi-level computational pathology pipeline that detects and annotates metastatic regions in digital pathology WSI of lymph node biopsy specimens, stages breast cancer according to standard guidelines (pN-staging), and provides detailed insights into nuclear morphology and tissue context. This integrated approach mirrors the comprehensive workflow of pathologists by combining both nuclei-level and tissue-level analyses. This goal is pursued through the following objectives:

1. **Robust Nuclei Segmentation and Classification:** Developing a dedicated deep-learning based framework that accurately segments and classifies individual nuclei in digital histopathology images. Nuclei provide essential diagnostic information (i.e., such as

histologic type, shape, size, chromatin distribution, and organizational patterns) that is critical for cancer diagnosis and characterization. Many existing models struggle with relatively high false positive/negative rates in detecting nuclei, over/under-segmentation in densely clustered regions, and misclassification of nuclei histologic types. By designing a dedicated nuclei segmentation and classification network, this objective aims to extract precise nuclear features that significantly enhance downstream tasks like tumor detection and characterization.

2. **Automated Tissue-Level Analysis to Differentiate between Normal and Tumor Patches:** Developing a deep-learning backbone for extracting distinctive tissue features from WSI patches for differentiating between normal (without metastasis) and tumor (with metastasis) patches, with acceptable generalizability across variations in staining and WSI scanner settings.
3. **Unified End-to-End Deep Learning Platform for Metastasis Detection, Localization, and pN-Staging:** Developing an end-to-end computational pathology platform that integrates nuclei-derived features with high-level tissue information. This objective leverage contextual information from the surrounding neighborhood of patches, filtering them based on relevance and similarity, to ensure that only the most informative and contextually consistent patches contribute to both patch-level classification and slide-level labeling. The combined feature representation will facilitate accurate delineation of tumor boundaries and capture subtle metastatic foci, enabling precise slide-level tumor mask generation and automated pN-staging at the patient level.
4. **Comprehensive Evaluation and Domain Generalization:** Rigorously evaluating the platform's performance in detecting and annotating lymph node metastasis and pN staging on publicly available multi-institutional datasets with out-of-distribution data for external validation. This evaluation will benchmark the proposed method against state-of-the-art

techniques, assess its domain generalization capabilities, and ensure robustness and scalability in diverse clinical environments.

1.8 Contributions

This thesis makes two primary contributions to computational pathology in breast cancer. First, it introduces a novel deep-learning model, TexSegNet, with an encoder–decoder architecture designed for high-precision nuclei segmentation and classification in histopathology images. TexSegNet employs multi-scale convolutions, spatial and channel recalibration, and multi-head self-attention mechanisms to capture fine-grained nuclear information and global tissue context, while incorporating mudslide boundary maps to refine segmentation in densely clustered or overlapping regions. Furthermore, the network integrates texture-aware feature extraction techniques that encode nuclear texture and morphological details critical for distinguishing cell types. A feedback mechanism further enhances the accuracy of nuclei classification by isolating and reintroducing segmented nuclear regions into the feature learning process. Benchmarking on the PanNuke dataset demonstrates that TexSegNet outperforms established models such as HoVer-Net, CellViT, and StarDist in both segmentation quality and classification accuracy, particularly for challenging and morphologically similar cell types.

The second and main contribution of this thesis project is the development of a novel selective neighborhood attention–based deep learning framework for breast cancer lymph node metastasis detection, localization and staging in digital pathology images. The proposed approach integrates the nuclei segmentation/classification outputs computed using TexSegNet with patch-level tissue embeddings, thereby capturing both detailed cellular-level information and global tissue context. A key innovation is the selective neighborhood attention mechanism, which adaptively focuses on the most relevant neighboring patches to refine patch-level classification. This fusion of nuclei-level information and tissue structure improves the sensitivity of tumor patch detection and the precision of tumor-normal tissue boundary delineation, resulting in generation of accurate

tumor probability maps that enable automated patient-level pN-staging. Experiments conducted on the CAMELYON16 and CAMELYON17 datasets demonstrate that the framework achieves state-of-the-art performance in patch-level, slide-level, and patient-level assessments, underscoring its potential for robust integration into clinical workflows.

Together, these contributions establish a comprehensive, end-to-end computational pathology pipeline that substantially enhances the accuracy and robustness of breast cancer lymph node analysis on H&E-stained WSIs. The developed platform offers a promising support tool in digital pathology that can potentially reduce pathologist workload, mitigate subjectivity in histopathology analysis, and improve diagnostic consistency in clinical practice.

1.9 Thesis Outline

This thesis is organized into five chapters as follows:

Chapter 1 introduces the thesis by providing relevant background on breast cancer, lymph node metastasis, TNM staging, and histopathology assessment for breast cancer diagnosis and staging. It describes the challenges clinicians face in manual pathology assessment of the large number of histopathology slides processed daily in the clinic. It identifies the need to automate routine processes such as tumor detection, annotation, and staging in WSIs, the tasks that are hindered by human subjectivity, staining variations, and the computational demands of gigapixel digital pathology images. The chapter also describes the research goal and objectives, contributions, and the thesis outline.

Chapter 2 provides a brief literature review, highlighting recent advances in computational pathology, particularly the roles of weakly supervised and multiple-instance learning methods, along with the growing importance of integrating spatial context and nuclei information to improve diagnostic accuracy.

Chapter 3 introduces TexSegNet, a texture-aware encoder–decoder network for nuclei segmentation and classification. By integrating multi-scale convolutions, attention modules, and a feedback loop that highlights segmented nuclei, TexSegNet effectively distinguishes various nuclear morphologies. The model is benchmarked against established methods such as HoVer-Net, CellViT, and StarDist using the PanNuke dataset, showing improved segmentation boundaries and reduced number of false positives/negatives, particularly in densely packed regions.

Chapter 4 centers on a selective-neighborhood attention–based framework for detecting, localizing, and staging breast cancer lymph node metastases in WSIs. This chapter details how WSIs are preprocessed, how both tissue-level and nuclei-level features are extracted, and how selective neighborhood attention refines patch-level and slide-level classifications. The method is evaluated on the CAMELYON16 and CAMELYON17 datasets, demonstrating robust performance across patch-level, slide-level, and patient-level metrics.

Chapter 5 concludes the thesis and outlines future work. It summarizes the key achievements of the thesis project, most notably TexSegNet's robust performance in nuclei segmentation and cell-type segmentation and the high accuracy of the selective-neighborhood attention framework in identifying metastatic tumor regions, and proposes directions for future research.

References

[1] World Health Organization. WHO Report on Cancer: Setting Priorities, Investing Wisely and Providing Care for All. Geneva: World Health Organization; 2020.

[2] World Health Organization. Breast Cancer. Fact Sheets. Geneva: World Health Organization; 2020. Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[3] Singletary SE. Anatomy and physiology of the breast. Cancer. 2003;97(S3):33–37. doi:10.1002/cncr.11141

[4] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–674. doi:10.1016/j.cell.2011.02.013

[5] Weigelt B, Geyer FC, Reis-Filho JS. Histological types of breast cancer: how special are they? Molecular Oncology. 2010;4(3):192–208. doi:10.1016/j.molonc.2010.04.004

[6] Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, et al., editors. SEER Cancer Statistics Review, 1975-2017. Bethesda, MD: National Cancer Institute; 2020. Available from: https://seer.cancer.gov/csr/1975_2017/

[7] Rakha EA, Ellis IO. Invasive lobular carcinoma of the breast: a special histological type with enduring challenges. Histopathology. 2017;70(1):7–18. doi:10.1111/his.13006

[8] Virnig BA, Tuttle TM, Shamliyan T, Kane RL. Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes. Journal of the National Cancer Institute. 2010;102(3):170–178. doi:10.1093/jnci/djp482

[9] Li CI, Malone KE, Daling JR. The risk of invasive breast cancer after ductal carcinoma in situ (DCIS): a population-based study. Cancer. 2005;103(5):900–909. doi:10.1002/cncr.20850

[10] Chan C, Law B, So W, Chow K, Waye MMY. Novel strategies on personalized medicine for breast cancer treatment: an update. International Journal of Molecular Sciences. 2017;18(11):2423. doi:10.3390/ijms18112423

[11] Tonelli M, Gorber SC, Joffres M, Dickinson J, Singh H, Lewin G, et al. Recommendations on screening for breast cancer in average-risk women aged 40–74 years. Canadian Medical Association Journal. 2011;183(17):1991–2001. doi:10.1503/cmaj.110334

[12] Shapiro S, Coleman EA, Broeders M, Codd MB, de Koning HJ, Fracheboud J, et al. Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. International Journal of Epidemiology. 1998;27(5):735–742. doi:10.1093/ije/27.5.735

[13] Lee CH, Dershaw DD, Kopans D, Evans P, Monsees B, Monticciolo D, et al. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. Journal of the American College of Radiology. 2010;7(1):18–27. doi:10.1016/j.jacr.2009.09.022

[14] Johns Hopkins Medicine. Breast Cancer Staging. Baltimore, MD: Johns Hopkins Medicine; n.d. Available from: https://www.hopkinsmedicine.org/breast_center/treatments_services/breast_cancer_staging.html

[15] Brierley JD, Gospodarowicz MK, Wittekind C, editors. TNM Classification of Malignant Tumours. 8th ed. Oxford: Wiley-Blackwell; 2017.

[16] Ping, J., Liu, W., Chen, Z., & Li, C. (2023). Lymph node metastases in breast cancer: Mechanisms and molecular imaging. Clinical Imaging, 94, 109985. https://doi.org/10.1016/j.clinimag.2023.109985

[17] Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. The Lancet Oncology. 2019;20(5):e261. doi:10.1016/S1470-2045(19)30154-8

[18] Chia S, Norris B, Speers C, Cheang MC, Gilks B, Gown AM, et al. Human epidermal growth factor receptor 2 overexpression as a prognostic factor in a large tissue microarray series of node-negative breast cancers. Journal of Clinical Oncology. 2008;26(35):5697–5704. doi:10.1200/JCO.2008.16.2096

[19] Cianfrocca M, Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. The Oncologist. 2004;9(6):606–616. doi:10.1634/theoncologist.9-6-606

[20] Amin MB, Edge SB, Greene FL, Byrd DR, Brookland RK, Washington MK, et al., editors. AJCC Cancer Staging Manual. 8th ed. New York: Springer; 2017

[21] Grace C, McKay M. Contemporary Issues in Postmastectomy Radiotherapy: A Brief Review. Journal of Clinical Medicine. 2024;13:7545. doi:10.3390/jcm13247545

[22] Fischer AH, Jacobson KA, Rose J, Zeller R. Hematoxylin and eosin staining of tissue and cell sections. Cold Spring Harbor Protocols. 2008;2008(5).prot4986. doi:10.1101/pdb.prot4986

[23] Jacobs L. Breast biopsy techniques. Current Opinion in Obstetrics and Gynecology. 1999;11(6):561–566. doi:10.1097/00001703-199912000-00012

[24] Qiu Y, Wu Y, Zeng L, He Z, Yu H, Yu Y, et al. Automated annotation of cancer pathology reports using deep learning. BMC Medical Informatics and Decision Making. 2021;21(Suppl 6):341. doi:10.1186/s12911-021-01693-w

[25] Saednia K, Tizhoosh HR, Thomas S, Salama P, Wang T, Shehata S. A hierarchical self-attention-guided deep learning framework to predict breast cancer response to chemotherapy using pretreatment tumor biopsies. Medical Physics. 2023;50(1):273–283. doi:10.1002/mp.15982

[26] Gambardella V, Tarazona N, Cejalvo JM, Lombardi P, Huerta M, Roselló S, et al. Personalized medicine: recent progress in cancer therapy. Cancers. 2020;12(4):1009. doi:10.3390/cancers12041009

Chapter 2

Literature Review

The digitalization of standard pathology slides presents a great opportunity for researchers to use existing image processing algorithms and develop new ones to improve the analysis of WSIs for different applications [1,2,3,4]. Historically, pathologists have relied on manual inspection of tissue samples under a microscope to diagnose and characterize the diseases. However, this manual approach is time-consuming and prone to subjective interpretation and human error [2]. The emergence of computational pathology can potentially address these limitations remarkably. By automating many of the histopathology assessment processes, computational pathology enables pathologists to concentrate on more complicated decision-making tasks. Computational pathology leverages quantitative features extracted from WSIs in conjunction with analytical or data-driven algorithms to assist in diagnosis [5,6,7], prognosis [8,9], identifying disease-specific biomarkers [10,11], and precise therapy planning for individual patients [12,13]. Despite its advantages, applying deep learning algorithms to WSI analysis introduces unique challenges, such as managing the vast size of these images and dealing with the weakly annotated nature of many disease datasets. These challenges underscore the need to develop specialized algorithms to analyze WSIs effectively for each specific application.

2.1 Weakly-Supervised Learning and Multiple-Instance Learning

A considerable portion of the recently proposed WSI classification algorithms fall under weakly supervised learning methods [11], in which labels are provided at the slide level rather than for individual ROIs within the slide. Multiple instance learning (MIL) has emerged as a critical technique for tackling this problem. In MIL, WSI classification typically involves several steps: A WSI is first divided into smaller patches (patching); these patches are then processed through

a feature extractor to generate feature vectors (feature extraction); finally, labels are assigned at the bag level (bag-level prediction). In this context, a WSI is considered a "bag" of instances (patches) [3,11,14], with each bag receiving a single label, a positive bag if it contains at least one positive instance. In contrast, a negative bag should contain only negative instances. **Figure 2.1** demonstrates the key concept of weakly supervised MIL, where a WSI is treated as a bag of patches, and the presence of at least one positive instance determines the slide-level label.

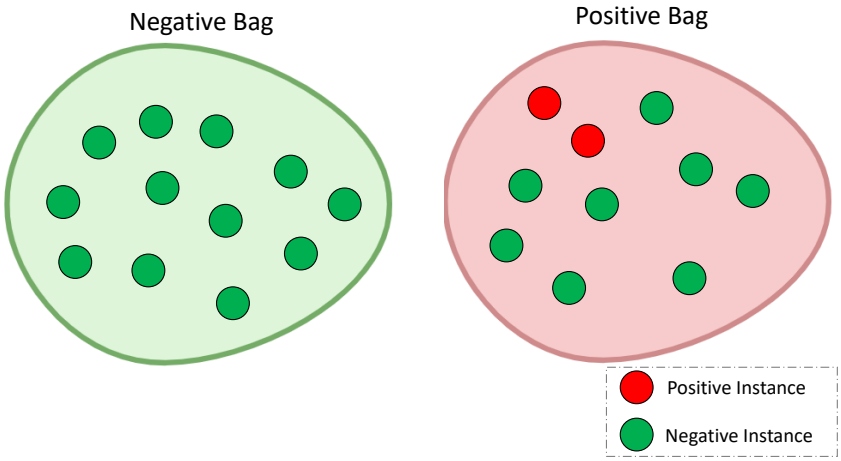

Figure 2.1: MIL principle for binary classification: the right bag is positive as it contains at least one positive instance (red), while the left bag is negative since all instances are negative (green).

Several MIL-based approaches have been developed for WSI. Campanella et al. proposed a method that uses a recurrent neural network (RNN) to aggregate patch-level predictions. They could achieve a high performance on three large datasets in different WSI classification applications including prostate cancer detection in biopsy specimens [3]. Campanella et al.'s approach involves an initial tissue segmentation step to discard non-tissue areas, followed by patch extraction and a full inference pass that ranks patches based on their predicted probabilities. The model is then updated using the top-ranked patches. Extensions like MIL-RNN further refine this process by updating model weights using the top-K ranked patches, thereby leveraging more instance-level information from each slide. Ilse et al. introduced attention-based multiple instance learning (AMIL), which identifies critical instances within a bag to receive higher weights in

decision-making (**Figure 2.2**), offering both high performance and interpretability [14]. Specifically, their attention mechanism computes attention scores for each patch embedding via a trainable transformation combined with a *softmax* normalization, and in some implementations a gating mechanism is added to capture more complex feature interactions. Lu et al. introduced CLAM, which gives an attention score to all patches in a WSI based on their relative importance in slide label prediction. Unlike traditional MIL, which uses max-pooling, CLAM uses attention to weigh the patch importance and their relationships. This allows it to achieve higher performance with comparatively fewer training samples. Additionally, CLAM extends the attention-based technique to handle multi-class classification tasks and incorporates instance-level clustering. This further improves data efficiency [11]. CLAM's integration of instance-level clustering helps to refine the learned feature space. Thus, these features are already a reduced abstract representation of instances/patches, making clustering robust in scenarios with limited training data. Most previous approaches focus on optimizing the aggregation and classification steps after feature extraction to improve performance. However, they neglect the refinement of feature vectors, resulting in sub-optimal outcomes. Recently, Tang et al. introduced the re-embedded regional transformer (R2T) module. This module emphasizes supervised fine-tuning of feature vectors before they are passed for downstream tasks, such as classification [9].

Despite these advancements, weakly-supervised models have a significant limitation: they only output slide-level classification labels. Accurate cancer staging, which requires annotation of tumor regions within each WSI to quantify their extent, cannot be fully addressed by these models.

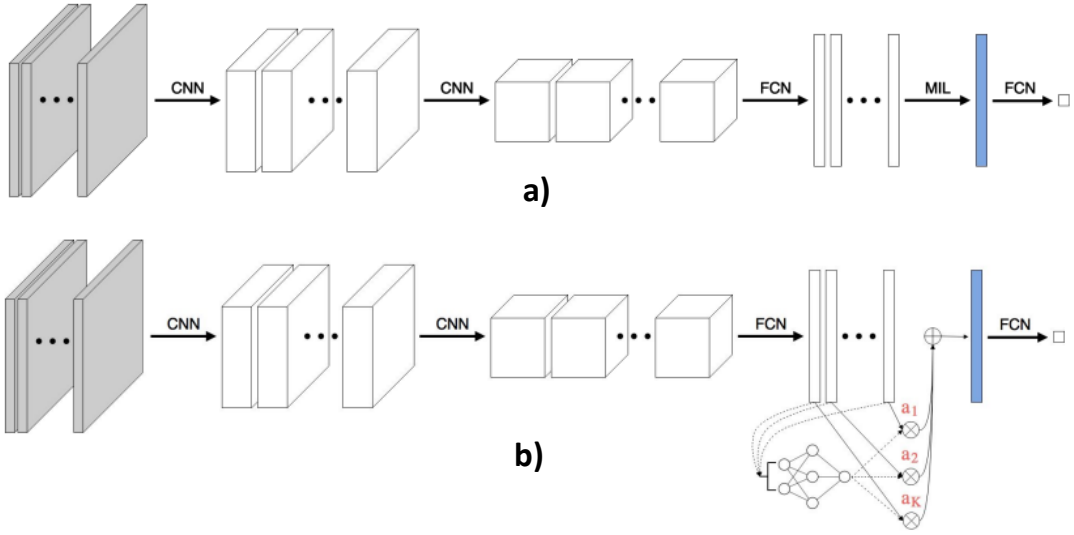

Figure 2.2: Comparison of MIL approaches for WSI classification. (a) Instance-based MIL aggregates patch features for slide-level prediction. (b) Attention-based pooling assigns weights (red) to instances before forming a weighted bag [14].

2.2 Spatial Context-Aware Approaches

Incorporating spatial context into histopathology image analysis has become a focus of recent research, recognizing the significance of capturing spatial relationships between different tissue regions and structures in a WSI for effective histopathology assessments. One approach for capturing both local and global information within a WSI is neural image compression (NIC), which compresses high-dimensional WSI patches into a lower-dimensional representation (**Figure 2.3**), facilitating the management of large-scale data while preserving critical spatial information [15]. For example, NIC compresses a high-dimensional WSI into a grid of lower-dimensional embeddings by encoding each patch using a trained encoder. A subsequent CNN is then employed to capture both local features and the spatial relationships among patches. However, NIC may struggle with slides containing small tumor lesions. Multi-scale techniques, such as multi-scale domain-adversarial MIL (MS-DA-MIL) proposed by Hashimoto et al., extract features at different magnification levels to capture various tissue structures and their spatial relationship across multiple scales [16]. Additionally, multi-scale approaches like MS-DA-MIL involve training separate feature extractors at different magnifications and integrating these multi-

scale features, often in conjunction with attention mechanisms to build a more robust slide-level representation. Some methods also incorporate domain adversarial normalization to address staining variability across different institutions. Moreover, for taking contextual information into effect Bejnordi et al. introduced a context-aware stacked convolutional neural network (CAS-CNN) for classification of breast histopathology images into normal/benign, DCIS, and IDC categories. Their approach first trains a CNN on high-resolution patches to capture cellular level information, and then stacks a second CNN that processes larger image patches to integrate global tissue context, thereby improving classification performance [17].

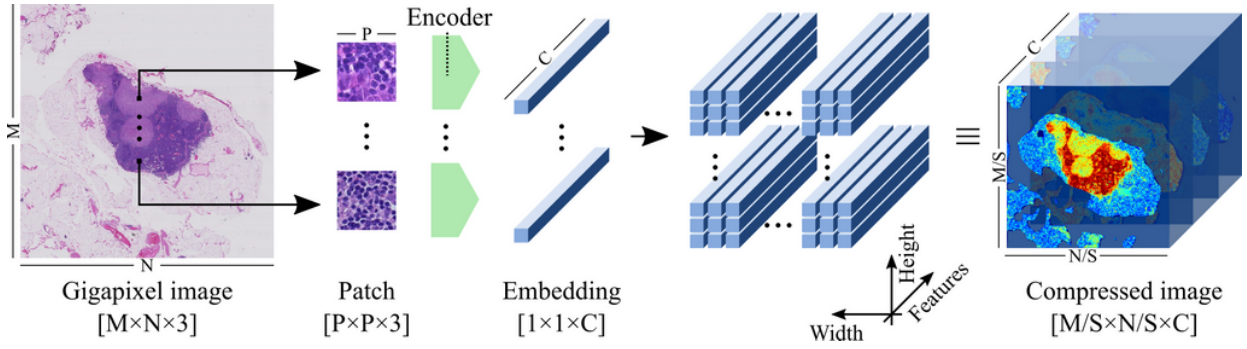

Figure 2.3: Overview of NIC framework [15].

Historically, the context-aware approaches have relied on either the output of a feature extractor [17] or nuclei segmentation model [18] for inferring context awareness. Feature extractor-based models target high-level features representing the tissue structures and overall cellular relationships. However, they usually overlook individual nuclei characteristics. On the other hand, in nuclei segmentation-based approaches, various hand-crafted features such as nuclei shape, size, and type are extracted from segmentation masks. These models take into consideration the spatial proximity of different nuclei and their interdependencies but lack information on the overall appearance, structural context, and large-range spatial relationships within the tissue. Neither of these approaches alone may represent the complex contextual interdependencies in tissues comprehensively. Current methods often fall short in this regard,

highlighting the need for improved approaches that balance and integrate critical aspects such as spatial features and cell nuclei information.

2.3 Feature Extraction Strategies

Feature extraction is an essential stage in the WSI analysis process that affects the quality and effectiveness of the classification models significantly. Specifically, the efficacy of feature extraction networks in deriving representative and distinctive features from the WSIs has a direct bearing on the accuracy and applicability of the data-driven models. Several methods have been compared in recent studies for improving the feature extraction modules [19]. According to research, fine-tuning pre-trained networks by transfer learning often leads to a better performance than using them as frozen feature extractors. Feature extraction performance may also be enhanced by incorporating multi-task learning, where models are pre-trained on many related tasks [19]. Self-supervised learning techniques like simple contrastive learning (SimCLR) [20], which uses contrastive learning to learn a representation of a dataset, have also demonstrated to improve feature extraction performance, frequently outperforming models pre-trained on ImageNet for pathology tasks [4]. Additional studies have confirmed that fine-tuning networks typically outperform fixed feature extraction, and that multi-task learning can further boost the discriminative power of the representations by leveraging auxiliary tasks [7]. Moreover, comparative analyses indicate that self-supervised approaches such as SimCLR yield superior feature representations for downstream tasks by learning invariant features from diverse augmentations [20], sometimes even surpassing the performance of models pre-trained on ImageNet [21]. To get more insight into the interdependencies of tissue regions and cell nuclei, a hierarchical graph representation learning framework was introduced by Li et al. [22]. Their framework models the relationships between tissue structures and cells to enhance classification accuracy. Moreover, by using hierarchical graph representations, each feature can be explicitly

linked to specific tissue components, permitting interpretations about the features driving the model's decisions.

Despite all these advancements, the current feature extraction techniques face limitations in generalization for new data and sometimes are prone to overfitting, especially with small datasets. As a result, these techniques may struggle with variations in staining and WSI scanner settings, leading to reduced performance when applied to new data from different sources.

2.4 Nuclei Segmentation and Classification

Cell nuclei segmentation is a crucial step in computational pathology, essential for accurately extracting and analyzing nuclear characteristics, and subsequently important for disease diagnosis and characterization [19,24]. It facilitates cell type classification, quantitative analysis, and high-throughput processing of histopathology images, serving as a foundation for advanced pathological assessments through machine learning models [25,26]. Recent developments in this area have substantially increased the accuracy and efficiency of nuclei detection and segmentation models. For example, HoVer-Net, trained and tested on the PanNuke dataset, is a model that utilizes distance maps to accurately segment the contours of nuclei. This model provides detailed information on nuclei's shape and type by concurrently performing intra-nucleus pixel-wise distance prediction, nuclei segmentation, and cell-type classification [7]. Another notable model is CellViT [23], a vision transformer-based network that also employs distance maps (similar to the approach undertaken by HoVer-Net) but incorporates the segment anything model (SAM) encoder [27]. This approach has achieved superior performance in multi-organ nuclei segmentation and classification [23], possibly due to its ability to capture long-range dependencies which is particularly beneficial for analyzing intricate tissue structures. These advanced segmentation techniques can be effectively integrated into graph-based networks for WSI analysis. For instance, Javed et al. have proposed a graph attention network that uses nuclei centroids as graph nodes, with edge weights based on nuclear characteristics and spatial proximity

[28]. This approach allows the model to capture the local cellular features and their spatial dependencies. Similarly, Wang et al. developed a hierarchical graph neural network that learns multi-scale tissue representations by creating multiple graph layers from nuclei segmentation findings [29]. These graph-based methods leverage nuclei segmentation/classification models to provide histopathological image representations that are more biologically relevant and informative, potentially enhancing the classification performance in applications like cancer grading and prognosis [18].

2.5 Summary

This chapter reviewed both the progress and persistent challenges in the digitization and computational analysis of WSIs in digital pathology. The transition from manual inspection toward automated computational pathology can drive notable improvements in disease diagnosis, prognosis, biomarker detection, and the planning of personalized therapies [3]. Nevertheless, challenges such as the enormous size of WSIs and the prevalence of weakly annotated datasets are a hurdle to effective analyses [3]. By automating histopathological assessments, computational pathology potentially minimizes human error [5,6,7]. It utilizes quantitative features extracted from WSIs to support diagnosis [2,19], prognosis [8], biomarker identification [11], and therapy planning [12].

In the context of WSI classification, weakly-supervised learning and MIL are extensively used [11]. MIL conceptualizes WSIs as “bags” of patches, where the presence of at least one positive instance determines the bag’s positive label [3,11,14]. Campanella et al. introduced an MIL-based strategy that leverages RNNs for aggregating patch-level information, thereby achieving high performance in prostate cancer detection [3]. Furthermore, approaches like AMIL and CLAM enhance both performance and interpretability by assigning weights to patch importance and incorporating instance-level clustering [11,14]. However, these models remain confined to slide-

level predictions and fall short of providing the precise tumor region annotations necessary for accurate cancer staging [9].

Spatial context-aware strategies, including NIC and multi-scale techniques like MS-DA-MIL, can capture both local and global spatial relationships within WSIs [15,16]. Additionally, stacked models such as CAS-CNN have been utilized to capture cellular level information as well as global tissue context to improve performance in applications like breast cancer pathology image classification [17]. Despite these advances, such methods often struggle to balance the detailed characteristics of local nuclei with the overarching tissue context [11,18].

Feature extraction is a critical component of WSI analysis. Techniques like fine-tuning pre-trained networks and employing self-supervised learning methods such as SimCLR have demonstrated to improve the model performance [4,19]. Moreover, frameworks based on multi-task learning and hierarchical graph representation learning further boost feature discriminability [7,22]. Nonetheless, these contemporary approaches continue to face challenges when generalizing to new datasets and dealing with variations in staining and scanner configurations [7,19].

Notable strides have also been made in nuclei segmentation and classification. Models like HoVer-Net and CellViT have achieved high accuracy in detecting and classifying nuclei [7,23]. These methods have been integrated into graph-based networks to capture local cellular features and spatial dependencies, thereby improving performance in applications such as cancer grading and prognosis [28,29].

In summary, although computational pathology has advanced remarkably, challenges remain in effectively integrating spatial context, refining feature extraction, and developing models capable of precise ROI annotations. Future research must address these limitations to further improve the accuracy and clinical applicability of automated WSI analysis.

References

[1] Li CI, Malone KE, Daling JR. The risk of invasive breast cancer after ductal carcinoma in situ (DCIS): a population-based study. Cancer. 2005;103(5):900–909. doi:10.1002/cncr.20850

[2] Pantanowitz L, Valenstein PN, Evans AJ, Kaplan KJ, Pfeifer JD, Wilbur DC, et al. Review of the current state of whole slide imaging in pathology. Journal of Pathology Informatics. 2011;2:36. doi:10.4103/2153-3539.83746

[3] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine. 2019;25(8):1301–1309. doi:10.1038/s41591-019-0508-1

[4] Zhuang X, Shen J, Tan J, Wong TY, Wang F. Self-supervised feature learning for 3D medical images by playing a Rubik's cube. In: Shen D, Liu T, Peters TM, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science, vol 11764. Springer; 2019. p. 420–428.

[5] American Joint Committee on Cancer. AJCC Cancer Staging Manual. 8th ed. New York: Springer; 2017.

[6] Wong B, Yi MY. Rethinking pre-trained feature extractor selection in multiple instance learning for whole slide image classification. arXiv preprint arXiv:2308.04245; 2023.

[7] Graham S, Vu QD, Raza SEA, Azam A, Tsang YW, Kwak JT, et al. HoVer-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical Image Analysis. 2019;58:101563. doi:10.1016/j.media.2019.101563

[8] Rakha EA, Ellis IO. Invasive lobular carcinoma of the breast: a special histological type with enduring challenges. Histopathology. 2017;70(1):7–18. doi:10.1111/his.13006

[9] Tang Z, Wang J, Li X, Gao J, Li L, Xie H, et al. Re-embedded Regional Transformer for Weakly Supervised Pathology Image Classification. arXiv preprint arXiv:2302.06104; 2023.

[10] Rao R, Euhus D, Mayo HG, Balch C. Axillary node interventions in breast cancer: a systematic review. JAMA. 2013;310(13):1385–1394. doi:10.1001/jama.2013.277804

[11] Lu MY, Chen TY, Williamson DFK, Zhao M, Shady M, Lipkova J, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering. 2021;5(6):555–570. doi:10.1038/s41551-020-00682-w

[12] World Health Organization. Breast Cancer. Fact Sheets. Geneva: World Health Organization; 2020. Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[13] Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. Medical Image Analysis. 2016;33:170–175. doi:10.1016/j.media.2016.06.037

[14] Ilse M, Tomczak JM, Welling M. Attention-based deep multiple instance learning. In: Dy J, Krause A, editors. Proceedings of the 35th International Conference on Machine Learning (ICML 2018). PMLR; 2018. p. 2127–2136.

[15] D. Tellez, G. Litjens, J. van der Laak and F. Ciompi, "Neural Image Compression for Gigapixel Histopathology Image Analysis.", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2021;43(2):567-578.

[16] Hashimoto K, Sato M, Tanaka T. Multi-scale domain-adversarial multiple instance learning for digital pathology. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI); 2020: 123–131. doi:10.1007/978-3-030-59716-3_14

[17] Bejnordi BE, Zuidhof G, et. all. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. J Med Imaging (Bellingham). 2017 Dec 14;4(4):044504. doi: 10.1117/1.JMI.4.4.044504.

[18] Chan TH, Cendra FJ, Ma L, Yin G, Yu L. Histopathology whole slide image analysis with heterogeneous graph representation learning. arXiv preprint arXiv:2301.07851; 2023.

[19] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical Image Analysis. 2017;42:60–88. doi:10.1016/j.media.2017.07.005

[20] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. E. (2020). A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning (Vol. 119, pp. 1597–1607). PMLR. https://proceedings.mlr.press/v119/chen20j.html

[21] Russakovsky, O., Deng, J., Su, H., et al. (2015). ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y.

[22] Li Y, Gu Y, Zhang H, Guo W, Lu C. Hierarchical graph representation learning for colorectal cancer grading. IEEE Transactions on Medical Imaging. 2022;41(12):3413–3425. doi:10.1109/TMI.2022.3194067

[23] Chen RJ, Chen C, Li Y, Chen TY, Trister AD, Krishnan RK, et al. CellViT: Vision transformers for precise cell segmentation and classification. Nature Methods. 2023;20(8):1026–1034. doi:10.1038/s41592-023-01916-4

[24] Xing F, Yang L. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. IEEE Reviews in Biomedical Engineering. 2016;9:234–263. doi:10.1109/RBME.2016.2572307

[25] Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. Nature Methods. 2019;16(12):1233–1246. doi:10.1038/s41592-019-0403-1

[26] Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. British Journal of Cancer. 2021;124(4):686–696. doi:10.1038/s41416-020-01122-x

[27] Kirillov A, Mintun E, Ravi N, et al. Segment Anything. arXiv. 2023 Apr 5; arXiv:2304.02643. doi:10.48550/arXiv.2304.02643

[28] Javed S, Mahmood A, Shah SAA, Kwon S. Cell graph convolutional neural networks for automated cancer diagnosis. IEEE Access. 2020;8:93774–93783. doi:10.1109/ACCESS.2020.2994907

[29] Wang Z, Yang X, Wang Y, Wang H, Zhang H, Wang J, et al. Weakly supervised pan-cancer segmentation and classification via hierarchical graph neural networks. Nature Communications. 2022;13(1):4619. doi:10.1038/s41467-022-32218-z

Chapter 3

TexSegNet: An Attention-Guided Feedback-Driven Texture-Aware Deep Learning Model for Nuclei Segmentation and Classification in Digital Pathology Images of Breast Tissues*

3.1 Introduction

Nuclei instance detection and segmentation are essential in digital pathology for accurate diagnosis and characterization of breast cancer, one of the most prevalent and deadly cancers worldwide [1]. Pathologists analyze nuclear attributes such as morphology, chromatin texture, size and mitotic activity to identify malignant cells, grade and subtype tumors, and evaluate their aggressiveness [2,3]. Classifying nuclei into histologic types offers insights into the tumor microenvironment, aiding in the evaluation of tumor heterogeneity, immune responses, and stromal interactions, all of which are crucial for determining treatment strategies and predicting patient outcomes [3]. Manual segmentation and classification of nuclei in digital pathology images are time-consuming, resource-intense, and prone to subjectivity, particularly when dealing with large datasets and subtle variations in nuclear morphology [4,5]. Automated nuclei instance segmentation and classification provide a scalable solution in computational pathology, improving diagnostic accuracy and supporting advanced research for biomarker discovery and development of targeted therapies [6].

* A version of the material presented in this chapter has been accepted for presentation in 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Copenhagen, Denmark, 2025.

In recent years, researchers have developed several methods to overcome traditional challenges in histopathological analysis across a range of tasks [7–9]. Deep learning, in particular, has had a significant impact for nuclei instance segmentation and classification, starting with the introduction of U-Net [10]. This framework, with its encoder-decoder architecture and skip connections, revolutionized biomedical image segmentation. Building upon this, advanced architectures such as ResNet [11] and vision transformer (ViT) [12] introduced sophisticated mechanisms for capturing intricate features and patterns, greatly enhancing model effectiveness beyond the capabilities of earlier techniques. The field has also progressed through the integration of multiple deep learning modalities. For example, combining convolutional neural networks (CNNs) with the U-Net structure or utilizing multi-branch networks like HoVer-Net [13] has improved both segmentation precision and robustness. More recently, models like CellViT [14], which leverage vision transformers for nuclei instance segmentation and classification, have achieved state-of-the-art results, demonstrating significant advancements in the accuracy and reliability of automated histopathological analysis. These ongoing innovations highlight the dynamic evolution of deep learning models, leading to effective and scalable solutions in digital pathology for breast cancer diagnostics and research.

Existing nuclei segmentation and classification models face critical limitations, including relatively high false positive and false negative rates in detecting nuclei, poor differentiation between cell types due to inadequate texture representation, and challenges in accurately delineating close or touching nuclei. To overcome these issues, this chapter introduces TexSegNet, a novel texture-aware feedback-driven framework for nuclei segmentation and classification in digital pathology images. This model leverages multi-scale convolutions, spatial and channel recalibration, and multi-head self-attention mechanisms to capture fine-grained nuclear texture

and morphological details, ensuring precise detection, segmentation, and classification of nuclei. To further enhance cell-type classification accuracy, TexSegNet employs a feedback mechanism that applies nuclei segmentation mask to the input pathology image to isolate the detected nuclei. This provides the classification branch with a focused view of regions to be classified. These isolated regions are then passed through convolutional layers, and the resulting features are reintegrated into the feature maps, enriching the network's ability to classify nuclei type. Additionally, TexSegNet incorporates mudslide maps [15] to mitigate challenges in contouring clustered nuclei through processing supplementary information on nuclei boundaries. By combining attention-guided texture extraction, feedback-enhanced feature refinement, and enhanced boundary delineation, TexSegNet could potentially offer a scalable solution for nuclei segmentation and classification in large digital pathology datasets.

3.2 Materials and Methods

3.2.1 Dataset

The model training and evaluation in this study were carried out using the PanNuke dataset [16], a comprehensive multi-organ digital pathology dataset with 7,904 histology images of size 256 × 256 pixels at 40× magnification. The dataset has been compiled from 2,000 whole slide images (WSIs) sourced from The Cancer Genome Atlas (TCGA) [17] and local hospitals, representing 19 distinct tissue types. The dataset contains 205,343 manually annotated nuclei, categorized into five classes: neoplastic, non-neoplastic epithelial, inflammatory (including lymphoid and macrophage), connective tissue, and other (including dead) cells.

The dataset was split into training (70%), validation (10%), and testing (20%) sets using a stratified random splitting to ensure a similar distribution of data samples from each tissue type across all

three sets, as presented in **Table 3.1**. The training strategy followed a two-stage process: initial training on all tissue types included in the PanNuke dataset to learn generalized morphological features and tissue structure in pathology images, followed by fine-tuning on the breast tissue subset. Throughout this process, strict separation of the test set was maintained to ensure independent evaluation. The experiments were repeated by training three models with different random splits of the training and validation data.

Table 3.1: Distribution of the histopathology images in the PanNuke dataset across 19 tissue types (Total). The dataset was split into the training (Train), validation (Val) and test sets using a stratified random splitting.

Tissue Types	Total	Train	Val	Test
Adrenal	437	314	35	88
Bile Duct	420	302	34	84
Bladder	146	104	12	30
Breast	2351	1692	188	471
Cervix	293	210	24	59
Colon	1440	1036	116	288
Esophagus	424	305	34	85
Head and Neck	384	276	31	77
Kidney	134	96	11	27
Liver	224	161	18	45
Lung	184	132	15	37
Ovarian	146	104	12	30
Pancreatic	195	140	16	39
Prostate	182	130	15	37
Skin	187	134	15	38
Stomach	146	104	12	30
Testis	196	140	16	40
Thyroid	226	162	18	46
Uterus	186	133	15	38

3.2.2 Oversampling and data augmentation

The PanNuke dataset exhibits class imbalances among different tissue types and nuclear categories. To address this, an oversampling technique [14] was utilized in which the sampling probabilities are adjusted based on the class frequencies in the dataset, thereby increasing the frequency of underrepresented categories by a factor of 2 to 4 during each training epoch. In addition, data augmentation was applied on the training data to enhance model generalization and mitigate potential overfitting. Specifically, each image underwent one or more of the following transformations with a 50% probability per transformation: random rotation (±15°), horizontal or vertical flipping, scaling within 0.8×–1.2×, Gaussian blurring (σ=1), additive Gaussian noise (±10 intensity units), color jittering (±10% in brightness/contrast/ saturation), superpixel segmentation (SLIC, region size=50) [18], zoom blur (1.1×), random cropping with 20% resizing, and elastic transformations (α=34, σ=4). These combined strategies yielded a more balanced and diverse dataset, strengthening the model's ability to generalize across variations in histopathological images.

3.2.3 Model Architecture

TexSegNet is a hybrid encoder–decoder architecture designed for nuclei segmentation and classification (**Figure 3.1**). Its encoder is based on the MaxViT backbone [19], pre-trained on ImageNet, with an input image size of 256 × 256 pixels × 3 channels. The encoder path processes the input images through four sequential stages, each containing eight MaxViT blocks. These blocks incorporate the mobile convolution (MBConv) module which applies squeeze-and-excitation (SE) attention, followed by local and global attention mechanisms to capture both localized details and broader contextual information.

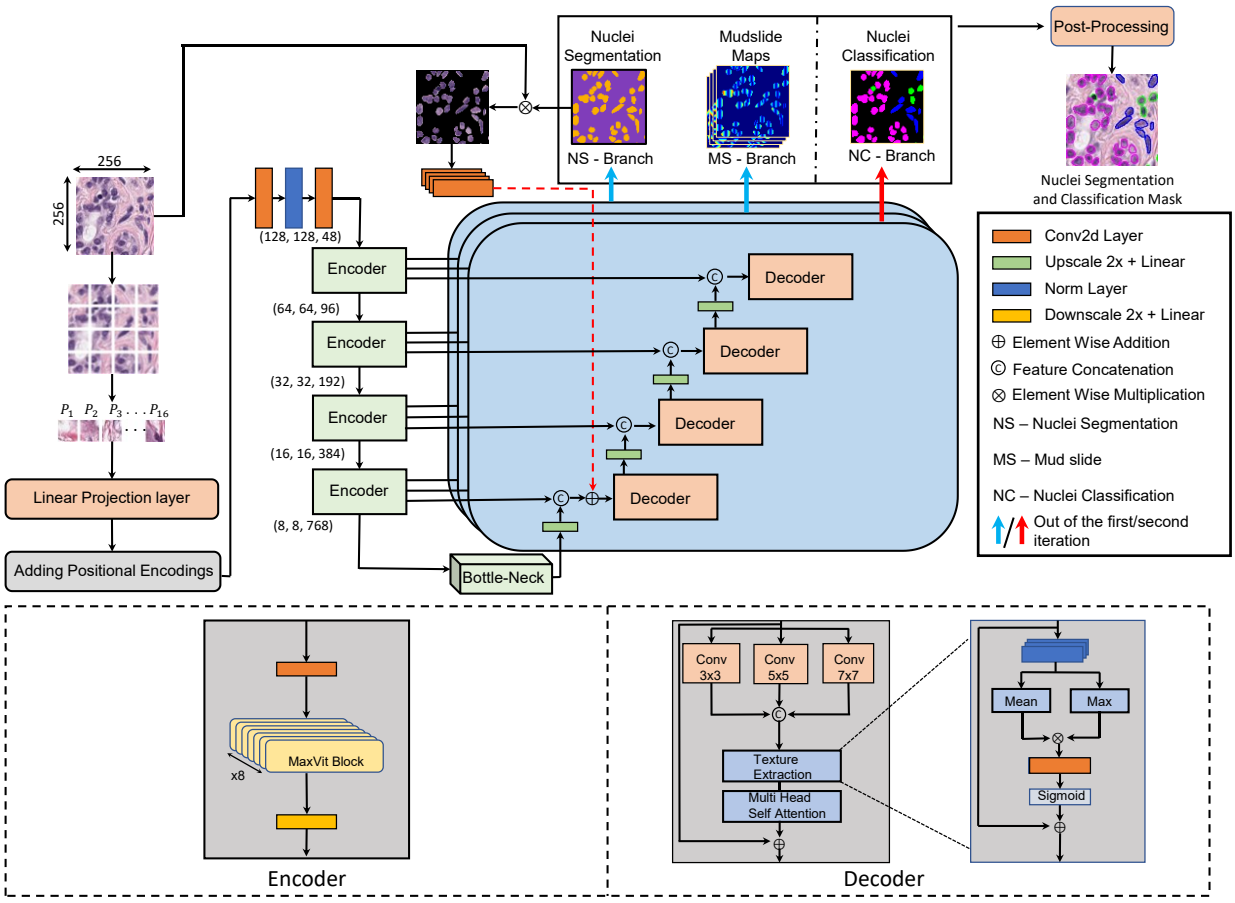

Figure 3.1: The proposed model architecture features a hybrid encoder-decoder design for nuclei instance segmentation and classification. Input images are divided into patches which are processed through a linear projection layer with positional encodings, followed by a four-stage MaxViT-based encoder. The bottleneck features are then passed through a series of four consecutive decoder blocks to generate segmentation/classification maps. The decoder includes three branches: one for generating nuclei segmentation probability maps, another for boundary maps, and a third one for nuclei classification probability maps. These maps are then refined through the post-processing block to generate the final nuclei segmentation and classification masks.

The patched input image I first goes through a linear projection layer, followed by the element-wise addition of positional encodings P:

$$X_0 = \text{Linear}(I) \oplus P. \tag{3.1}$$

The subsequent encoder stages $\{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4\}$ produce feature maps F_j:

$$F_j = \mathcal{E}_j(F_{j-1}), \quad j = 1, \ldots, 4, \tag{3.2}$$

with $F_0 = X_0$. The final stage outputs F_4 is the bottleneck representation $F_{\text{bottleneck}}$. Within each MaxViT block, convolution filters extract spatial details, while SE attention, applied after global average pooling, reweights channel responses to yield discriminative feature maps. This is followed by the block attention and grid attention modules to capture the short- and long-range dependencies, respectively.

The decoder path comprises three parallel branches for processing the encoded features, each designed for specific outputs: the nuclei segmentation probability map, boundary maps, and classification probability maps. Each decoder branch includes four decoder blocks $\{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4\}$. Before each decoder block, the feature maps from the previous block undergo bilinear upsampling to double their spatial dimensions, followed by a linear projection that reduces the number of channels by half. These feature maps are then concatenated with those from the encoder block at the corresponding upper level. In the decoder block, a parallel set of convolutions with 3×3, 5×5, and 7×7 kernel sizes is applied to the input features $F \in R^{H\times W\times C}$ to extract multi-scale feature maps:

$$F_{multi-scale} = \text{Concat}\big(\text{Conv}_{3\times3}(F), \text{Conv}_{5\times5}(F), \text{Conv}_{7\times7}(F)\big). \tag{3.3}$$

A 1×1 projection is used to map the number of channels in the concatenated features back to C, yielding $\tilde{F}_{\text{multi-scale}}$. These feature maps go through a texture extraction block that enhances texture feature representation by processing combined mean-pooled and max-pooled feature maps through a convolution layer followed by a sigmoid activation function, yielding $\tilde{F}^{\text{texture}}_{\text{multi-scale}}$. Multi-head self-attention (MSA) is then employed to capture long-range interactions:

$$\text{MSA}(Q, K, V) = \text{Softmax}\left(\frac{Q\,K^{\top}}{\sqrt{d_k}}\right)V, \tag{3.4}$$

where $Q, K, V \in R^{(HW)\times d_k}$ are reshaped projections of $\tilde{F}^{\text{texture}}_{\text{multi-scale}}$. Subsequently, the spatial attention (SA) module produces a location-wise gating mask that highlights the salient regions. The resulting spatially-attended feature map F_S undergoes SE attention to recalibrate the feature channels:

$$\mathrm{F}_{decoder} = \sigma\left(W_2\,\delta\big(W_1\,\mathrm{GAP}(F_{\mathrm{S}})\big)\right) \odot F_{\mathrm{S}}, \tag{3.5}$$

where σ is the sigmoid function, δ is *ReLU* function, $\mathrm{GAP}(\cdot)$ denotes global average pooling, and $\odot$ shows element-wise multiplication. This yields the final decoder output F_{decoder} which is refined both spatially and channel-wise to emphasize representation of important nuclear structures. In the decoder's first branch, a nuclei segmentation probability map $\hat{S}_{\text{seg}}$ is generated to distinguish nuclei regions from the background. In parallel, the second branch outputs four mudslide boundary maps $\hat{S}_{\text{boundary}}$ [15], which are used to refine the segmentation of densely packed nuclei.

The segmentation maps are binarized ($\hat{S}^{bin}_{seg}$) and multiplied element-wise with the input image to isolate nuclear regions:

$$\mathrm{I}_{nuclei} = \hat{S}^{bin}_{seg} \odot I. \tag{3.6}$$

The feature representation of the isolated nuclei is then projected to ($R^{16\times16\times384}$) for computational efficiency and fused with the input to the first decoder block. This fused representation is processed by the third decoder branch to generate classification probability maps $\hat{S}_{\text{class}}$ for the five nuclei categories: neoplastic, inflammatory, epithelial, connective tissue, and other cells. **Algorithm 3.1** summarizes the overall process of the decoder path in the proposed model. Maps $\hat{S}_{\text{seg}}$, $\hat{S}_{\text{boundary}}$, and $\hat{S}_{\text{class}}$ undergo a post-processing step where the mudslide maps are

applied to mitigate overlapping nuclei boundaries and generate refined nuclei segmentation and classification masks, along with class probabilities for each nucleus.

Algorithm 3.1: Pseudo Code Decoder with Feedback Mechanism for Nuclei-Type Mask Generation.

Input: $F^{(i)}$: Feature maps from encoder blocks; $F^{(0)} = X_0$

$F^{(i)}_{\{\text{map type}\}}$: Feature maps from decoder blocks; $F^{(0)}_{\{.\}} = F_{bottleneck}$

Output: $\hat{S}_{seg}, \{\hat{S}^j_{boundary}\}_{j=1}^{4}, \{\hat{S}^j_{class}\}_{j=1}^{5}$

1. **Decoder Block Definition:** ▷ Generic Decoder Block

$$\mathcal{D}(F) = \text{SE}\Big(\text{SA}\Big(\text{MSA}\big(\text{Conv}_{\text{multi}}(F)\big)\Big)\Big)$$

2. **Branch 1: Nuclei Segmentation**
3. for $i = 1$ to 4 do
4. $\quad F^{(i)}_{seg} \leftarrow \mathcal{D}\big(Concat(UP(F^{(i-1)}_{seg}), F^{(4-i)})\big)$
5. end for
6. $\hat{S}_{seg} \leftarrow \sigma\big(F^{(4)}_{\text{seg}}\big)$ ▷Nuclei Segmentation Probability Maps
7. **Branch 2: Mudslide Boundary Maps**
8. for i = 1 to 4 do
9. $\quad F^{(i)}_{boundary} \leftarrow \mathcal{D}\big(Concat(UP(F^{(i-1)}_{boundary}), F^{(4-i)})\big)$
10. end for
11. $\{\hat{S}^j_{boundary}\}_{j=1}^{4} \leftarrow \sigma\big(F^{(4)}_{boundary}\big)$ ▷ Boundary Maps
12. **Feedback Loop: Isolate Nuclei**
13. $\quad I_{nuclei} \leftarrow \hat{S}^{bin}_{seg} \odot I$

▷ Element-wise multiplication to isolate nuclei regions

14. **Branch 3: Nuclei-Type Classification**
15. $\quad F^{(1)}_{class} \leftarrow \mathcal{D}\big(Concat(UP(F^{(0)}_{class}), F^3) \oplus I_{nuclei}\big)$
16. for i = 2 to 4 do
17. $\quad F^{(i)}_{class} \leftarrow \mathcal{D}\big(Concat(UP(F^{(i-1)}_{class}), F^{(4-i)})\big)$
18. end for
19. $\quad \{\hat{S}^j_{class}\}_{j=1}^{5} \leftarrow \text{Softmax}\big(F^{(4)}_{class}\big)$

▷ Classification Probability Maps

Return: $\hat{S}_{seg}, \{\hat{S}^j_{boundary}\}_{j=1}^{4}, \{\hat{S}^j_{class}\}_{j=1}^{5}$

3.2.4 Model Training

TexSegNet utilizes a weighted combination of loss functions during training to achieve precise segmentation and classification of nuclei. The nuclei segmentation branch combines binary cross-entropy (BCE) and dice loss functions [20]; the boundary prediction branch employs mean squared error (MSE) and gradient loss functions [21]; and the classification branch uses cross-entropy (CE)

and focal Tversky loss functions [22]. The focal Tversky loss helps address class imbalance, while the Dice loss supports overlap accuracy. The total loss, expressed as $\mathcal{L}_{Total} = w_s\mathcal{L}_{\text{Segmentation}} + w_b\mathcal{L}_{\text{Boundary}} + w_c\mathcal{L}_{\text{Classification}}$, is regulated by loss weights as the hyperparameters to optimize performance across all branches.

Model training was conducted on NVIDIA L40 GPUs for 130 epochs, using the same dataset and experimental settings across all benchmarked models. The training process was divided into two phases: the first 30 epochs froze the encoder weights, training only the decoder, while the remaining epochs fine-tuned the entire model. The Adam optimizer was used with a learning rate of 10^{-3}, a cosine annealing schedule, and a batch size of 16. The loss weights were tuned to 0.5, 0.2, and 0.3 for the segmentation, boundary prediction, and classification components, respectively. Early stopping with a patience of 15 epochs monitored the validation loss, and the best-performing model checkpoint was retained for testing.

3.3 Evaluation Metrics

An evaluation scheme was developed to categorize each nucleus instance based on an intersection over union (IoU) threshold of 0.7, ensuring a comprehensive assessment of detection performance. Let $G = \{N_g\}$ be the set of ground truth nuclei and $P = \{\widehat{N}_p\}$ be the set of nuclei detected by the model. A categorization function $\Omega: G \times P \mapsto \{\text{CD,ED,ND}\}$ was defined to map pairs of ground truth and detected nuclei to one of three categories: correctly detected (CD), not detected (ND), or extra detected (ED) nuclei. Specifically, Ω is defined as follows:

$$\Omega\left(N_g, \widehat{N}_p\right) = \begin{cases} CD & if\ \ IoU\left(N_g, \widehat{N}_p\right) > 0.7 \\ ND & if\ \max\limits_{\widehat{N}_p \in P} \text{IoU}\left(N_g, \widehat{N}_p\right) \leq 0.7 \\ ED & if\ \max\limits_{N_g \in G} \text{IoU}\left(N_g, \widehat{N}_p\right) \leq 0.7 \end{cases} \tag{3.7}$$

This categorization scheme was introduced to account for both over-detection and under-detection errors, providing a more thorough evaluation of nuclei detection performance. It may be noted that even the correctly detected nuclei may still be misclassified. To account for this, a confusion matrix was generated using the predicted nuclei masks and ground truth masks. The rows of this matrix correspond to "extra detected" and ground-truth nucleus types, while the columns correspond to "not detected" and predicted nucleus types. This confusion matrix was then used to compute the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for each class [23], from which standard metrics such as precision, recall, F1-score, and accuracy were calculated.

3.4 Results

The performance of the proposed model for nuclei segmentation and classification were benchmarked against CellViT [14], HoVer-Net [13], and StarDist [24] using various evaluation metrics. The quantitative performance metrics are presented in **Table 3.2**, with the corresponding confusion metrices shown in **Figure 3.2**. The results in **Table 3.2** demonstrate that the proposed model has achieved the highest overall accuracy of 81.4%, outperforming CellViT (77.5%), HoVer-Net (75.2%), and StarDist (70.7%). Additionally, the proposed model has delivered the best F1-scores for neoplastic (89.3%), inflammatory (88.9%), connective (84.3%), and epithelial (91.1%) nuclei, with high precision and recall in each category. These results highlight the proposed model's effectiveness in detecting and classifying various types of nuclei, including underrepresented categories like inflammatory and connective tissue cells, while maintaining a decent balance between precision and recall.

The confusion matrices in **Figure 3.2** demonstrate that the proposed model, as compared to the other models, minimizes errors associated with the "extra detected" and "not detected" nuclei,

with higher true positive rates for all major classes. Specifically, the neoplastic, inflammatory, connective, and epithelial nuclei show fewer off-diagonal misclassifications as compared to the other models. CellViT demonstrates relatively good performance but exhibits higher false negative rates in nuclei detection and lower true positive rates in classification. HoVer-Net and StarDist display a broader distribution of errors, including more frequent misclassifications between similar-looking nucleus types and background artifacts (extra detected). The reduced number of false positives and false negatives in **Figure 3.2** illustrates the robustness of the proposed model in handling clustered nuclei with complex structures in histopathological images.

Table 3.2: Performance of the proposed model (TexSegNet) compared to CellViT, HoVer-Net, and StarDist in detecting and classifying different types of nuclei on the breast test set. All values are accompanied by standard deviations across the three repeated experiments (mean ± std).

Class	**Metric**	**Proposed Model (%)**	**CellVit (%)**	**HoVer-Net (%)**	**StarDist (%)**
Neoplastic	Precision	**93.5 ± 0.2**	93.0 ± 0.1	86.0 ± 0.4	86.1 ± 0.4
	Recall	**85.4 ± 0.8**	81.6 ± 1.2	83.1 ± 0.5	79.4 ± 1.3
	F1-score	**89.3 ± 0.4**	86.9 ± 0.7	84.5 ± 0.4	82.6 ± 0.7
Inflammatory	Precision	**90.0 ± 0.8**	88.7 ± 0.9	83.7 ± 1.8	80.2 ± 0.9
	Recall	**87.7 ± 1.0**	83.6 ± 2.1	84.6 ± 0.6	79.8 ± 1.9
	F1-score	**88.9 ± 0.8**	86.1 ± 1.4	84.1 ± 0.8	80.0 ± 1.3
Connective	Precision	**91.9 ± 0.2**	89.8 ± 0.3	86.7 ± 0.7	84.2 ± 1.0
	Recall	**77.9 ± 0.6**	73.0 ± 2.0	75.6 ± 0.9	70.5 ± 3.0
	F1-score	**84.3 ± 0.3**	80.5 ± 1.4	80.8 ± 0.8	76.7 ± 2.1
Epithelial	Precision	**97.6 ± 0.2**	97.4 ± 0.1	95.8 ± 0.5	94.7 ± 0.7
	Recall	**85.4 ± 0.7**	81.9 ± 2.3	82.8 ± 0.3	76.1 ± 3.0
	F1-score	**91.1 ± 0.5**	88.9 ± 1.4	88.8 ± 0.1	84.4 ± 2.1
Other	Precision	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0
	Recall	66.7 ± 0.0	66.7 ± 0.0	66.7 ± 0.0	66.7 ± 0.0
	F1-score	80.0 ± 0.0	80.0 ± 0.0	80.0 ± 0.0	80.0 ± 0.0
All	Accuracy	**81.4 ± 0.4**	77.5 ± 0.5	75.2 ± 0.2	70.7 ± 1.6

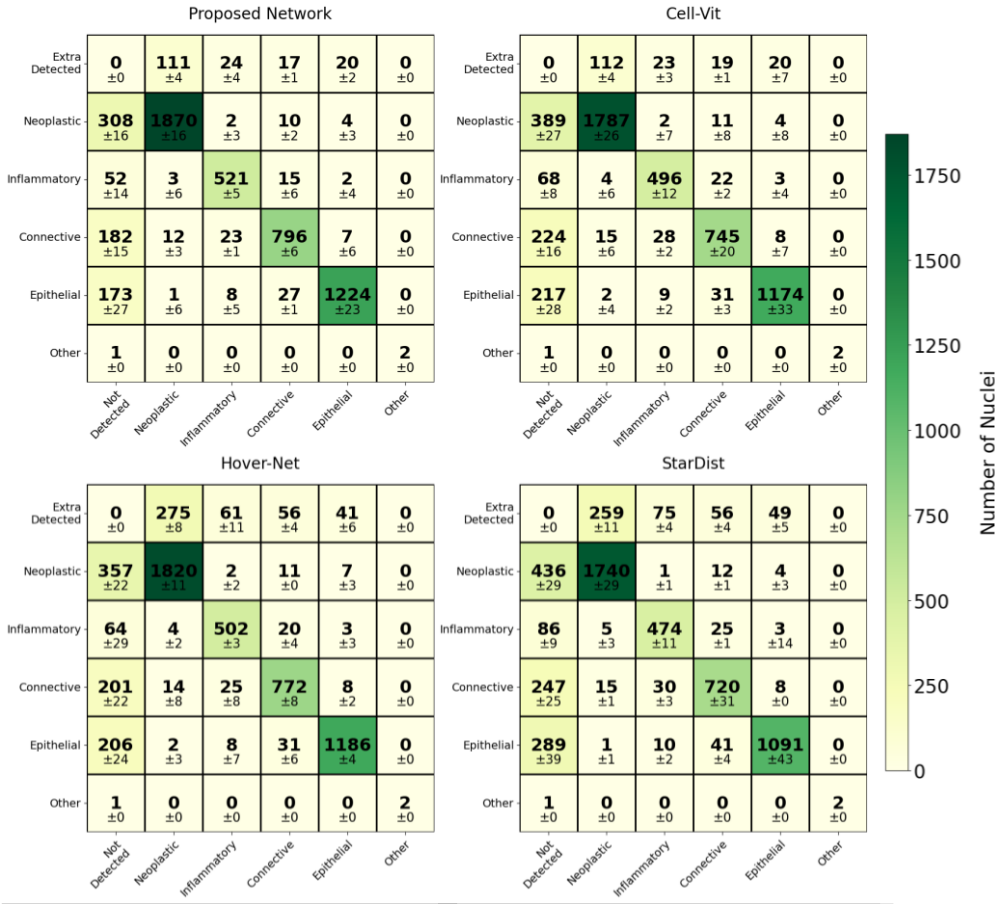

Figure 3.2: Confusion matrices illustrating per-class detection and classification counts for the four benchmarked models. Rows represent ground-truth classes, and columns represent predicted classes. The "Extra Detected" row represents model predictions that do not match any ground-truth nuclei (false positives), while the "Not Detected" column indicates ground-truth nuclei missed by the model (false negatives).

Figure 3.3 visualizes the nuclei segmentation and classification masks generated by different models for a representative histopathology image. The qualitative results in this figure demonstrate the proposed model's ability to accurately delineate nuclear boundaries and reduce false negative/ positive rates. Its boundary refinement facilitates clear segmentation even in densely packed regions, while the texture-focused decoder, coupled with the feedback mechanism, minimizes the misclassification of detected nuclei. Compared to CellViT, HoVer-Net, and StarDist, the proposed

model captures finer morphological details, achieving better segmentation precision and reducing errors in distinguishing touching nuclei, in addition to minimizing nuclei misclassifications.

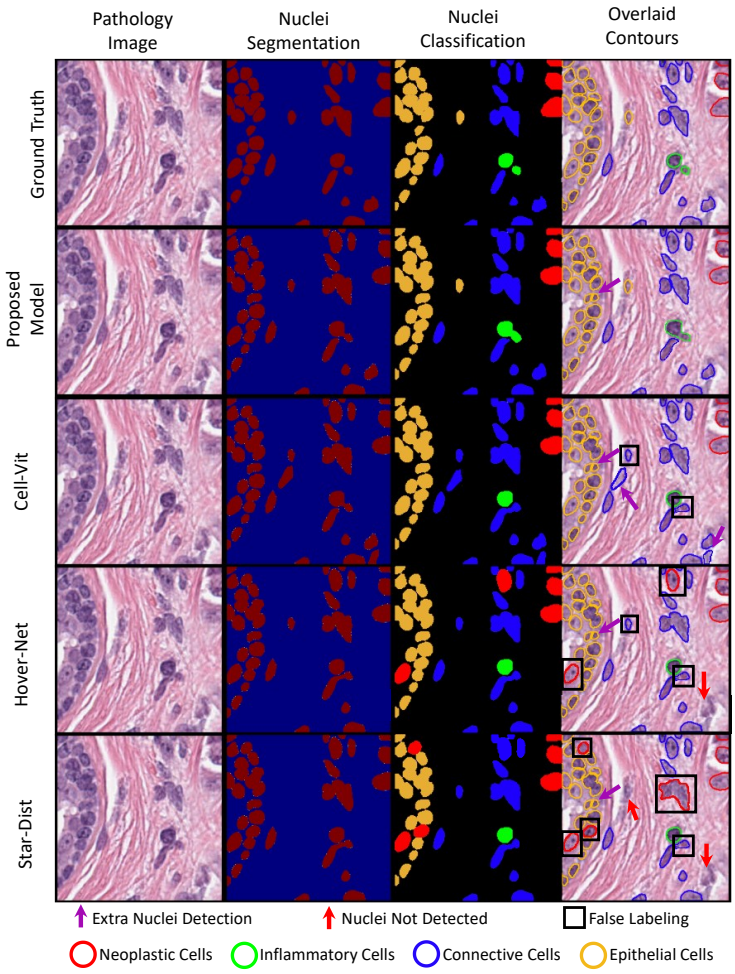

Figure 3.3: Qualitative segmentation results for a representative histopathology image, comparing the performance of the proposed model (TexSegNet) with CellViT, HoVer-Net, and StarDist.

To further evaluate the segmentation and classification performance, the accuracy and macro-averaged F1-score were analyzed across various IoU thresholds, as shown in **Figure 3.4**. In terms of accuracy, the proposed model consistently outperforms its counterparts across all IoU thresholds, even as the threshold becomes more stringent. Competing methods show a noticeable drop, particularly at higher thresholds, indicating challenges with clustered or morphologically complex nuclei. For the macro F1-score, the proposed model also demonstrates superior

performance, retaining a 3–5% advantage over the CellVit and an even larger margin over the other models as the IoU threshold increases.

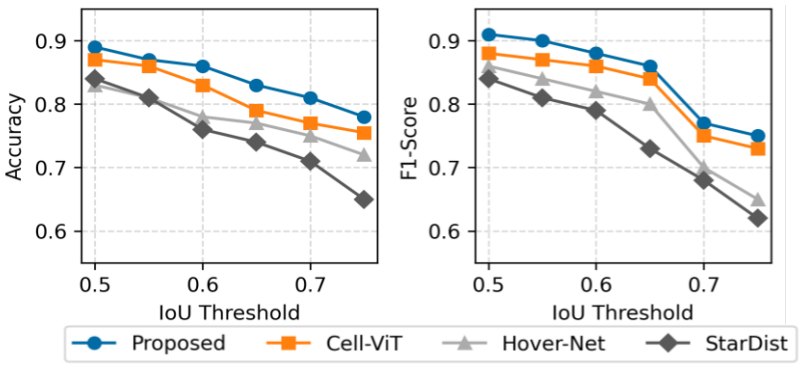

Figure 3.4: Comparison of the accuracy and macro F1-score across different IoU thresholds.

3.5 Discussion and Conclusion

In this paper, we introduced TexSegNet, a novel hybrid encoder–decoder framework to address persistent challenges in breast histopathology image analysis. By integrating multi-scale convolutions, improved texture feature representation, attention mechanisms, mudslide boundary maps, and a feedback-based classification branch, TexSegNet achieves considerable improvements in nuclei segmentation and classification on the breast test set of the PanNuke dataset. Experimental comparisons with established methods such as CellViT, HoVer-Net, and StarDist highlight TexSegNet's superior performance, particularly in accurately delineating clustered or touching nuclei and reducing errors in nuclei detection. Although occasional mis-segmentation issues remain unresolved, the strong quantitative and qualitative results underscore its promise as a robust and scalable tool for digital pathology research, and a valuable decision-support tool in breast cancer diagnostics. Future work will focus on refining augmentation approaches, conducting more extensive domain-specific fine-tuning, and validating TexSegNet in clinical workflows to enhance and further assess its performance and generalizability.

References

[1] Veta, M., Pluim, J. P. W., van Diest, P. J., & Viergever, M. A. (2014). Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering, 61*(5), 1400–1411.

[2] Cireşan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. In *Lecture Notes in Computer Science* (pp. 411–418). Springer. https://doi.org/10.1007/978-3-642-40763-5_51.

[3] Tafavvoghi, M., Sildnes, A., Rakaee, M., Shvetsov, N., Bongo, L. A., Busund, L.-T. R., et al. (2025). Deep learning-based classification of breast cancer molecular subtypes from H&E whole-slide images. *Journal of Pathology Informatics, 16*, 100410.

[4] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88.

[5] Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA, 318*(22), 2199–2210.

[6] Priego-Torresa, B. M., Lobato-Delgado, B., Atienza-Cuevas, L., & Sanchez-Morillo, D. (2023). Deep learning-based instance segmentation for the precise automated quantification of digital breast cancer immunohistochemistry images. *Expert Systems with Applications, 213*, 116471.

[7] Kadaskar, M., & Patil, N. (2023). Image analysis of nuclei histopathology using deep learning: A review of segmentation, detection, and classification. *SN Computer Science, 4*(6), 698. https://doi.org/10.1007/s42979-023-02115-2.

[8] Basu, A., Senapati, P., Deb, M., Rai, R., & Dhal, K. G. (2024). A survey on recent trends in deep learning for nucleus segmentation from histopathology images. *Evolving Systems, 15*, 203–248.

[9] Irshad, H., Veillard, A., Roux, L., & Racoceanu, D. (2014). Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential. *IEEE Reviews in Biomedical Engineering, 7*, 97–114.

[10] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90.

[12] Kolesnikov, A., Dosovitskiy, A., Weissenborn, D., Heigold, G., Uszkoreit, J., Beyer, L., Minderer, M., Dehghani, M., Houlsby, N., Gelly, S., Unterthiner, T., & Zhai, X. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *Proceedings of the International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.2010.11929.

[13] Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y. W., Kwak, J. T., et al. (2019). HoVer-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis, 58*, 101563.

[14] Hörst, F., Rempe, M., Heine, L., Seibold, C., Keyl, J., Baldini, G., et al. (2023). CellViT: Vision transformers for precise cell segmentation and classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1–12). IEEE. https://doi.org/10.48550/arXiv.2306.15350.

[15] Wang, J. (2024). Mudslide: A universal nuclear instance segmentation method. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11673–11682). IEEE. https://doi.org/10.1109/CVPR52688.2024.01121.

[16] Gamper, J., Alemi Koohbanani, N., Benet, K., Khuram, A., & Rajpoot, N. (2019). PanNuke: An open pan-cancer histology dataset for nuclei instance segmentation and classification. In C. Reyes-Aldasoro, A. Janowczyk, M. Veta, P. Bankhead, & K. Sirinukunwattana (Eds.), *Digital Pathology. ECDP 2019. Lecture Notes in Computer Science* (Vol. 11435). Springer, Cham. https://doi.org/10.1007/978-3-030-23937-4_2.

[17] Aaltonen, L. A., Abascal, F., Abeshouse, A., Aburatani, H., Adams, D. J., Agrawal, N., et al. (2020). Pan-cancer analysis of whole genomes. *Nature, 578*(7793), 82–93. https://doi.org/10.1038/s41586-020-1969-6

[18] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(11), 2274–2282.

[19] Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., et al. (2022). MaxViT: Multi-axis vision transformer. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 459–479). Springer. https://doi.org/10.1007/978-3-031-20053-3_27.

[20] Galdran, A., Carneiro, G., & Ballester, M. A. G. (2023). On the optimal combination of cross-entropy and soft dice losses for lesion segmentation with out-of-distribution robustness. In M. H. Yap, C. Kendrick, & B. Cassidy (Eds.), *Diabetic Foot Ulcers Grand Challenge. DFUC 2022. Lecture Notes in Computer Science* (Vol. 13797). Springer, Cham. https://doi.org/10.1007/978-3-031-26354-5_4.

[21] Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., & Ayed, I. B. (2021). Boundary loss for highly unbalanced segmentation. *Medical Image Analysis, 67*, 101851.

[22] Abraham, N., & Khan, N. M. (2019). A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. In *2019 IEEE 16th Int. Symposium on Biomedical Imaging (ISBI 2019)* (pp. 683–687). IEEE. https://doi.org/10.1109/ISBI.2019.8759329.

[23] Alom, Z., Asari, V. K., Parwani, A., et al. (2022). Microscopic nuclei classification, segmentation, and detection with improved deep convolutional neural networks (DCNN). *Diagnostic Pathology, 17*, 38.

[24] Weigert, M., & Schmidt, U. (2022). Nuclei instance segmentation and classification in histopathology images with StarDist. In *Proceedings of the IEEE International Symposium on Biomedical Imaging* (pp. 1–4). IEEE. https://doi.org/10.1109/ISBIC56247.2022.9854534.

Chapter 4

Detection, Localization, and Staging of Breast Cancer Lymph Node Metastasis in Digital Pathology WSIs Using Selective Neighborhood Attention-Based Deep Learning*

4.1 Introduction

Breast Cancer is the most common malignancy diagnosed in women globally, with around 2.3 million new cases and 670,000 deaths reported in 2022 [1]. Lymph node metastasis is a critical diagnostic and prognostic factor in breast cancer, significantly influencing treatment decisions and patient outcomes [2]. This impact is evident in five-year survival rates: patients with localized breast cancer have a five-year survival of 99%, which drops to 86% with regional lymph node involvement and falls sharply to 29% in cases with distant metastasis [3].

Traditionally, pathologists analyze stained histopathology slides with multiple tissue sections from each biopsied lymph node under a microscope and annotate them for diagnosis and staging of regional lymph node metastasis (known as the pN-staging). However, this process is labor-intensive and is prone to inter-observer variability and cognitive fatigue [4]. Studies have shown that pathologists may experience discordance rates as high as 24% in detecting micro-metastases, underscoring the importance of standardized automated detection systems [5].

The emergence of digitally scanned and annotated histopathology image datasets has created new opportunities for the development of automated analysis systems. Recent progress in deep learning

* A version of the material presented in this chapter has been submitted for publication in Scientific Reports, Nature, 2025.

networks have shown considerable potential in examining these histopathological images [6,7]. In controlled environments, these networks have achieved performance levels comparable to those of experienced pathologists [7]. However, these approaches face considerable limitations when scaled to clinical practice. Whole slide images (WSIs) in pathology present unique challenges due to their enormous size (often exceeding 100,000 × 100,000 pixels), heterogeneous tissue structures, staining variations, and complex morphological patterns. These factors render traditional deep learning approaches both computationally intractable and medically inconsistent [8,9]. Moreover, existing models often fail to generalize across different institutions and preparation protocols, showing marked degradation in performance on out-of-distribution (OOD) samples [10].

In computational pathology, classifying WSIs poses significant challenges due to their gigapixel size and the scarcity of detailed annotations. Multiple instance learning (MIL) has emerged as a popular approach to handle WSI classification by treating each slide as a bag of instances (patches) and assigning labels at the bag level. Traditional MIL methods, however, often struggle with the imbalance of positive instances and their inability to capture the most informative regions within a WSI [11]. To address these limitations, the attention-based MIL (ABMIL) method has introduced an attention mechanism to weigh the importance of each instance, allowing models to focus on salient regions for classification [12]. Building upon this, the clustering-constrained attention MIL (CLAM) approach has incorporated clustering constraints to enhance instance discrimination by encouraging the model to learn more diverse and representative features [13]. Recent approaches have further explored advanced mechanisms to improve MIL for WSI classification. The masked hard instance mining MIL (MHIM-MIL) method has proposed a framework that implicitly mines hard instances by masking out easy-to-classify instances based

on attention scores [14]. This method focuses on challenging instances near the decision boundary, enhancing the discriminative power of the model without requiring instance-level supervision. Similarly, the spatially aware MIL (SAM-MIL) method incorporates spatial information by embedding the spatial coordinates of patches into the learning process [15]. By leveraging the spatial context, SAM-MIL improves the model's ability to understand tissue structures and spatial relationships within a WSI for its classification. In the context of cancer staging, in addition to WSI-level classification, annotation of metastatic tumor regions within each WSI is required to quantify their extent. While the MIL models optimize slide-level classification by weighing numerous tumor patches through attention mechanisms, they often underperform in patch-level classification, a critical factor for accurate cancer staging. This limitation arises because slide-level tasks benefit from abundant tumor patches, allowing the model to prioritize relevant regions. In contrast, patch-level classification requires refined distinctive patch-level features that MIL models typically do not capture. Consequently, MIL models may struggle to annotate individual tumor regions accurately, necessitating advanced feature enhancement and processing modules focused on patch-level analysis.

Deep learning approaches that utilize contextual neighborhood information to improve patch-level classification have recently gained attention. For instance, the nearby patch contrastive learning (NearbyPatchCL) introduced a technique that leverages neighborhood information along with a decoupled contrastive loss to enhance patch-level classification [16]. Similarly, the gated graph attention (GGAT) networks have been used to model spatial relationships between patches by treating them as nodes in a graph, capturing both local and global tissue structures [17]. Incorporating the spatial neighborhood around a target patch provides essential contextual information for patch-level classification. This extra information allows feature maps to be re-

embedded for easier classification, enhancing the model's ability to infer when the target patch's features are unclear. However, if the surrounding patches are predominantly of an opposing class, this approach can lead to misclassification. During training, the model may overly rely on neighborhood context regardless of its similarity to the central (target) patch's features. Since central patches are typically surrounded by patches of the same class, the model may tend to assign the majority class of the surrounding context to the central patch. This behavior could minimize the loss function with minimal computational overhead, potentially causing the optimizer to converge prematurely to a suboptimal local minimum. While using neighborhood context can attenuate overall misclassification, it creates significant challenges at the boundaries between tumors and normal tissues, where precise classification is crucial for accurate staging.

Nuclei segmentation has been extensively used in various WSI analysis tasks such as tumor grading and subtyping [18,19]. By accurately segmenting nuclei, methods can extract meaningful features such as size, shape, density, and organization, which are crucial for pathological assessments [20,21]. These features provide interpretable and biologically relevant information that potentially enhances the model's discriminative capabilities. However, the features extracted from segmented nuclei may not fully capture the tissue context or complex cytoplasmic patterns, leading to inaccurate results across different tissue types.

To address these challenges, we revisit the foundational principles of WSI annotation, and the methodologies employed by pathologists. Studies indicate that nuclei attributes serve as the primary features used in manual annotations, with factors such as histologic type, morphology, heterogeneity, and distribution playing pivotal roles [22]. Recognizing this, we propose a novel framework that leverages both the nuclei and tissue-level features with an attention-based neighbor patch selection mechanism. This approach incorporates contextual information from dynamically

identified relevant regions for the detection, localization and staging of breast cancer metastasis in WSIs of lymph node biopsy specimens. By combining the nuclei and tissue-level features, we leverage the complementary information from nuclei histologic type, morphology, texture, and distribution, as well as high-level tissue structures to create a comprehensive patch representation, thereby enhancing classification accuracy and robustness. To mitigate issues arising from over-reliance on surrounding context and to improve classification accuracy at boundaries of tumors and normal tissues, we introduce an attention-based neighbor patch selection mechanism. This mechanism calculates the similarity between the central patch and its neighboring patches, selecting only those neighbors with the highest attention scores based on their relevance to the target patch (**Figure 4.1**). This selective attention approach mitigates the risk of misclassification caused by dissimilar neighboring patches and maintains the integrity of classification at critical tissue boundaries, which is essential for the localization and staging of lymph node metastasis.

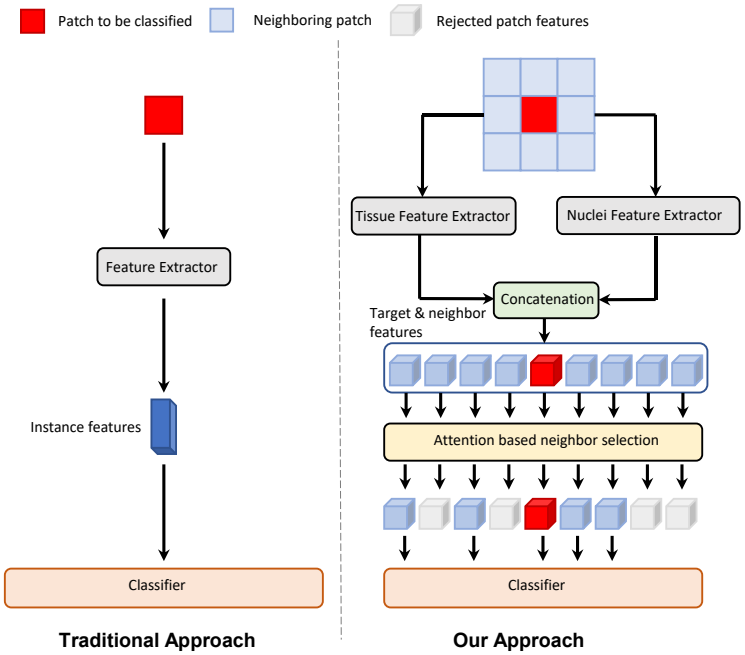

Figure 4.1: Left: Traditional approach for histopathology patch classification using single patch features, often disregarding neighboring context. Right: The proposed method integrating nuclei and tissue features, attention-based neighbor selection, and contextual information from dynamically-identified relevant regions for enhanced patch classification.

4.2 Materials and Methods

4.2.1 Dataset

For the development and evaluation of the proposed framework, we used the CAMELYON grand challenge datasets (CAMELYON16 and CAMELYON17) [23,24], which have become established benchmarks in computational pathology for metastasis detection in lymph node [25]. The CAMELYON16 dataset comprises 398 WSIs of lymph node sections sourced from two Dutch medical centers: Radboud University Medical Center and University Medical Center Utrecht. The WSIs have a pixel size of 0.24 μm at 40× magnification, and their size ranges from (45,000 × 36,000) to (216,000 × 1,110,000) pixels. The CAMELYON16 dataset contains separate training (269 WSI) and test (129 WSI) sets at the patient level. According to the dataset documentation, 20 WSIs in the training set include tissue sections that have not been exhaustively annotated. To maintain the integrity of the training data, the normal tissue regions with incomplete annotations within these WSIs were excluded from the training set. In this study, the CAMELYON16 training set was used for training and optimization of the models, and its test set was used for evaluation of model performance on patch and WSI levels. Detailed statistics of the applied dataset, including the number of normal and metastatic WSIs along with the distribution of micro and macro metastases, are provided in **Table 4.1**.

Table 4.1: Summary of dataset splits for the CAMELYON16 [1] dataset, detailing the number of WSIs, the count of normal and metastatic cases, and the breakdown of micro and macro metastases across the training and test sets based on the ground-truth annotations available for all WSIs.

Dataset Split	Total WSIs	Normal WSIs	Metastatic WSIs		Number of Tumor regions	
			Micro Metastases	Macro Metastases	Micro	Macro
Training	269	158	72	39	345	81
Test	129	79	30	20	185	40

CAMELYON17, the successor to CAMELYON16, comprises 1,000 WSIs collected from five medical centers: Radboud University Medical Center, University Medical Center Utrecht, Canisius-Wilhelmina Hospital, Rijnstate Hospital, and Haga Hospital. The ground-truth information is available for 500 WSIs sourced from a total of 100 patients (five lymph node slides per patient) [24]. Detailed lesion-level annotations for metastasis (tumor) regions are provided in the dataset for 50 WSIs, with each hospital contributing 10 annotated slides. The Radboud University Medical Center and University Medical Center Utrecht are the medical centers which contributed to both CAMELYON16 and CAMELYON17. Consequently, the 30 annotated images from the other three medical centers were utilized for out-of-distribution (OOD) evaluation of the model performance on patch and WSI levels. The patient-level pN-stage is available in the dataset for the 100 patients. All the 500 WSIs were used in this study for evaluating the model performance in patient-level pN-staging.

4.2.2 Data Preprocessing and Augmentation

A multi-step tissue mask generation and patch extraction pipeline was used to prepare the WSIs for analysis. The OTSU thresholding method [26] was applied at 20× magnification to generate high-resolution binary tissue masks and extract tissue regions from the background. The extracted tissue regions in each WSI were divided into (256 × 256)-pixel patches at 40× magnification [13]. Patches with less than 30% overlap with ground-truth tumor regions were labeled as 0 (normal), and those with more than 30% overlap were labeled as 1 (tumor). After tissue masking and patching, approximately 14.38 million image patches were generated from the training set WSIs, with 13.8 million labeled as normal and 0.58 million as tumor. Each slide generated an average of 90,000 patches, while metastatic slides produced an average of 2,000 tumor patches per slide.

The percentage of tumor patches per metastatic slide ranged from 0.01% to 70%, with a median of 2%.

To address the relative scarcity of tumor patches in training the tissue-level feature extractor model, data augmentation techniques were applied specifically to the tumor class in the training set. Each tumor patch was flipped vertically and horizontally, and rotated by increments of 90°, resulting in eight distinct orientations (seven augmented samples). During each training epoch, the training data included all augmented tumor patches and a random subset of normal patches. The size of the normal patch subset was adjusted to match the number of tumor patches, ensuring class balance within each epoch. The random selection of the normal patches was implemented such that each patch in the entire training set had an equal chance for inclusion in the subset, while all normal patches were presented to the model in approximately the same number of training epochs. For training the patch/slide-level classification models, due to incorporation of the neighborhood attention, all patches associated with the training WSIs were presented to the models during each epoch, along with their corresponding patch/slide-level labels. To enhance model robustness against staining and imaging variabilities, random color variations were performed on input images during all model trainings. Brightness was varied in the range of [-55, 55] levels, contrast and saturation were modified by factors in the range of [-0.6, 0.6] and [-0.25, 0.25], respectively, and hue was shifted in the range of [-6, 6] degrees. While all images in the CAMELYON16 and CAMELYON17 datasets have been acquired at 40× nominal magnification, the header file of some images indicate a magnification of 37.2×. To improve model generalization to slight magnification differences in images acquired from various sources, a random scaling factor between 0.93 to 1.08 was applied to input images during all model trainings.

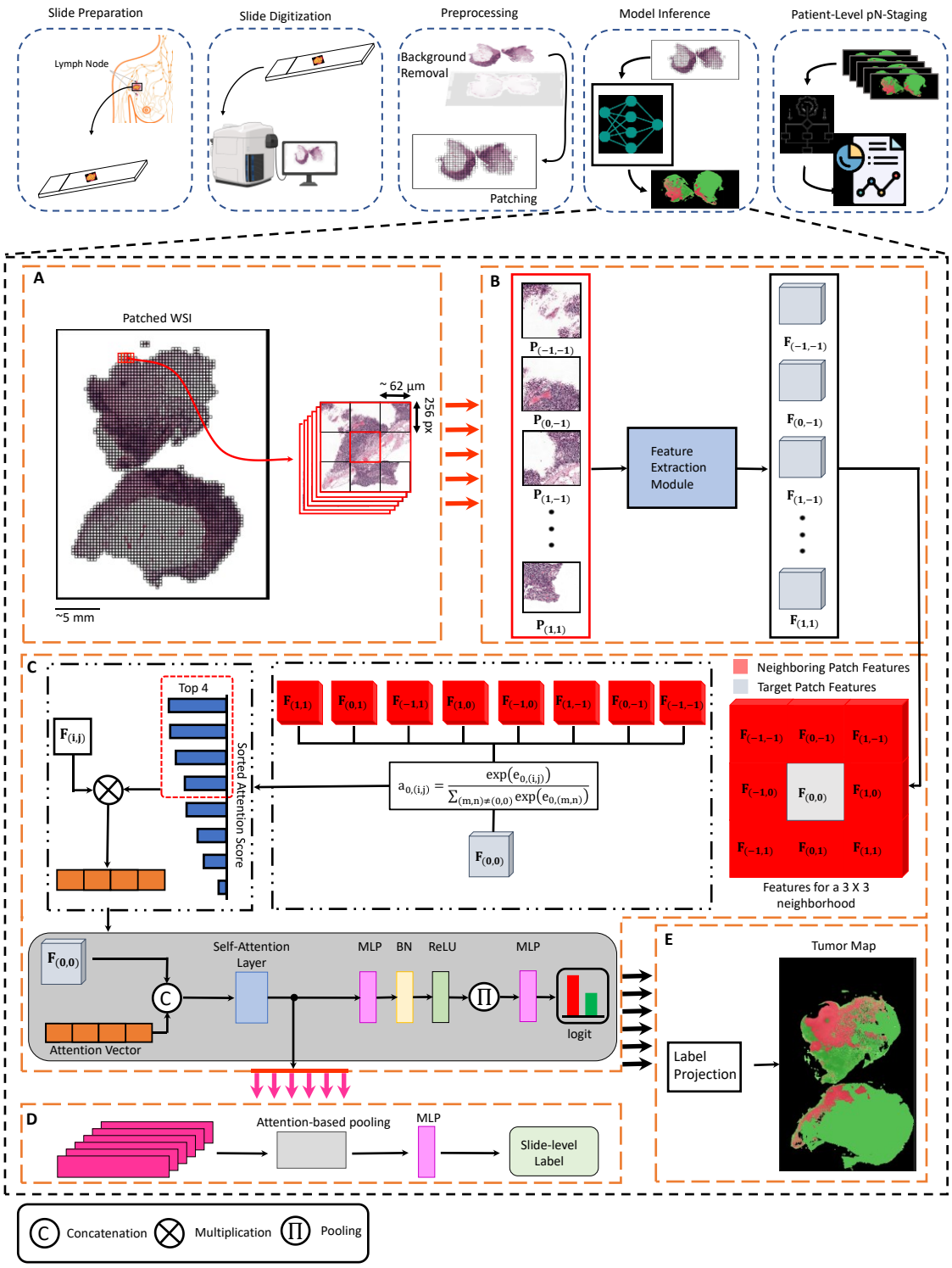

Figure 4.2: The proposed framework that processes the patches associated with each WSI (A) through the feature extraction module (B; further details in Figure 4.3). The feature arrays for patches within a 3×3 neighborhood grid are processed through the selective attention mechanism (C), followed by patch-level classification, slide-level classification (D), and tumor map generation for the WSI (E). The generated tumor maps for each patient are then processed to determine the patient-level pN-stage automatically.

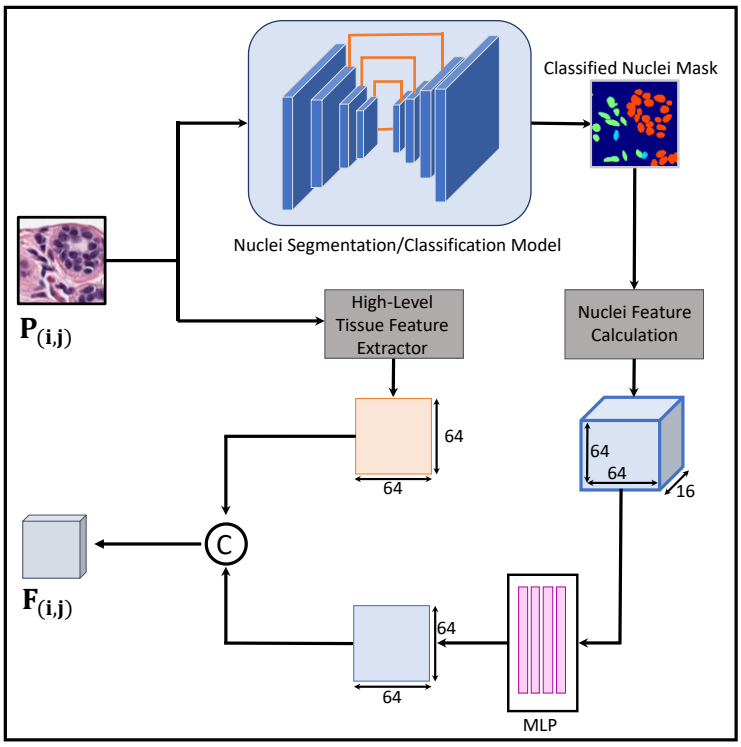

Figure 4.3: Feature extraction Module: High-level tissue features and nuclei features are combined to create a comprehensive representation for each patch.

4.2.3 Model Architecture

The proposed model (**Figure 4.2**) consists of a feature extraction module, a selective attention patch classification module, and modules for slide-level classification, tumor map generation, and patient-level pN-staging. These modules are described below.

Feature Extraction Module: The patches $P_{(i,j)}$ extracted from each WSI are passed through the feature extraction module (**Figure 4.3**). This module consists of two parallel branches: The first branch processes patches through a nuclei segmentation and classification model to generate nuclei mask for different cell types and derive the nuclei features, while the second branch extracts high-level tissue features. The outputs from these branches are combined into a fixed-size feature vector $F_{(i,j)}$.

Nuclei Feature Extractor: The framework utilizes a nuclei segmentation and classification model to extract the nuclei features from the input patches. Comprehensive details of the applied model are provided in **Chapter 3** of the thesis. This model generates two types of masks for each input patch, both of the same size: one indicating the segmented nuclei and the other providing the class probabilities of segmented nuclei with respect to the five cell types, i.e., neoplastic, inflammatory, connective tissue, epithelial, or other (including dead) cells. Using these masks, 16-channel nuclei feature representation is constructed. The segmented nuclei mask values at each channel indicate the corresponding feature values (described below). The five cell type masks provide five distinct feature channels, each representing the probability of the nuclei belonging to the corresponding class. To capture the nuclei shape attributes, seven Hu moment invariants [27] are computed and assigned to individual feature channels. These Hu moments effectively encapsulate complex shape features of nuclei, including subtle morphological variations, and abnormalities in nuclear shape such as irregular contours, elongation, and asymmetry, which are important factors in histopathology analysis [28,29]. Two additional feature channels are assigned to the mean and standard deviation of RGB color intensities within each nucleus, capturing staining variations that reflect differences in chromatin density [30]. Texture features are quantified using entropy and local binary patterns, each assigned to a separate channel, to characterize chromatin organization and heterogeneity [31]. Entropy is computed by measuring the randomness in the intensity distribution of nuclear regions, providing a scalar value that reflects the complexity of chromatin texture [32,33]. LBP captures local texture patterns by analyzing the intensity relationship between each pixel and its neighbors, serving as a potential descriptor of chromatin spatial heterogeneity [34,35]. Once these features are assigned to their respective channels, the nuclei feature maps are down-sampled from 256×256 pixels to 64×64 pixels using average pooling to preserve spatial

distribution while reducing computational complexity. The 64×64×16 feature maps are then processed through a multi-layer perceptron (MLP) block to compress the number of channels to a single output channel.

High-Level Tissue Feature Extractor: Our high-level tissue feature extractor processes 256×256 pixel patches to produce 64×64 spatial feature maps. Inspired by the CTransPath encoder [36], the proposed framework adopts a similar architecture combining CNN and transformer components. The architecture begins with convolutional and pooling layers to extract low-level features and reduce spatial dimensions. These layers are followed by transformer encoder blocks, which employ self-attention to gather global contextual information. The unsupervised training approach of CTransPath has been adapted into a supervised training strategy for the classification task, utilizing a weighted combination of binary cross-entropy and contrastive loss functions. This supervised training enhances the model's ability to differentiate between the normal and tumor patches by driving the two classes in opposite directions within the embedding space. After training, the classification head is discarded, and the backbone is used for tissue-level feature extraction.

The nuclei and tissue feature maps are concatenated to produce a feature array of size 64×64×2. This array is subsequently flattened into an 8192-element vector and passed through a linear transformation with a learnable weight matrix $W_{\text{concat}} \in R^{1024\times 8192}$, yielding a resized feature vector with 1024 elements.

Selective Attention Patch Classification Module: To enhance the classification accuracy of each patch $P_{0,0}$ within a WSI, a neighborhood contextual representation is constructed in the framework. A 3×3 patch grid centered around $P_{0,0}$ is considered, incorporating its eight immediate neighbors $P_{i,j}$, where $i, j \in \{-1, 0, 1\}$ and $(i, j) \neq (0,0)$. Each patch $P_{i,j}$ is

represented by a feature vector $F_{i,j}$. To quantify the contextual relevance of each neighboring patch relative to the central patch, a gated attention mechanism is employed. This approach introduces a gating function $g(\cdot)$ that dynamically modulates attention scores based on feature relevance, enhancing the model's ability to focus on relevant neighboring patches while mitigating the influence of less pertinent or noisy ones. The attention score $e_{0,(i,j)}$ between the central patch and a neighboring patch is computed as:

$$e_{0,(i,j)} = g\left(F_{0,0}, F_{i,j}\right) \cdot \frac{\left(W_o F_{0,0}\right)^{\top}\left(W_p F_{i,j}\right)}{\sqrt{d}} \tag{4.1}$$

where W_o and W_p are learnable weight matrices that transform the central ($F_{0,0}$) and neighboring patch features ($F_{i,j}$) into the query and key representations, respectively. The variable d represents the dimensionality of the query and key matrices, and is used to normalize the attention scores. The gating function $g\left(F_{0,0}, F_{i,j}\right)$ is a sigmoid-based scalar multiplier that modulates attention based on the combined features of the central and neighboring patches. The attention coefficients $\alpha_{0,(i,j)}$ are then obtained by applying the *softmax* function to the attention scores as follows:

$$a_{0,(i,j)} = \frac{\exp\left(e_{0,(i,j)}\right)}{\sum_{m,n\in\{-1,0,1\},(m,n)\neq(0,0)} \exp\left(e_{0,(m,n)}\right)} \tag{4.2}$$

Focusing on the most relevant neighbors, the top $K = 4$ patches with the highest $\alpha_{0,(i,j)}$ values are selected to reduce the risk of misclassification due to dissimilar surrounding context. For these selected neighboring patches, their attention-enhanced features H_k are computed by scaling their feature vectors $F_{i,j}$ with their attention coefficients $\alpha_{0,(i,j)}$, reflecting their relative significance:

$$H_k = \alpha_{0,(i_k,j_k)} \cdot F_{i_k,j_k} \quad \text{where} \quad (i_k, j_k) \sim top\ K\left\{\alpha_{0,(i,j)}\right\}, \quad \forall\, k = 1,.,.,K \tag{4.3}$$

This weighted aggregation ensures that more relevant neighbors contribute proportionally to the central patch's classification, enhancing the model's discriminative power by utilizing the most pertinent contextual information. To incorporate this relevant neighborhood context into the feature representation, the selected attention-scored features are concatenated with the central feature vector:

$$X = \text{Concat}\left(F_{(0,0)}, H_1, H_2, \dots, H_k\right) \quad \forall\, k = 1, .., K \tag{4.4}$$

The concatenated feature vector X encapsulates a comprehensive representation of the local tissue context by combining both intrinsic properties of central patch with the relevant features from neighboring patches. It is then passed through a self-attention (SA) block to identify and capture intricate interdependencies and synergistic relationships within the local tissue context. The self-attention of vector X is computed using Equation (5), resulting in β based on the following expression:

$$\beta = \text{SelfAttention}(X) = \text{Softmax}\left(\frac{\left(W_Q X\right)(W_K X)^{\top}}{\sqrt{d_k}}\right)(W_V X) \tag{4.5}$$

where W_q, W_k and W_V are learnable weight matrices. The attention output R is computed as:

$$R = \sigma(W_x \beta X) \tag{4.6}$$

where W_x is a learnable weight matrix and σ is the non-linear *ReLU* activation function. The attention-enhanced output R is subsequently processed by a network to synthesize more abstract feature representations. This network comprises several components: an MLP for nonlinear transformation, batch normalization (BN) to stabilize training, *ReLU* activation, and a pooling layer to consolidate feature information. The output from the pooling layer is then passed to

another MLP, followed by the *softmax* function to generate class probabilities for tumor versus non-tumor (normal) patch classification.

Slide-Level Classification: In the slide-level classification process, the model aggregates individual patch representations within a WSI to generate a comprehensive feature vector and determine whether the WSI contains metastasis. This aggregation is achieved by assigning the normalized self-attention weight β_i, to each of the N patches in the WSI. Specifically, the slide-level feature vector S_{slide} is computed as a weighted sum of the patch feature vectors X_i based on the following expression:

$$S_{slide} = \sum_{i=1}^{N} \beta_i X_i \tag{4.7}$$

The normalized attention weights β_i represent the importance of each patch, assigning higher weights to patches that are more indicative of tumor presence. This enables the most relevant regions to have a greater influence the final representation, while less informative patches contribute minimally. Once S_{slide} is obtained, it is passed through a classification layer with learnable parameters W_s, followed by a sigmoid activation function σ, to produce the classification probability $\hat{y}_{slide}$ as follows:

$$\hat{Y}_{slide} = \sigma(W_s S_{slide}) \tag{4.8}$$

The output probabilities $\hat{Y}_{slide}$ are then binarized to classify WSIs as positive (with metastasis) and negative (without metastasis) slides.

Slide-Level Tumor Map Generation: Once the network has classified each patch within the WSI, these predictions are spatially mapped back onto the slide to form a binary tumor mask. The classification output (tumor/normal) for each patch is projected onto its corresponding region on

the WSI, resulting in a mask that outlines the tumorous regions while excluding normal tissue. This tumor mask is then used for subsequent analyses in pN-staging.

Patient-Level pN-Staging: Following slide-level classification and tumor map generation, the framework performs an automated pN-staging at patient level. For slides classified as positive, the area of each region of metastasis detected by the model is calculated through $A = n \times (pixel\ size)\ \mu m^2$, where n represents the number of pixels in the detected tumor region. The tumor regions are classified into three categories, i.e., isolated tumor cells (ITC) with $A \leq 200\ \mu m^2$, micro metastasis with $200\ \mu m^2 < A \leq 2.0\ mm^2$, and macro metastasis with $A > 2.0\ mm^2$. The patient-level pN-stage is determined based on the size of detected tumor regions and the number of positive lymph nodes (slides): $pN0$ (no metastases), $pN0(i+)$ (only ITCs detected), $pN1_{mi}$ (only micro-metastases detected), $pN1$ (1–3 metastatic lymph nodes with at least one macro-metastasis), and $pN2$ (4–9 metastatic lymph nodes with at least one macro-metastasis).

4.2.4 Training Configuration

A 6-fold cross-validation strategy was applied on the CAMELYON 16 training set for model training and optimization. The training process was conducted in two stages: first, training the high-level tissue feature extractor, and then training the selective attention classification modules.

The high-level tissue feature extractor was trained for 25 epochs using a weighted combination of binary cross-entropy (BCE) and contrastive loss. The weights were tuned to 0.7 and 0.3, respectively, through a grid search using cross-validation on the training set. A learning rate of 0.001 with a cosine annealing schedule was used, alongside a batch size of 16. To prevent overfitting, the $L2$ weight decay with a coefficient of 1×10^{-4} and a dropout rate of 0.3 were applied. Early stopping was implemented based on the model's performance on validation set,

halting training if the validation loss did not improve for 15 consecutive epochs. After training, the feature extractor's weights were locked to preserve the learned representations for classification model training. As part of the ablation study, the classification performance of this standalone high-level tissue feature extractor was compared with the complete model's performance, thereby quantifying the improvement gained by integrating the additional components.

The patch-level and slide level classification modules in the framework were trained together, where the shared components in the selective attention mechanism were trained with a combined hierarchical loss function while the exclusive components relied on individual losses. The adapted hierarchical loss function combines the attention-regularized patch-level (instance) and slide-level (bag) losses. The total loss function is formulated as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{instance} + \lambda_2 \mathcal{L}_{attention} + \lambda_3 \mathcal{L}_{bag} \tag{4.9}$$

where $\lambda_1 = 0.3$, $\lambda_2 = 0.2$, and $\lambda_3 = 0.5$ are weighting coefficients. The instance-level loss, $\mathcal{L}_{instance}$, is computed using the binary cross-entropy on individual patch predictions for each patch i in a bag containing N patches:

$$\mathcal{L}_{instance} = -\frac{1}{\mathrm{N}} \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \tag{4.10}$$

where y_i is the ground-truth label and p_i is the predicted probability. This component ensures that the framework learns discriminative features associated with individual patches for tumor detection. The instance-level supervision guides local feature processing in the model and provides a strong foundation for the attention mechanism.

To prevent the attention mechanism from becoming too focused on a single patch or region, an attention regularization term, $\mathcal{L}_{attention}$, is incorporated. The attention mechanism loss, $\mathcal{L}_{attention}$, prevents over-concentration of attention weights:

$$\mathcal{L}_{attention} = \left|\alpha - \frac{1}{K}\right|_1 + \gamma|\alpha|_2 \tag{4.11}$$

where α represents the attention weights, K is the number of neighboring patches, and $\gamma = 0.01$ is a weighting factor. The first term encourages a uniform distribution of attention when appropriate, while the second term prevents extreme attention values.

The bag-level classification loss, L_{bag}, aims at optimizing the model for whole-slide image classification. A weighted binary cross-entropy loss function is applied to address potential class imbalance in the dataset:

$$\mathcal{L}_{bag} = -\mathrm{w_p} Y_{bag} \log\left(\hat{Y}_{bag}\right) - w_n\left(1 - Y_{bag}\right)\log\left(1 - \hat{Y}_{bag}\right) \tag{4.12}$$

where Y_{bag} is the slide-level ground-truth label, $\hat{Y}_{bag}$ is the predicted probability for being a positive slide, and w_p and w_n are class-specific weights calculated as the inverse of class frequencies in the training set.

The model training was performed using an Adam optimizer with an initial learning rate of 0.0005, batch size of 2, and a weight decay of 5×10^{-4}. Gradient clipping was implemented with a threshold of 1.0 to prevent gradient exploding. A ReduceLROnPlateau scheduler with a factor of 0.5 and patience of 10 was employed. Training was carried out for 80 epochs with an early stopping patience of 20 epochs based on the validation AUC. All training procedures were executed using the PyTorch deep learning framework on a cluster of 4 NVIDIA L40s GPUs with 48 GB of memory.

4.2.5 Model Evaluation and Comparison

Model performance was comprehensively evaluated on the CAMELYON16 test set and CAMELYON17 dataset across multiple levels using metrics tailored to each stage. At the patch level, the accuracy, sensitivity, specificity, precision, recall, and F1-score were applied to evaluate the model's ability to classify individual patches as tumor or non-tumor. To measure the overlap, spatial precision, and detection accuracy of identified tumor regions within the WSIs compared to the ground truth, the slide-level evaluation incorporated the DICE score [37], aggregated Jaccard index (AJI) [38], mean distance (D_{mean}) [39], Hausdorff distance ($H_{distance}$) [39], slide-level area under the ROC curve (AUC) [23], and the lesion-level free-response ROC (FROC) score [23]. For the patient-level pN-staging, accuracy, macro F1-score, and weighted AUC, were utilized to evaluate the agreement between predicted and actual pN-stages, in addition to a five-class quadratic-weighted Cohen's kappa [24] that accounts for the severity of misclassifications. These metrics collectively provided a multifaceted assessment of the model's effectiveness in accurately detecting, annotating, and staging nodal metastasis within histopathological images.

The proposed model was comprehensively benchmarked against several state-of-the-art models, including TransMIL [40], DSMIL [41], DTFD-MIL [42], SAM-MIL [15], AB-MIL [12], CLAM [13], and Bayes-MIL [43] using the same dataset and experimental setup.

4.3 Results

Benchmarked against several state-of-the-art models, the proposed model demonstrated competitive performance, achieving superior or comparable values in key evaluation metrics such as patch-level F1-score, lesion-level FROC score, slide-level AUC, and patient-level Cohen's kappa, underscoring its effectiveness in accurately detecting and annotating tumor regions within

histopathological images. **Figure 4.4** presents the performance of the proposed and benchmark models across key evaluation metrics at the patch, lesion, slide and patient levels. More comprehensive quantitative evaluation results are presented in **Table 4.2**.

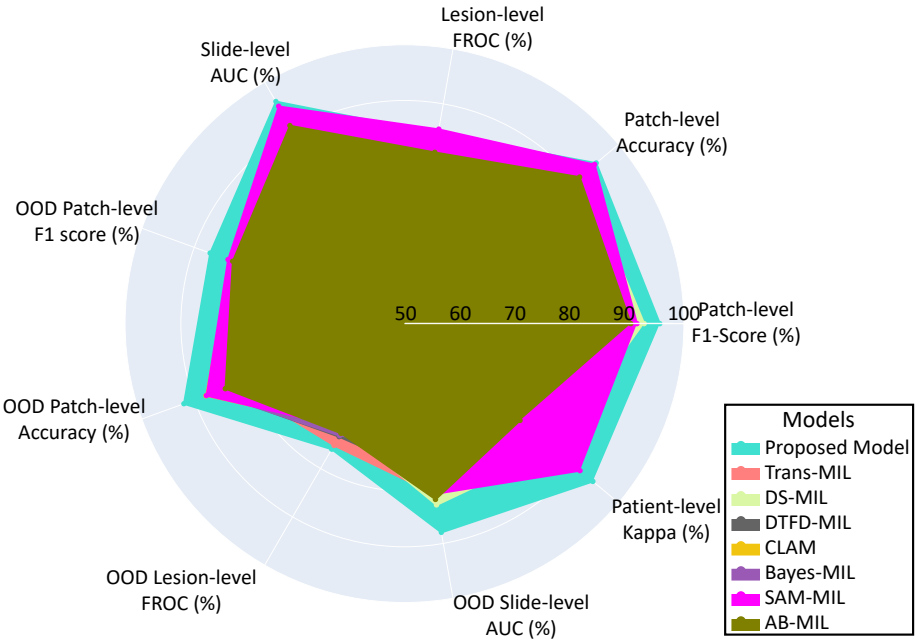

Figure 4.4: Radar chart comparing the performance of different models across key evaluation metrics.

The proposed model demonstrated outstanding performance on the CAMELYON16 and CAMELYON17 dataset, achieving a patch-level sensitivity of 96.2%, precision of 95.3%, and F1-score of 95.7% in CAMELYON16 test set. These results surpass those of established methods like TransMIL and DSMIL, highlighting our model's superior ability to detect tumor-positive patches. The high DICE score of 90.5% and Jaccard Index of 82.6% further emphasize its precise annotation of tumor regions within a WSI, while the lesion-level FROC score of 84.6% and slide-level AUC of 0.96 underscores its robust tumor detection and localization capability. On the CAMELYON17 dataset, used as the out-of-distribution (OOD) benchmark, the proposed model achieved a patch-level F1-score of 87.0%, and a slide-level AUC of 0.88. This reflects the model's ability to generalize across different data distributions with staining and imaging variability. At

the patient level, the proposed model excelled in pN-staging, achieving the highest kappa score on the CAMELYON 17 dataset. Specifically, our model achieved a kappa score of 0.94, which exceeds the kappa values of competing models such as TransMIL (0.89) and SAM-MIL (0.91).

Table 4.2: Performance evaluation of the model on CAMELYON16 test set, CAMELYON17 out of distribution (OOD) set, and the complete CAMELYON17 dataset across various patch-level, slide-level, and patient-level metrics. Results are presented as mean ± standard deviation.

Metrics \ Models	Proposed Model	Trans-MIL	DS-MIL	DTFD-MIL	CLAM	Bayes-MIL	SAM-MIL	AB-MIL
CAMELYON 16 Test Set								
Patch-Level Metrics								
Sensitivity (%)	**96.2±1.5**	91.1±2.8	94.0±1.7	92.1±1.4	89.2±4.0	89.8±2.4	91.0±1.9	91.0±1.7
Specificity (%)	94.5±4.6	94.5±0.9	91.3±3.0	90.8±1.1	94.1±0.7	91.8±2.2	**94.8±0.9**	90.9±1.9
Precision (%)	**95.3±2.4**	90.0±2.0	92.3±2.4	89.9±1.5	90.9±3.2	87.3±0.9	92.8±2.5	90.3±3.1
Accuracy (%)	**94.8±4.4**	94.2±0.9	91.6±2.9	90.9±1.0	93.6±0.7	91.6±2.1	94.4±0.9	90.9±1.8
F1-Score (%)	**95.7±3.1**	89.3±0.6	92.9±1.9	90.6±0.7	89.7±2.2	88.3±3.1	91.6±3.2	90.5±2.4
Slide-Level Metrics								
DICE Score (%)	**90.5±2.0**	86.5±1.5	88.3±0.8	88.3±1.4	87.3±1.2	85.6±1.7	87.5±5.3	87.9±2.0
Jaccard Index (%)	**82.6±0.8**	76.2±1.7	79.1±3.2	79.1±4.1	77.5±4.9	74.8±6.6	77.8±1.6	78.5±2.4
Mean Dist. (µm)	**11.2±0.7**	13.2±1.3	17.1±1.5	12.9±1.5	13.2±1.2	15.2±1.7	13.2±1.2	14.6±1.9
Hausdorff Dist. (µm)	35.6±7.0	33.9±2.8	36.9±4.3	**32.8±4.2**	32.9±3.2	34.5±2.7	37.1±4.3	38.8±3.4
Lesion-Level FROC Score (%)	84.6±2.8	82.4±1.9	83.3±2.7	81.5±2.7	80.7±1.2	79.7±4.2	**85.4±2.7**	81.2±5.4
Slide-Level AUC	**0.96±0.01**	0.91±0.01	0.92±0.02	0.93±0.02	0.94±0.01	0.89±0.02	0.95±0.02	0.91±0.03
CAMELYON 17 (OOD)								
Patch-Level Metrics								
Sensitivity (%)	92.6±2.2	87.1±4.8	86.2±2.5	88.5±4.1	92.6±2.0	85.8±2.9	**92.7±1.6**	92.4±3.8
Specificity (%)	**91.9±1.9**	81.7±1.6	83.3±3.5	85.7±4.3	87.3±3.3	86.5±2.0	87.1±2.6	83.2±2.3
Precision (%)	**82.0±1.4**	74.0±2.3	74.1±2.7	77.7±2.3	75.4±3.2	76.8±2.8	76.2±3.9	74.9±3.5
Accuracy (%)	**92.0±1.8**	82.2±1.5	83.6±3.3	86.0±4.1	87.8±3.1	86.4±1.9	87.7±2.5	84.1±2.2
F1-Score (%)	**87.0±1.8**	80.0±4.3	79.7±2.3	82.7±2.8	83.1±4.5	81.1±5.0	83.6±1.4	82.7±2.7
Slide-Level Metrics								
DICE Score (%)	**87.3±4.1**	84.1±2.9	86.8±2.1	85.0±4.3	85.7±1.9	83.6±1.9	85.9±1.3	86.7±3.0
Jaccard Index (%)	**77.5±1.7**	72.5±8.6	76.7±4.3	73.9±2.2	75.0±2.4	71.8±2.9	75.3±1.6	76.5±2.0
Mean Dist. (µm)	**15.2±1.3**	17.4±0.7	18.6±5.3	17.0±1.9	17.3±0.2	17.2±2.8	16.9±6.8	16.7±3.4
Hausdorff Dist. (µm)	42.1±4.6	**36.6±6.2**	39.0±6.8	37.6±3.8	41.3±3.0	38.4±5.6	41.7±6.1	38.5±3.8
Lesion-Level FROC Score (%)	**76.0±1.4**	75.2±3.8	71.7±3.4	73.4±2.2	71.4±1.1	73.2±2.8	71.5±1.7	72.7±3.0
Slide-Level AUC	**0.88±0.03**	0.81±0.02	0.83±0.02	0.78±0.02	0.77±0.03	0.72±0.09	0.81±0.04	0.82±0.02
CAMELYON 17 (Complete Dataset)								
Patient-Level Metrics								
Accuracy (%)	**91.3±2.7**	84.3±3.2	77.9±3.3	85.2±2.9	78.2±2.6	74.2±2.8	87.4±2.7	68.6±2.9
Macro F1-Score (%)	**89.4±3.0**	81.8±3.1	74.9±3.0	82.5±2.9	76.1±2.3	69.7±2.7	84.8±2.7	66.1±2.7
Weighted AUC	**0.92±0.02**	0.87±0.03	0.86±0.04	0.87±0.03	0.83±0.05	0.81±0.02	0.90±0.02	0.78±0.02
Kappa Score	**0.94±0.02**	0.89±0.02	0.85±0.02	0.89±0.02	0.85±0.02	0.79±0.02	0.91±0.02	0.77±0.03

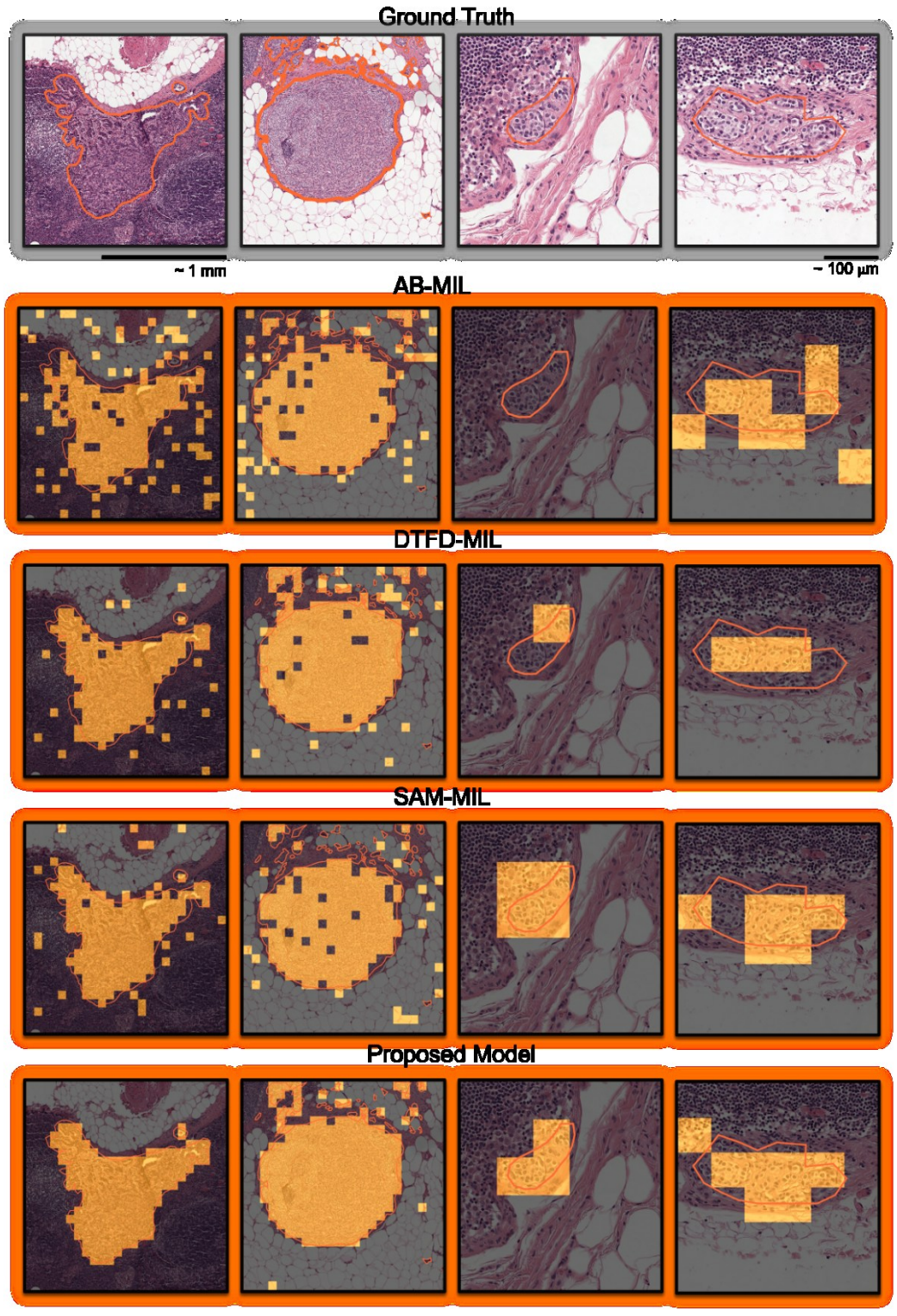

Figure 4.5: Performance of different models in detecting and annotating metastatic tumor regions on representative pathology images. The first two columns represent samples with large tumor sizes (0.67 and 1.0 mm^2), while the third and fourth columns correspond to smaller tumors (8×10^{-3} and 25×10^{-3} mm^2). Orange contours show the ground-truth tumor region boundaries, and highlighted areas indicate tumor regions detected by the models.

Figure 4.5 demonstrates the model performance in detecting and annotating metastatic tumor regions of different sizes on four representative pathology images. Visual comparison of model outputs highlights our model's superior capability to localize tumor regions across varying tumor sizes, outperforming state-of-the art methods like AB-MIL, DTFD-MIL, and SAM-MIL. For large tumors, the model's selective attention mechanism facilitates precise and continuous segmentation, closely aligning with ground truth annotations, while other methods often show fragmented outputs. In the case of small tumors, the model excels by leveraging segmented/classified nuclei features to accurately detect regions with very few tumor cells, effectively capturing boundaries that other methods tend to miss.

4.3.1 Ablation Study

To comprehensively evaluate the contributions of each component within the proposed model, we conducted a series of ablation experiments. These experiments systematically remove or alter specific elements of the framework to assess their impact on overall performance. First, the effect of nuclei segmentation and classification model was evaluated on feature embeddings and is illustrated in **Figure 4.6.** The visualization was generated by extracting patch-level features from the model for a total of 30,877 patches, collected from 10 positive (containing tumor) and 10 negative (normal) WSIs in the test set. Among these, 28,437 patches were normal, and 2,880 were tumor patches. Each patch was further categorized into five classes (neoplastic, inflammatory, connective, epithelial, and other cells) based on the dominant type of nuclei present in the patch, as detailed in **Figure 4.6.C**. The extracted patches were processed through the model, and their features were obtained at two stages: after the tissue feature extractor and after the self-attention module. The high-dimensional feature vectors were then projected to two dimensions using a t-distributed stochastic neighbor embedding (t-SNE) for visualization on a 2D plane [44], as

presented in **Figure 4.6.A**. Features from the tissue feature extractor include no information about nuclei and rely solely on general tissue-level features, leading to a substantial overlap between the normal and tumor patches. In contrast, features from the self-attention module incorporate nuclei-specific attributes along with a neighborhood attention mechanism, which selectively assigns weights to neighboring patches based on their relevance. This combination results in clearer clustering of the patch-level features and better separation between the normal and tumor patches in the embedding space, enhancing the model's ability to distinguish between them. **Figure 4.6.B** visualizes the influence of different cell types on clustering patch-level features in the embedding space, with patches labeled according to their dominant nuclei type. Neoplastic patches form a compact cluster, essential for tumor identification, while inflammatory and connective patches are more dispersed, reflecting biological interactions such as immune cell infiltration within tumor regions. **Figure 4.6.D** visualizes individual cell-type clusters in the embedding space. Neoplastic cells present a dense cluster due to their distinct attributes in the derived feature space, whereas other cell types, such as inflammatory and connective, exhibit more variability, consistent with their biological diversity. Overall, the integration of nuclei segmentation features can enhance the model's ability to identify different tissue structures and regions, improve classification performance at critical tumor-normal tissue boundaries, and potentially addresses dataset imbalances by providing richer and more biologically meaningful representations.

Figure 4.7 illustrates the model performance at different stages of analysis for a representative WSI. **Figure 4.7.A** demonstrates the ground truth tumor boundaries, with magnified views of two representative boundary regions providing a closer look at the tissue structures. **Figure 4.7.B** presents the output of the nuclei segmentation and classification model. While this model demonstrates an overall good performance in segmenting and classifying different cell types, some

normal and tumor regions are misclassified based solely on detected nucleus types. Two such examples are magnified in pink boxes.

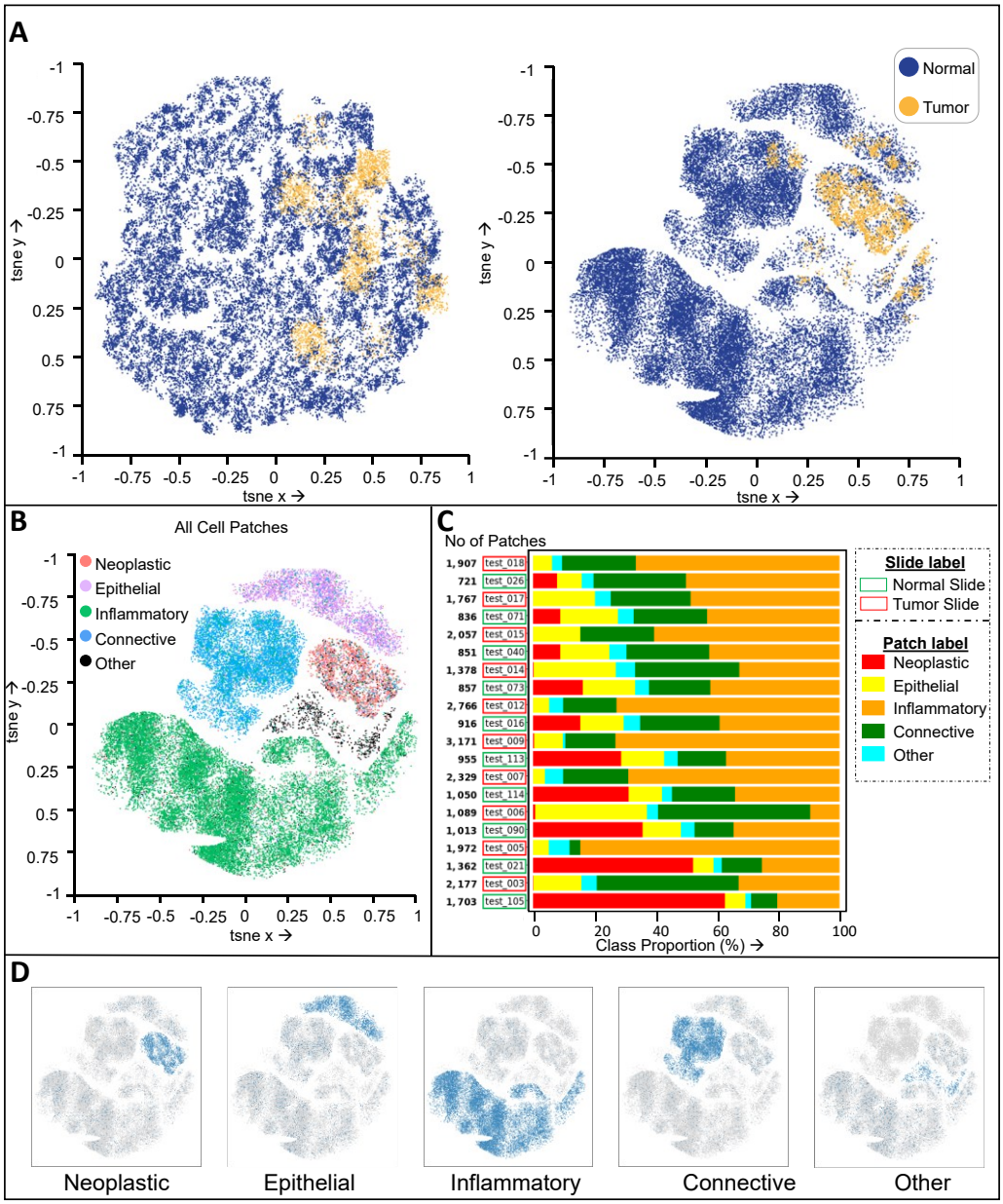

Figure 4.6: Panel A compares the t-SNE map of patch-level features extracted from two stages of the model: the tissue features extractor (left) and the self-attention module (right). The left plot shows overlapping normal (blue) and tumor (orange) patches due to sole reliance on general tissue-level features, while the right plot demonstrates improved separation by incorporating nuclei-specific features and attention mechanisms. Panel B highlights clustering in the t-SNE space based on dominant nucleus types, with colors representing different cell types. Panel C displays the distribution of these cell types across patches for a number of representative WSIs in the test set, with the bar length and colors corresponding to the patch counts and cell types, respectively. The t-SNE maps in Panel D visualize individual cell-type clusters.

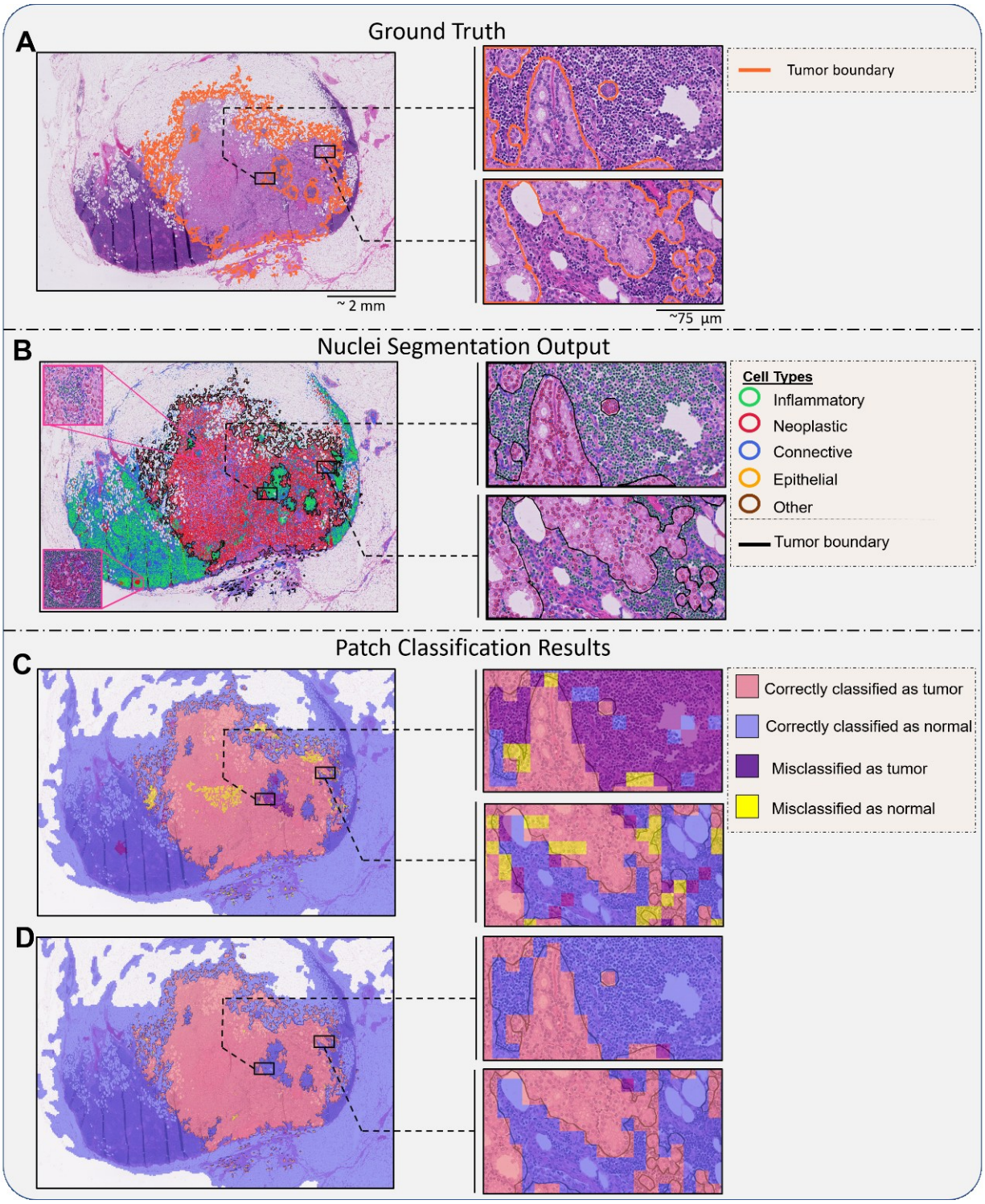

Figure 4.7: Inference results on a representative WSI. A: WSI at low magnification with two boundary regions at high magnification. The ground-truth tumor boundaries are contoured in orange. B: Nuclei segmentation and classification output, categorizing cells into five types. Two regions with incorrect classification based on the detected nucleus types are highlighted in pink boxes. C and D: Slide-level tumor maps generated using the patch-level classification results obtained from the tissue feature extractor and the proposed model, respectively. Magnified views on the right illustrate the performance of the classification models on representative tumor-normal tissue boundaries.

Figure 4.7.C shows the slide-level tumor maps generated using patch-level classification by the tissue feature extractor model without nuclei segmentation. The absence of detailed nuclei information leads to misclassification of normal and tumor patches, particularly at boundary areas. **Figure 4.7.D** demonstrates the results obtained by the proposed model, which incorporates nuclei information and selective neighborhood attention, improving performance at complex tumor-normal interfaces and reducing overall misclassification.

Table 4.3: Performance comparison of the tissue feature extractor model and the complete model across various metrics. Results are presented as mean ± standard deviation.

Metric/Model	**Tissue Feature Extractor Model**	**Complete Model**
Patch-Level Metrics		
Sensitivity (%)	88.4±2.5	**96.2±1.5**
Specificity (%)	91.0±2.6	**94.5±4.6**
Precision (%)	87.8±2.0	**95.3±2.4**
Accuracy (%)	88.5±0.8	**94.8±4.4**
F1-Score (%)	87.3±2.9	**95.7±3.1**
Slide-Level Metrics		
DICE Score (%)	80.5±2.4	**90.5±2.0**
Jaccard Index (%)	75.5±1.4	**82.6±0.8**
Lesion-Level FROC Score (%)	79.0±2.5	**84.6±2.8**
Slide-Level AUC	0.91±0.01	**0.96±0.01**

Table 4.3 compares the performance of the tissue feature extractor model with the proposed model in terms of different patch and slide-level evaluation metrics on CAMELYON16 test set. As shown in the table, the model using only tissue features exhibit lower patch-level sensitivity (88.4%) and specificity (91.0%) as compared to the complete model, which incorporates the nuclei information and selective neighborhood attention, which achieved 96.2% sensitivity and 94.5% specificity. Precision and F1-score have also improved substantially, from 87.8% and 87.3% to 95.3% and 95.7%, respectively. Accuracy has increased from 88.5% to 94.8%, while the DICE score and Jaccard Index increased from 80.5% and 75.4% to 90.5% and 82.6%, respectively. These improvements underscore the benefits of incorporating nuclei features and neighborhood

attention, particularly in analyzing complex tissue patterns and refining tumor-normal boundaries identified by the model. The slide-level AUC and FROC metrics further confirm the superior performance of the complete model in WSI classification.

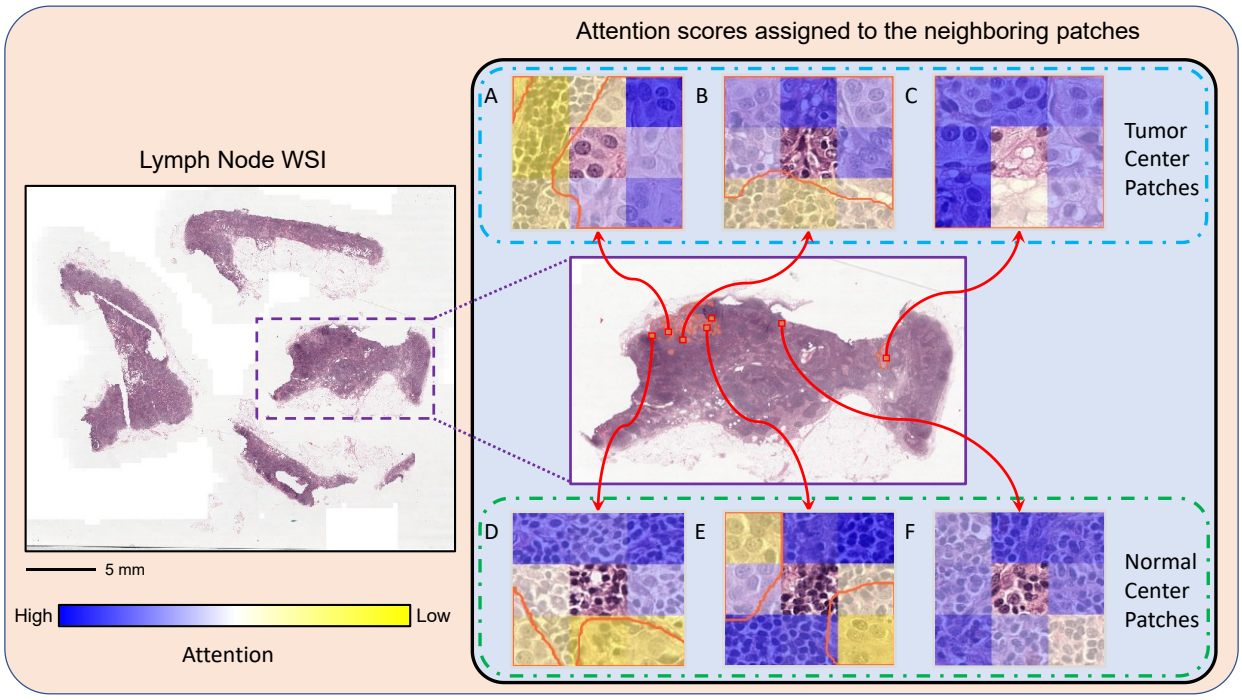

Figure 4.8: A representative lymph node WSI with attention scores on six 3×3 patch neighborhood grids from different boundary regions (A-F). The ground-truth tumor-normal tissue boundaries are shown by orange contours. In panels A-C, the center patch (classification target) is a tumor patch, with neighboring normal patches (on the other side of the boundary) receiving low attention. In panels D-F, the center patch is a normal patch, with neighboring tumor patches receiving low attention.

4.3.2 Attention Visualization

To better understand how the proposed neighborhood attention mechanism contributes to patch-level classification and boundary delineation, the neighborhood attention space in the proposed model is visualized for different regions within a representative WSI in **Figure 4.8.** This figure demonstrates the distribution of attention scores assigned to neighboring patches of different center (target) patches near the tumor-normal-tissue boundaries. For tumor center patches (A, B, C), neighboring tumor patches have received higher attention scores, emphasizing relevant

features and aiding accurate tumor localization. In contrast, for normal center patches (D, E, F), attention scores are higher for neighboring normal patches, ensuring consistent classification within normal regions. An interesting insight from **Figure 4.8** is the model's ability to effectively handle tumor-normal boundaries. For example, in patch (B), the model assigns higher attention to tumor-like regions while reducing focus on less relevant areas (other side of the boundary.) This dynamic attention distribution highlights the model's capability to adapt to mixed regions, which is crucial for precise tumor boundary delineation.

4.4 Discussion and Conclusion

In this study, a selective neighborhood attention-based deep learning framework was developed and evaluated for the automated detection and localization of metastatic tumor regions in lymph node WSIs. The proposed approach leverages detailed nuclei segmentation features with high-level tissue embeddings and selective neighborhood attention to enhance classification at the patch and slide levels. The model was tested on the CAMELYON16 and CAMELYON17 datasets, demonstrating strong potential to support clinical practice by accurately identifying tumor regions and providing reliable patient-level staging.

On the CAMELYON16 test set, the proposed model demonstrated a high patch-level classification performance, with a sensitivity of 96.2±1.5%, precision of 95.3±2.4%, and an F1-score of 95.7±3.1%. These metrics highlight the model's ability to accurately detect tumor-positive patches while minimizing false positives. The model's effectiveness in handling challenging tumor–normal boundaries was reflected in its Dice score of 90.5±2.0% and Jaccard Index of 82.6±0.8%, both indicating precise delineation of metastatic lesions within a WSI. Its robustness in detecting and localizing metastatic lesions across large WSIs was further demonstrated by achieving a lesion-level FROC score of 84.6±2.8% and a slide-level AUC of 0.96±0.01%.

Evaluation on the CAMELYON17 dataset OOD samples from multiple clinical centers demonstrated the model's generalizability. The patch-level F1-score of 87.0±1.8% and slide-level AUC of 88.0±3.0% indicated stable performance across variations in sample preparation, staining, and scanning protocols. Notably, the model achieved a kappa score of 0.94±0.02 in patient-level staging on the CAMELYON17 dataset, reflecting near-expert performance in distinguishing between the five pN stages of lymph node metastasis. This underscores its clinical utility in determining the extent of metastasis, which is critical for tailoring subsequent treatment strategies.

Central to the model's efficacy is the integration of nuclei-level features with higher-level tissue embeddings. By incorporating nuclei segmentation and classification outputs, the model can more effectively distinguish subtle morphological cues, particularly around tumor–normal tissue boundaries where nuclear shape, organization, and chromatin texture often vary markedly for different cells. The selective neighborhood attention mechanism further refines classification by emphasizing patches with similar local features, thereby reducing errors caused by distant or dissimilar patches influencing the context. Visual comparisons of tumor maps, with and without nuclei features and selective neighborhood attention, demonstrated more accurate boundary delineations and fewer misclassifications, particularly for small metastases.

The ablation studies highlighted the contribution of individual design components. Removing nuclei segmentation features or eliminating the selective neighborhood attention resulted in noticeable declines in accuracy and an increase in boundary errors, particularly in complex tissue regions. Without nuclei-derived inputs, the model occasionally struggled to detect smaller clusters of malignant cells or distinguish borderline patches. Meanwhile, the absence of selective attention deteriorated the over-segmentation of healthy tissue near tumors.

The hierarchical loss function, adapted for training the proposed model, captures both instance-level and slide-level patterns necessary for accurate tumor detection in the WSIs while maintaining stable training dynamics. The instance-level loss provides critical supervision for feature learning, the attention regularization term ensures a robust attention weight distribution, and the bag-level loss guides the model toward accurate whole-slide classification. The weighted combination of these components enables our model to effectively leverage both local and global information through a stable training process. Ablation studies confirmed that this three-component loss function outperforms single-component alternatives.

The proposed model demonstrated strong performance on CAMELYON datasets, which include WSIs of lymph node biopsy samples acquired from multiple institutions. However, further validation on diverse histopathological datasets and broader tumor types is needed to confirm the model's robustness and generalizability in the clinic.

Overall, the proposed attention-based approach offers clear advantages in detecting and annotating metastatic tumor regions, with particularly compelling performance at tumor–normal boundaries, a common challenge for standard patch-level classification methods. By combining nuclei-level and tissue-level information with the selective neighborhood attention strategy, the model closely mimics the detailed workflow of pathologists, reducing inaccuracies in automated WSI annotation. These findings suggest that the proposed framework could serve as a valuable support tool in computational pathology, potentially extending beyond metastasis detection and staging to diverse applications such as tumor grading [45], subtype classification [46,47], and therapy outcome prediction [48].

References

[1] World Health Organization. Breast Cancer. World Health Organization; 2024.

[2] Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991;19:403–410. https://doi.org/10.1111/j.1365-2559.1991.tb00229.x

[3] American Cancer Society. Survival Rates for Breast Cancer. American Cancer Society; 2024.

[4] Fitzgibbons P, Washington MK, Murphy D. Clinical practice patterns and cost-effectiveness of human epidermal growth receptor 2 testing strategies in breast cancer patients. Cancer 2010;116:3980–1; author reply 3981. https://doi.org/10.1002/cncr.25186.

[5] Vestjens JHMJ, Pepels MJ, de Boer M, Borm GF, van Deurzen CHM, van Diest PJ, et al. Relevant impact of central pathology review on nodal classification in individual breast cancer patients. Ann Oncol 2012;23:2561–6. https://doi.org/10.1093/annonc/mds072.

[6] Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: A survey. Med Image Anal 2023;88:102802. https://doi.org/10.1016/j.media.2023.102802.

[7] Hekler A, Utikal JS, Enk AH, Solass W, Schmitt M, Klode J, et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. Eur J Cancer 2019;118:91–6. https://doi.org/10.1016/j.ejca.2019.06.012.

[8] Sharma Y, Shrivastava A, Ehsan L, Moskaluk CA, Syed S, Brown DE. Cluster-to-Conquer: A Framework for End-to-End Multi-Instance Learning for Whole Slide Image Classification. Proc Mach Learn Res 2021;143:682–98.

[9] Boyd J, Liashuha M, Deutsch E, Paragios N, Christodoulidis S, Vakalopoulou M. Self-Supervised Representation Learning using Visual Field Expansion on Digital Pathology. 2021 IEEE/CVF Int Conf Comput Vis Workshops (ICCVW) 2021;639–47. https://doi.org/10.1109/ICCVW54120.2021.00077.

[10] Sikaroudi M, Hosseini M, Gonzalez R, Rahnamayan S, Tizhoosh HR. Generalization of vision pre-trained models for histopathology. Sci Rep 2023;13:6065. https://doi.org/10.1038/s41598-023-33348-z.

[11] Wang J, Mao Y, Guan N, Xue CJ. Advances in Multiple Instance Learning for Whole Slide Image Analysis: Techniques, Challenges, and Future Directions. arXiv: 2408.09476, 2024.

[12] Ilse M, Tomczak J, Welling M. Attention-based Deep Multiple Instance Learning. Proc 35th Int Conf Mach Learn 2018;80:2127–36. https://proceedings.mlr.press/v80/ilse18a.html.

[13] Lu MY, Williamson DFK, Chen TY, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng 2021;5:555–70. https://doi.org/10.1038/s41551-020-00682-w.

[14] Tang W, Huang S, Zhang X, Zhou F, Zhang Y, Liu B. Multiple Instance Learning Framework with Masked Hard Instance Mining for Whole Slide Image Classification. Proc IEEE/CVF Int Conf Comput Vis (ICCV) 2023;4078–87.

[15] Fang H, Huang S, Tang W, Huangfu L, Liu B. SAM-MIL: A Spatial Contextual Aware Multiple Instance Learning Approach for Whole Slide Image Classification. Proc 32nd ACM Int Conf Multimedia 2024;6083–92. https://doi.org/10.1145/3664647.3681534.

[16] Le GB, Nguyen VT, Le TN, Tran MT. NearbyPatchCL: Leveraging Nearby Patches for Self-Supervised Patch-Level Multi-Class Classification in Whole-Slide Images. MMM 2024 - 30th Int Conf Multimedia Model 2024. https://doi.org/10.1007/978-3-031-53308-2_18

[17] Alzoubi I, Zhang L, Zheng Y, Loh C, Wang X, Graeber MB. PathoGraph: An Attention-Based Graph Neural Network Capable of Prognostication Based on CD276 Labelling of Malignant Glioma Cells. Cancers (Basel) 2024;16:750. https://doi.org/10.3390/cancers16040750.

[18] Drioua WR, Benamrane N, Sais L. Breast Cancer Histopathological Images Segmentation Using Deep Learning. Sensors 2023;23(17):7318. https://doi.org/10.3390/s23177318.

[19] Addo D, Zhou S, Sarpong K, Nartey OT, Abdullah MA, Ukwuoma CC, et al. A hybrid lightweight breast cancer classification framework using the histopathological images. Biocybern Biomed Eng 2024;44:31–54. https://doi.org/10.1016/j.bbe.2023.12.003.

[20] Sirinukunwattana K, Ahmed Raza SE, Yee-Wah Tsang, Snead DRJ, Cree IA, Rajpoot NM. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. IEEE Trans Med Imaging 2016;35:1196–206. https://doi.org/10.1109/TMI.2016.2525803.

[21] Graham S, Vu QD, Raza SEA, Azam A, Tsang YW, Kwak JT, Rajpoot N. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal 2019;58:101563. https://doi.org/10.1016/j.media.2019.101563.

[22] Abels E, Pantanowitz L, Aeffner F, Zarella MD, van der Laak J, Bui MM, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. J Pathol 2019;249:286–94. https://doi.org/10.1002/path.5331.

[23] CAMELYON16: Grand Challenge on Cancer Metastasis Detection in Lymph Nodes. Grand Challenge. 2016. Available from: https://camelyon16.grand-challenge.org/

[24] CAMELYON17: Grand Challenge on Cancer Metastasis Detection in Lymph Nodes. Grand Challenge. n.d. Available from: https://camelyon17.grand-challenge.org/

[25] Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA 2017;318:2199–210. https://doi.org/10.1001/jama.2017.14585.

[26] Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans Syst Man Cybern 1979;9:62–6. https://doi.org/10.1109/TSMC.1979.4310076.

[27] Ming-Kuei Hu. Visual pattern recognition by moment invariants. IEEE Trans Inf Theory 1962;8:179–87. https://doi.org/10.1109/TIT.1962.1057692.

[28] Zink D, Fischer AH, Nickerson JA. Nuclear structure in cancer cells. Nat Rev Cancer 2004;4:677–87. https://doi.org/10.1038/nrc1430.

[29] Singh I, Lele TP. Nuclear Morphological Abnormalities in Cancer: A Search for Unifying Mechanisms. Results Probl Cell Differ 2022;70:443–67. https://doi.org/10.1007/978-3-031-06573-6_16.

[30] Bianconi F, Kather JN, Reyes-Aldasoro CC. Experimental Assessment of Color Deconvolution and Color Normalization for Automated Classification of Histology Images Stained with Hematoxylin and Eosin. Cancers (Basel) 2020;12(11):3337. https://doi.org/10.3390/cancers12113337

[31] Zhao S, Chen D-P, Fu T, Yang J-C, Ma D, Zhu X-Z, et al. Single-cell morphological and topological atlas reveals the ecosystem diversity of human breast cancer. Nat Commun 2023;14:6796. https://doi.org/10.1038/s41467-023-42504-y.

[32] Yogesan K, Jørgensen T, Albregtsen F, Tveter KJ, Danielsen HE. Entropy-based texture analysis of chromatin structure in advanced prostate cancer. Cytometry 1996;24:268–76. https://doi.org/10.1002/(SICI)1097-0320(19960701)24:3<268::AID-CYTO10>3.0.CO;2-O.

[33] Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Trans Syst Man Cybern 1973;SMC-3:610–21. https://doi.org/10.1109/TSMC.1973.4309314.

[34] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002;24:971–87. https://doi.org/10.1109/TPAMI.2002.1017623.

[35] Mäenpää T, Pietikäinen M. Texture Analysis with Local Binary Patterns. In: Handbook of Pattern Recognition and Computer Vision. World Scientific; 2005. p. 197–216. https://doi.org/10.1142/9789812775320_0011.

[36] Wang X, Yang S, Zhang J, Wang M, Zhang J, Yang W, et al. Transformer-based unsupervised contrastive learning for histopathological image classification. Med Image Anal 2022;81:102559. https://doi.org/10.1016/j.media.2022.102559.

[37] Shamir RR, Duchin Y, Kim J, Sapiro G, Harel NY. Continuous Dice Coefficient: A Method for Evaluating Probabilistic Segmentations. bioRxiv 2018. https://doi.org/10.1101/306977

[38] Xie X, Chen J, Li Y, Shen L, Ma K, Zheng Y. Instance-Aware Self-supervised Learning for Nuclei Segmentation. Med Image Comput Comput Assist Interv – MICCAI 2020; 2020:341–50. https://doi.org/10.1007/978-3-030-59722-1_33.

[39] Eelbode T, Bertels J, Berman M, Vandermeulen D, Maes F, Bisschops R, Blaschko MB. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. IEEE Trans Med Imaging 2020;39(11):3679–90. https://doi.org/10.1109/TMI.2020.3002417.

[40] Shao Z, Bian H, Chen Y, Wang Y, Zhang J, Ji X, et al. TransMIL: Transformer-Based Correlated Multiple Instance Learning for Whole Slide Image Classification. Adv Neural Inf Process Syst 2021;34:2136–47.

[41] Li B, Li Y, Eliceiri KW. Dual-Stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-Supervised Contrastive Learning. Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR) 2021;14318–28. https://doi.org/10.1109/CVPR46437.2021.01409

[42] Zhang H, Meng Y, Zhao Y, Qiao Y, Yang X, Coupland SE, Zheng Y. DTFD-MIL: Double-Tier Feature Distillation Multiple Instance Learning for Histopathology Whole Slide Image Classification. Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR) 2022;18802–12. https://doi.org/10.1109/CVPR52688.2022.01824

[43] Hegde P, Yıldız Ç, Lähdesmäki H, Kaski S, Heinonen M. Variational Multiple Shooting for Bayesian ODEs with Gaussian Processes. Proc 38th Conf Uncertain Artif Intell 2022;180:790–9. Available from: https://proceedings.mlr.press/v180/hegde22a.html.

[44] van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res 2008;9:2579–605. Available from: http://jmlr.org/papers/v9/vandermaaten08a.html.

[45] Lagree A, Shiner A, Alera MA, Fleshner L, Law E, Law B, Lu FI, Dodington D, Gandhi S, Slodkowska EA, Shenfield A, Jerzak KJ, Sadeghi-Naini A, Tran WT. Assessment of Digital Pathology Imaging Biomarkers Associated with Breast Cancer Histologic Grade. Curr Oncol 2021;28(6):4298–316. https://doi.org/10.3390/curroncol28060366.

[46] Saednia K, Tran WT, Sadeghi-Naini A. Automatic characterization of breast lesions using multi-scale attention-guided deep learning of digital histology images. Comput Methods Biomech Biomed Eng Imaging Vis 2022;11(1):103–11. https://doi.org/10.1080/21681163.2022.2058415.

[47] Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, Perou CM, Troester MA, Niethammer M. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. NPJ Breast Cancer 2018;4:30. https://doi.org/10.1038/s41523-018-0079-1.

[48] Saednia K, Tran WT, Sadeghi-Naini A. A hierarchical self-attention-guided deep learning framework to predict breast cancer response to chemotherapy using pre-treatment tumor biopsies. Med Phys 2023;50(12):7852–64. https://doi.org/10.1002/mp.16574.

Chapter 5

Summary, Conclusion and Future Work

5.1 Summary & Conclusion

This thesis introduces two novel deep learning frameworks. The primary framework is a selective neighborhood attention–based model designed for detecting and localizing metastatic tumor regions in whole-slide images (WSIs). The model was trained and evaluated on the CAMELYON16 dataset (n = 269 WSIs for training and n = 129 WSIs for testing) and further assessed on the CAMELYON17 dataset for out-of-distribution performance (n = 30 WSIs) and patient-level pN-staging (n = 500 WSIs from 100 patients) [1,2]. The second framework, TexSegNet, is a multi-branch network for nuclei segmentation and classification. TexSegNet was developed using the PanNuke dataset [3].

In the framework developed for detection, localization, and staging of breast cancer lymph node metastases, the input WSIs are divided into smaller patches, which are processed through a ViT–based image feature extractor [4]. The extracted features are combined with nuclei features derived from segmentation and classification of the input patches. These combined features, together with information from the eight neighboring patches, are refined by a selective neighborhood attention module that emphasizes the local context surrounding each target patch. The filtered features are then input to a classification network that assigns labels to the patches. The integration of this attention module mitigates challenges inherent in MIL approaches, such as the imbalance of positive patches and difficulties in accurately delineating the boundaries of tumor and normal tissues [5]. The results on the CAMELYON16 test set indicate a patch-level sensitivity of 96.2%,

a precision of 95.3%, and an F1-score of 95.7%, with an AUC of 0.96 for slide-level classification. When evaluated on CAMELYON17, the model achieved a patch-level F1-score of 87.0% and a slide-level AUC of 0.88, demonstrating robustness against variations in staining, scanner differences, and data from new medical centers. Further, for 500 WSIs from 100 patients in the CAMELYON17 dataset, the framework demonstrated a Cohen's kappa of 0.94 in patient-level pN-staging, indicating near-expert concordance in detecting and classifying the extent of nodal metastasis.

A separate yet complementary objective was the development of TexSegNet for nuclei segmentation and classification. Utilizing a pretrained MaxViT encoder and a custom-designed decoder, TexSegNet employs an encoder-decoder architecture with texture extraction blocks to capture and quantify features from nuclei [6]. A key innovation is the integration of the nuclei segmentation masks with raw images, which is subsequently incorporated into the decoder's input of the classification branch of the decoder's input. In comparative evaluations against leading methods such as HoVer-Net, CellViT, and StarDist, TexSegNet achieved an overall accuracy of 81.4% in detecting and classifying nuclei on the breast test set of the PanNuke dataset, outperforming CellViT (77.5%), HoVer-Net (75.2%), and StarDist (70.7%). Class-specific F1-scores were 89.3% for neoplastic, 88.9% for inflammatory, 84.3% for connective tissue, and 91.1% for epithelial nuclei [7,8,9]. TexSegNet's performance improvements stem from its incorporation of multi-scale convolutions, mudslide boundary maps, texture-aware representations, and a feedback loop of nuclei segmentation maps. These enhancements refine segmentation boundaries and reduce false positives and negatives, particularly in densely clustered regions.

The broader implications of this research suggest that the proposed AI-driven pipeline can support and enhance traditional pathology practices. Data-driven automated methods of pathology image analysis are capable of detecting intricate diagnostic and prognostic cues from routine H&E slides, flagging suspicious tumor areas for closer examination, and systematically highlighting critical features such as micro-metastases [10]. Moreover, the integration of nuclei segmentation models provides valuable morphological and textural details that support advanced downstream tasks, including nuclear morphology assessment for grading, evaluation of tumor microenvironment changes, and identification of specific cell types. This multi-layered approach holds promise for deriving robust biomarkers for patient prognosis and treatment stratification and could potentially be adapted to other cancer types using large clinical datasets [11].

In summary, this thesis introduced and investigated a multi-level computational pathology model that leverages selective neighborhood attention for metastasis detection, annotation, and staging in digital WSIs, alongside a texture-aware nuclei segmentation and classification network. Results on the CAMELYON16 and CAMELYON17 datasets confirm the model's ability to accurately localize and quantify tumor regions while maintaining clinically essential specificity, reducing both missed and falsely identified lesions. Additionally, the integrated nuclei segmentation and classification scheme provides important morphological and textural details, resulting in improved delineation of tumor boundaries and enhanced characterization of individual cell types. The overall pipeline offers multiple pathways for integration into clinical workflows, including use as a triaging assistant in pathology laboratories, ensuring consistent and standardized reporting, and providing a foundation for the development of new AI-driven biomarkers.

5.2 Future Work

While the outcomes of this study are promising, there remain several avenues for future research and development to further improve and translate this work into routine clinical practice:

Algorithmic Refinement and Explainability: Enhancing the model's architecture and transparency should be a key priority moving forward. Future versions of the algorithm could explore advanced network designs, such as graph neural networks, to capture even more complex tissue patterns. Additionally, improving interpretability is crucial. For example, integrating more advanced attention visualization or explainable AI techniques would enable the model to highlight histological features influencing its decisions. This would not only potentially boost performance but also foster clinician trust, as the lack of explainability is a well-known challenge in pathology AI deployments [12]. By providing heatmaps or textual explanations for its predictions, the model could evolve into a support tool that pathologists and oncologists can comfortably rely on for insights [13].

Dataset Expansion and Generalization: Expanding the dataset is crucial to ensure the model generalizes effectively across diverse settings. We plan to train and evaluate the model on larger, multi-institutional datasets that encompass a larger variety of scanners, sample preparation protocols, and patient populations. This approach will address the current limitation of domain specificity and help the model learn invariant features that are robust to staining differences or image artifacts. Additionally, techniques like domain adaptation [14] and federated learning [15] could be utilized to allow the model to adapt to new data from hospitals while preserving patient privacy. Strengthening generalization is essential, as many AI models in computational pathology experience significant performance degradation on out-of-distribution datasets [13].

Clinical Integration and Workflow Validation: Transitioning from research to clinical practice will require addressing practical implementation challenges. Collaboration with pathologists is vital to integrate the AI system into existing pathology workflows, for example, as a plug-in for digital pathology software that flags ROI or provides second-opinion classifications. Extensive validation through prospective pilot studies and clinical trials will be essential to assess the impact of the tool on diagnostic accuracy, efficiency, and patient outcomes. Concurrently, addressing regulatory and standardization issues is crucial as AI adoption in pathology has lagged behind other fields, such as radiology, partly due to the need for clear guidelines and regulatory approval for diagnostic algorithms [16].

References

[1] Bejnordi BE, Veta M, van Diest PJ, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA; 2017;318(22):2199–2210. https://doi.org/10.1001/jama.2017.14585.

[2] Litjens G, Bandi P, Bulten W, et al. Analysis of the CAMELYON17 Challenge for Metastasis Detection in Lymph Nodes. IEEE Trans Med Imaging; 2018;37(5):1049–1059. https://doi.org/10.1109/TMI.2018.2803308.

[3] Kumar N, Verma R, Sharma S, et al. PanNuke: An Image Dataset for Nuclei Instance Segmentation and Classification in Histology Images. IEEE Trans Med Imaging; 2019;38(10):2357–2368. https://doi.org/10.1109/TMI.2019.2931432.

[4] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: International Conference on Learning Representations; 2020. Retrieved from https://openreview.net/forum?id=YicbFdNTTy.

[5] Ilse M, Tomczak JM, Welling M. Attention-based Deep Multiple Instance Learning. In: Proceedings of the 35th International Conference on Machine Learning; 2018:2127–2136. https://doi.org/10.1007/978-3-319-70112-1_132.

[6] Beltagy I, Hendrycks D, Song C. MaxViT: Multi-Axis Vision Transformer. arXiv preprint arXiv:2104.01411; 2021. Retrieved from https://arxiv.org/abs/2104.01411.

[7] Graham S, Vu QD, Raza SE, et al. Hover-Net: Simultaneous Segmentation and Classification of Nuclei in Multi-Tissue Histology Images. Med Image Anal; 2019;58:101563. https://doi.org/10.1016/j.media.2019.101563.

[8] Kumar A, Singh R, Mehta P, et al. CellViT: A Vision Transformer Approach for Cellular Image Analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022:1234–1243. https://doi.org/10.1109/CVPR52688.2022.01824.

[9] Schmidt U, Weigert M, Broaddus C, Myers G. Cell Detection with Star-Convex Polygons. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018; Lecture Notes in Computer Science; 2018:265–273. https://doi.org/10.1007/978-3-030-00934-2_31.

[10] Campanella G, Hanna MG, Geneslaw L, et al. Clinical-Grade Computational Pathology Using Weakly Supervised Deep Learning on Whole Slide Images. Nat Med; 2019;25(8):1301–1309. https://doi.org/10.1038/s41591-019-0508-1.

[11] Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng; 2017;19:221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442.

[12] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference

on Knowledge Discovery and Data Mining; 2016:1135–1144. https://doi.org/10.1145/2939672.2939778.

[13] Tjoa E, Guan C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. IEEE Trans Neural Netw Learn Syst; 2020;31(11):4619–4631. https://doi.org/10.1109/TNNLS.2020.2988276.

[14] Mormont M, Ghesu FC, Georgescu B, et al. Domain Adaptation for Medical Image Analysis: A Review of Recent Advances. IEEE Access; 2020;8:12345–12353. https://doi.org/10.1109/ACCESS.2020.2976230.

[15] Chang K, Balachandar N, Lam C, Yi D, Brown J, Beers A, Rosen B, Rubin DL, Kalpathy-Cramer J. Distributed deep learning networks among institutions for medical imaging. Journal of the American Medical Informatics Association. 2018;25(8):945–954. doi:10.1093/jamia/ocy017

[16] McMahan HB, Moore E, Ramage D, Hampson S, Arcas BA. Communication-Efficient Learning of Deep Networks from Decentralized Data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics; 2017:1273–1282. https://doi.org/10.48550/arXiv.1602.05629.