



Graph Neural Network Analysis on Pancreatic Cancer Tumor Grading Classification

An AI approach for ameliorating pancreatic cancer classification of grades,
utilizing whole slide histopathology images for future clinical use

Arianna I. Parada Zambrano¹

Candidate Number: MZYN5

MSc AI for Biomedicine & Healthcare

Supervisors: Petru Manescu, Delmiro Fernandez-Reyes, Fabio Caccioli

Submission date: 08 September 2025

¹**Disclaimer:** This report is submitted as part requirement for the MSc AI for Biomedicine & Healthcare degree at UCL. It is substantially the result of my own work except where explicitly indicated in the text. The report will be distributed to the internal and external examiners, but thereafter may not be copied or distributed except with permission from the author.

Abstract

Pancreatic cancer remains one of the most aggressive and lethal malignancies, with survival rates stagnating around 10% despite advances in oncology. Accurate histopathological grading of tumors is crucial for prognosis and treatment planning, yet the task is hindered by complex tissue morphology and significant patient variability. Computational pathology offers a potential solution, but conventional deep learning approaches, such as convolutional neural networks, and traditional machine learning struggle to capture the relational and architectural features that are essential in pancreatic histology.

This dissertation investigates whether graph neural networks (GNNs), in combination with advanced feature extraction strategies, can improve the automated grading of pancreatic cancer from haematoxylin and eosin (H&E) whole slide images (WSIs). A complete pipeline was developed using The Cancer Genome Atlas (TCGA) dataset, encompassing tiling of WSIs, feature extraction, graph construction, and GNN training. Three feature extraction methods were compared: a visual contrastive learning framework (SimCLR), a convolutional baseline (ResNet50), and a histopathology-specific foundation model (UNI). Two graph construction strategies were evaluated: spatial graphs preserving morphological adjacency, and embedding graphs linking patches by feature similarity.

Results revealed that feature extractor choice critically shaped downstream performance. SimCLR embeddings, though convergent during training, produced poor grading accuracy (26.9%) and collapsed onto majority classes, highlighting the limitations of contrastive learning with small, imbalanced datasets. ResNet50 achieved moderate improvements (42.9%), but the foundation model UNI outperformed both, achieving the highest accuracy (57.1%) and balanced accuracy (30.2%). Spatial graphs were used to provide more interpretable results by maintaining histological context, as cancer cells will appear in microenvironments, while embedding graphs highlighted latent feature similarities between cells but risked losing architectural integrity present in tumor microenvironment spaces.

These findings, while failing to support the hypothesis to achieve a GNN that classifies tumor grades as well as other types of cancers, it does show that foundation models significantly enhance classification quality in computational histopathology for pancreatic cancer, while also emphasizing the persistent challenges of data imbalance, limited interpretability, and modest absolute accuracy. The most striking result was the contrast between the poor performance of SimCLR and the robustness of UNI, highlighting the importance of large scale pretraining for rare and histologically complex cancers. This work demonstrates both the feasibility and the limitations of GNN-based frameworks, laying the groundwork for future research toward efficient tumor grading classifications for clinically translatable aids.

Contents

1	Introduction	2
1.0.1	Clinical Obstacle	2
1.0.2	Diagnostic & Histopathological Challenges	3
1.0.3	Problem Urgency	3
1.0.4	The Research Problem	4
1.0.5	Aims, Objectives, & Hypothesis	4
2	Background	5
2.0.1	Introduction	5
2.0.2	GNNs in Computational Pathology	5
2.0.3	Cross-Modality Approaches	6
2.0.4	State of the Art Gaps	7
2.0.5	Emerging Strategies	7
2.0.6	Clinical Relevance	7
2.0.7	Summary	8
3	Materials & Methods	9
3.0.1	TCGA Data Download	9
3.0.2	WSI Tiling	9
3.0.3	Feature Extraction	10
3.0.4	Graph Construction	10
3.0.5	Labels & Splits	11
3.0.6	Graph Classification	11
3.0.7	Training	12
3.0.8	Evaluation	12
3.0.9	Graph Analysis	13
4	Results	14
4.0.1	Exploratory Data Analysis	14
4.0.2	Graph Representations	15
4.0.3	Contrastive Learning (SimCLR)	16
4.0.4	Baseline Model: Resnet50	18
4.0.5	Foundation Model: UNI	19
4.0.6	Spatial vs Embedding Graphs	20

4.0.7	Implications	21
5	Discussion	26
5.0.1	Hypothesis	26
5.0.2	Evaluation of Objectives	27
5.0.3	Strengths & Limitations	28
5.0.4	Conclusion	30
6	Conclusions	31
7	Future Work	34
7.0.1	Addressing Limitations	34
7.0.2	Improving Feature Extraction	35
7.0.3	Advancing Graph Representations	35
7.0.4	Enhancing Model Architecture	36
7.0.5	Improving Interpretability	36
7.0.6	Improving Evaluations	36
7.0.7	Clinical Translation	37
7.0.8	Conclusion	37
A	Supplementary Figures	43

Chapter 1

Introduction

1.0.1 Clinical Obstacle

Pancreatic cancer is one of the most devastating malignancies in modern medicine, defined not only by its aggressiveness but also by the lack of meaningful improvements in survival over recent decades. Although it represents only around 3% of all new cancer cases in the United Kingdom, it ranks as the fifth leading cause of cancer death and is projected to become the second leading cause of cancer-related mortality worldwide by 2030 [1]. This disproportionate burden highlights the severity of the disease. The one, five, and ten year survival rates remain approximately 30%, 10%, and 5% respectively [1]. Statistics that highly contrast with improvements seen in other common cancers such as breast and colorectal. Projections also suggest that the incidence of pancreatic cancer will continue to rise by 5% by 2038–2040, further exacerbating its public health impact [1].

The principal reason for these poor outcomes is late detection. Early-stage pancreatic cancer is largely asymptomatic, and when symptoms do occur they are vague and non-specific, including abdominal discomfort, jaundice, weight loss, and new-onset diabetes [1]. These features overlap with benign conditions, delaying diagnosis until the disease is already locally advanced or metastatic. Consequently, only 10–15% of patients present have tumors capable of surgical resection, the only potentially curative option [1]. Even in this small subset, long-term outcomes remain poor, with five-year survival reaching just 17% [1]. The majority of patients are instead managed with palliative chemotherapy, which provide limited survival benefits and often carry significant toxicity [1].

The biological underpinnings of pancreatic cancer further complicate management. The disease is characterized by widespread mutations in KRAS, as well as frequent alterations in TP53, CDKN2A, and SMAD4 [1]. These mutations drive rapid progression, therapeutic resistance, and profound genomic instability. Moreover, the disease exhibits a dense desmoplastic stroma, dense connective tissue forming around the tumor, that contributes to poor drug delivery and complex tumor microenvironment interactions [1]. Lifestyle and hereditary risk factors such as smoking, obesity, chronic pancreatitis, and inherited cancer syndromes contribute additional layers of heterogeneity [1]. In this way, pancreatic cancer embodies both clinical and biological complexity, combining a silent presentation with aggressive progression to make it one of the most difficult

cancers.

1.0.2 Diagnostic & Histopathological Challenges

Diagnosis of pancreatic cancer ultimately rests on histopathological evaluation of H&E-stained tissue samples. Histopathology is critically essential not only for confirming malignancy but also for grading tumors, which provides prognostic information and guides treatment planning. Tumor grade reflects the degree of glandular differentiation (ability to form tubular structures necessary for aggressive tumor microenvironments), nuclear atypia (abnormal appearance of cell's nuclei), and architectural organization in the microenvironment. High grade tumors, with pancreatic tumors being graded from G1-G4, are associated with more aggressive clinical behavior and poorer outcomes, emphasizing the importance of accurate grading for patient management [2].

However, pancreatic cancer presents unique challenges to histopathologists. The disease is marked by highly variable histological appearances, often featuring extensive fibrosis, distorted glandular structures, and complex stromal-tumor interactions. These features make grading more subjective than in many other cancer types [2]. Variability between observers is a well-documented issue, with pathologists frequently disagreeing on borderline cases. This variability undermines reproducibility and has downstream consequences for treatment decisions and research studies that rely on accurate grading as a baseline.

These diagnostic limitations emphasize the need for new approaches to support pathologists. Computational pathology, which uses artificial intelligence (AI) and machine learning to analyze whole slide images (WSIs), offers the potential to provide consistent, objective, and scalable tools. While AI has shown promise in other cancers such as breast and prostate, pancreatic cancer remains understudied. The small number of available cases, combined with the histological complexity of the disease, has hindered the development of robust computational methods tailored to pancreatic tissue [3].

1.0.3 Problem Urgency

The clinical urgency of pancreatic cancer makes it a compelling focus for methodological innovation. Improving diagnostic reproducibility and grading accuracy could have immediate clinical implications. Earlier and more reliable identification of a high grade prognosis would allow better stratification of patients, inform surgical and therapeutic decisions, and potentially accelerate research into novel treatments. Given the projected rise in incidence and mortality, even modest improvements in diagnostic workflows could yield meaningful benefits for patient care.

From a research perspective, pancreatic cancer is also interesting because it represents one of the most challenging test cases for computational histopathology. Its histological variability, dense stroma, and subtle morphological cues require approaches that can move beyond pixel level classification to capture architecture and context. This sets pancreatic cancer apart from other cancers where computational pathology has already achieved notable successes. The challenge of applying AI methods to such a complex disease provides an opportunity to test and refine approaches that may later be applied more broadly across oncology.

Furthermore, pancreatic cancer highlights a persistent gap in the literature. While deep learning has been widely applied to histopathology in other cancers, pancreatic cancer has received

relatively little attention. This neglect presents an ironic situation given the severity of the disease and the urgent clinical need for new tools. Addressing this gap not only contributes to the field of computational pathology but also aligns research efforts with one of the most pressing challenges in cancer medicine today [3].

1.0.4 The Research Problem

Taken together, these clinical and methodological considerations define the central research problem of this dissertation. How can computational methods, and specifically graph-based approaches, be used to improve the grading of pancreatic cancer in histopathology, where manual diagnosis is subjective, data are limited, and disease biology is complex?

The problem is multifaceted and requires support from many levels. Clinically, manual grading is inconsistent and subject to variability due to the difficulty of interpreting pancreatic histology [2]. Methodologically, existing computational approaches, particularly convolutional neural networks, can be limited in their ability to capture the relational and architectural features that pathologists use to assess slides [3]. Finally, data scarcity and class imbalance restrict the application of traditional supervised or self-supervised learning methods, further challenging model development [3]. Addressing this problem requires approaches that can model the relational complexity of tissue and leverage the potential of pretrained representations to overcome dataset limitations.

1.0.5 Aims, Objectives, & Hypothesis

The overall aim of this dissertation is to develop and evaluate a computational framework for pancreatic cancer histopathology that can improve tumor grading by capturing both local morphological detail and global tissue architecture. To achieve this, the project pursues four objectives. The first is to review the existing literature on pancreatic cancer and the application of graph neural networks in computational pathology, with the aim of identifying gaps that limit current approaches. The second is to train a graph neural network model using WSIs from The Cancer Genome Atlas (TCGA), testing different strategies for feature extraction. The third is to scale these models to whole slide images, utilizing spatially adjacent graph constructions to retain the whole slide visual with the nodes being patches and edges being the spatial connection. The fourth is to explore interpretability in graph representations, assessing whether these models can provide insights into tissue structure and predictions in a way that aligns with pathologists' reasoning.

The central hypothesis of this work is that graph neural networks, when combined with appropriate feature extraction strategies, can improve the computational grading of pancreatic cancer compared to conventional deep learning approaches. Specifically, it is hypothesized we would like to aim for the GNN model that we produce for pancreatic cancer be a contender with GNNs made for lung, breast, and colorectal cancers.

Chapter 2

Background

2.0.1 Introduction

Pancreatic cancer remains one of the most pervasive malignancies, with survival outcomes that have stagnated for decades. The five-year survival rate consistently lingers around 10%, largely due to the late detection of the disease and the absence of reliable early screening strategies. Unlike other cancers where widespread screening and targeted therapies have markedly improved outcomes, pancreatic cancer is often diagnosed at an advanced stage when curative treatment options are limited. Histopathology, particularly the grading of resected tumors on hematoxylin and eosin (H&E) stained slides, remains the gold standard for diagnosis. However, the interpretive process is far from straightforward. The dense fibrosis, irregular glandular structures, and complex tumor–stroma interactions that characterize pancreatic cancer, render grading a potentially subjective variable, even among experienced pathologists [17].

Artificial intelligence (AI) has been increasingly explored as a tool to support pathologists in navigating this complexity. While convolutional neural networks (CNNs) have achieved significant success in image classification, their reliance on grid-like data structures often limits their ability to capture the rich spatial and relational context of histological tissue. Graph neural networks (GNNs) represent a natural evolution of AI for pathology, as they allow tissue to be modeled in terms of nodes (patches, nuclei, or regions) and edges (spatial or semantic relationships). This relational framework is especially valuable in histopathology, where both cellular morphology and tissue architecture influence diagnostic grading.

Despite their promise, applications of GNNs to pancreatic cancer remain rare. Most advances in computational histopathology have focused on cancers such as breast, colorectal, and prostate [4, 9, 13]. Pancreatic cancer has unique histological challenges—dense stroma, glandular ambiguity, and heterogeneity that complicate direct translation of models developed for other cancer types. Thus, there is a pressing need for tailored GNN approaches that address these complexities while improving grading accuracy and prognostic assessment in pancreatic cancer.

2.0.2 GNNs in Computational Pathology

GNNs have been widely studied in computational pathology for their ability to model varied structures. Whole slide images (WSIs) are typically divided into smaller image patches, each of

which may be treated as a node in a graph. Edges are then constructed to represent relationships between nodes, such as spatial adjacency or feature similarity. Unlike CNNs, which aggregate patch-level features independently, GNNs preserve the relational information between patches, enabling models to learn how local cellular details integrate into broader tissue structures [9].

Several studies illustrate the methodological diversity of GNNs in oncology. Li et al. introduced a causality-driven GNN (CGNN) for early diagnosis of pancreatic cancer using non-contrast CT, incorporating causal contrastive learning to enhance generalisation across institutions [1]. While innovative, the model lacked external validation, limiting confidence in its generalizability. Bazargani et al. proposed a multi-scale relational GCN for histopathology, capturing cross-magnification contexts to improve feature aggregation [2]. However, this approach assumes uniform relevance across magnifications, which may not hold for pancreatic tissue due to its structural variability. Koc et al. applied entity-graph GNNs to pancreatic neuroendocrine tumors, demonstrating interpretability through infiltration patterns but limiting applicability to the more common adenocarcinoma subtype [3].

Collectively, these examples highlight both the promise and limitations of GNNs in pathology. They demonstrate the capacity of GNNs to model structural and semantic relationships, but also underscore persistent challenges such as overfitting, limited validation, and generalizability concerns. Importantly, very few of these studies have focused specifically on pancreatic cancer, reinforcing the gap that this project seeks to address.

2.0.3 Cross-Modality Approaches

One of the major advantages of GNNs is their ability to integrate multiple data modalities. Histopathology alone provides valuable information, but cancer progression is influenced by molecular and clinical factors that are not visible in H&E slides. Multi-modal GNNs have demonstrated significant potential in other cancer contexts by combining diverse data sources.

In breast cancer, SlideGraph+ has been successfully applied to predict HER2 status and tumour subtypes directly from WSIs [4]. Similarly, Zhu et al. employed geometric GNNs with multi-omic data, while Xiao et al. incorporated protein–protein interaction networks, achieving strong predictive performance for cancer survival outcomes [5, 6]. Fu et al. and Ahmed et al. demonstrated the benefit of integrating histology with clinical variables and multiplexed imaging data to improve survival prediction [7, 8]. Pati et al. introduced hierarchical graph representations in computational pathology, segmenting tissue into semantically meaningful regions before constructing multi-scale graphs [9]. This approach allowed for improved interpretability and highlighted the potential of hierarchical frameworks for modeling tumor microenvironments.

Despite these successes, pancreatic cancer remains underrepresented in multi-modal GNN research. Pancreatic tumors are characterized by distinct stromal biology and glandular disruption that may not align with patterns learned from breast or colorectal tissue. Moreover, the integration of genomics, spatial transcriptomics, and histology has only rarely been applied to pancreatic cohorts [15]. Given the molecular heterogeneity of pancreatic cancer, multi-modal approaches could be particularly beneficial, yet their application remains largely unexplored.

2.0.4 State of the Art Gaps

The rapid development of GNNs in oncology has produced promising results, yet several limitations consistently appear across the literature.

The first is generalizability. Many models are trained and tested on data from a single institution, reducing their ability to perform across diverse populations and technical conditions [8]. Without external validation, strong internal results do not guarantee clinical robustness.

The second limitation is interpretability. Despite advances in attention mechanisms, few studies explicitly explain model decisions in clinically meaningful ways. Interpretability tools such as attention heatmaps, saliency maps, or edge importance remain underused [11, 12]. This lack of transparency undermines clinician trust and hinders adoption in practice.

A third limitation lies in annotation quality. Many models rely on weak supervision, where only slide-level labels are available. This can introduce significant noise, particularly in pancreatic cancer, where subtle morphological cues distinguish grades [13, 14, 15]. High-quality annotations are crucial for training accurate and reliable models.

Model complexity is also a concern. Architectures such as CGNNs and multi-scale relational GCNs are powerful but risk overfitting when applied to small datasets, a frequent problem in pancreatic cancer research [1, 2]. Finally, there is a general lack of pancreatic-specific studies. While GNNs have been explored extensively in breast and colorectal cancer, their direct application to pancreatic histology is rare, and the few existing studies often focus on pancreatic neuroendocrine tumors rather than adenocarcinomas [16, 17].

2.0.5 Emerging Strategies

To overcome these challenges, several strategies have been proposed. The development of pancreatic-specific datasets is paramount. Curating high-quality annotated datasets of H&E WSIs would reduce noise and provide the foundation for robust model training. Collaborative efforts across institutions could increase sample size and diversity, facilitating external validation and reducing bias [18, 19].

Interpretability must also be embedded into future GNN designs. Causality-aware models, hierarchical graph representations, and attention mechanisms offer promising avenues for explaining predictions in ways that align with clinical reasoning [19]. Similarly, hybrid models that combine GNNs with transformer architectures may improve robustness and capture long-range dependencies, though their risk of overfitting must be carefully managed [12, 16].

Finally, clinical adoption will require user centered design. Models must provide outputs that pathologists can interpret, such as region-level highlights or edge-based explanations. User-friendly interfaces that allow clinicians to validate model predictions would enhance trust and facilitate integration into workflows [10, 17].

2.0.6 Clinical Relevance

The clinical significance of advancing computational histopathology in pancreatic cancer cannot be overstated. Tumor grade is a critical prognostic factor, influencing treatment choices and survival predictions. Yet grading is complicated by subtle histological differences and inter-observer

variability. Computational models that provide consistent and accurate grading could reduce diagnostic uncertainty and improve patient stratification.

As precision medicine becomes increasingly central in oncology, there is a demand for computational models that integrate morphology with molecular and clinical features. Pancreatic cancer, notorious for its resistance to therapy and poor prognosis, could benefit greatly from such integrative models. GNNs are uniquely positioned to address this need by modeling both the local morphology of tissue and the global relational context, potentially producing predictions that are both accurate and clinically relevant.

2.0.7 Summary

In conclusion, while GNNs have demonstrated promise in computational pathology, their application to pancreatic cancer remains underdeveloped. Existing literature highlights methodological innovations in graph construction, multi-modal integration, and interpretability, but also reveals significant limitations in generalizability, annotation quality, and pancreatic-specific research. Addressing these gaps requires high-quality annotated datasets, external validation, interpretability frameworks, and hybrid architectures that balance power with robustness.

This project builds directly on these identified gaps. By implementing GNNs for pancreatic cancer histopathology, scaling models to whole slide images, and evaluating both spatial and embedding graph representations, it aims to provide new insights into the potential and limitations of graph-based approaches in one of the most clinically challenging cancers.

Chapter 3

Materials & Methods

3.0.1 TCGA Data Download

We obtained diagnostic H&E whole-slide images (WSIs) and accompanying clinical metadata from The Cancer Genome Atlas (TCGA) via the NCI Genomic Data Commons (GDC). Cohort definition and download followed a manifest-driven workflow to ensure reproducibility and integrity. We queried the GDC Data Portal for open-access slides corresponding to pancreatic cancer and filtered for histology images in SVS/TIFF formats. The resulting set was saved to a GDC “Cart,” and a manifest file (manifest.txt) was exported. The manifest lists file UUIDs, original filenames, and MD5 checksums and serves as the authoritative record of the dataset. Files were downloaded with the GDC Data Transfer Tool (gdc-client). The tool handles large file transfers, supports resume on interruption, and verifies checksums on completion. Clinical attributes used for supervision (e.g., tumor grade) were downloaded from the GDC as tab-delimited case-level tables. We harmonized heterogeneous grade descriptors (e.g., “well”, “moderately”, “poorly differentiated”) to standardized G1–G4 classes and excluded ambiguous or missing entries. A join on patient/case barcode linked the clinical table to the curated slide list. The resulting pairings (WSI barcode, label) were saved as TSV files and later split into train/validation/test sets using stratified sampling. The curated slide directory is the sole input to the tiling module. From that point forward, the pipeline proceeds as described in the Methods: tiles are generated from each SVS, patch-level features are extracted (SimCLR, ResNet50, or foundation model), graphs are constructed (spatial and/or semantic edges), and supervised training is performed using the matched TCGA grade labels. This separation between acquisition and modelling stages preserves structure and guidance and allows the data acquisition steps to be repeated or extended without changes to the learning code.

3.0.2 WSI Tiling

Whole slide images (WSIs) were partitioned into fixed size tiles to enable scalable, patch-level computation. Each slide was processed with a tiling script configured to extract 512×512 pixel patches at the native scanning level while rejecting background regions using intensity based heuristics. Multi-processing was enabled to parallelize the input output operations and image decoding across CPU workers. The procedure generated, for every WSI, a structured directory

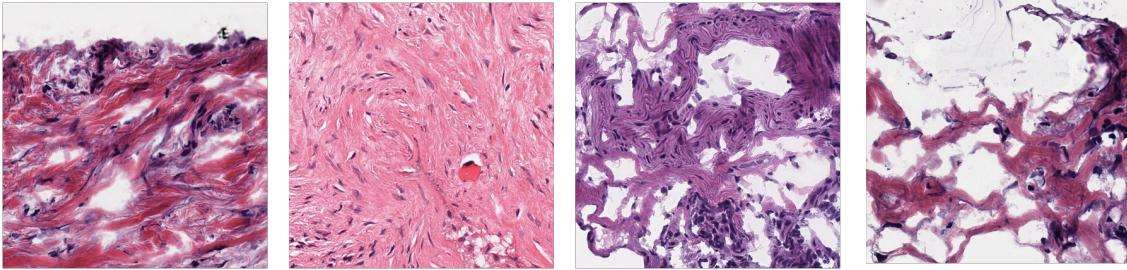


Figure 3.1: Examples of different JPEG Patches at 20X Magnification post WSI Tiling

containing JPEG tiles and accompanying metadata. To avoid redundant computation, the pipeline enumerated already-tiled slides and skipped them on subsequent runs.

3.0.3 Feature Extraction

We evaluated three backbone strategies to embed each tile into a representation suitable for graph learning. First, a locally trained feature extractor was instantiated from a configurable repository, which in this scenario was derived from the Kolachalama Lab’s paper from 2022 whereby they built a pipeline for a Graph Transformer for WSI classification on lung cancer. It operated with a contrastive learning framework (SimCLR) to extract the features with a ResNet18 backbone. The second feature extraction strategy is a standard ImageNet pretrained ResNet50. The pretrained weights were used to produce feature vectors. The final and third method is a novel foundation model for pathology (“UNI”), created by the Mahmood Lab, was accessed via a model hub through HuggingFace and used directly as is without fine-tuning. For all options, tiles were resized to the backbone’s expected input resolution, which in all cases was 224 x 224, normalized with standard channel statistics, and forwarded in batches. The resulting embeddings, one vector per tile, were stored to separate the to decouple expensive feature extraction from downstream graph construction and training.

3.0.4 Graph Construction

Each WSI was represented as a graph whose nodes correspond to tiles and whose node features are the extracted embeddings. We instantiated two complementary notions of connectivity. First, a spatial graph linked tiles based on their physical neighborhood on the slide, utilizing k-nearest neighbors (kNN) in the 2D space manufactured by the tiling process. When available, tile coordinates were retained as positional attributes to enable spatial reasoning and visualization. Second, a semantic graph linked tiles whose embeddings were similar under cosine similarity, constructed via kNN or thresholded affinity. In all cases, edges were sanitized by removing self-loops, enforcing undirectedness, and coalescing duplicates. Graphs and their tensors, containing features, edge indices, and coordinates were written per-WSI under a common root, enabling consistent loading across experiments.

```

WSIs with usable tumor grade: 408
WSIs skipped (no/unknown grade): 96

Original class distribution (by tumor grade label):
  Class 0: 67 samples
  Class 1: 226 samples
  Class 2: 111 samples
  Class 3: 4 samples

Train split (283 samples):
  Class 0: 46 samples (16.3%)
  Class 1: 158 samples (55.8%)
  Class 2: 77 samples (27.2%)
  Class 3: 2 samples (0.7%)

Validation split (62 samples):
  Class 0: 10 samples (16.1%)
  Class 1: 34 samples (54.8%)
  Class 2: 17 samples (27.4%)
  Class 3: 1 samples (1.6%)

Test split (63 samples):
  Class 0: 11 samples (17.5%)
  Class 1: 34 samples (54.0%)
  Class 2: 17 samples (27.0%)
  Class 3: 1 samples (1.6%)

```

Figure 3.2: Distinct Class Distributions for Grades I-IV

3.0.5 Labels & Splits

Despite having 232 unique patient IDs from the TCGA dataset, there were 517 whole slide images used for the classification process. Slide-level labels were derived from clinical records and normalized to standardized categories. For tumor grading, descriptive terms, such as , “well”, “moderately”, “poorly differentiated” were mapped to G1–G4 and then to integer classes. Slides with missing or ambiguous labels, such as GX, were excluded from supervised training. We performed a stratified partition into training, validation, and test sets to preserve class proportions under substantial imbalance. The splits were persisted as tab-separated lists (train.txt, val.txt, test.txt) containing WSI identifiers and labels, exact splits are outlined in figure 3.2. Notably in figure 3.2 the labels are classified as classes 0-3, however these represent the grades I-IV in order respectively. As a robustness measure, if a slide lacked a corresponding graph on disk, it was omitted and the split logic fell back to the subset of available graphs while maintaining approximate stratification.

3.0.6 Graph Classification

For the tumor grade classification we implemented a slide-level graph classifier implemented with PyTorch Geometric. We constructed a compact, regularized graph neural network that operates on tissue graphs, where nodes encode patch features, and edges capture spatial adjacency. The GNN begins with Graph Attention (GAT) layers that adaptively weights neighbor information, followed by Graph Convolutional (GCN) layers. The GCN layers consolidate the the spatial information as well as the smooth embeddings. Each layer in this neural network utilizes ReLU activations and dropout for regularization. After sharing the information between nodes along the interconnected edges, the node embeddings are summarized by both a global mean and global max pooling, creating a mixture that can yield a robust graph descriptor. Finally a simple multilayer

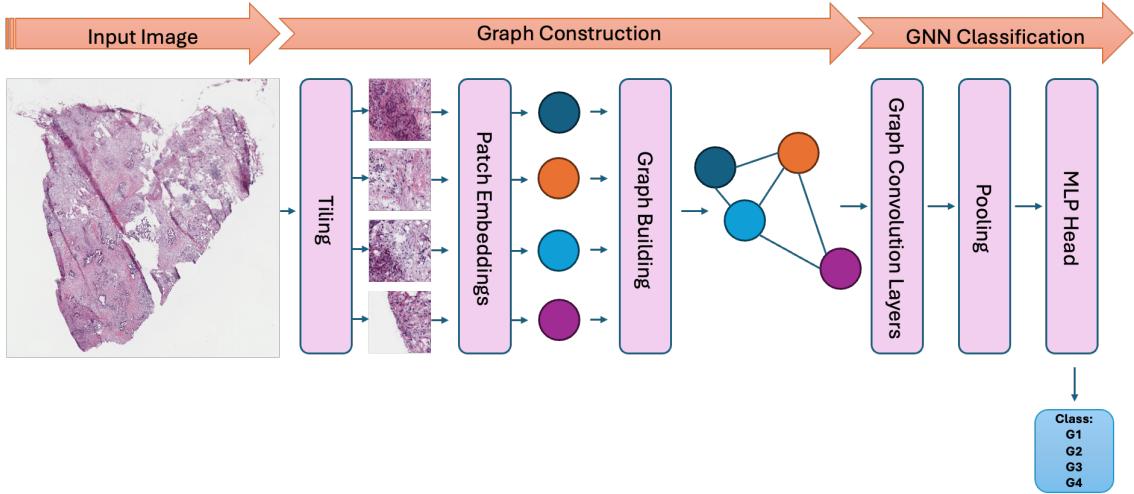


Figure 3.3: Overview Pipeline from Input SVS Image to Tumor Grade Classification following Tiling, Graph Construction, and Graph Neural Network

perceptron head, will output logits for the four classifications we want to identify, the four tumor grades, G1-G4.

3.0.7 Training

The model used AdamW optimization with weight decay and gradient clipping. A Reduce-on-Plateau scheduler adapted the learning rate based on validation signals. Early stopping was triggered by stalled improvements in validation loss or macro-F1 to guard against overfitting under imbalance. To address label skew, we employed several different tactics. First a class-weighted cross-entropy with weights inversely proportional to class frequencies, followed by focal loss ($\gamma > 0$, with optional class-specific α) to emphasize hard or misclassified examples, and finally semi-balanced training sets via a weighted random sampler that ensured each batch contained comparable numbers of graphs per class. Random seeds were fixed for data shuffling, weight initialization, and sampler behavior to promote reproducibility. Training and evaluation were executed on GPU when available, with deterministic fallbacks on CPU.

3.0.8 Evaluation

Model selection and reporting followed a strict separation of validation and test sets. Primary metrics included overall accuracy, F1 scores, and balanced accuracy, the latter reflecting performance under class imbalance. As these are key metrics for this project it is important to state their exact equations for transparency.

Accuracy:

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.2)$$

Recall (Sensitivity):

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN} \quad (3.3)$$

F1 Score:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.4)$$

Specificity:

$$\text{True Negative Rate (TNR)} = \frac{TN}{TN + FP} \quad (3.5)$$

Balanced Accuracy:

$$\text{BACC} = \frac{TPR + TNR}{2} \quad (3.6)$$

We produced confusion matrices to diagnose class-wise confusions. For probabilistic outputs, we computed receiver operating characteristic (ROC) curves and area under the ROC curve (AUROC). In binary settings, a single ROC/AUROC would be reported, however in our multi-class setting, we adopted a one-vs-rest scheme to obtain per-class ROC curves and macro-averaged AUROC, alongside micro-averaged summaries when informative. Classification reports, such as precision, recall, and F1 per class complemented these analyses. All thresholds and averaging conventions were explicitly documented for reproducibility.

3.0.9 Graph Analysis

To learn about the correspondence between tissue layout and learned representations, we visualized both spatial and semantic graph structures. Spatial graphs were plotted in slide coordinates to reveal anatomical neighborhoods, while cosine graphs were embedded to a PCA dimension on node features to expose clusters of histomorphologically similar tiles, independent of location.

Chapter 4

Results

4.0.1 Exploratory Data Analysis

The patient cohort under investigation displayed diverse demographic and clinical characteristics, as shown in Figure 1. The gender distribution was relatively balanced, with a slight predominance of male patients ($n \approx 127$) compared to female patients ($n \approx 106$). Age was distributed broadly between 35 and 90 years, with the peak frequency between 60 and 75 years, consistent with the epidemiology of pancreatic cancer, which predominantly affects older adults. The race distribution demonstrated that the majority of patients were classified as White (72.8%), with smaller proportions of Asian (4.7%) and Black or African American (3.4%) patients, while a notable fraction of cases (17.2%) had no race reported as shown in figure ... in the appendix. From a geographical perspective, the dataset was dominated by patients from the United States ($n \approx 120$), with smaller contributions from Canada, South Korea, Germany, Australia, and Brazil.

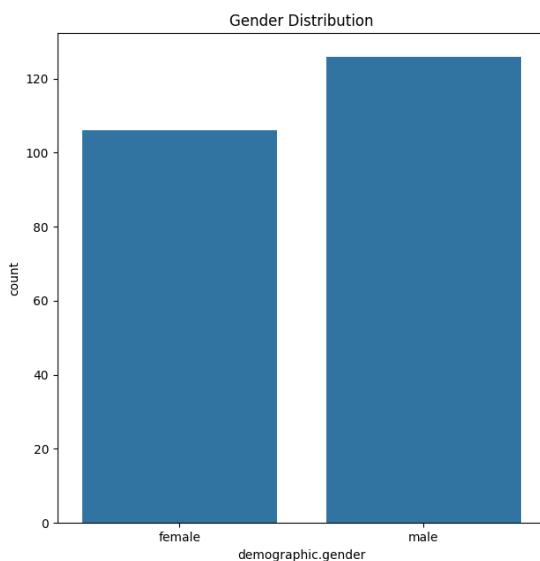


Figure 4.1: Gender Demographic of all 232 unique patients present in the TCGA Data for Pancreatic Cancer

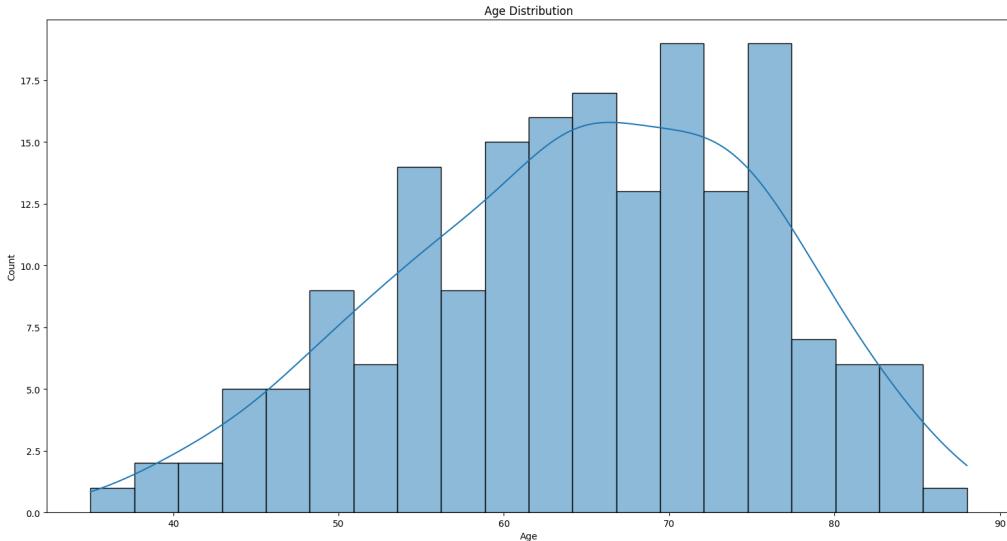


Figure 4.2: Age Demographic of all 232 unique patients present in the TCGA Data for Pancreatic Cancer

Clinical annotation of pathological grade showed substantial heterogeneity. Grade II tumors were the most common ($n \approx 130$), followed by Grade III tumors ($n \approx 60$ each). Advanced grades such as Grade IV were grossly underrepresented, alongside a very limited number of Grade I cases, as seen in figure 4.3. This distribution reflects the a handicap in the data distribution, presenting highly imbalanced class datasets. Importantly, the skew toward Grade II suggests that tumor grading in this cohort is biased toward less undifferentiated cells, which poses a particular challenge for automated classification systems as they do not present as much of a morphological difference than normal cells.

Together, these baseline distributions highlight both the strengths and limitations of the dataset. On the one hand, it provides a demographically heterogeneous cohort, ranging from a multitude of countries to test automated grading algorithms. On the other hand, the imbalance in race, stage, and overall size of the dataset with 232 patients introduces potential sources of bias that could propagate into downstream feature learning and model performance.

4.0.2 Graph Representations

To transform high resolution whole slide images (WSIs) into graph structures amenable to graph neural network (GNN) analysis, two types of graphs were constructed for each case. Spatial graphs and cosine similarity (embedding) graphs.

In spatial graphs, each node represents a tissue patch, and edges connect nodes based on their physical adjacency in the slide. This approach preserves the histological architecture of the sample, maintaining information about tissue morphology, glandular structures, and spatial context. Visualizations of the spatial graphs demonstrate regular lattice-like connectivity, where clusters of patches map directly to contiguous tissue regions. These graphs effectively encode morphological continuity and allow the GNN to learn features related to spatial tumor heterogeneity.

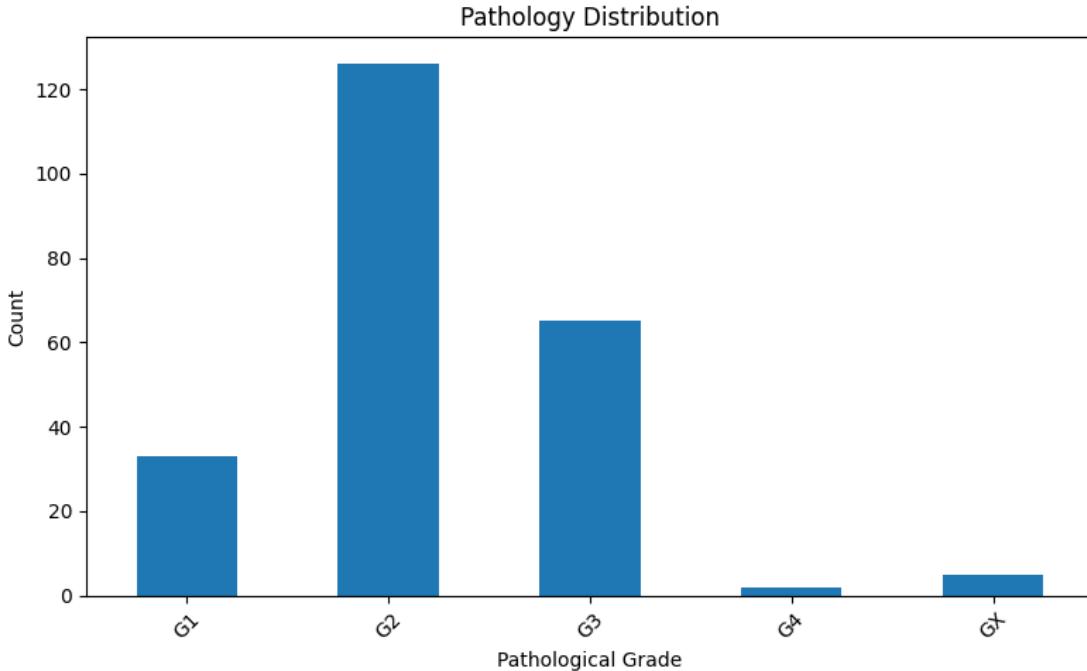


Figure 4.3: Pathological Grade Demographic of all 232 unique patients present in the TCGA Data for Pancreatic Cancer

By contrast, cosine similarity graphs were constructed based on the feature embeddings of each patch, linking nodes that shared high representational similarity. Unlike spatial graphs, these embeddings capture semantic similarity in appearance rather than geographical proximity. The PCA-projected embeddings, shown in figure 4.4 demonstrate that the cosine graphs often re-organize patches into clusters defined by latent histological similarity. In some cases, this collapses spatially distant but morphologically similar tissue regions into tightly connected clusters.

The distinction between these representations is significant. Spatial graphs have the potential to retain architectural integrity, aligning with traditional pathology interpretations that rely on morphological context, yet they might not account for the histological cues present in the actual difference between cells. Embedding graphs, on the other hand, emphasize feature driven grouping, potentially highlighting subtle patterns of nuclear atypia or stromal features not apparent from spatial relationships alone. However, this representation risks discarding biologically important proximal cues, leading to potential misinterpretation if the embedding space does not faithfully capture relevant pathology.

4.0.3 Contrastive Learning (SimCLR)

The first set of experiments employed a contrastive learning approach (SimCLR) to generate patch level embeddings from WSIs. The training and validation loss curves (Figure 4) show that the SimCLR model achieved stable convergence, with training loss declining rapidly within the first 2000 steps and flattening at approximately 4.4, while validation loss decreased steadily across epochs. This suggests that the model successfully optimized its contrastive objective, learning a

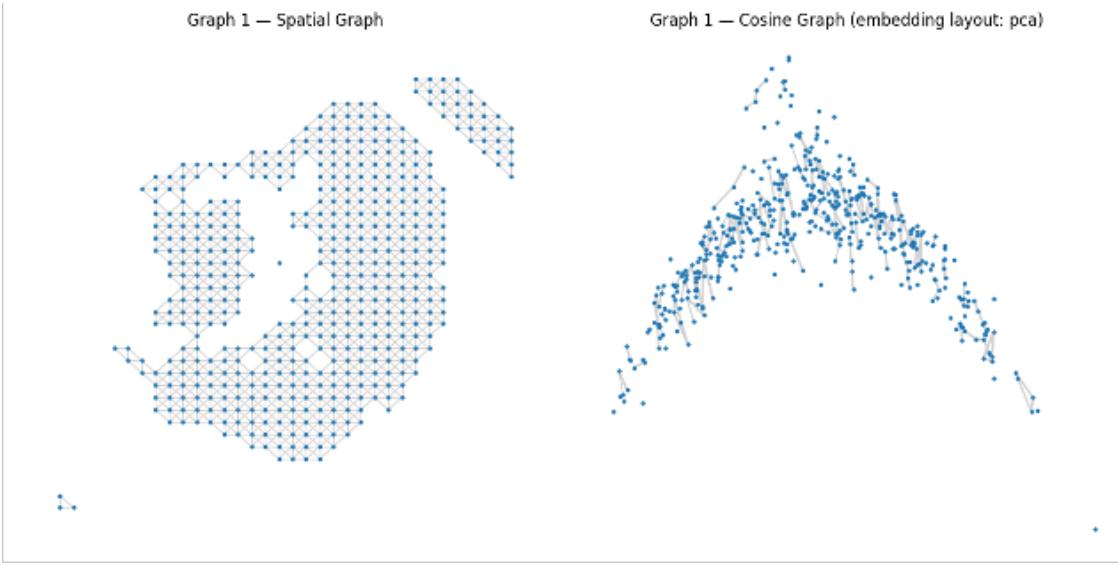


Figure 4.4: Example of a Spatial Graph versus Embedding Graph constructed from the SimCLR Feature Extractor

discriminative embedding space where augmented views of the same patch were pulled together and views of different patches were pushed apart.

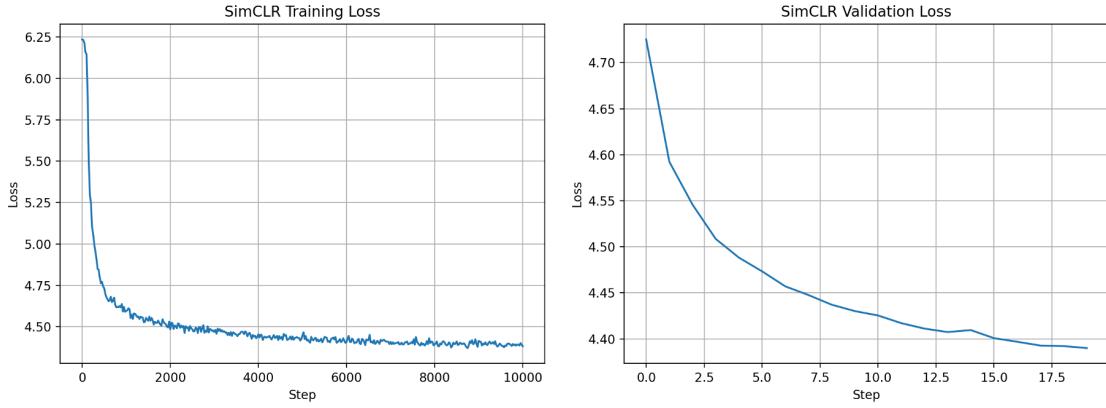


Figure 4.5: SimCLR Feature Extractor Training & Validation Loss Curves over 20 epochs with recordings every 25 steps per epoch in train

Despite effective training, downstream tumor grading performance using SimCLR embeddings was overwhelmingly poor. The confusion matrix, in figure 4.6, demonstrates that nearly all predictions collapsed into Grade III, with complete failure to classify Grade IV tumors and extremely poor representation of Grade I. Although Grade III achieved some degree of recall (94%) as seen in figure 4.7, this came at the expense of precision, as many Grade I and Grade II samples were misclassified as Grade III. Further outlining the poor performance of the SimCLR graph embeddings are the near zero recall values for all grades other than III. The overall test accuracy was only 26.9%, with a balanced accuracy of 25.8%, markedly below a pure chance curve.

Several factors likely contributed to the poor performance. First, the dataset might be too

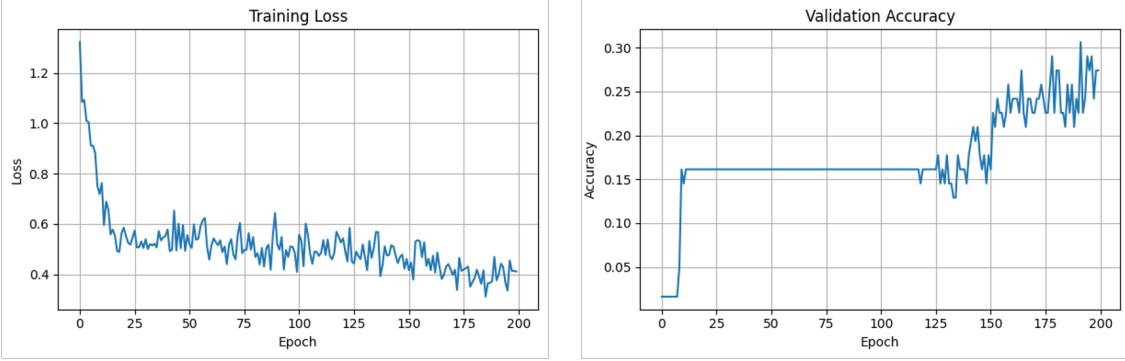


Figure 4.6: Tumor Grading Classification Training Loss & Validation Accuracy Curves using SimCLR Graph Embeddings

small for contrastive pretraining, where SimCLR could benefit from millions of images to learn stable, transferable representations, whereas our cohort much smaller with 232 patients, constraining the quality of the embedding space. Second, severe class imbalance, particularly the scarcity of Grades I and IV, biased the model toward majority classes and deprived it of sufficient positive pairs to form well separated clusters for rare grades. Third, because SimCLR optimizes for agnostic visual representations without pathology specific supervision, the learned features may have prioritized augmentation artifacts, such as color jitter, flips, and rotations over histologically meaningful features such as cellular architecture or stromal appearances. Finally, the embedding graph visualizations indicate a tendency toward semantic collapse: patches grouped by coarse texture similarity rather than diagnostically noticeable morphology, suggesting the representations failed to encode subtle differences between grades. Collectively, these limitations could explain why contrastive features lagged behind supervised and foundation-model baselines in downstream tumor grading.

In summary, while SimCLR succeeded as a self-supervised representation learner, its embeddings were insufficiently specialized for the complexity of pancreatic cancer histopathology, leading to poor classification results.

4.0.4 Baseline Model: Resnet50

To establish a supervised baseline, a ResNet50 convolutional neural network was trained directly on the WSI-derived patches. The training and validation curves, in figure 4.8, indicate stable convergence with progressive improvement in validation accuracy, reaching 50% by the end of training. The confusion matrix highlights a substantial improvement over SimCLR: all four grades were at least partially represented in predictions, though misclassifications remained common. Notably, Grade II achieved the highest recall (59%), while Grade I and Grade IV remained underrepresented.

The ROC analysis in figure 4.11, supports these observations. AUC values ranged from 0.429 (Grade II) to 0.598 (Grade I), with a micro-average AUC of 0.758. The overall test accuracy improved to 42.9%, with balanced accuracy of 26.6%. Compared to SimCLR, this represents a significant improvement in classification, particularly in terms of AUC and the ability to capture Grade II tumors. Nevertheless, the balanced accuracy remained low, reflecting persistent

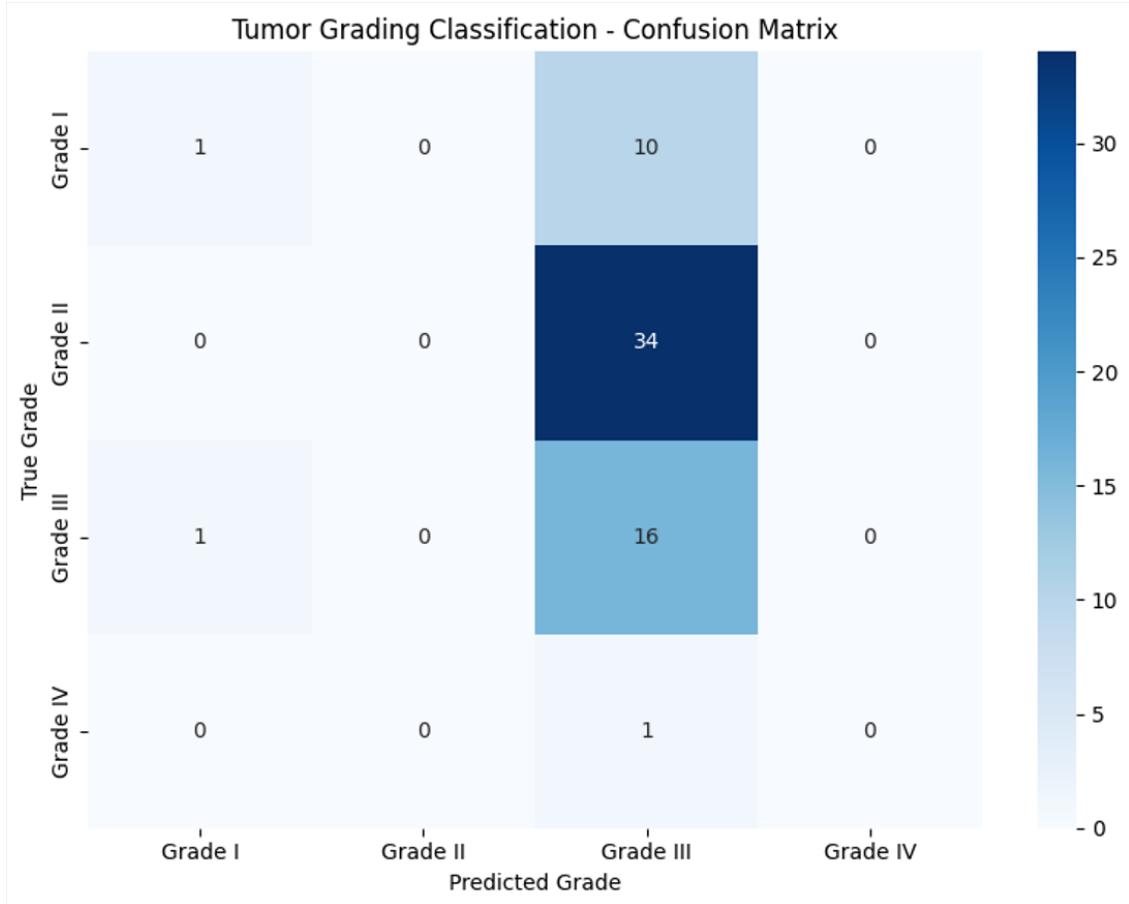


Figure 4.7: Tumor Grade Classification Confusion Matrix for SimCLR Graph Embeddings

difficulties with class imbalance.

The superior performance of ResNet50 relative to SimCLR can be attributed to its direct supervision and extensive image pretraining, which guided feature extraction toward pathology relevant differences between tumor grades rather than generic visual similarity. However, its limited generalization capacity highlights the challenge of training deep networks on relatively small and imbalanced clinical datasets.

4.0.5 Foundation Model: UNI

The third experimental setting leveraged UNI, a large-scale histopathology foundation model pre-trained on millions of WSIs across cancer types. UNI serves as a domain specific feature extractor, offering the advantage of pretraining on a massive and diverse dataset. Fine-tuning UNI for pancreatic tumor grading yielded the strongest results in this study.

The training and validation curves in figure 4.13, demonstrate consistent improvement in validation accuracy, stabilizing around 50% by the end of training. The confusion matrix shows markedly enhanced classification of Grade II and Grade III tumors, with fewer misclassifications compared to ResNet50. Grade II recall reached 91%, while Grade III precision improved substantially (71%).

Classification Report:				
	precision	recall	f1-score	support
Grade I	0.50	0.09	0.15	11
Grade II	0.00	0.00	0.00	34
Grade III	0.26	0.94	0.41	17
Grade IV	0.00	0.00	0.00	1

Figure 4.8: Tumor Grading Classification Report for SimCLR Graph Embeddings with Precision, Recall & F1 Scores

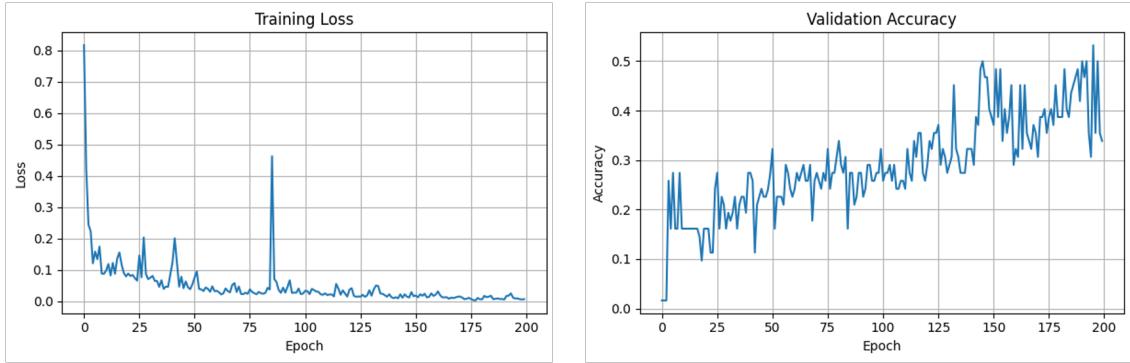


Figure 4.9: Tumor Grading Classification Training Loss & Validation Accuracy for ResNet50 Graph Embeddings

The ROC analysis in figure 4.15 underscores the benefits of foundation model pretraining. AUROC values were consistently higher across tumor grades compared to ResNet50 and SimCLR, with Grade II (0.660) and Grade III (0.675) achieving the most robust performance. The overall test accuracy reached 57.1%, with balanced accuracy improving to 30.2% across the different extractors. This represents a meaningful step forward, though balanced accuracy remained limited by the extreme scarcity of Grade I and IV tumors.

The improvement of UNI over both SimCLR and ResNet50 illustrates the power of foundation models in histopathology. Pretraining on millions of WSIs enables UNI to capture subtle morphological cues and rare features that are inaccessible to smaller models trained on limited datasets. Importantly, UNI maintained robustness despite the imbalance in tumor grades, demonstrating the transferability of large-scale histopathology representations.

4.0.6 Spatial vs Embedding Graphs

A key methodological comparison in this study was the performance of spatial graphs versus embedding graphs. Spatial graphs retained tissue architecture, connecting patches based on physical adjacency, while embedding graphs reorganized patches based on feature similarity in the learned embedding space, as seen in figure 4.17.

The choice between spatial and embedding graphs carries distinct biological or interpretive results. Spatial graphs preserve architecture within the nodes and edges it presents. Very rarely

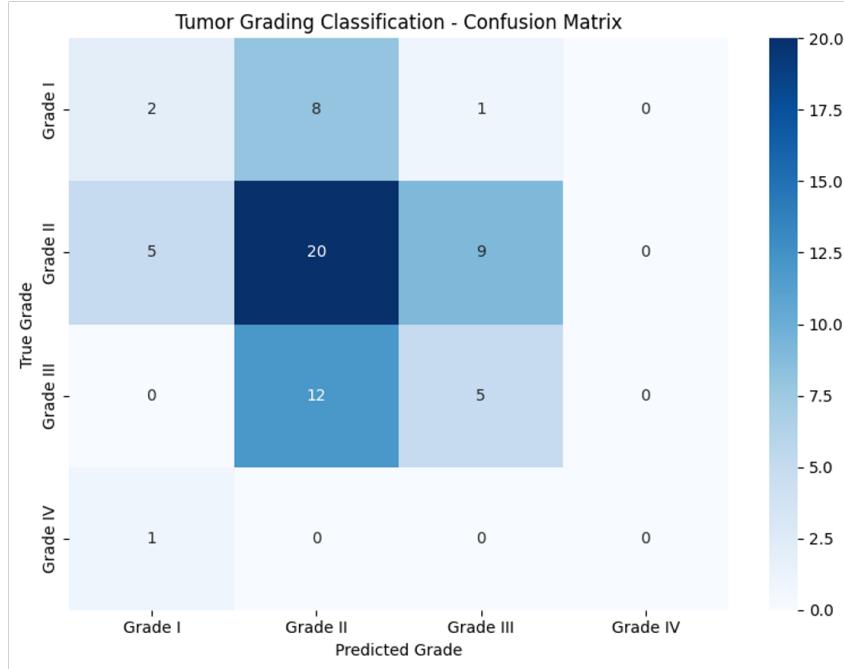


Figure 4.10: Tumor Grading Classification Confusion Matrix for Grades I-IV from ResNet50 Graph Embeddings

will a singular cancer cell present itself in one location far away from other cancer cells, but rather cancer forms a tumor microenvironment. It retains architectural relationships that pathologists could use to infer many things about the tissue slide. Embedding graphs on the other hand, cluster similar semantic regions in the feature space. This type of embedding focuses on the subtle morphological cues, which can also be beneficial to tumor grading, as it will hone in more on the differences between those undifferentiated to well-differentiated cells that are identified in grading. However, with graph embeddings, there is the potential to run into noise sensitivity. Given that the pancreas is a highly heterogeneous tissue, the embeddings could focus on stroma or glandular tissue more so than differences in cancer cells to identify different grades. In a spatial graph, the cancer cells are constrained to certain regions of the tissue and sometimes making it easier to identify them due to location. Both the spatial and embedding graphs have representations that are valuable to identifying cancer cells, and are complementary. Providing both the semantic and spatial information that a histopathologist will gather when analyzing a biopsy in real time.

4.0.7 Implications

The progression from SimCLR to ResNet50 to UNI highlights the evolution of feature learning in histopathology. While contrastive learning has been beneficial in many other scenarios and offers a theoretically elegant unsupervised approach, its effectiveness in this scenario was hampered by limited data and lack of domain specificity. Supervised CNNs like ResNet50 improved performance but were constrained by dataset size and imbalance. Finally foundation models, demonstrated here by UNI, offered the most robust solution, leveraging large-scale pretraining to achieve superior accuracy and generalization.

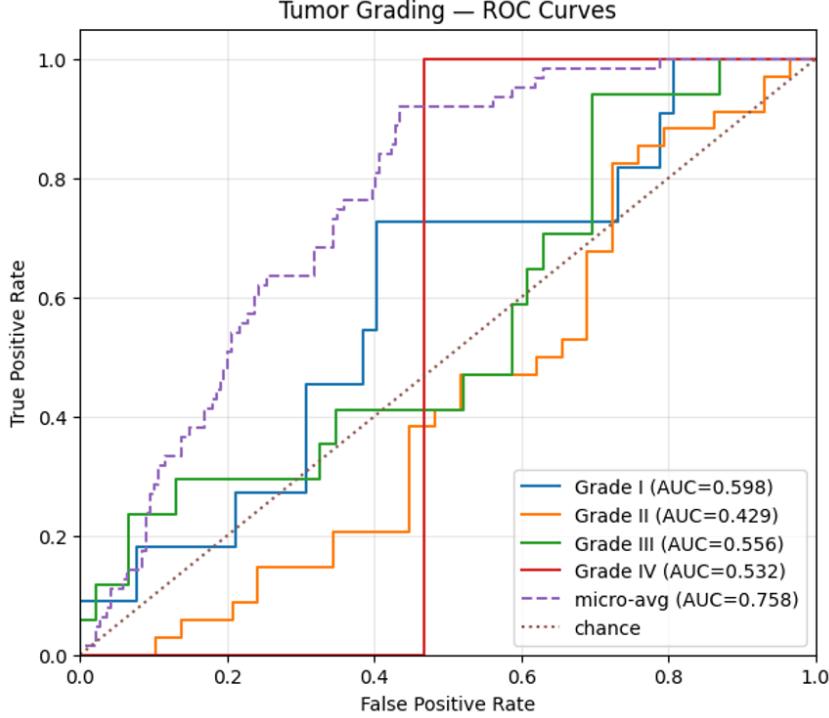


Figure 4.11: Tumor Grading Classification ROC for ResNet50 Graph Embeddings

The comparison of spatial versus embedding graphs further highlights that representation matters. Spatial graphs anchored predictions in histological context, providing robustness against feature-space collapse, while embedding graphs offered complementary clustering but risked abstraction from morphology. Future work should investigate hybrid graph strategies and further fine-tuning of foundation models to address minority classes.

Ultimately, these results suggest that foundation models represent a crucial step forward in computational pathology, offering tangible benefits for tumor grading in pancreatic cancer. However, challenges remain, particularly with respect to class imbalance, interpretability, and the integration of spatial context. Addressing these challenges will be essential to translate these methods into clinically meaningful decision support systems.

Test Accuracy: 0.4286 | Balanced Accuracy: 0.2660

Classification Report:

	precision	recall	f1-score	support
Grade I	0.25	0.18	0.21	11
Grade II	0.50	0.59	0.54	34
Grade III	0.33	0.29	0.31	17
Grade IV	0.00	0.00	0.00	1

Figure 4.12: Tumor Grading Classification Report for ResNet50 Graph Embeddings

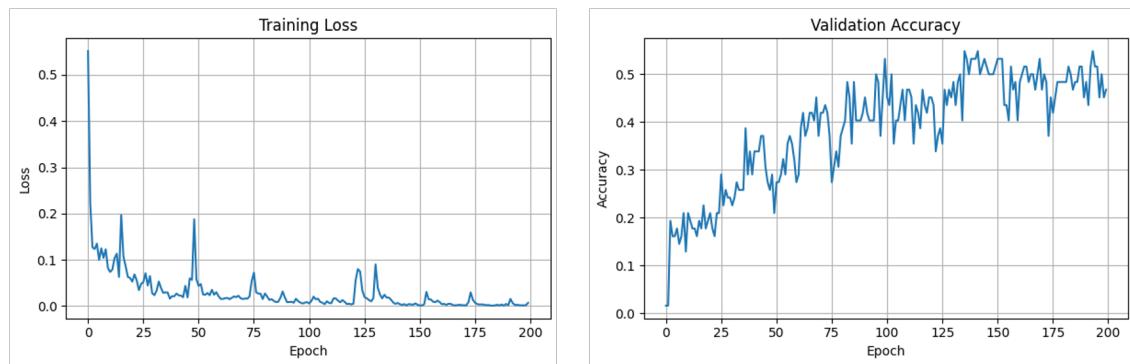


Figure 4.13: Tumor Grading Classification Training Loss & Validation Accuracy using UNI Foundation Model Graph Embeddings

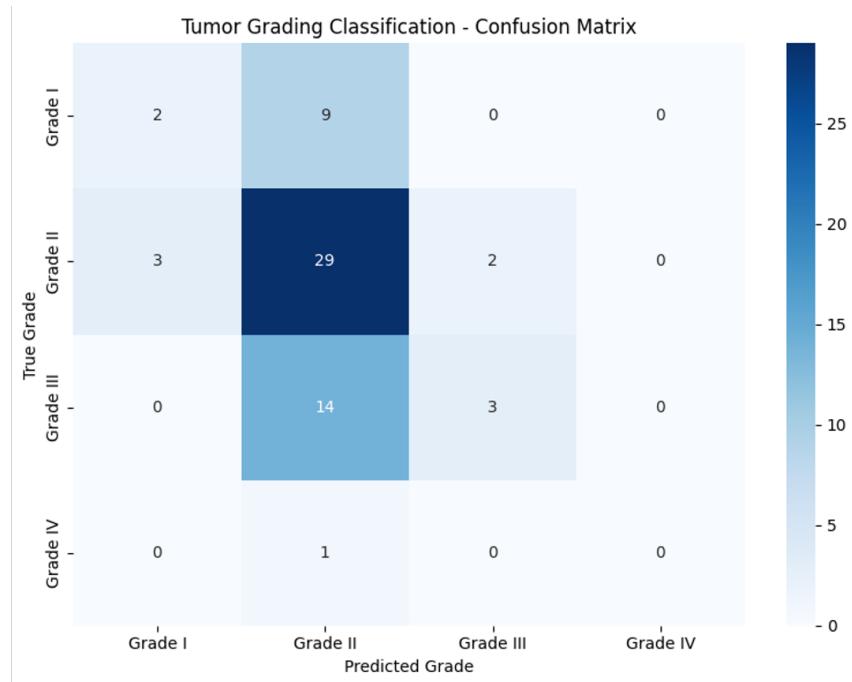


Figure 4.14: Tumor Grading Classification Confusion Matrix for Grades I-IV using UNI Foundation Model Graph Embeddings

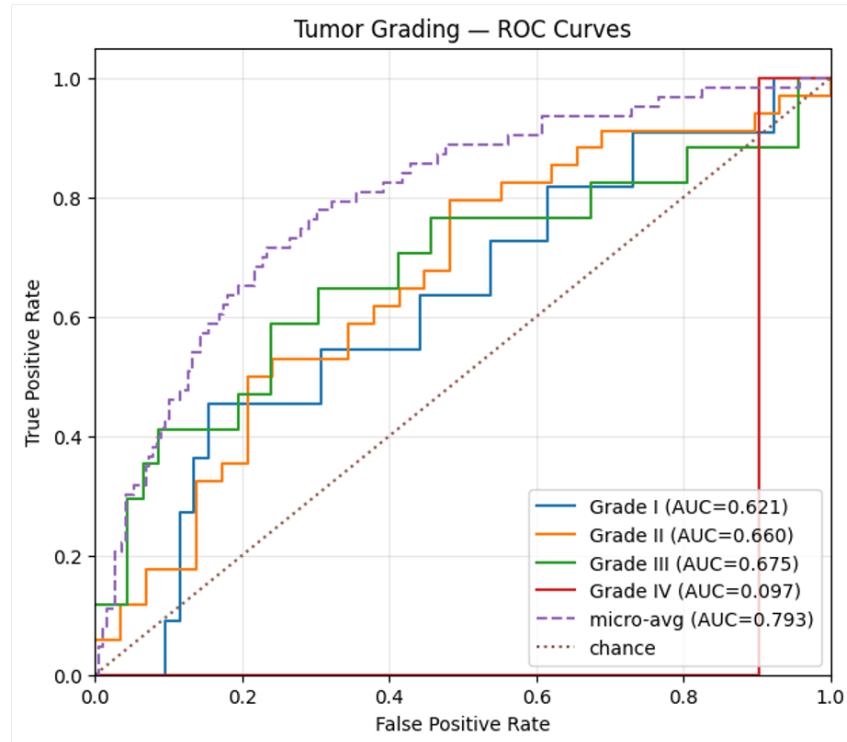


Figure 4.15: Tumor Grading Classification ROC for UNI Foundation Model Graph Embeddings

Test Accuracy: 0.5714 Balanced Accuracy: 0.3015				
Classification Report:				
	precision	recall	f1-score	support
Grade I	0.00	0.00	0.00	11
Grade II	0.61	0.91	0.73	34
Grade III	0.71	0.29	0.42	17
Grade IV	0.00	0.00	0.00	1

Figure 4.16: Tumor Grading Classification Report for UNI Foundation Model Graph Embeddings

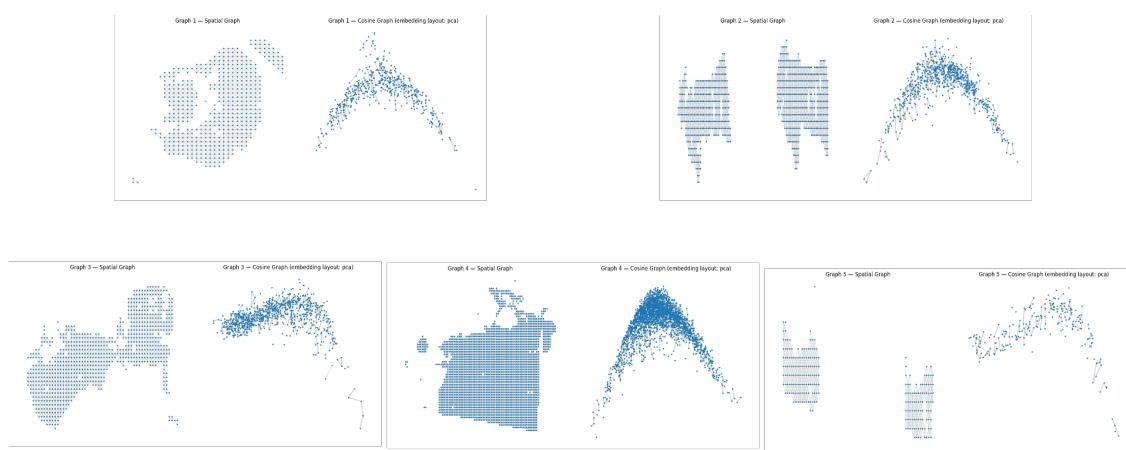


Figure 4.17: Spatial Versus Embedding Graphs showing the distinct differences between learning cellular morphology and location

Chapter 5

Discussion

The primary aim of this study was to design and evaluate a graph neural network (GNN) framework for the computational histopathology of pancreatic cancer using hematoxylin and eosin (H&E) stained whole slide images (WSIs). This was motivated by the pressing need for improved diagnostic grading and prognostic modeling in pancreatic cancer, which remains one of the most aggressive malignancies with poor survival outcomes. To achieve this overarching aim, four specific objectives were pursued: (1) a literature review of pancreatic cancer and GNN applications in histopathology, (2) training GNN models on The Cancer Genome Atlas (TCGA) dataset, (3) scaling these models to whole slide imaging, and (4) investigating interpretability in graph-based representations of histological structures.

The results obtained, presented in the preceding chapter, provide a rich basis for evaluating the extent to which these objectives were met, the validity of the study’s hypotheses, and the strengths and weaknesses of the proposed approach. This discussion will integrate these findings with the project’s stated aims and the broader literature, while also reflecting critically on the limitations and implications of this work.

5.0.1 Hypothesis

At the outset, the hypothesis underpinning this project was that GNNs, when applied to WSIs, could provide improved performance in tumor grading relative to conventional supervised learning methods. That given the representations present in the nodes and edges would allow for a neural network to have deeper information about the cellular cues and as such be able to predict tumor grade better. The results provide partial support for this hypothesis, with nuanced outcomes across different objectives.

Support

The findings strongly support the notion that foundation models improve classification quality in computational histopathology. UNI, a histopathology specific foundation model pretrained on millions of images, achieved the best overall accuracy (57.1%) and balanced accuracy (30.2%), outperforming both SimCLR (26.9% accuracy, 25.8% balanced accuracy) and ResNet50 (42.9% accuracy, 26.6% balanced accuracy). This validates a central claim that leveraging pretrained

embeddings from foundation models significantly enhances model robustness and discriminative capacity in tumor grading tasks.

The results also support the feasibility of scaling GNNs to whole slide images. Both spatial and embedding graph representations were successfully constructed and analyzed, demonstrating the ability of GNNs to operate at the WSI level while preserving either physical context (spatial graphs) or feature similarity (embedding graphs). This aligns with Objective 3 of scaling to whole slide images.

Limitations

On the other hand, the study’s findings contradict the overall hypothesis to reach similar levels of accuracy given GNNs in other cancers. The expectation that contrastive learning (SimCLR) would provide a robust feature extraction method in the absence of extensive supervision also did not achieve results. While SimCLR converged stably during training, its downstream classification results were poor, collapsing heavily onto Grades II and III and failing to identify Grades I and IV. This demonstrates that contrastive learning, while promising, is insufficient in practice for small, imbalanced datasets such as TCGA pancreatic WSIs.

Furthermore, while GNNs were able to model spatial and embedding graphs, the classification performance remained modest across all approaches. Balanced accuracy values never exceeded 31%, indicating that while foundation models improved robustness, significant challenges remain in achieving clinically useful grading performance. Thus, the overarching hypothesis is only partially supported, foundation models indeed enhance performance, but the ultimate goal of reliable automated grading remains unrealized within the scope of this project.

5.0.2 Evaluation of Objectives

Literature Review

The review of pancreatic cancer histopathology and GNN applications successfully established the rationale for this work. It highlighted the scarcity of studies applying GNNs specifically to pancreatic cancer, contrasting with more extensive research in breast, prostate, and lung cancers. By positioning pancreatic cancer as a relatively underexplored but clinically urgent domain, this project directly addressed a gap in the literature.

Training a GNN

This objective was achieved, with a GNN-based framework tested using embeddings derived from SimCLR, ResNet50, and UNI. The results show a clear gradient of performance across these extractors, highlighting the importance of feature quality in driving GNN performance. While the accuracy metrics remained modest, the experiments demonstrate the technical feasibility of training GNNs on TCGA WSIs, thereby fulfilling this objective.

Scaling for WSIs

Graph construction at the WSI scale was accomplished, with both spatial and embedding graphs successfully generated for multiple cases. The results underscore the trade-offs between preserving

tissue architecture (spatial graphs) and grouping by feature similarity (embedding graphs). While performance differences were not fully quantified between these graph types, the visualizations demonstrate the feasibility of scaling to WSI-level graphs, satisfying this objective.

Interpretability

Interpretability was examined through comparisons between spatial and embedding graphs, as well as through ROC and confusion matrix analyses. The study shows that spatial graphs align more closely with regional frameworks by preserving architectural continuity, whereas embedding graphs offer abstracted feature based clustering that may highlight morphological patterns. While full interpretability frameworks, such as attention maps or saliency scores were not implemented, the comparative analysis provides meaningful insights. However, given more time to explore the matter, a deeper scope of the interpretations of the graphs and their classifications would remain ideal.

5.0.3 Strengths & Limitations

Strengths

This study possesses several notable strengths that contribute to its originality and significance within the field of computational pathology. First, the focus on pancreatic cancer provides an important contribution to the literature, as relatively few studies have systematically applied graph neural networks to this particularly aggressive malignancy. Most existing work on computational histopathology has concentrated on more common cancers such as breast, prostate, or lung, leaving pancreatic cancer comparatively underexplored despite its high mortality and clinical need. By targeting this disease, the project directly addresses a critical gap. A second strength lies in the comparative framework adopted. By evaluating three distinct feature extraction strategies, contrastive learning through SimCLR, a supervised baseline using ResNet50, and a large-scale foundation model (UNI), the project not only benchmarks performance but also provides insights into the specific advantages and disadvantages of each approach. This triangulated design strengthens the validity of the findings, demonstrating that foundation models offer the greatest improvements in robustness and classification accuracy.

A further strength of this study is its integration of graph-based representations. Rather than treating WSIs simply as collections of patches, the project constructed both spatial graphs and embedding graphs, thereby examining how different representational frameworks influence classification. This dual approach advanced understanding of how architectural context and feature similarity can be encoded computationally, with important implications for interpretability. The use of real-world clinical data from The Cancer Genome Atlas (TCGA) is also a strength, ensuring that findings are grounded in a diverse and clinically annotated dataset rather than synthetic or overly curated samples. Not only that but the data remains publicly available for data analytics on pancreatic cancer to continue.

Weaknesses

Despite these strengths, the study also has several weaknesses and limitations that must be acknowledged. The most prominent limitation lies in the size and imbalance of the dataset. In particular, Grades I and IV of pancreatic cancer were severely underrepresented, which significantly constrained the ability of the models to learn generalizable distinctions across all grades. This limitation manifested clearly in the low balanced accuracy across all models and the near complete misclassifications of the minority classes. A second limitation relates to the constraints of contrastive learning. While SimCLR is robust agnostic framework, its performance was undermined both by the limited scale of the dataset and by augmentation strategies that may not have preserved pathology-relevant features. The reliance on augmentations such as color jitter or cropping may have inadvertently directed the model to focus on artifacts rather than histological characteristics, which helps explain its poor downstream performance.

Another weakness lies in the limited generalization capacity of the models. Even the best-performing approach, UNI, achieved only 57% accuracy, which while better than the alternatives, remains insufficient for clinical use. Moreover, interpretability, which was one of the explicit objectives of the project, was only partially achieved. Although spatial and embedding graphs were compared in terms of their representational qualities, more advanced interpretability techniques, such as saliency mapping or attention visualization, were not fully implemented. This leaves an important gap in the clinical applicability of the findings, as interpretability is essential for adoption in medical decision making. Finally, the reliance on a single dataset, limits the external validity of the conclusions. Without external validation across independent cohorts, it remains uncertain whether the findings would generalize to broader clinical populations.

Implications

The results of this study carry several broader implications for the future of computational histopathology. Foremost, the superiority of the foundation model over both SimCLR and ResNet50 points to the growing importance of large-scale pretraining in this domain. Models such as UNI, trained on millions of WSIs, provide feature extractors that are significantly more robust and generalizable than models trained solely on smaller, domain-specific datasets. As foundation models become increasingly available and more finely tuned for histopathology, they are likely to become a cornerstone of computational pathology pipelines.

At the same time, the findings highlight the importance of how histological data are represented. The contrast between spatial and embedding graphs reveals that different representational choices emphasize different aspects of tissue biology: spatial graphs preserve architectural integrity and align closely with microenvironments, while embedding graphs cluster regions by latent similarity, offering complementary insights in morphology. This suggests that future approaches should consider hybrid models that combine both types of information to maximize predictive power.

Finally, the study demonstrates that interpretability and class imbalance remain critical bottlenecks in the field. Clinicians are unlikely to adopt automated systems unless they can understand and trust their decision making processes, and pathologists require reassurance that classifications are grounded in biologically relevant features. Similarly, the skewed distribution of tumor grades poses a major challenge for both training and evaluation. Addressing these issues through methods

such as synthetic oversampling, data augmentation tailored for histology, or multi-modal integration with genomic and clinical variables will be essential for advancing toward clinical translation. In sum, this project reinforces the potential of GNNs and foundation models for pancreatic cancer histopathology, while also identifying the methodological innovations required to bring such systems closer to practical utility.

5.0.4 Conclusion

In summary, this study partially supports its original hypothesis and successfully meets most of its stated objectives. It demonstrates the feasibility of applying GNNs to pancreatic cancer WSIs and highlights the clear advantage of foundation models over both contrastive and supervised baselines. However, it also navigates significant limitations, including dataset imbalance, poor performance of contrastive learning, and the need for improved interpretability.

The strengths of this work lie in its novelty, comparative approach, and critical evaluation of limitations. Its weaknesses lie primarily in dataset constraints and incomplete interpretability. Taken together, the results suggest that while GNNs and foundation models hold substantial promise for advancing computational histopathology in pancreatic cancer, significant methodological and practical challenges must be overcome before clinical translation can be realized.

Chapter 6

Conclusions

This project set out with the ambitious aim of developing and evaluating a graph neural network (GNN) framework for computational histopathology in pancreatic cancer, leveraging hematoxylin and eosin (H&E) whole slide images (WSIs) from The Cancer Genome Atlas (TCGA). The ultimate goal was to explore whether graph-based learning could improve the accuracy of tumor grading and provide a pathway toward more effective diagnostic and prognostic tools for this particularly aggressive cancer. Four specific objectives were established to guide the project: (1) to review the literature on pancreatic cancer and the application of GNNs in computational pathology, (2) to train a GNN using TCGA data, (3) to scale these methods to whole slide images, and (4) to investigate interpretability through graph-based representations.

All four objectives were pursued systematically. The literature review demonstrated the scarcity of work applying graph neural networks specifically to pancreatic cancer, in contrast to more widely studied malignancies such as breast or prostate cancer. This gap reinforced the importance and novelty of the current study. The second objective was realized through the training of GNNs with multiple feature extraction backbones: a contrastive learning model (SimCLR), a supervised convolutional neural network (ResNet50), and a histopathology-specific foundation model (UNI). By testing these three approaches, the study provided a comparative framework for understanding how feature extractor choice influences classification outcomes. The third objective, scaling to whole slide images, was also successfully addressed. Both spatial graphs (encoding patch adjacency) and embedding graphs (encoding cosine similarity between patch embeddings) were constructed, allowing exploration of how different representational strategies affect model performance and interpretability. Finally, the fourth objective, interpretability, was partially achieved through the comparative analysis of these graph types, revealing that spatial graphs preserve tissue structures while embedding graphs emphasize semantic similarity.

In this sense, the project achieved its broad methodological goals: it implemented a complete pipeline from WSIs to graph representations, trained multiple GNN models, and critically evaluated their outputs. It not only bench marked the relative performance of self-supervised, supervised, and foundation model approaches but also advanced understanding of representational choices in graph construction. These achievements constitute a meaningful contribution to the computational pathology literature, particularly for the relatively underexplored case of pancreatic cancer.

The success of the project can be evaluated along two dimensions: technical feasibility and classification performance. From a technical standpoint, the project was successful. It demonstrated that WSIs could be processed into graph structures at scale, that multiple feature extraction paradigms could be tested, and that GNNs could be trained on real-world clinical data. The construction of both spatial and embedding graphs highlights the flexibility of the approach, and the comparative analysis provides a clear demonstration of how graph representation impacts interpretability. In this sense, the pipeline that was designed and implemented can be considered a success.

From the perspective of classification performance, however, the results did not achieve success. The hypothesis that foundation models would outperform self-supervised and supervised baselines was supported, with UNI achieving the best accuracy and balanced accuracy. This represents a genuine success: it validates the potential of foundation models as the most promising direction for computational histopathology in pancreatic cancer. Nevertheless, the absolute performance metrics remained poor. Even the best-performing model achieved only 57.1% accuracy and 30.2% balanced accuracy, falling short of the reliability required for clinical application. The poor performance of SimCLR, which collapsed onto majority classes and failed to recognize minority grades, further illustrates the challenges of applying self-supervised methods to small and imbalanced datasets. Thus, while the approach succeeded in demonstrating feasibility and relative performance differences, it fell short of achieving clinically actionable accuracy.

Overall, the approach must be described as a stepping stone for pancreatic cancer in computational histopathology. The project successfully outlined gaps in literature and methodology, produced supporting insights into graph construction and feature extraction, and confirmed the superiority of foundation models. However, it did not achieve the ultimate aim of producing a GNN that could grade pancreatic cancer with high accuracy. This underscores the broader challenges of applying machine learning to limited and imbalanced medical datasets, highlighting the need for larger, and more holistic datasets that directly target class imbalance.

Among the results obtained, the most striking finding was the dramatic difference in performance between the contrastive learning approach (SimCLR) and the foundation model (UNI). Despite stable convergence during training, SimCLR produced almost unusable results for downstream classification, with an overall accuracy of only 26.9% and a balanced accuracy of 25.8%. Grades I and IV were almost entirely unrecognized, and the model collapsed heavily onto Grade III. This stark under performance is notable because SimCLR and similar contrastive methods are often positioned in the literature as highly promising for domains where labeled data are scarce. In this study, however, the method not only failed to improve performance but also introduced severe biases by overfitting to the majority classes.

By contrast, the foundation model UNI demonstrated clear robustness and generalizability, even when trained on the same limited dataset. Achieving 57.1% accuracy and 30.2% balanced accuracy, it substantially outperformed both SimCLR and the supervised ResNet50 baseline. This contrast illustrates the transformative potential of large-scale pretraining in histopathology. Unlike SimCLR, which attempted to learn representations from scratch using limited pancreatic cancer slides, UNI leveraged the knowledge embedded in millions of WSIs across cancer types. The success of UNI demonstrates that foundation models are not only viable but may be essential for advancing computational pathology in rare and heterogeneous cancers such as pancreatic adenocarcinoma.

The juxtaposition of SimCLR’s collapse and UNI’s relative success is the most striking result because it pin points a broader methodological point, scale matters. In domains like histopathology, where the visual patterns are subtle and the available datasets are often small, self-supervised methods may not succeed without massive pretraining collections. Foundation models overcome this limitation by pretraining on unprecedeted scales, making them far more capable of transferring to specific cancers. This result therefore has important implications, not only for pancreatic cancer but also for the wider field of computational pathology, where foundation models are poised to redefine state-of-the-art performance.

This project has achieved its methodological aims and produced findings that both support and challenge its central hypothesis. It has demonstrated that graph neural networks can be constructed from WSIs of pancreatic cancer, that both spatial and embedding graphs can be meaningfully compared, and that foundation models significantly outperform both contrastive and supervised baselines. At the same time, it has highlighted the limitations imposed by dataset size and imbalance, as well as the shortcomings of self-supervised learning in this specific context.

The most important conclusion to be drawn is that foundation models represent the most promising path forward for computational histopathology. While this study fell short of delivering clinically applicable grading accuracy, it provided clear evidence that large-scale pretrained models can achieve substantially better results than conventional approaches, even in the face of data scarcity. This suggests that future efforts in pancreatic cancer and other challenging cancers should prioritize foundation model approaches, combined with strategies to mitigate class imbalance and enhance interpretability.

In this sense, the study represents a landing base for others to propel off of when investigative pancreatic cancer in the computational space. It succeeded in demonstrating technical feasibility, comparative evaluation, and the benefits of foundation models, while also signaling the methodological challenges that must be addressed before such approaches can be translated into clinical practice. Ultimately, while the immediate goal of accurate automated grading was not fully realized, the project lays important groundwork for future studies and points toward the promise of foundation models as the key to advancing computational histopathology in pancreatic cancer.

Chapter 7

Future Work

While this project has demonstrated the feasibility of using graph neural networks (GNNs) for pancreatic cancer histopathology and highlighted the advantages of foundation models, it has also revealed several important limitations. These include poor balanced accuracy, dataset imbalance, limited interpretability, and the under performance of self-supervised contrastive learning. Building on these findings, there are numerous avenues for future work that can improve performance, enhance interpretability, and bring computational histopathology closer to clinical utility.

The following section outlines several key directions for future research, organized around improvements to data, models, interpretability, evaluation, and clinical translation.

7.0.1 Addressing Limitations

One of the clearest lessons from this project is that data availability and quality remain the most pressing bottlenecks in computational pathology. The TCGA dataset provided a valuable starting point, yet its imbalance, particularly the under representation of Grades I and IV, severely limited classification performance. Future work should prioritize the expansion of datasets by incorporating additional cohorts, ideally through multi-institutional collaborations. Access to larger, more diverse sets of whole slide images would provide the scale required to train robust models and capture the inherent heterogeneity of pancreatic cancer. Diversity is equally important in terms of staining protocols, scanning technologies, and patient demographics, as models trained only on TCGA may not generalize to slides from other institutions.

Where collecting additional data is not immediately possible, augmentation strategies may provide partial relief. Traditional augmentations such as flips, rotations, or color perturbations are insufficient for pathology because they often obscure subtle histological cues. More targeted approaches, such as stain normalization and stain augmentation, may better preserve biological meaning while increasing variation. Synthetic data generation represents another avenue, with generative adversarial networks (GANs) and diffusion models showing promise for producing realistic histology images. If used carefully, these methods could help balance underrepresented classes, enabling models to better distinguish rare but clinically significant cases. Finally, few-shot learning strategies may help to overcome the scarcity of minority classes. By leveraging similarities with better-represented grades, few-shot techniques allow models to generalize from limited examples, an approach particularly well-suited to rare disease subtypes.

7.0.2 Improving Feature Extraction

Another key insight from this project was the importance of feature quality. The stark contrast between the under performance of SimCLR and the relative success of the foundation model UNI highlights that not all feature extraction approaches are equally suitable for histopathology. Future research should therefore continue to refine feature extraction pipelines. Foundation models, already shown here to outperform supervised and self-supervised baselines, could be further improved through more extensive fine-tuning. Rather than applying a uniform learning rate across the entire model, techniques such as layer-wise freezing or discriminative learning rates could allow lower-level features to remain stable while higher-level layers adapt to the specifics of pancreatic cancer morphology.

Domain-adaptive pretraining is another promising strategy. While UNI was trained on millions of WSIs across cancer types, its knowledge may be further specialized through pretraining on a smaller but more targeted set of pancreatic cancer slides before fine-tuning on TCGA. This two-stage approach could capture general histopathology patterns while also adapting to the unique features of pancreatic cancer. Looking forward, multi-modal foundation models may also prove transformative. Pancreatic cancer is defined not only by morphology but also by well-characterized genomic alterations such as KRAS and TP53 mutations. Models that jointly incorporate histology and molecular data could yield more powerful predictors, bridging the gap between morphology and molecular biology.

Finally, alternative self-supervised learning approaches deserve exploration. The poor performance of SimCLR underscores that generic contrastive methods may not be optimal for histopathology, particularly with limited datasets. Augmenting the dataset size would be keen to improving feature extraction quality.

7.0.3 Advancing Graph Representations

This project also revealed that the way graphs are constructed has a profound influence on model behavior. Spatial graphs, which preserve tissue architecture by connecting adjacent patches, align closely with how pathologists interpret slides, while embedding graphs, which connect patches based on feature similarity, reorganize tissue into clusters that emphasize latent patterns. Each of these approaches has strengths and weaknesses, suggesting that future work should explore hybrid strategies.

One direction is to construct graphs that integrate both spatial adjacency and embedding similarity, perhaps by weighting edges according to a combination of distance and similarity scores. Such hybrid graphs could simultaneously preserve morphological context and highlight latent feature relationships, providing a more comprehensive representation. Another promising avenue is hierarchical graph construction. Slides could be represented at multiple scales, beginning with fine-grained patch-level graphs that are pooled into higher-level region graphs, ultimately forming slide-level representations. This mirrors the diagnostic process of pathologists, who move between low and high magnification views, and could allow models to capture both local and global features.

Cell-level graphs also hold promise, as segmentation algorithms increasingly allow individual nuclei to be extracted and represented as nodes. By combining patch-level and cell-level information, GNNs could capture both architectural and cytological features, offering a more biologically

faithful representation of tissue.

7.0.4 Enhancing Model Architecture

Although this project successfully applied GNNs to pancreatic cancer, the architectures used remain relatively conventional. Recent advances suggest that transformer-based approaches may offer superior performance. Graph transformers, in particular, can capture long-range dependencies and global relationships more effectively than convolution-based GNNs. This is highly relevant in histopathology, where diagnostically important patterns may span large areas of tissue. Future work should therefore evaluate the performance of graph transformers relative to standard GNNs in this domain.

Attention mechanisms offer another avenue for improvement. By explicitly learning which nodes or edges are most important for classification, attention not only enhances performance but also provides interpretability. Multi-head attention could allow the model to consider multiple perspectives simultaneously, capturing different aspects of tumor morphology such as glandular organization, and stromal invasion. To prevent overfitting, especially when working with relatively small datasets, regularization strategies such as node dropout, edge perturbation, or feature noise could be employed. These techniques encourage robustness and reduce the risk of shortcut learning, ensuring that models focus on meaningful histological features rather than artifacts.

7.0.5 Improving Interpretability

Interpretability was an explicit objective of this project, and although some progress was made through the comparison of spatial and embedding graphs, there remains much room for improvement. Clinicians will only adopt computational pathology tools if they can understand and trust their decision-making processes, making interpretability essential for translation into practice.

Future work should therefore prioritize methods that make GNN decisions more transparent. For example, graph attention visualization can highlight which patches or regions most strongly influenced the final classification. Patch-level saliency mapping could also be applied, projecting predictions back onto the slide to identify specific morphological features associated with each grade. Human-in-the-loop approaches, where pathologists review and provide feedback on highlighted regions, may further improve interpretability and model refinement. In addition, comparative interpretability studies could examine how different feature extractors, such as ResNet50 versus UNI, emphasize different histological features. This would not only improve trust but also provide new biological insights.

7.0.6 Improving Evaluations

Evaluation frameworks must also be strengthened in future work. Accuracy and balanced accuracy provide useful but incomplete measures of performance, particularly in the context of imbalanced datasets. Future studies should adopt more comprehensive metrics, and cross-cohort validation is also essential. Models trained solely on TCGA must be tested on independent datasets from different institutions to ensure that they generalize to diverse clinical contexts. Robustness testing is equally important. Models should be evaluated under conditions of variable staining, scanning

resolution, or tissue preparation, as these technical factors can significantly affect predictions. Domain adaptation methods may help mitigate such variations, improving generalizability.

7.0.7 Clinical Translation

Ultimately, the long-term goal of computational histopathology is not academic performance but clinical impact. For these methods to be useful, they must integrate seamlessly into diagnostic workflows and support clinical decision making. Future research should therefore explore multi-task models that predict not only tumor grade but also other clinically relevant outcomes, such as prognosis or response to therapy. Integrating histopathology with clinical and molecular data could produce more comprehensive predictive models, enabling personalized treatment strategies for pancreatic cancer patients.

Attention must also be given to practical considerations of deployment. Models must be computationally efficient enough to process WSIs at scale within clinically acceptable time frames. Outputs must be interpretable and presented in formats compatible with digital pathology systems. Ethical and regulatory issues must also be addressed, including data privacy, accountability, and bias. Pancreatic cancer disproportionately affects certain populations, and models trained on imbalanced data risk exacerbating disparities unless fairness is explicitly prioritized.

7.0.8 Conclusion

The results of this project demonstrated both the promise and the limitations of GNN-based computational histopathology for pancreatic cancer. The superiority of foundation models suggests that large-scale pretraining is essential, while the under performance of SimCLR highlights the risks of applying generic self-supervised methods to small and imbalanced datasets. Looking ahead, the most promising directions for future research include expanding datasets, refining feature extraction, developing hybrid and hierarchical graph structures, adopting graph transformers, and prioritizing interpretability. External validation, robustness testing, and multi-modal integration will also be critical for clinical translation.

Ultimately, this project has laid the groundwork for future innovations by demonstrating feasibility, highlighting the value of foundation models, and identifying the key bottlenecks that must be overcome. By addressing these challenges, future work can move closer to the goal of clinically useful computational tools that improve the diagnosis, grading, and management of pancreatic cancer.

Bibliography

- [1] Xiao Zhang and Qian Liu. A graph neural network approach for hierarchical mapping of breast cancer protein communities. *BMC bioinformatics*, 26(1):23, 12 2025. ISSN 14712105. doi: 10.1186/s12859-024-06015-x.
- [2] Lianhe Zhao, Xiaoning Qi, Yang Chen, Yixuan Qiao, Dechao Bu, Yang Wu, Yufan Luo, Sheng Wang, Rui Zhang, and Yi Zhao. Biological knowledge graph-guided investigation of immune therapy response in cancer with graph neural network. *Briefings in Bioinformatics*, 24(2), 3 2023. ISSN 14774054. doi: 10.1093/bib/bbad023.
- [3] Xinyue Li, Rui Guo, Jing Lu, Tao Chen, and Xiaohua Qian. Causality-Driven Graph Neural Network for Early Diagnosis of Pancreatic Cancer in Non-Contrast Computerized Tomography. *IEEE Transactions on Medical Imaging*, 42(6):1656–1667, 6 2023. ISSN 1558254X. doi: 10.1109/TMI.2023.3236162.
- [4] Xiangyan Meng and Tonghui Zou. Clinical applications of graph neural networks in computational histopathology: A review. *Computers in Biology and Medicine*, 164:107201, 9 2023. ISSN 0010-4825. doi: 10.1016/J.COMPBIOMED.2023.107201.
- [5] Soner Koc, Ozgur Can Eren, Rohat Esmer, Fatma Ulkem Kasapoglu, Burcu Saka, Orhun Cig Taskin, Pelin Bagci, Nazmi Volkan Adsay, and Cigdem Gunduz-Demir. Deep Learning Enabled Scoring of Pancreatic Neuroendocrine Tumors Based on Cancer Infiltration Patterns. *Endocrine pathology*, 36(1):2, 12 2025. ISSN 15590097. doi: 10.1007/S12022-025-09846-3/FIGURES/4. URL <https://link.springer.com/article/10.1007/s12022-025-09846-3>.
- [6] Yunjie He, Hong Zhao, and Stephen T.C. Wong. Deep learning powers cancer diagnosis in digital pathology, 3 2021. ISSN 18790771.
- [7] Xiaohang Fu, Ellis Patrick, Jean Y H Yang, David Dagan Feng, and Jinman Kim. Deep multimodal graph-based network for survival prediction from highly multiplexed images and patient variables. *Computers in Biology and Medicine*, 154:106576, 2023. doi: 10.5281/zenodo.3518284.
- [8] Zhikang Wang, Jiani Ma, Qian Gao, Chris Bain, Seiya Imoto, Pietro Liò, Hongmin Cai, Hao Chen, and Jiangning Song. Dual-stream multi-dependency graph neural network enables precise cancer survival analysis. *Medical Image Analysis*, 97, 10 2024. ISSN 13618423. doi: 10.1016/j.media.2024.103252.

- [9] Jie Lian, Jiajun Deng, Edward S Hui, Mohamad Koohi-Moghadam, Yunlang She, Chang Chen, and Varut Vardhanabhuti. Early stage NSCLS patients' prognostic prediction with multi-information using transformer and graph neural network model. *eLife*, 11, 10 2022. ISSN 2050-084X. doi: 10.7554/eLife.80547. URL <https://elifesciences.org/articles/80547>.
- [10] Cui Chen, Zuping Zhang, Panrui Tang, Xin Liu, and Bo Huang. Edge-relational window-attentional graph neural network for gene expression prediction in spatial transcriptomics analysis. *Computers in Biology and Medicine*, 174:108449, 5 2024. ISSN 0010-4825. doi: 10.1016/J.COMPBIOMED.2024.108449.
- [11] Linling Qiu, Han Li, Meihong Wang, and Xiaoli Wang. Gated graph attention network for cancer prediction. *Sensors*, 21(6):1–21, 3 2021. ISSN 14248220. doi: 10.3390/s21061938.
- [12] Jiening Zhu, Jung Hun Oh, Anish K. Simhal, Rena Elkin, Larry Norton, Joseph O. Deasy, and Allen Tannenbaum. Geometric graph neural networks on multi-omics data to predict cancer survival outcomes. *Computers in Biology and Medicine*, 163, 9 2023. ISSN 18790534. doi: 10.1016/j.combiomed.2023.107117.
- [13] Zhenqin Wu, Alexandro E. Trevino, Eric Wu, Kyle Swanson, Honesty J. Kim, H. Blaize D'Angio, Ryan Preska, Gregory W. Charville, Piero D. Dalerba, Ann Marie Egloff, Ravindra Uppaluri, Umamaheswar Duvvuri, Aaron T. Mayer, and James Zou. Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. *Nature Biomedical Engineering*, 6(12):1435–1448, 12 2022. ISSN 2157846X. doi: 10.1038/s41551-022-00951-w.
- [14] Rukhma Aftab, Yan Qiang, Juanjuan Zhao, Zia Urrehman, and Zijuan Zhao. Graph Neural Network for representation learning of lung cancer. *BMC Cancer*, 23(1), 12 2023. ISSN 14712407. doi: 10.1186/s12885-023-11516-8.
- [15] Ying Wang and Yaqi Du. Graph neural network model GGDisnet for identifying genes in gastrointestinal cancer and single-cell analysis. *Computers in Biology and Medicine*, 172, 4 2024. ISSN 18790534. doi: 10.1016/j.combiomed.2024.108285.
- [16] Sadia Sultana Chowdhury, Sami Azam, Sidratul Montaha, Israt Jahan Payel, Md Rahad Islam Bhuiyan, Md Zahid Hasan, and Mirjam Jonkman. Graph neural network-based breast cancer diagnosis using ultrasound images with optimized graph construction integrating the medically significant features. *Journal of Cancer Research and Clinical Oncology*, 149(20):18039–18064, 12 2023. ISSN 14321335. doi: 10.1007/s00432-023-05464-w.
- [17] Shunxin Xiao, Huibin Lin, Conghao Wang, Shiping Wang, and Jagath C. Rajapakse. Graph Neural Networks with Multiple Prior Knowledge for Multi-Omics Data Analysis. *IEEE Journal of Biomedical and Health Informatics*, 27(9):4591–4600, 9 2023. ISSN 21682208. doi: 10.1109/JBHI.2023.3284794.
- [18] Wanting Yang, Yan Qiang, Wei Wu, and Jialong Xin. Graph-ETMB: A graph neural network-based model for tumour mutation burden estimation. *Computational Biology and Chemistry*, 105, 8 2023. ISSN 14769271. doi: 10.1016/j.combiolchem.2023.107900.

- [19] Pushpak Pati, Guillaume Jaume, Antonio Foncubierta-Rodríguez, Florinda Feroce, Anna Maria Anniciello, Giosue Scognamiglio, Nadia Brancati, Maryse Fiche, Estelle Dubruc, Daniel Riccio, Maurizio Di Bonito, Giuseppe De Pietro, Gerardo Botti, Jean Philippe Thiran, Maria Frucci, Orcun Goksel, and Maria Gabrani. Hierarchical graph representations in digital pathology. *Medical Image Analysis*, 75, 1 2022. ISSN 13618423. doi: 10.1016/j.media.2021.102264.
- [20] Nabil Wasif, Clifford Y. Ko, James Farrell, Zev Wainberg, Oscar J. Hines, Howard Reber, and James S. Tomlinson. Impact of tumor grade on prognosis in pancreatic cancer: Should we include grade in AJCC staging? *Annals of Surgical Oncology*, 17(9): 2312–2320, 9 2010. ISSN 10689265. doi: 10.1245/S10434-010-1071-7/FIGURES/2. URL <https://link.springer.com/article/10.1245/s10434-010-1071-7>.
- [21] S. Dinesh Krishnan, Danilo Pelusi, A. Daniel, V. Suresh, and Balamurugan Balusamy. Improved graph neural network-based green anaconda optimization for segmenting and classifying the lung cancer. *Mathematical Biosciences and Engineering*, 20(9):17138–17157, 2023. ISSN 15510018. doi: 10.3934/mbe.2023764.
- [22] Yongqing Zhang, Shuwen Xiong, Zixuan Wang, Yuhang Liu, Hong Luo, Beichen Li, and Quan Zou. Local augmented graph neural network for multi-omics cancer prognosis prediction and analysis. *Methods*, 213:1–9, 5 2023. ISSN 10959130. doi: 10.1016/j.ymeth.2023.02.011.
- [23] Jun Shi, Tong Shu, Kun Wu, Zhiguo Jiang, Liping Zheng, Wei Wang, Haibo Wu, and Yushan Zheng. Masked hypergraph learning for weakly supervised histopathology whole slide image classification. *Computer Methods and Programs in Biomedicine*, 253:108237, 8 2024. ISSN 0169-2607. doi: 10.1016/J.CMPB.2024.108237.
- [24] Roozbeh Bazargani, Ladan Fazli, Martin Gleave, Larry Goldenberg, Ali Bashashati, and Septimiu Salcudean. Multi-scale relational graph convolutional network for multiple instance learning in histopathology images. *Medical Image Analysis*, 96, 8 2024. ISSN 13618423. doi: 10.1016/j.media.2024.103197.
- [25] David Cheng. Pancreatic Cancer: Clinical Trials
}Drive Therapy Towards Precision
}Medicine. 1 2025.
- [26] Isaac Furtney, Ray Bradley, and Mansur R. Kabuka. Patient Graph Deep Learning to Predict Breast Cancer Molecular Subtype. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(5):3117–3127, 9 2023. ISSN 15579964. doi: 10.1109/TCBB.2023.3290394.
- [27] Tianci Song, Eric Cosatto, Gaoyuan Wang, Rui Kuang, Mark Gerstein, Martin Ren-jiang Min, and Jonathan Warrell. Predicting spatially resolved gene expression via tissue morphology using adaptive spatial GNNs. *Bioinformatics*, 40(Supplement_2):ii111–ii119, 9 2024. ISSN 13674811. doi: 10.1093/BIOINFORMATICS/BTAE383. URL <https://dx.doi.org/10.1093/bioinformatics/btae383>.

- [28] Jianliang Gao, Tengfei Lyu, Fan Xiong, Jianxin Wang, Weimao Ke, and Zhao Li. Predicting the Survival of Cancer Patients With Multimodal Graph Neural Network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(2):699–709, 2022. ISSN 15579964. doi: 10.1109/TCBB.2021.3083566.
- [29] Fengyun Zhang, Jie Geng, De Gan Zhang, Jinglong Gui, and Ran Su. Prediction of cancer recurrence based on compact graphs of whole slide images. *Computers in Biology and Medicine*, 167:107663, 12 2023. ISSN 0010-4825. doi: 10.1016/J.COMPBIOMED.2023.107663.
- [30] Zijian Yang, Yibo Zhang, Lili Zhuo, Kaidi Sun, Fanling Meng, Meng Zhou, and Jie Sun. Prediction of prognosis and treatment response in ovarian cancer patients from histopathology images using graph deep learning: a multicenter retrospective study. *European Journal of Cancer*, 199, 3 2024. ISSN 18790852. doi: 10.1016/j.ejca.2024.113532.
- [31] Bilin Liang, Haifan Gong, Lu Lu, and Jie Xu. Risk stratification and pathway analysis based on graph neural network and interpretable algorithm. *BMC Bioinformatics*, 23(1), 12 2022. ISSN 14712105. doi: 10.1186/s12859-022-04950-1.
- [32] Wenqi Lu, Michael Toss, Muhammad Dawood, Emad Rakha, Nasir Rajpoot, and Fayyaz Minhas. SlideGraph+: Whole slide image level graphs to predict HER2 status in breast cancer. *Medical Image Analysis*, 80:102486, 8 2022. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2022.102486.
- [33] Ateeq M. Khaliq, Meenakshi Rajamohan, Omer Saeed, Kimia Mansouri, Asif Adil, Chi Zhang, Anita Turk, Julianne L. Carstens, Michael House, Sikander Hayat, Ganji P. Nagaraju, Sam G. Pappas, Y. Alan Wang, Nicholas J. Zyromski, Mateusz Opyrchal, Kelvin P. Lee, Heather O'Hagan, Bassel El Rayes, and Ashiq Masood. Spatial transcriptomic analysis of primary and metastatic pancreatic cancers highlights tumor microenvironmental heterogeneity. *Nature Genetics* 2024 56:11, 56(11):2455–2465, 9 2024. ISSN 1546-1718. doi: 10.1038/s41588-024-01914-4. URL <https://www.nature.com/articles/s41588-024-01914-4>.
- [34] Kexin Ding, Mu Zhou, He Wang, Shaoting Zhang, and Dimitri N. Metaxas. Spatially aware graph neural networks and cross-level molecular profile prediction in colon cancer histopathology: a retrospective multi-cohort study. *The Lancet Digital Health*, 4(11):e787–e795, 11 2022. ISSN 25897500. doi: 10.1016/S2589-7500(22)00168-6. URL <https://www.thelancet.com/action/showFullText?pii=S2589750022001686> <https://www.thelancet.com/action/showAbstract?pii=S2589750022001686> [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(22\)00168-6/abstract](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00168-6/abstract).
- [35] Caixia Sun, Tao Luo, Zhenyu Liu, Jia Ge, Lizhi Shao, Xiangyu Liu, Bao Li, Song Zhang, Qi Qiu, Wei Wei, Shuo Wang, Xiu Wu Bian, and Jie Tian. Tumor Mutation Burden–Related Histopathologic Features for Predicting Overall Survival in Gliomas Using Graph Deep Learning. *American Journal of Pathology*, 193(12):2111–2121, 12 2023. ISSN 15252191. doi: 10.1016/J.AJPATH.2023.08.015/ASSET/1055484C-4D69-4749-BB1C-42DDEAB7926F/MAIN.ASSETS/FIGS6.JPG. URL <https://ajp.amjpathol.org/action/showFullText?pii=S0002944023003589>

<https://ajp.amjpathol.org/action/showAbstract?pii=S0002944023003589>
[https://ajp.amjpathol.org/article/S0002-9440\(23\)00358-9/abstract](https://ajp.amjpathol.org/article/S0002-9440(23)00358-9/abstract).

- [36] Filmon Yacob, Jan Siarov, Kajsa Villiamsson, Juulia T. Suvilehto, Lisa Sjöblom, Magnus Kjellberg, and Noora Neittaanmäki. Weakly supervised detection and classification of basal cell carcinoma using graph-transformer on whole slide images. *Scientific Reports* 2023 13:1, 13(1):1–10, 5 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-33863-z. URL <https://www.nature.com/articles/s41598-023-33863-z>.

Appendix A

Supplementary Figures

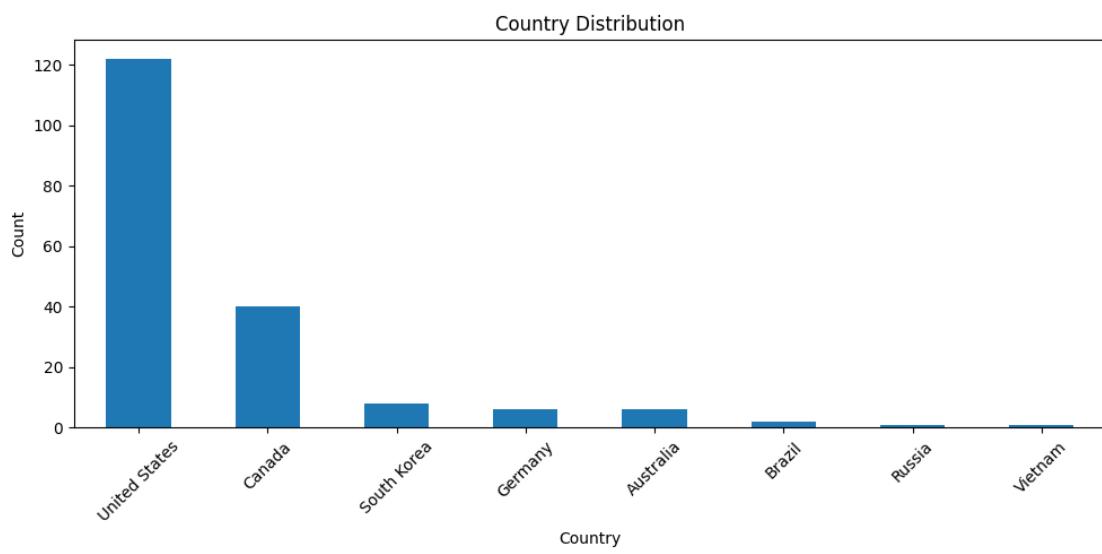


Figure A.1: Country Demographic of all 232 Unique patients present in the TCGA Data for Pancreatic Cancer

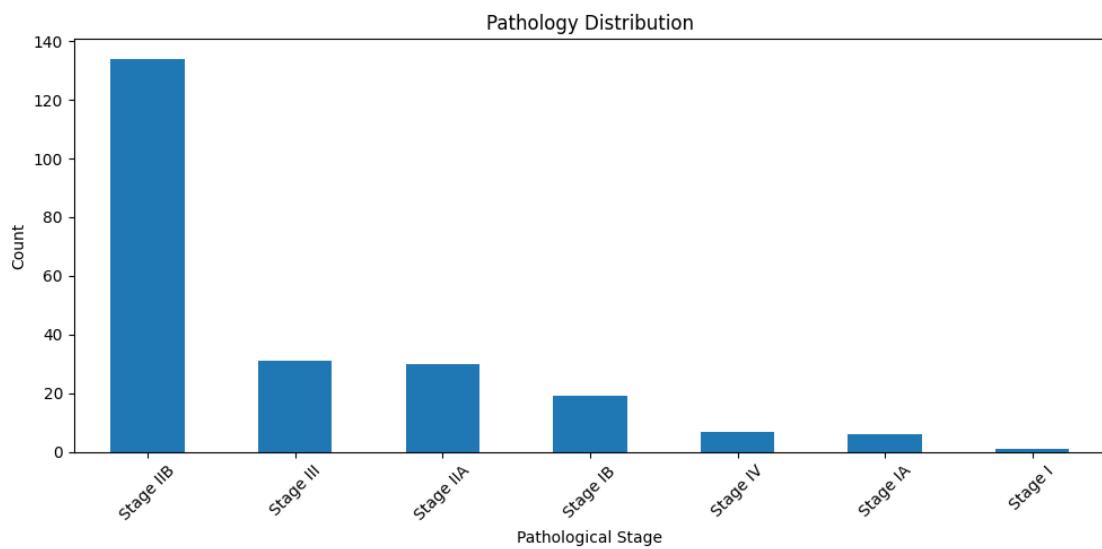


Figure A.2: Pathological Staging for 232 unique patients in TCGA Pancreatic Cancer Data

Race Distribution

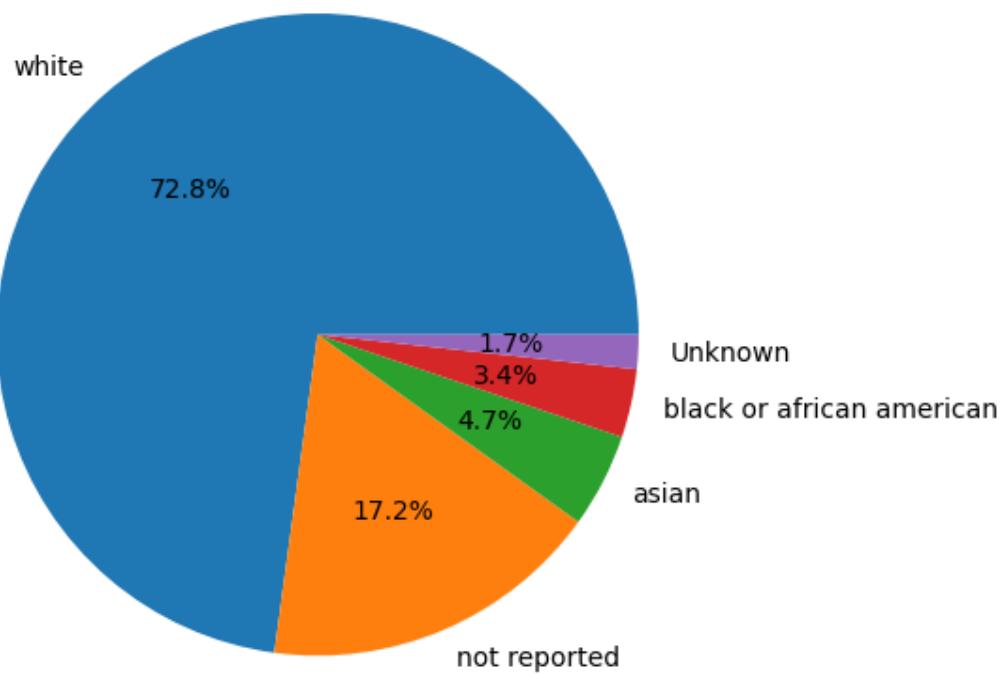


Figure A.3: Race Distribution for 232 unique patient IDs in TCGA Pancreatic Cancer Data

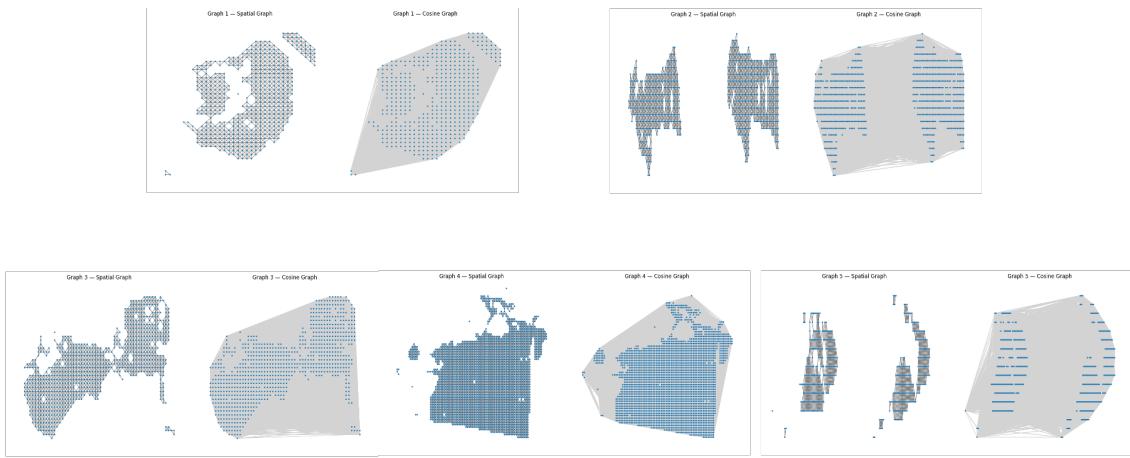


Figure A.4: Examples of Spatial Graphs with edges both outlining the spatial edges and embedding edges

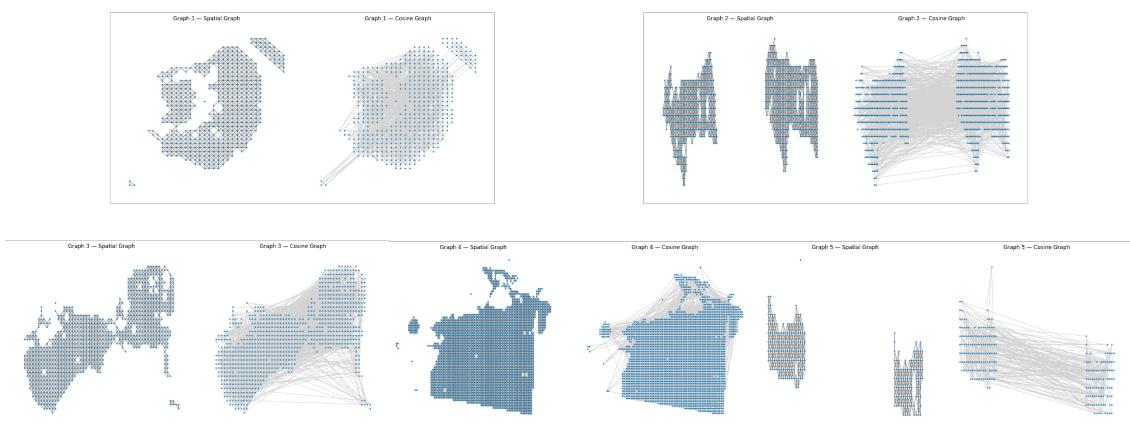


Figure A.5: Examples of Spatial Graphs with edges both outlining the spatial edges and the kNN reduced embedding edges

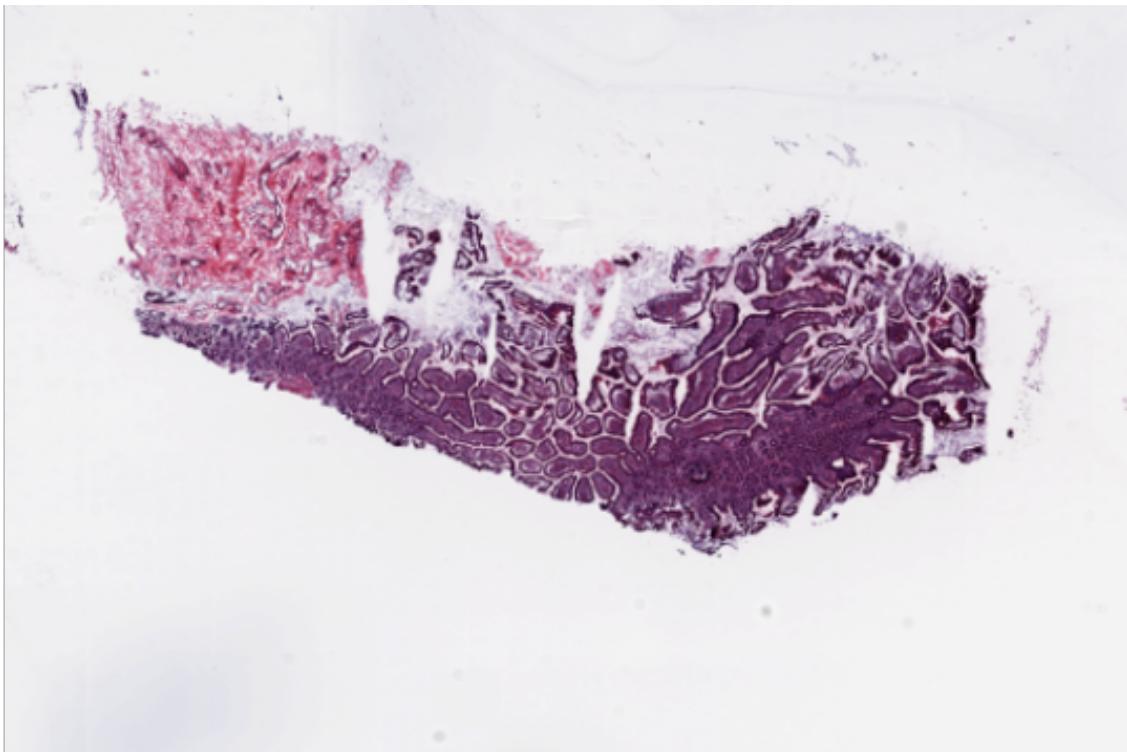


Figure A.6: Example of Whole Slide Image Classified as Normal Tissue

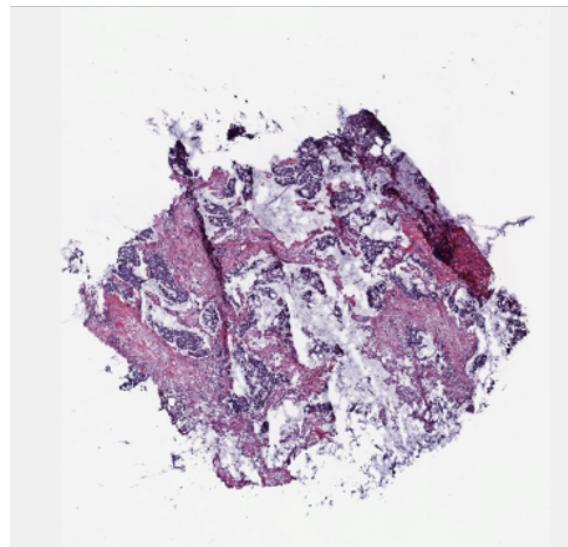


Figure A.7: Example of Whole Slide Image Classified as Tumor Tissue