Deep Learning Based Analysis of Cancer Similarities and Subtyping Using Histopathological Whole Slide Images

Thesis submitted in partial fulfilment
of the requirements for the degree of

*Master of Science in **Computer Science and Engineering** by Research*

by

Piyush Singh
2020701031
piyush.singh@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
October 2025

International Institute of Information Technology

Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Deep Learning Based Analysis of Cancer Similarities and Subtyping Using Histopathological Whole Slide Images" by Piyush Singh, has been carried out under my supervision and is not submitted elsewhere for a degree.

_____ _____

Date Advisor: Prof. C. V. Jawahar

_____ _____

Date Advisor: Prof. P. K. Vinod

To the cherished and esteemed memory of my father

Acknowledgments

I would like to extend my sincerest gratitude to my advisors, Prof. C. V. Jawahar and Prof. P. K. Vinod, for their invaluable guidance, encouragement, and support throughout my research journey. This work would not have been possible without their insightful feedback, steadfast patience, and constant motivation. I am deeply grateful for the countless discussions, constructive criticisms, and the freedom they provided me to explore my ideas while ensuring I remained on the right track. Their mentorship has been instrumental in shaping both my academic growth and personal development.

I would also like to sincerely thank my project mate, Ashish Menon, for his significant contributions to parts of this thesis. Together, we explored research ideas, discussed methods and results, and co-authored a research paper. I am grateful to my other project mates, Ekansh, Amit, and Vivek, for collaborating on related research problems in the medical imaging domain and for their valuable insights and discussions.

I am equally thankful to my seniors at CVIT, Avijit and Aditya, and my colleagues Rudrabha, Madhav, Siddhant, Kiran, Sheshadri, Kushal and Harsh, with whom I have had countless discussions spanning technical, philosophical, and personal topics. I also want to thank Karan for his assistance in obtaining images from our dataset for this thesis.

I am deeply grateful to Dr. Shantveer G. Uppin at NIMS Hyderabad for his instrumental role in curating the IPD-Lung dataset. His expertise and years of experience in histopathology were invaluable in carefully annotating the dataset and in explaining the pipeline for cancer diagnosis in a clear and thorough manner.

I would also like to thank the CVIT and IIIT staff members Rohitha, Cecilia, Nagaraju, Manasa, and Sakshi, who assisted me at various times with administrative and other issues.

Lastly, I wish to express my heartfelt gratitude to my family for their unwavering love and emotional support, which kept me motivated even during the most challenging times.

Abstract

Whole slide images (WSIs) contain rich information for computational pathology, yet systematic evaluations of cross-organ and multi-cohort generalization remain limited. This thesis addresses these challenges through two complementary studies. In the first study, patch-level convolutional neural networks (CNNs) were trained on 9,792 slides from The Cancer Genome Atlas (TCGA), spanning 11 cancer subtypes across seven organs, to distinguish cancerous from normal tissue. Both within-organ and cross-organ inference were evaluated, revealing that cancers such as breast, colorectal, and liver can be reliably detected by models trained on other organs. Strong transferability was also observed between subtypes within an organ, such as kidney and lung. To investigate these patterns, feature similarity, overlap in high-attention regions, and nuclear geometry were analyzed, all of which showed positive correlations with cross-organ transfer performance. The second study focuses on lung cancer in the Indian population through the introduction of IPD-Lung, a curated dataset of adenocarcinoma and squamous carcinoma cases. Benchmark evaluations were established using multiple instance learning (MIL) models, with additional experiments exploring model transferability from publicly available TCGA-Lung dataset. Stain normalization methods, particularly Macenko, reduced domain discrepancies to some extent but did not fully bridge performance gaps, especially for squamous carcinoma. A domain adaptation strategy based on a gradient reversal layer (GRL) similarly yielded limited improvements. In contrast, models trained directly on IPD-Lung achieved substantial performance gains, with further enhancements obtained by training on expert-annotated regions of interest (RoIs). Together, these studies demonstrate that deep learning models can uncover meaningful cross-organ similarities in cancer histopathology. At the same time, they highlight the difficulty of generalizing across cohorts, even for the same cancer subtype, underscoring the influence of dataset-specific factors such as scanners and staining protocols. Finally, the findings show that expert annotations can substantially improve performance in small and imbalanced datasets, providing a practical pathway to more robust and clinically relevant computational pathology models.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Cancer is a major global health risk, representing a significant cause of morbidity and mortality worldwide. Nearly 20 million new cases and 9.7 million deaths were attributed to various types of cancer worldwide in 2022. [1]. In terms of lifetime risk estimates, globally about one in five people will develop cancer during their lifetime, and about one in nine men and one in twelve women will eventually die from it. These statistics highlight both the high burden of cancer on healthcare systems around the world and the substantial lifetime risk it poses to populations.

Lung cancer stood out in 2022 as both the most frequently diagnosed cancer with almost 2.5 million new cases accounting for 12.4% of all cancers globally—and the leading cause of cancer-related death, responsible for 1.8 million deaths or 18.7% of all cancer mortality. After lung cancer, the most common cancers were those of the female breast (11.6%), colorectum (9.6%), prostate (7.3%), and stomach (4.9%). In terms of mortality, the hierarchy was lung, colorectal (9.3%), liver (7.8%), female breast (6.9%), and stomach (6.8%) cancers.

Characterized by the uncontrolled growth and spread of abnormal cells, cancer arises when the normal regulatory mechanisms that control cell proliferation, differentiation, and death are disrupted. This unregulated growth allows malignant cells to invade surrounding tissues and, in many cases, spread to distant organs through the blood or lymphatic systems, a process known as metastasis. The resulting disruption to normal tissue architecture and organ function can impair vital physiological processes, leading to progressive morbidity and, if untreated, death. Because these malignant transformations can occur in virtually any cell type, cancer has the potential to affect almost every organ system in the human body. Cancers are caused by a mix of genetic, environmental and lifestyle factors. Clinicians use TNM grading system to define the anatomical extent of cancer by coding the primary tumor (T), regional lymph nodes (N) and metastasis (M). These are further aggregated into overall stage groups (I to IV) to

guide treatment and standardize reporting. Although cancer may spread to distant sites, it is identified and named based on the primary organ where it first develops.

1.1 Motivation

Histopathology, the microscopic examination of tissue to study disease, remains a cornerstone of medical diagnosis, particularly in identifying and characterizing cancers. Accurate analysis of tissue samples provides essential insights into disease progression, informs treatment decisions, and can significantly impact patient outcomes. The process involves a sequence of steps given below and summarized in figure 1.1.

1. Tissue is obtained via biopsy, depending on the location, size, and type of the suspected lesion, and preserved in formalin to prevent degradation.

2. It is embedded in paraffin wax to provide structural support for thin sectioning.

3. The paraffin block is sliced into very thin sections, typically 4–5 µm, using a microtome.

4. Sections are mounted on glass slides and passed through a graded series of ethanol solutions of decreasing concentration to remove the paraffin and rehydrate the tissue.

5. Slides are stained, most often with hematoxylin and eosin solutions (H&E), where hematoxylin colors cell nuclei purple and eosin stains the cytoplasm and extracellular matrix pink.

6. The prepared slides are examined under a microscope by a pathologist to assess tissue architecture, cellular morphology, and abnormalities.

However, performing this process and extracting insights reliably often requires clinicians to have expertise in histopathology which can be only be obtained with several years, sometimes decades of training and experience in the domain. Finding such experts can be difficult in remote places. Furthermore, the diagnostic process is time-intensive and prone to inter-observer variability, where different experts may offer differing opinions, as well as intra-observer variability, where the same pathologist may reach inconsistent conclusions at different times. Additionally, difficulties in maintaining consistent training standards pose a systemic barrier in ensuring uniform diagnostic quality across institutions

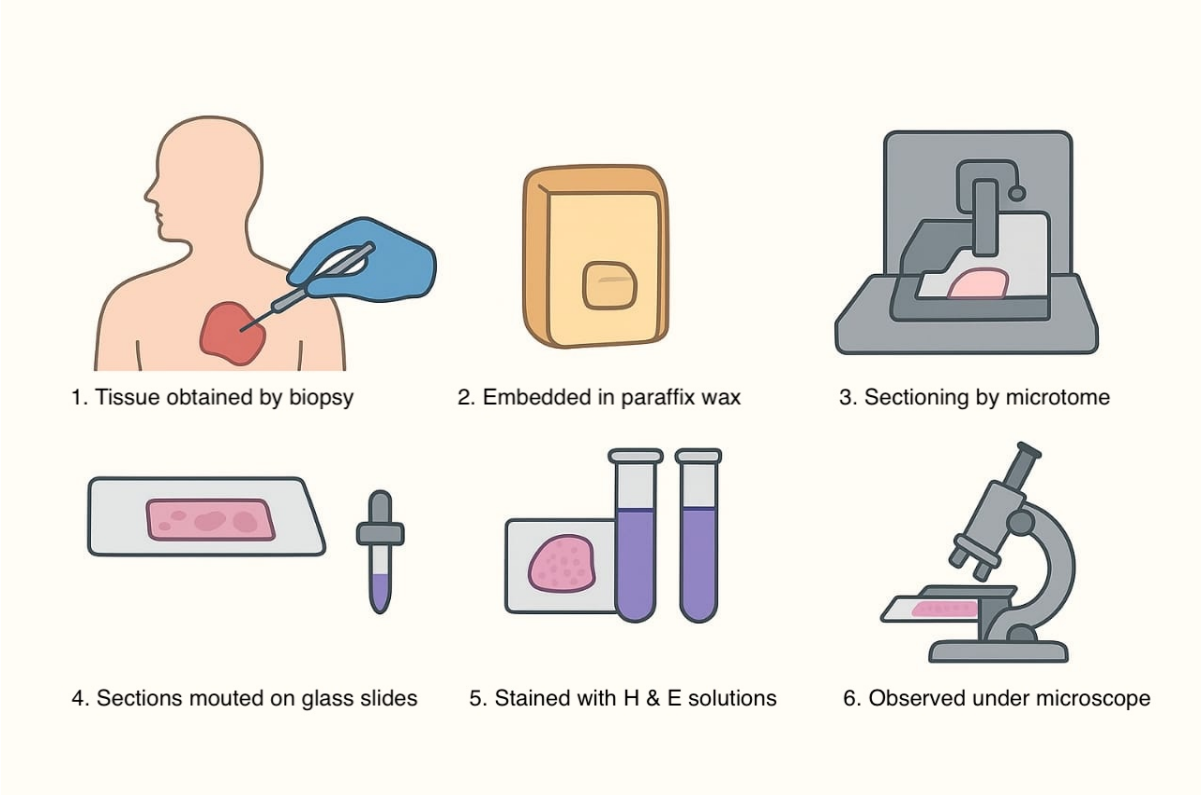

Figure 1.1: Sequence of steps in histopathology

and regions. These challenges present an opportunity for technological intervention, enabling the development of tools that can augment the diagnostic process by improving consistency, efficiency, and accessibility across diverse clinical settings.

The digital transformation of pathology offers a promising solution. By digitizing histopathology slides, large, high-resolution images can be stored, analyzed, and shared across healthcare networks, creating a new field of digital pathology. These digital slides, popularly called as whole slide images or WSIs are unique due to their size and complexity, offering vast amounts of information at multiple zoom levels—enabling detailed cellular examination at high magnifications and a holistic view at lower ones. This multiscale nature of histopathological images requires specialized storage solutions, often utilizing a pyramidal TIFF file structure to facilitate efficient data retrieval and zooming capabilities. Digital images can also be transmitted for collaborative diagnosis, allowing multiple specialists to review and discuss cases regardless of location, enhancing diagnostic accuracy and facilitating expert consultation. These systems also enable training and evaluation of machine learning based automated

diagnosis systems which will fill in the gaps in traditional pathology discussed earlier. Computational or digital pathology enables a wide range of essential tasks for cancer diagnosis, prognosis, and personalized treatment planning.

One fundamental task is classifying whether or not a patient has cancer. Following this is the task of classifying the subtype of cancer. Further downstream tasks follow from here such as region detection, where models identify malignant cells or abnormal regions in tissue samples, helping pathologists pinpoint areas of concern efficiently and segmentation, distinguishing tumor boundaries and subregions like stroma, epithelium, and necrotic areas, which aids in understanding tumor growth patterns and microenvironment interactions.

1.2 Thesis structure

We observed that models trained to detect cancer on one organ can often detect cancer in other organs surprisingly well, suggesting that these cancers have similarities that deep learning models are able to identify. The first part of the thesis delves deeper into these similarities. We study these similarities among cancers at the model accuracy level and then dive deeper into feature-level similarities. Further we examine the overlapping areas between regions of interest identified by models trained on different cancers and then finally we study the similarities between the nuclear geometry of identified regions of interest using statistical analysis.

The second part of the thesis is about lung cancer in Indian context. As we saw in earlier section, lung cancer is the most commonly occurring cancer and also the cause of maximum deaths worldwide. In India, respiratory cancers (including trachea, bronchus, and lung) account for 5.8% of new cancer cases and 8.2 % of deaths which rank 4th in terms of incidence and mortality both, below breast, oral and cervical cancers. [2]. We curated a dataset of histopathology images of Indian lung cancer patients which we call as IPD-Lung Dataset (Indian Pathology Dataset - Lungs). We discuss the salient features of this dataset and train and evaluate baseline classification models for subtype classification.

Chapter 2

Deep learning based exploration of cancer similarities in histological images

The availability of large-scale public datasets, combined with the rapid growth of GPU-powered computing capable of efficiently performing matrix algebra, triggered a new era of machine learning driven by multi-layered neural networks, known as deep learning. Within deep learning, convolutional neural networks (CNNs) emerged as a breakthrough architecture, where the convolutional layer enabled automatic extraction of spatial features, making them particularly effective for image recognition and other related tasks. Their immense potential was clearly demonstrated in 2012, when AlexNet [3], a deep CNN, achieved a landmark result on the ImageNet Large Scale Visual Recognition Challenge [4], catalyzing the widespread adoption of CNNs in computer vision. Subsequently, CNNs became a cornerstone of modern AI, driving progress in healthcare through medical image analysis and disease diagnosis, in autonomous systems through object detection and navigation, in security through facial recognition and surveillance, in multimedia through video classification, content recommendation, and image enhancement, in scientific research through satellite image interpretation and climate modeling, and in industry through defect detection, robotics, and quality control, among several other application areas.

Convolutional neural networks have become indispensable in medical imaging and AI-assisted diagnostic workflows. Extensive reviews illustrate that CNN-based methods have excelled in critical tasks such as segmentation, disease classification, anomaly detection, computer-aided diagnosis, and retrieval across diverse imaging modalities—including MRI, CT, X-ray, ultrasound, PET, and histopathology [5, 6]. Several studies [7, 8, 9, 10, 11, 12] have demonstrated the value of using ImageNet-pretrained CNN models for histopathological image analysis. Multiple works have adopted the pipeline of patch

extraction and filtering out background patches, then fine-tuning an ImageNet-pretrained CNN model such as Inception-v3 [13] to classify tissue patches as cancerous or non-cancerous and in some cases further predict the cancer subtype, genetic mutation or survival. Coudray et al. [10] presented this pipeline for lung cancer classification and mutation prediction, Kwok et al. [14] for breast cancer classification and Tabibu et al. [9] for renal cell carcinomas for the task of cancer detection and subtype classification at patch-level and aggregated at slide-level. Wang et al. [11] performed background segmentation using HSV color space-based thresholding and used a GoogLeNet [15] model pretrained on ImageNet to identify regions of metastatic breast cancer. Xu et al. [8] applied CNN-based feature extraction and fully connected networks for classification and segmentation tasks in brain and colon pathological slides. Pan-cancer studies using deep learning are emerging as well. Fu et al. [16] utilized CNN-derived features from cancer vs normal classification models to predict genomic, molecular, and prognostic features across organs. Cheerla et al. [17] used a multimodal deep learning approach to predict patient survival by combining genetic data and histopathological image features from multiple cancer types. Noorbakhsh et al. [18] studied cross-organ correlations using slide-wise ROC-AUC scores, computed from patch-level predictions using a thresholded fraction of cancerous patches. They used Inception-v3 [13] as a feature extractor and fine-tuned only the final fully connected layer to classify slides. Their work included hierarchical clustering of slide-wise ROC-AUC values and correlation analysis of model logits across organs to suggest shared tumor morphology. We build on these ideas to conduct a more in-depth quantitative and qualitative analysis of cross-organ morphological correlations using deep learning models.

2.1 Dataset

In this study, we made use of the publicly available whole slide image (WSI) collection from the TCGA project [19], comprising 9,792 formalin-fixed paraffin-embedded (FFPE) slides spanning seven organ systems. FFPE sections were chosen because they provide higher fidelity in preserving tissue morphology compared to frozen slides, which often contain artifacts that compromise diagnostic accuracy [20]. The methodology to create these slides is given in Chapter 1.

The dataset covered 11 cancer types from 7 organs as follows:

1. Breast: Breast invasive carcinoma (BRCA)

2. Colorectal: Colon adenocarcinoma (COAD), Rectum adenocarcinoma (READ)

3. Kidney: Renal clear cell carcinoma (KIRC), Renal papillary cell carcinoma (KIRP), Chromophobe renal carcinoma (KICH)

4. Liver: Liver hepatocellular carcinoma (LIHC)

5. Lung: Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC)

6. Prostate: Prostate adenocarcinoma (PRAD)

7. Stomach: Stomach adenocarcinoma (STAD)

A detailed distribution of the number of slides and images for each cancer type is shown in Figure 2.1.

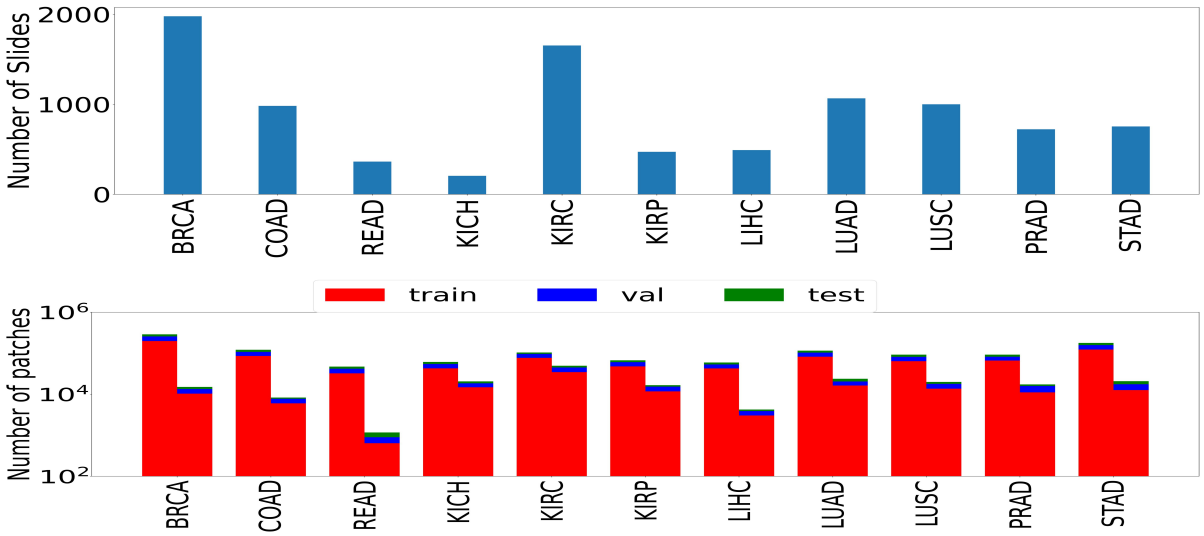

Figure 2.1: The number of slides (top) and patches(bottom) used in the study. Number of patches belonging to both classes (left bar represents cancer samples and right bar represents normal samples) are shown in the form of two rectangular bar plots.

2.2 Implementation details

This work followed a stepwise pipeline, beginning with pre-processing and data preparation, followed by model training and optimization. The resulting models were then used to investigate cross-organ similarities through performance metrics, feature analyses, and morphological comparisons.

2.2.1 Pre-processing

Whole-slide images (WSIs) used in this study are hematoxylin and eosin (H&E) stained and often in gigapixel scale, making it infeasible to process them in their entirety by ingesting them directly into CNNs. Also downsampling to smaller resolutions has been shown to degrade fine cellular and subcellular structures, thereby reducing diagnostic accuracy [21]. To address this, WSIs were divided into smaller image patches (or tiles), a widely used strategy in computational pathology [22].

Following the approach of Coudray et al. [10], we extracted non-overlapping 512 × 512 patches at 20× magnification. Background and low-information tiles were removed using the patch filtering pipeline described in [9]. In addition, we implemented an additional filtering step to exclude patches with fractal-like noise by requiring at least ten connected components in the binarized mask. As patch-level ground-truth was unavailable for TCGA, each patch inherited the slide-level label, a strategy validated in prior works [10, 9].

The dataset was split into training, validation, and test subsets with a 70–20–10 ratio, ensuring class balance at the patient level. To improve generalization, we applied common data augmentation techniques including random cropping and horizontal flipping. All patches were normalized channel-wise using statistics computed from the training set.

2.2.2 Model Training

For the classification task, we adopted the ResNet-18 architecture [23], initialized with weights pre-trained on ImageNet. Residual networks have demonstrated strong performance in medical imaging applications [24], and ResNet-18 provided a good balance between representational capacity and computational efficiency. While deeper models such as ResNet-34, ResNet-50, and ResNet-101 were also considered, we found only marginal performance improvements at the expense of significantly more parameters. The final fully connected layer of ResNet-18, originally designed for 1,000 ImageNet classes, was replaced with a fully connected network (FCN) with a binary classification head. The schematic flow diagram is shown in Figure 2.2 for the classification task.

All model parameters were optimized using backpropagation with cross-entropy loss. To address class imbalance between tumor and normal samples, we employed weighted cross-entropy combined with stratified sampling. The training batch size was fixed at 256.

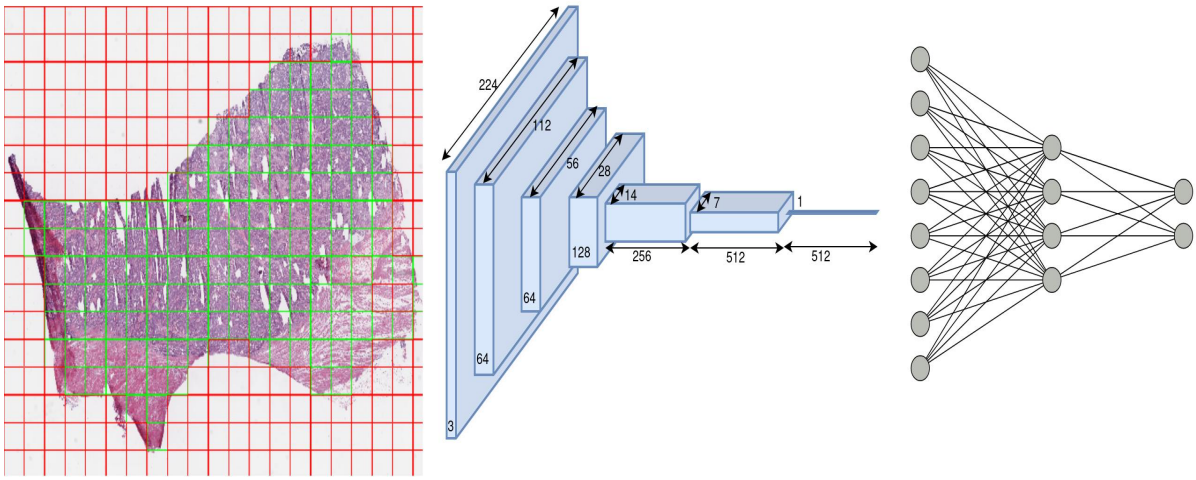

Figure 2.2: Overview of architecture used in our work: Patch extraction (left): red shows rejected background patches and green shows patches used for training model, ResNet-18 architecture (middle) and Fully connected network (right).

2.2.3 Hyperparameter tuning

Hyperparameter optimization was conducted using the Optuna framework [25]. The search space included:

- Optimizer: Adam, RMSProp, or SGD (categorical).
- Learning rate: log-uniform between $[1e-05, 1e-01]$.
- Dropout rate: uniform between 0.2 and 0.5.
- Fully connected layers: 1–3.
- Neurons per layer: 4–128.

We executed 20 trials per organ, training each configuration for 20 epochs, and selected the best-performing set of hyperparameters based on validation accuracy. Final models were trained for 50 epochs using these optimal settings. To test their utility, hyperparameter optimization was carried out on datasets from four different organs. This analysis demonstrated clear gains across multiple performance indicators, such as classification accuracy, AUC, and F1-score. Given these consistent improvements, the same optimization protocol was systematically applied to the other organs. A summary of the hyperparameter tuning and its effect on validation performance is illustrated in figure 2.3

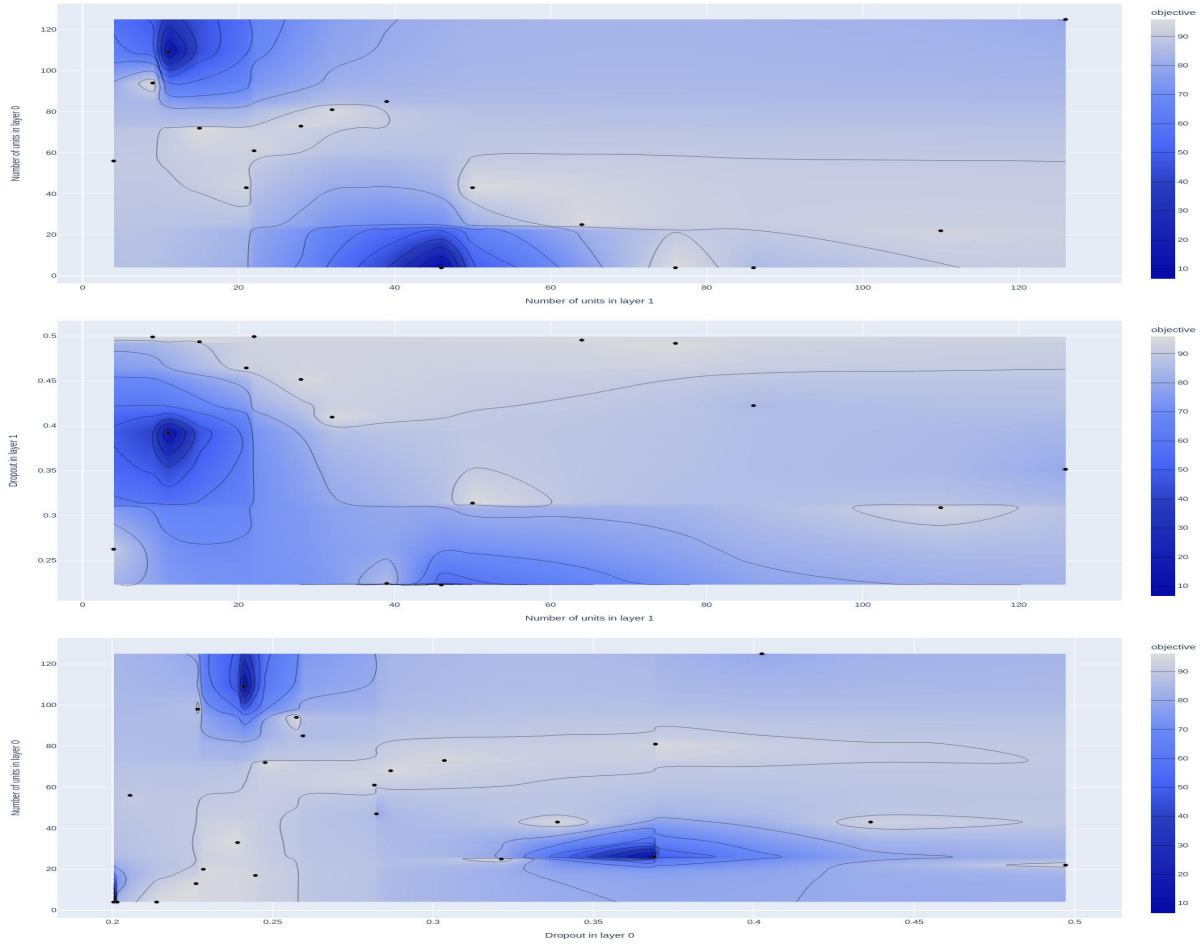

Figure 2.3: Example contour plot obtained during hyperparameter tuning.

2.3 Experiments and Results

In this study, we focus on exploring similarities among different cancer types through a deep learning–based approach. The dataset comprises of 11 cancer subtypes, each partitioned into training, validation, and test sets. The training and validation sets were utilized for model development, including initial training and fine-tuning of hyperparameters. Once we had the 11 models, we utilized the test sets which we had reserved exclusively for downstream analysis of exploring similarities.

Thus our framework is organized into two components:

- Intra-subtype evaluation: This analysis quantifies how well the model captures discriminative features specific to each cancer subtype. For this we run inference of each model on the test set of the same subtype that it was trained on.

- Cross-organ inference: To investigate relationships and potential similarities across cancers, each trained model was also evaluated on the test sets of all other cancers. This design yields a cross-inferencing matrix of size 11 × 11 (121 combinations), where diagonal entries correspond to within-subtype performance and off-diagonal entries reflect transferability across cancer types.

2.3.1 Intra-subtype evaluation

A high classification performance was observed for most models in terms of accuracy, F1 score and ROC-AUC. (Figure 2.4).

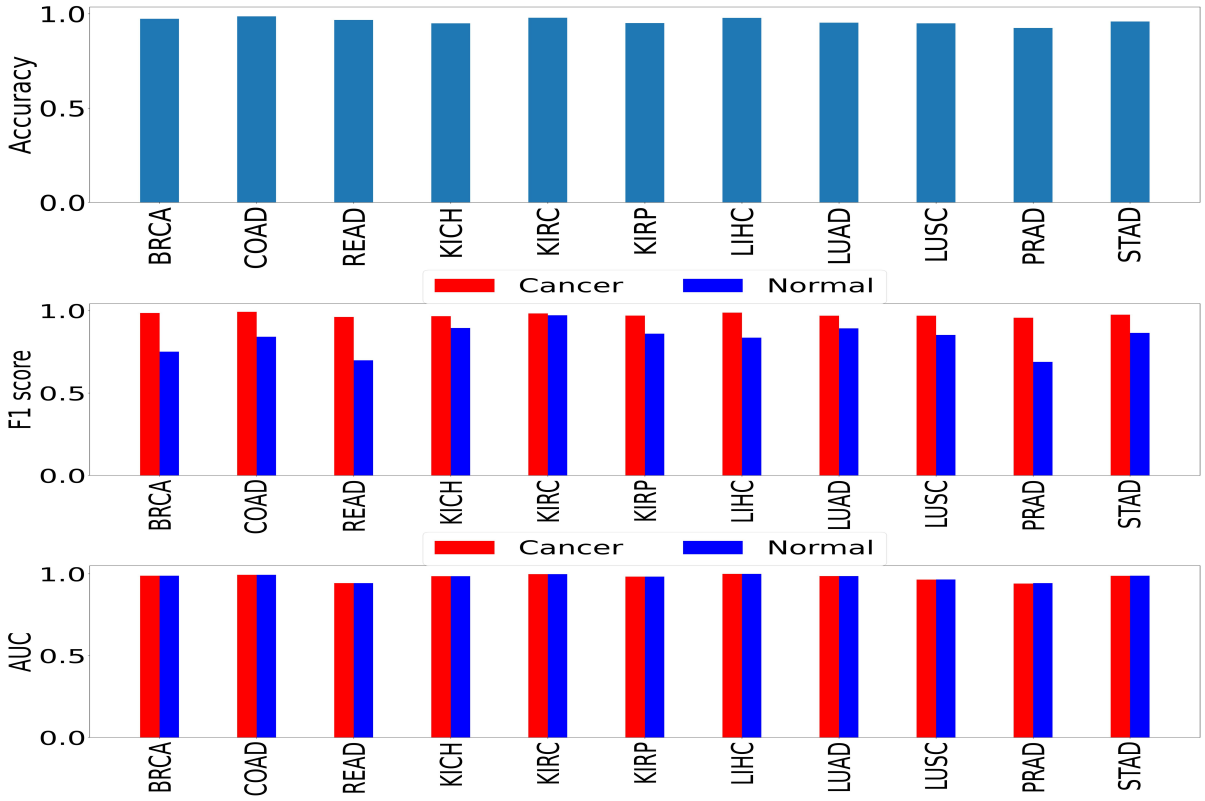

Figure 2.4: Intra-subtype inference showing the performance obtained using models trained on each cancer and tested on the held out test set of the same cancer.

2.3.2 Cross-organ inference

For this experiment, we run the inference of each of these models against the held-out test sets of all of the other cancers. As a result, we get the matrix given in Figure 2.5

Figure 2.5: cross-organ inference results: Accuracies obtained using models trained on the organs along the rows and tested on the organs along the column are shown.

The cross-organ inference matrix highlights instances where a model trained on organ A exhibits strong predictive performance on organ B. Such cases suggest potential histological similarities between the two cancers. To further examine this, we summarize in Table 2.1 the best- and worst-performing organ pairs according to ROC-AUC and F1 scores. Subsequent analyses focus on one high-performing (good) pair and one low-performing (bad) pair in order to investigate the factors underlying their differences.

Model	F1 score				AUC			
	best		worst		best		worst	
BRCA	READ	0.9443	KICH	0.6600	READ	0.9837	KIRC	0.7815
COAD	READ	0.9799	KIRC	0.5287	READ	0.9981	KIRC	0.6246
KICH	COAD	0.9294	STAD	0.7519	KIRP	0.9783	STAD	0.8163
KIRC	KIRP	0.9423	PRAD	0.7678	KICH	0.9881	PRAD	0.8069
KIRP	KICH	0.9490	PRAD	0.6893	READ	0.9840	PRAD	0.6692
LIHC	READ	0.9442	KIRC	0.6157	READ	0.9893	KICH	0.7203
LUAD	LUSC	0.9381	KIRC	0.5675	LUSC	0.9831	KIRC	0.6256
LUSC	BRCA	0.9251	KIRC	0.5769	LUAD	0.9683	KIRC	0.5998
PRAD	BRCA	0.9422	KICH	0.5632	LIHC	0.9453	KICH	0.5481
READ	COAD	0.9680	KIRC	0.4987	LIHC	0.9507	KIRC	0.5246
STAD	READ	0.9410	KICH	0.6921	BRCA	0.9822	KICH	0.8062

Table 2.1: Cross-organ inference indicating the quantitative results of best and worst inference of individually trained models when tested on other unseen organs.

We observed that cross-subtype inference was notably higher when models were evaluated on BRCA, LIHC, COAD, and READ (Figure 2.5), indicating the presence of morphological traits that may be shared across these cancers. Among them, the two colorectal cancers (COAD and READ) displayed strong mutual predictability and also showed affinity with BRCA and LIHC. These results are in agreement with earlier pan-cancer clustering studies, which reported groupings of gynecological and gastrointestinal malignancies based on histological patterns [18].

In contrast, models exhibited limited transferability to kidney subtypes (KIRC, KIRP, KICH) and lung cancers (LUAD, LUSC), suggesting that the morphological signatures of these organs are highly organ-specific and less generalizable across cancer types. The distinctiveness of renal tumors has also been highlighted in molecular studies, where kidney cancers demonstrated unique expression profiles compared to other cancers [26]. Interestingly, the relationship among kidney subtypes was not symmetric: while KICH and KIRP models performed poorly on KIRC, the reverse was not observed, implying that KIRC contains features that are subtype-specific and less represented in other kidney cancers.

Another intriguing observation was that despite both READ and STAD being gastrointestinal tumors, the READ-trained model did not achieve high performance on STAD, highlighting subtle but meaningful differences in morphology. Similarly, inference patterns across adenocarcinomas (LUAD, COAD, PRAD, READ, STAD) were not consistent, further underscoring that while adenocarcinomas of different organs might have some shared morphological traits owing to their grandular architecture, they also retain organ-specific differentiation which deep models are finding difficult to generalize.

2.3.3 Feature similarity analysis

The t-SNE embedding was obtained for different model-organ pairs. t-SNE was performed using a perplexity of 30, 1,500 optimization iterations, a learning rate of 200, and PCA initialization. Feature vectors were extracted from the penultimate layer of each trained model using samples from the held-out test sets and were L2-normalized prior to dimensionality reduction. Figure 2.6 shows t-SNE plots for all of the 11 models. The subplot title denotes the organ and subtype used to train the model and the plotted points denote sample images in the feature space, dot for cancerous and cross for normal whereas color denotes the label of the test set from which the samples have been taken. For visualization purposes, 100 random samples from each test set were selected for plotting (50 cancerous and 50 normal). This balanced sampling was adopted to ensure clear visualization of both classes while avoiding excessive crowding in the scatterplot.

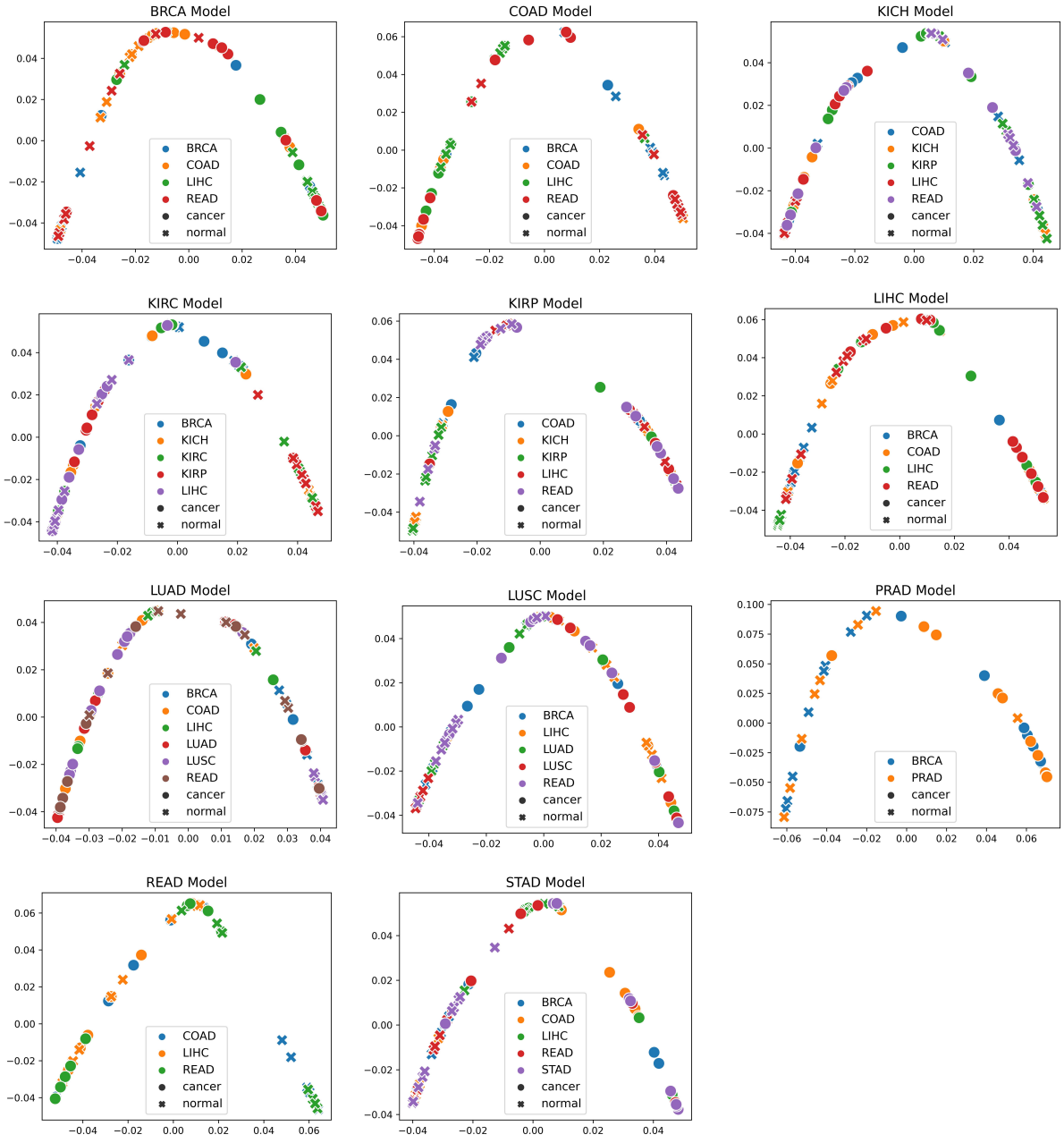

Figure 2.6: t-SNE embeddings of the trained models (mentioned in the title of each figure) helping to visualize the separability of cancer and normal embeddings of organs unseen by the trained models.

The t-SNE visualizations demonstrate that the trained models can effectively separate cancerous and normal tissue patches in the learned feature space, particularly for the cancer subtype they were optimized on and for those subtypes/organs where cross-inference accuracy exceeded 90%. All organs with accuracy greater than 90% were projected jointly so that the scatter plot is in the same 2D space. The overlap of cancer samples across organs reflects shared learned features. At the same time, the embeddings also reveal regions where cancer and normal samples appear in close proximity within the two-dimensional projection. This overlap may arise due to several factors, such as residual classification errors of the models, information loss or distortion during dimensionality reduction into 2D space, and the simplifying assumption that every patch extracted from a cancer slide necessarily corresponds to malignant tissue.

2.3.4 RoI overlap analysis

We have already seen that models with good cross-performance are able to place cancerous samples together and normal samples together which then shows separability between cancerous and normal samples across organs and cancer subtypes. The same does not hold for models with bad cross-performance. Now we seek to examine if this trend is also reflected in terms of region of interest (RoI) overlap. For this we utilized GradCAM [27] and followed the following steps:

1. Heatmap generation: GradCAM was applied to patches to identify high-saliency regions that contributed most strongly to the model's decision.

2. Color space transformation: Each heatmap was converted to HSV space to facilitate thresholding, since the hue channel provides a reliable basis for separating the high-activation regions.

3. Segmentation of salient regions: A thresholding operation was applied to isolate the most activated areas, corresponding to the regions of high model attention.

4. Contour detection and refinement: Canny edge detection was performed on the thresholded maps, followed by polygonal approximation to obtain closed contours.

5. Bounding box extraction: From the refined contours, rectangular bounding boxes were computed to produce a compact representation of each region of interest.

6. Quantification of overlap: To examine whether different organ-trained models attended to similar histological structures, we measured the similarity between detected regions using (i) Intersection over Union (IoU) of bounding boxes, and (ii) the Jaccard index of the thresholded pixel maps.

Figure 2.7 shows the degree of overlap between GradCAM outputs using the IoU of the bounding boxes and Jaccard index of the thresholded pixelmaps described above. Each subplot shows 3 or 4 high overlap pairs and 2 low overlap pairs. We can see that models which can achieve high performance on another cancer subtype are also able to localize the tumor to a region with high overlap with the model trained on the other cancer. This can be seen in the IoU and Jaccard index values. Conversely, models with poor cross performance have low overlap with the models trained on the other cancer. For instance, the BRCA model achieves the highest cross-organ accuracy, as well as the highest IoU and Jaccard index on COAD whereas low cross-organ accuracy as well as low IoU and Jaccard Index on KICH and KIRC.

Figure 2.8 illustrates the GradCAM-based visualizations generated by the BRCA model for COAD and KICH subtypes. In these visualizations, green regions correspond to high attention, red regions to moderate attention, and blue regions to low or no attention during the classification inference. Ground truth attention maps for a given patch were obtained using a model trained on the same organ. By comparing the overlap between the model-specific and cross-organ visualizations, we assessed similarities in tumor morphology across different cancer types.

2.3.5 Statistical analysis of nuclear geometry in RoIs

Previous research has established a strong connection between nuclear morphology and tumor characteristics, including clinical outcomes and molecular profiles [9, 28, 29, 30]. Based on this, we hypothesized that regions highlighted by high model attention across different cancer types may share similar nuclear shapes and spatial arrangements, reflecting common aspects of the tumor microenvironment.

To investigate this, we analyzed the high-attention regions obtained from GradCAM and extracted two types of features:

- Nuclear shape features, capturing individual nucleus geometry such as area, convex area, perimeter, eccentricity, and major and minor axes.
- Inter-nuclear density features, describing spatial organization, including the number of neighboring nuclei within predefined radii.

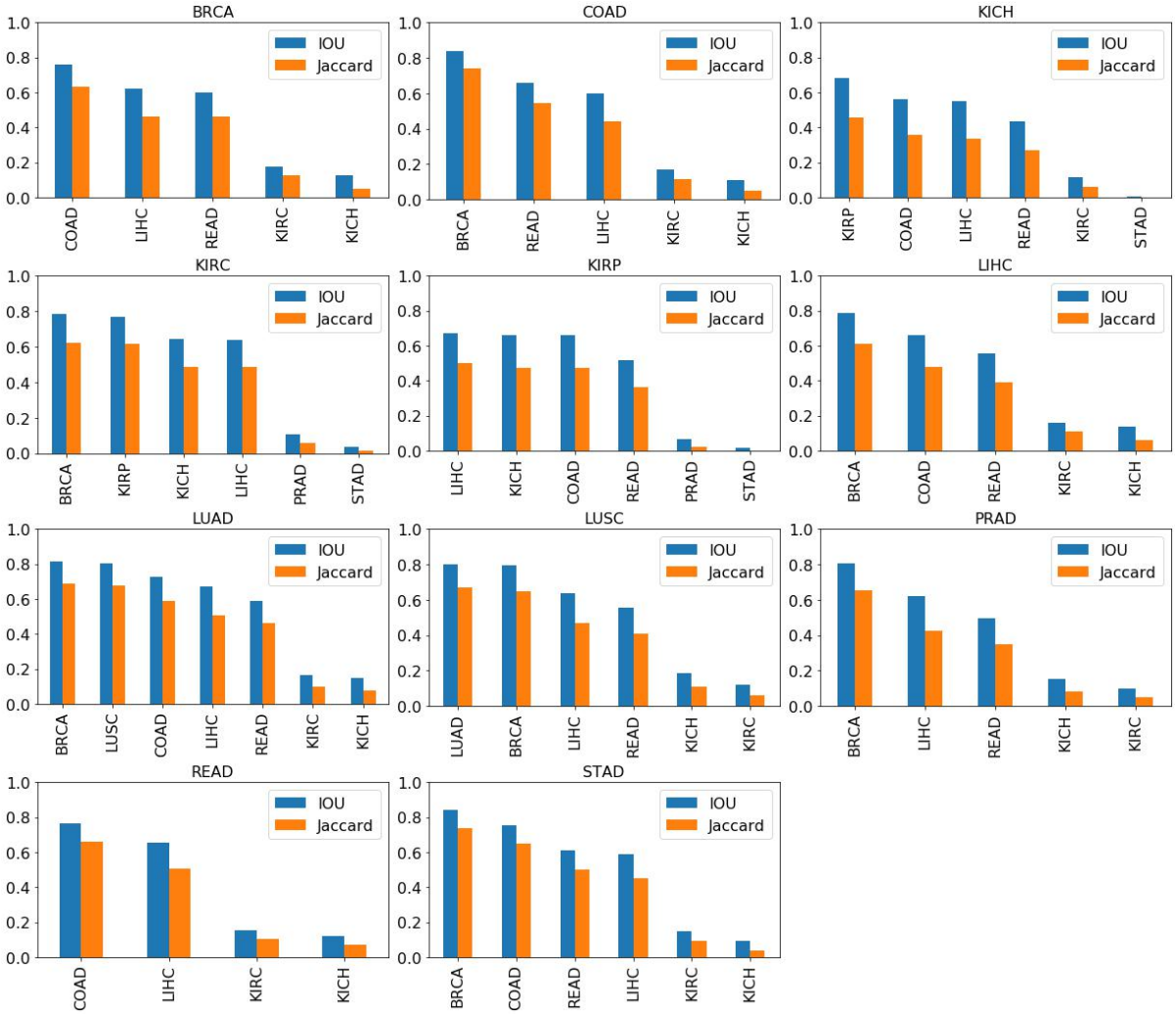

Figure 2.7: Cross-organ GradCAM results showing IoU and Jaccard index of high attention regions. The model used for visualization is indicated on the title of each plot and the subtypes used are indicated on the x axis.

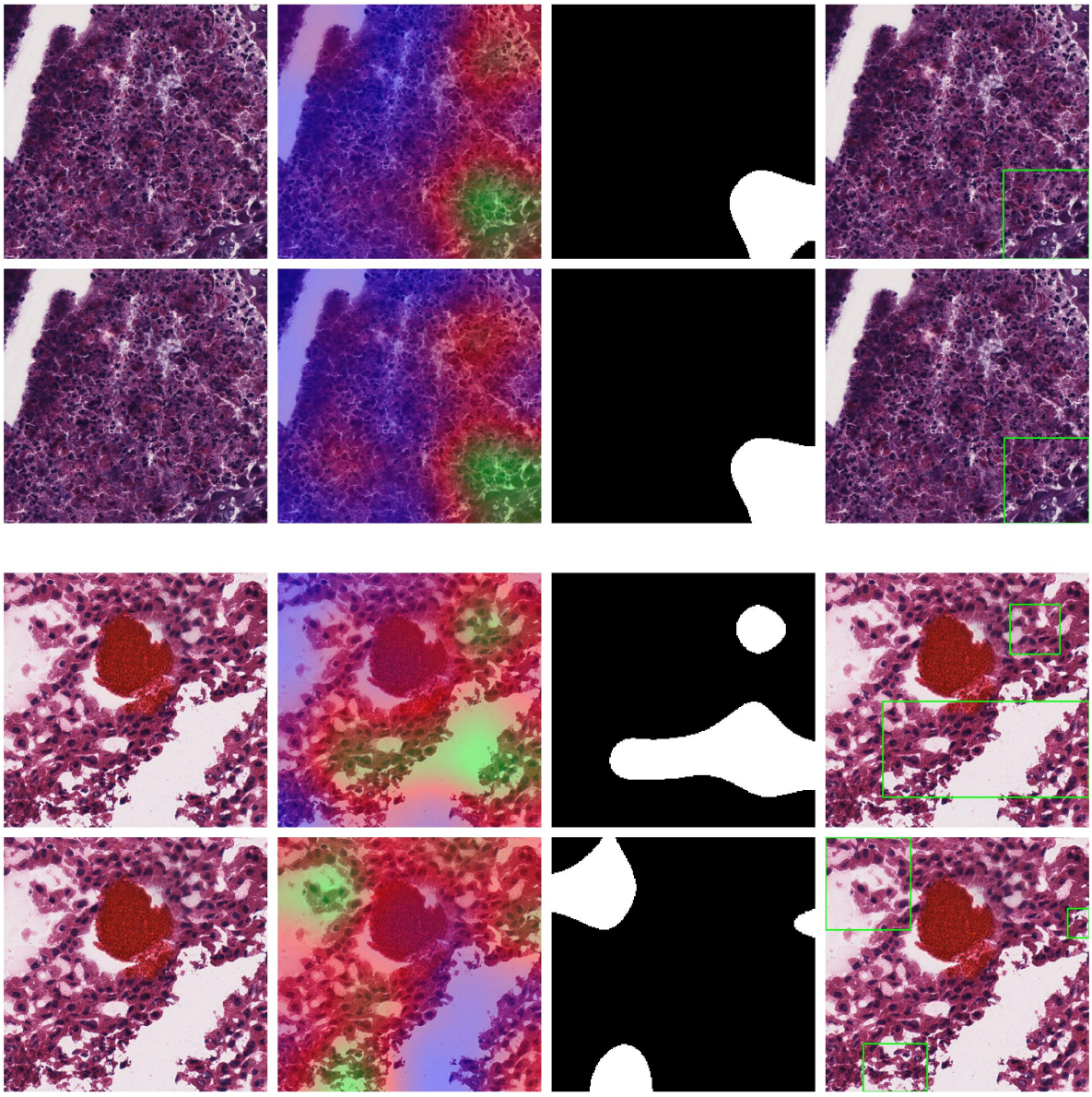

Figure 2.8: Cross Organ GradCAM visualization of BRCA model on COAD, and KICH cancer patches. Columns show the input patch, GradCAM output, GradCAM thresholded and GradCAM with bounding box respectively. Top 2 rows show COAD input patches and visualization using BRCA model (1st row) and COAD model (2nd row). Bottom 2 rows show KICH input patches and visualization using BRCA model (3rd row) and KICH model (4th row). Top 2 form a good pair whereas bottom 2 form a bad pair.

The analysis workflow is summarized in Figure 2.9 and involved the following stages:

- For each patch, regions of high attention were obtained from the model trained on the organ of interest and from a model trained on a different organ.
- Three distinct regions were considered: overlapping regions shared by both models, and organ-specific regions unique to each model. Overlapping areas were computed using a logical AND operation on thresholded GradCAM maps, while organ-specific regions were obtained by subtracting the overlap from each map.
- Each extracted region was processed using a hierarchical multilevel thresholding method to segment individual nuclei [31]. This ensured precise delineation of nuclei, even in dense clusters.
- Geometric features were calculated for each segmented nucleus, while inter-nuclear spatial relationships were captured using graph-based techniques [32].
- Probability distributions were computed for these features for all 3 regions and statistical tests were conducted to validate our hypothesis.

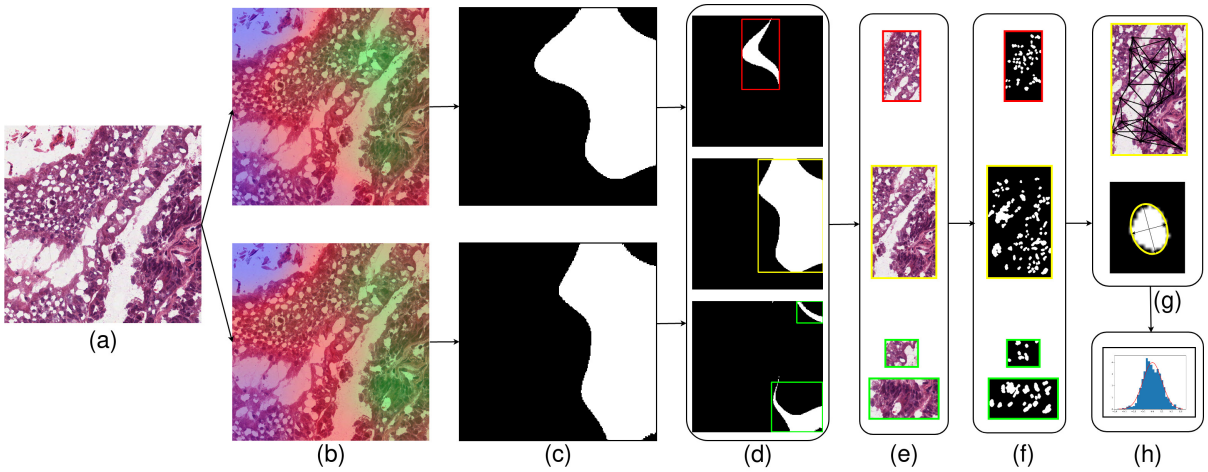

Figure 2.9: Nuclei segmentation workflow involved in segmenting nuclei from the specific regions of a sample patch: (a) COAD sample patch, (b) GradCAM outputs of BRCA model (top) and COAD model (bottom), (c) thresholded GradCAM mask, (d) BRCA-specific mask (top), overlapping mask (middle) and COAD-specific mask (bottom), (e) Masked regions of BRCA-specific(top), overlap(middle) and COAD-specific(bottom), (f) Nuclei segmented regions of BRCA-specific(top), overlap(middle) and COAD-specific(bottom), (g) Obtaining the nuclei shape and graphical features for each region, (h) Distributions of these features.

To characterize nuclear morphological patterns, we analyzed two groups based on cross-organ inference performance: one showing strong performance (BRCA and COAD) and another showing poor performance (BRCA and KICH). For this analysis, we selected high-confidence patches with predicted cancer probability greater than 0.98 (as per models trained on themselves respectively) from COAD and KICH samples.

Comparing the distributions of nuclear geometric features revealed notable differences between the groups. For example, features such as main region extent and solidity in the regions highlighted by BRCA and COAD models on COAD patches showed similar and correlated distributions. In contrast, the same features exhibited distinct and uncorrelated patterns when comparing BRCA and KICH models on KICH patches (Figure 2.10).

Statistical analysis further supported these observations. The Kolmogorov–Smirnov (KS) test was used to compare the feature distributions between model pairs in which the null hypothesis assumes that the two distributions are identical and the test statistic measures the maximum difference between their empirical cumulative distribution functions. Throughout this analysis, a p-value ≤ 0.05 was considered statistically significant, indicating that the observed distribution differences are unlikely to be explained by random variation alone and that the null hypothesis should be rejected. Conversely, p-values > 0.05 were interpreted as statistically insignificant, suggesting that any observed differences in the distributions can reasonably be attributed to random variation and that the distributions are indeed similar.

Eight nuclear shape features—including total area, main extent, and main region solidity—and three inter-nuclear density features, corresponding to neighbor counts within radii of 10, 20, and 30 pixels, were associated with the observed similarities in tumor morphology (p-values larger than 0.05; Table 2.2). As shown in Table 2.2, the BRCA–COAD comparison yields predominantly high p-values (mostly greater than 0.05), indicating that the majority of nuclear shape and inter-nuclear density features are statistically similar between these two cancer types. In contrast, the BRCA–KICH comparison produces predominantly very low p-values (often several orders of magnitude smaller than 0.05), indicating statistically significant differences in nuclear geometry and spatial organization. This difference in statistical behavior aligns with the observed contrast in cross-organ inference performance between the two pairs.

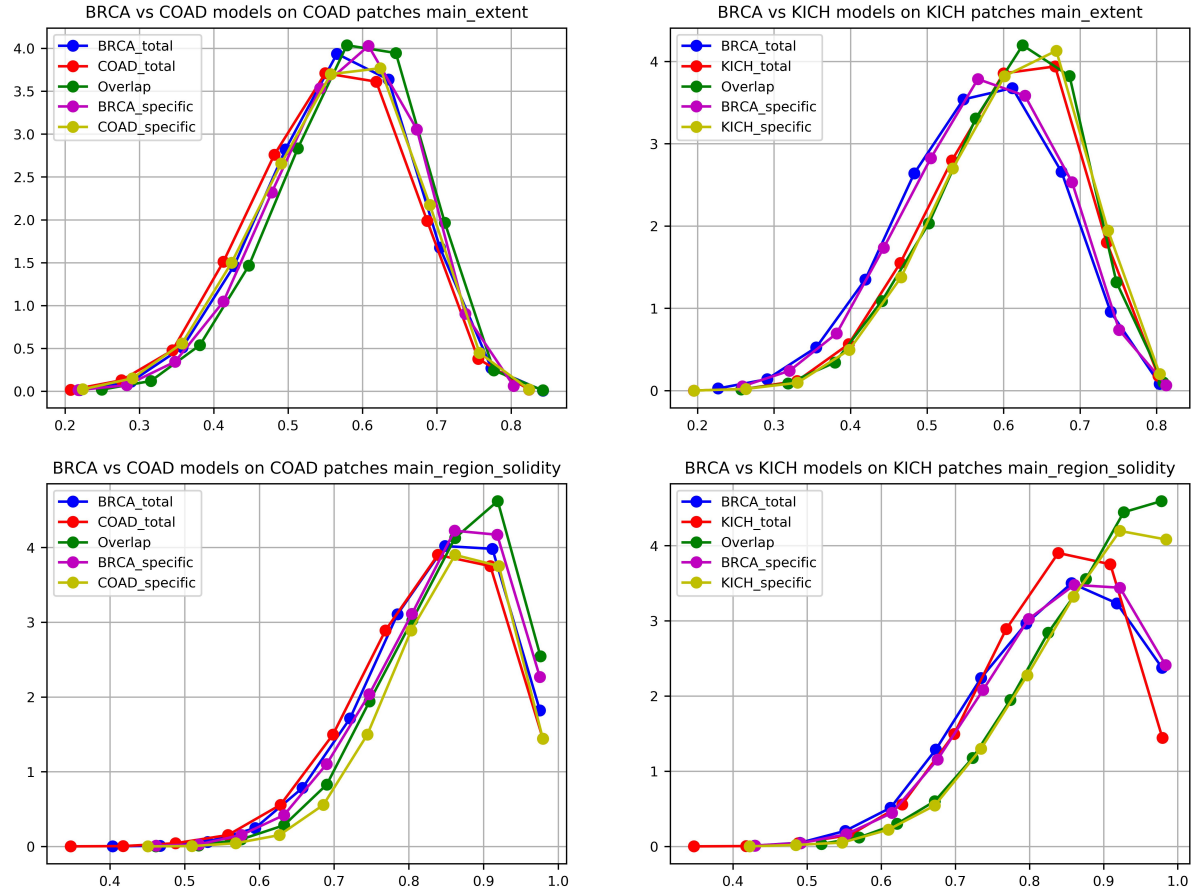

Figure 2.10: Distribution of nuclear shape features for BRCA and COAD models applied to COAD patches (left) and BRCA and KICH models applied to KICH patches (right). The X-axis indicates the feature values, while the Y-axis represents the probability density function. For each subplot, ”Total” corresponds to the entire high-attention region identified by the model, ”Overlap” indicates the shared high-attention region between the two models, and ”Specific” represents the portion of the total region excluding the overlap.

Feature Type	Features	p-value (BRCA and COAD)	p-value (BRCA and KICH)
Nuclei Shape Features	Total area	0.0736	1.0711E-132
	Total convex area	0.0823	4.6482E-147
	Total perimeter	0.0004	3.3526E-240
	Total filled area	0.0738	9.6456E-133
	Total major axis	0.0004	3.3976E-228
	Total minor axis	0.0002	1.8502E-194
	Total peri by area	0.0768	1.9164E-235
	Main region area	0.0171	2.0913E-246
	Main region convex area	0.0182	1.1829E-228
	Main region eccentricity	0.1387	4.3698E-18
	Main extent	0.1002	5.0677E-37
	Main region solidity	0.0583	2.4308E-34
	Main region perimeter	0.0007	0
	Main region angle	0.091	0.4033
	Main region peri by area	0.0142	1.4962E-203
	Main region major axis	0.0004	0
	Main region minor axis	0.0014	0
	Total diameter	0.0003	7.524E-225
Inter Nuclei Density Features	Neighbour Count within 10 pixels radius	0.5974	0.0009
	Neighbour Count within 20 pixels radius	0.6044	6.6978E-07
	Neighbour Count within 30 pixels radius	0.1945	1.0807E-19

Table 2.2: Nuclei Features Statistical Analysis: The table showing the p-values obtained after performing the KS-test on two pairs of groups (BRCA-COAD) and (BRCA-KICH). Higher p-value indicates similarity and lower p-value indicates differences. The high p-values of the BRCA-COAD group goes in agreement with the observed distribution plot

2.4 Conclusion

In this study, we investigated tumor similarities across multiple organs using deep learning. Our key findings are:

- Cross-Organ Cancer Feature Generalization: Models trained on one cancer type achieved high performance when tested on another, indicating that learned features capture patterns shared across cancers from different organs.
- Feature Space Separability: Visualization using t-SNE demonstrated clear separation between normal and cancerous patches even though model trained on another organ was used for feature extraction, further supporting the presence of shared discriminative features of multiple cancers
- GradCAM Attention Analysis: Models with strong cross-organ performance exhibited significant overlap in attention regions whereas models with poor cross-organ performance exhibit lower overlap, suggesting that common morphologically salient areas drive classification across cancers.
- Nuclei-Level Morphology: Statistical analysis of geometric and spatial features of nuclei in overlapping and unique attention regions confirmed morphological similarities at the level of nuclear geometry.
- Overall, this work provides a computational proof-of-principle that deep learning can reveal shared tumor morphology across organs, laying the foundation for further experimental validation in future studies.

Chapter 3

MIL framework for classification of lung cancer subtypes in Indian population

Among the various types of cancer, lung cancer stands out due to its high incidence and lethality worldwide, largely driven by risk factors such as tobacco use, environmental pollutants, and genetics. In India, respiratory cancers—including those of the trachea, bronchus, and lung—constitute approximately 5.8% of all new cancer cases and account for 8.2% of all cancer-related deaths, ranking fourth in both incidence and mortality, following breast, oral, and cervical cancers [2]. However, gender-specific patterns show that among males, respiratory cancers rank second in both incidence (8.6%) and mortality (11.6%), following oral cancer [2]. In contrast, female cancer burden is dominated by breast and cervical cancers, with respiratory cancers contributing far less to overall incidence and deaths. This gender disparity can be attributed to higher occupational exposure to carcinogenic materials such as asbestos, higher rates of tobacco consumption among men, and greater contact with industrial pollutants [33].

3.1 Biology of Lung Carcinomas

Lung carcinomas are broadly classified into small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC). NSCLC accounts for the majority of lung cancer cases. NSCLC is further classified into two subtypes: lung adenocarcinoma and lung squamous cell carcinoma. These subtypes are defined by their histological patterns and cellular differentiation observed under the microscope. Adenocarcinoma is more common in non-smokers and often linked to genetic mutations, whereas squamous cell carcinoma is strongly associated with smoking.

Adenocarcinomas are characterized by glandular differentiation and/or mucin production. They commonly display acinar, papillary, micropapillary, lepidic, or solid growth patterns and are often lo-

cated in peripheral lung regions. Squamous cell carcinomas show squamous differentiation, including keratinization and intercellular bridges, and are more frequently found as centrally located bronchial tumors.

While histology is the primary basis for classification, some cases are difficult to resolve by histology alone. In such situations, immunohistochemistry (IHC) provides reliable distinction between subtypes. IHC uses antibodies to detect subtype-specific proteins in tissue, producing a color change where the protein is present. Adenocarcinomas typically stain positively for TTF-1 and Napsin-A, and negatively for p40 and p63. Squamous cell carcinomas show the reverse profile, with positivity for p40 and p63 and negativity for TTF-1 and Napsin-A. Taken together, histological features and IHC markers define the gold-standard labels used in subtype classification.

Figure 3.1 shows representative histological patterns of adenocarcinoma and squamous cell carcinoma. The images have been adapted from [34] and [35].

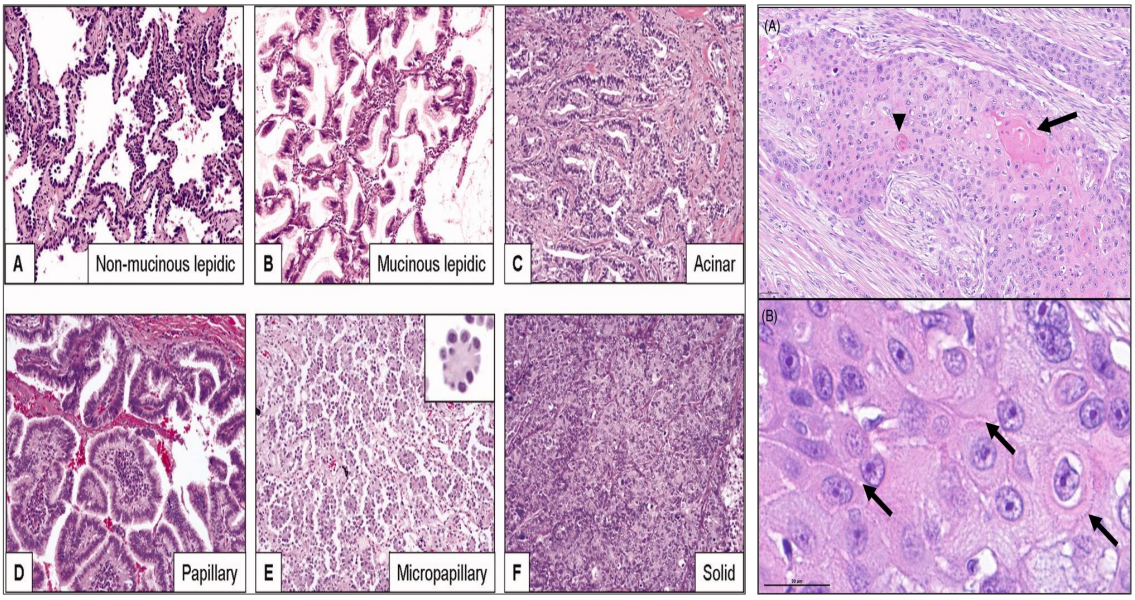

Figure 3.1: Histological patterns of lung carcinoma. Left: lepidic (A,B), acinar (C), papillary (D), micropapillary (E) and solid (F) patterns of adenocarcinoma. Right: keratin pearls (A) and intercellular bridging (B) patterns of squamous cell carcinoma.

3.2 Dataset Details

We use two datasets in our experiments: TCGA-Lung and IPD-Lung. TCGA-Lung is the lung subset of The Cancer Genome Atlas (TCGA), a large public repository of cancer whole-slide images (WSIs) [19]. It contains 1026 cases, comprising 522 lung adenocarcinoma and 504 lung squamous cell carcinoma cases. IPD-Lung, or Indian Pathology Dataset - Lung, is curated in-house. It consists of 345 lung cancer cases obtained from the Nizam's Institute of Medical Sciences (NIMS) hospital archive, with some patients contributing two slides, for a total of 358 WSIs. All cases are from the past 3–4 years, and the histopathological slides have been carefully preserved. Hematoxylin and Eosin (H&E) re-staining when necessary. Subtype labels were assigned by pathologists based on the histological markers described in Section 3.1.

We conducted a review process to ensure image quality, and WSIs with poor quality were re-scanned. In diagnostically challenging cases, slides where histology suggested a subtype and IHC was used only for confirmation were included, while those with no histological clue requiring IHC for diagnosis were excluded. After review and filtering, the IPD-Lung dataset comprises 262 adenocarcinoma and 77 squamous cell carcinoma cases. The imbalance in subtype distribution reflects the higher incidence of adenocarcinoma in the Indian population.

Table 3.1: IPD-Lung dataset summary

	Adenocarcinoma	**Squamous**	**Total**
Whole-slide images	271	87	358
Cases before review	262	83	345
Cases after review	262	77	339

3.3 Model Architecture

In whole-slide images (WSIs), slide-level labels correspond to specific regions within large gigapixel images, which consist of thousands of patches whose numbers vary across samples. The challenge is to aggregate patch-level features into a slide-level representation that accurately reflects diagnostic or prognostic labels.

A naive approach is to average patch features along each dimension:

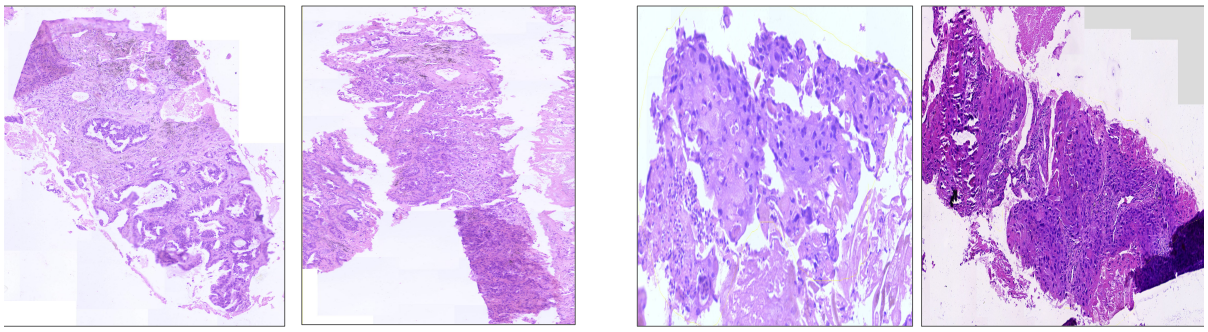

Figure 3.2: WSI examples from IPD-Lung: the two on the left are lung adenocarcinoma, and the two on the right are lung squamous cell carcinoma.

$$z = \frac{1}{K} \sum_{k=1}^{K} h_k$$

However, not all patches are equally informative, and averaging can dilute discriminative information by giving equal weight to irrelevant patches.

Attention-based multiple instance learning (AMIL) [36] overcomes this limitation by introducing a learnable weighting mechanism:

$$z = \sum_{k=1}^{K} a_k h_k$$

where h_k is the feature embedding of the k-th patch and a_k is its attention weight, learned through a small two-layer neural network:

$$a_k = \frac{\exp(w^T \tanh(V h_k^T))}{\sum_{j=1}^{K} \exp(w^T \tanh(V h_j^T))}$$

This mechanism allows the network to emphasize patches with strong morphological signatures of cancer (e.g., glandular structures or keratin pearls) while suppressing irrelevant tissue regions such as stroma. The aggregated embedding z is then fed into a downstream classifier or regressor for slide-level prediction.

A variant, called gated attention, uses two parallel attention pathways with a multiplicative gate:

$$a_k = \frac{\exp\left(w^T \left[\tanh(V h_k^T) \odot \sigma(U h_k^T)\right]\right)}{\sum_{j=1}^{K} \exp\left(w^T \left[\tanh(V h_j^T) \odot \sigma(U h_j^T)\right]\right)}$$

This allows the network to model more complex, non-linear interactions between patch features. In practice, we observed that gated attention did not yield a significant performance gain over standard attention.

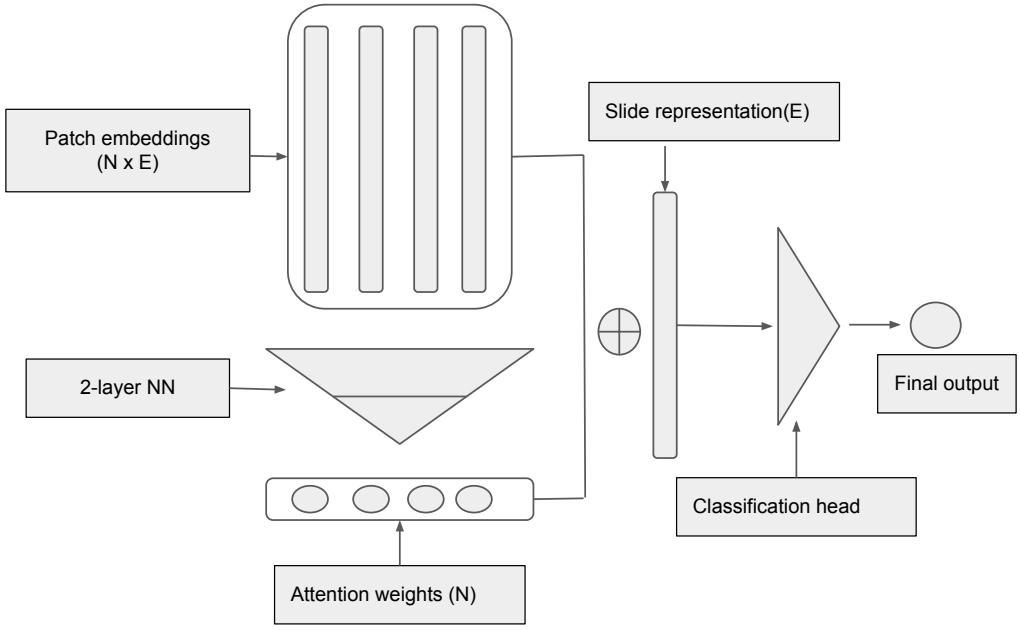

Figure 3.3: Block diagram of attention-based MIL.

Lu et al. [37] proposed Clustering-constrained Attention MIL (CLAM), which extends Ilse's framework by introducing instance-level clustering. The top K highest-ranked patches are encouraged to match the slide label, while the lowest-ranked K patches are encouraged not to match. This is achieved by increasing the depth of the attention network and adding an auxiliary patch-level classification task with a soft SVM loss. A value of $K = 8$ is typically effective for most cases.

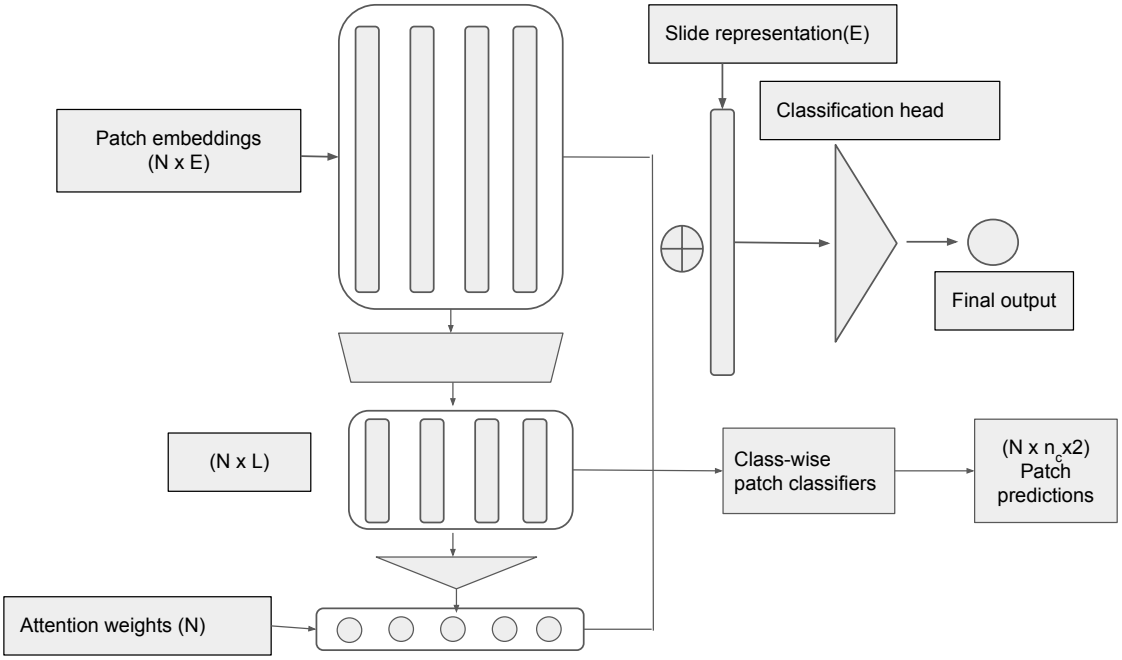

Figure 3.4: Block diagram of clustering-constrained attention-based MIL (CLAM).

3.4 Experiments and Results

We split both the IPD-Lung and TCGA-Lung datasets into training and test sets in 80:20 ratio. Table 3.2 summarizes the number of cases for each subtype in the train and test sets.

Dataset	Subtype	Train	Test	Total
IPD-Lung	Adenocarcinoma	210	52	262
	Squamous carcinoma	62	15	77
TCGA-Lung	Adenocarcinoma	418	104	522
	Squamous carcinoma	403	101	504

Table 3.2: Train-test split of IPD-Lung and TCGA-Lung datasets.

3.4.1 Inference using TCGA-trained model

For our first experiment, we trained MIL-based models on the TCGA-Lung dataset for the subtype classification task and evaluated them on both TCGA-Lung and IPD-Lung test sets. The IPD-Lung dataset differs from TCGA-Lung in scanners, tissue preparation, and H&E staining. To mitigate

these differences, we applied stain normalization to standardize color distributions across slides of both datasets. Specifically, we experimented with Macenko, Reinhard, and Vahadane methods [38, 39, 40]. Macenko uses color deconvolution to align stain intensity distributions, Reinhard adjusts the mean and standard deviation of LAB channels to match a reference image, and Vahadane decomposes the image into stain concentration and color matrices to generate a normalized image using a reference stain matrix. We found Macenko to yield the best performance and incorporated it into our feature extraction pipeline.

Table 3.3 reports the performance of TCGA-trained models, including AMIL, single-branch CLAM (CLAM-SB), and multi-branch CLAM (CLAM-MB), on both test sets.

Test Set	**Model**	**LUAD**			**LUSC**			**Overall**	
		P	**R**	**F1**	**P**	**R**	**F1**	**Acc (%)**	**Macro F1**
TCGA-Lung	AMIL	0.96	0.83	0.89	0.81	0.95	0.88	0.88	0.89
	CLAM-SB	0.97	0.84	0.90	0.82	0.95	0.89	0.89	0.90
	CLAM-MB	0.96	0.83	0.89	0.81	0.95	0.88	0.88	0.89
IPD-Lung	AMIL	0.82	0.97	0.89	0.75	0.19	0.30	0.76	0.60
	CLAM-SB	0.83	0.98	0.90	0.76	0.20	0.31	0.77	0.61
	CLAM-MB	0.82	0.97	0.89	0.75	0.19	0.30	0.76	0.60

Table 3.3: Performance metrics of TCGA-Lung trained models on the TCGA-Lung test set (top three rows) and IPD-Lung test set (bottom three rows).

From Table 3.3, the following observations can be made:

- Models trained on TCGA-Lung performed well on the TCGA-Lung test set, with CLAM-SB achieving the best results.

- On the IPD-Lung test set, models show good performance for adenocarcinoma but poor performance for squamous carcinoma, particularly with very low recall for the squamous class.

- All models exhibit similar performance trends within each test set.

3.4.2 Domain Adaptation Experiment

To investigate the potential for domain adaptation, we implemented a MIL-GRL model by extending the Gradient Reversal Layer (GRL) framework [41] to the MIL setting. Our goal was to improve the performance of a TCGA-Lung-trained model on the IPD-Lung test set without using labels from the training set of target domain. GRL introduces an auxiliary domain classifier head alongside the main task-specific subtype classification head. During the forward pass, this domain head predicts the domain of each sample, while in the backward pass, its gradient is multiplied by $-\lambda$, encouraging the model to learn domain-invariant features.

However, the MIL-GRL approach did not lead to significant improvements. A likely reason is that MIL models primarily perform feature aggregation, so if the patch-level features exhibit substantial variability between domains, enforcing domain invariance becomes challenging. Moreover, the differences between TCGA-Lung and IPD-Lung—arising from scanners, tissue preparation, and staining—may be too complex for a simple GRL-based approach, indicating that more advanced domain adaptation strategies may be necessary for this dataset pair.

3.4.3 Model training and testing on IPD-Lung

In this set of experiments, models were trained on the IPD-Lung training set and evaluated on the held-out IPD-Lung test set. In addition to training on entire WSIs, we obtained expert annotations of regions of interest (RoIs) from a highly experienced pathologist.

The highest performance was observed when training was restricted to these annotated RoIs. Since squamous carcinoma was more challenging to classify, RoI annotation efforts focused specifically on this subtype. Figure 3.5 shows examples of these annotated regions. The figure was generated using QuPath [42].

Table 3.4 summarizes model performance across different training settings, including variations in patch size (20×256 and 40×512) and the area used for feature extraction (full WSI or RoIs). The first column encodes three key details: patch size, MIL model used, and the WSI area for feature extraction. For example, "20×256 CLAM-SB (all)" indicates 256-pixel patches at 20× magnification, using the CLAM-SB model, with features extracted from the entire WSI.

Training on annotated RoIs substantially improved classification performance on squamous carcinoma while maintaining strong performance on adenocarcinoma. Specifically, for 20×256 patching

with CLAM-SB, precision for squamous carcinoma increased from 63% to 74% and recall improved from 49% to 66% leading to improvement in F1 score from 55% to 70%.

Setting	Adeno			Squamous			Overall	
	P	R	F1	P	R	F1	Acc	Macro avg F1
20 x 256 CLAM-SB (all)	0.90	0.94	0.92	0.63	0.49	0.55	0.87	0.74
40 x 512 CLAM-SB (all)	0.91	0.94	0.93	0.63	0.56	0.59	0.87	0.76
20 x 256 CLAM-SB (ROI)	0.90	0.93	0.92	0.74	0.66	0.70	0.87	0.81
40 x 512 CLAM-SB (ROI)	0.89	0.93	0.91	0.71	0.60	0.65	0.86	0.78

Table 3.4: Performance metrics for different settings with CLAM-SB model with 20x and 40x patching

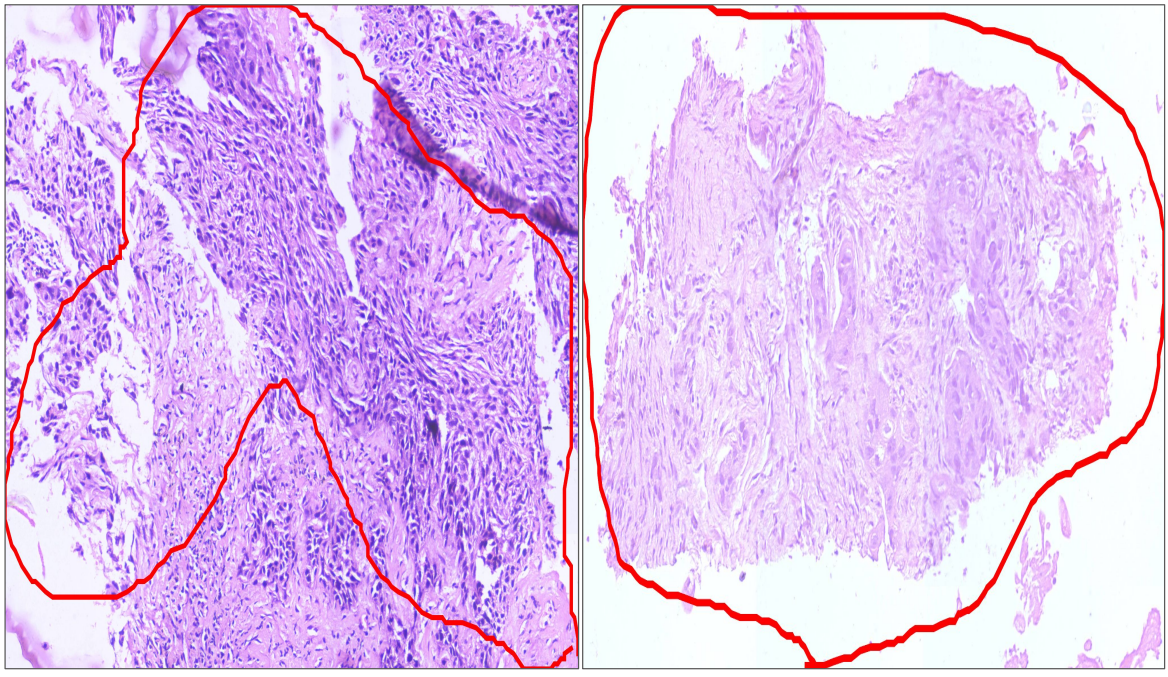

Figure 3.5: Examples of annotated regions of interest (RoIs) of squamous class used for experiment.

3.5 Conclusion

The experiments conducted in this study highlight several important insights regarding MIL-based subtype classification of lung cancer in diverse cohorts.

First, models trained on the TCGA-Lung dataset achieved strong performance when evaluated on the TCGA-Lung test set, particularly with the CLAM-SB model, confirming the effectiveness of MIL approaches for WSI-based subtype classification. However, when applied to the IPD-Lung test set, performance decreased notably for squamous carcinoma, with very low recall values, while adenocarcinoma classification remained relatively robust. This drop in performance can be attributed to differences in the two datasets with respect to scanners, tissue preparation, and H&E staining between the two datasets.

We attempted to mitigate this domain shift using stain normalization techniques, including Macenko, Reinhard, and Vahadane methods. Among these, Macenko normalization provided the best results and was incorporated into the feature extraction pipeline. Nevertheless, even with stain normalization, the TCGA-trained models struggled to perform well on the squamous subtype class of the IPD-Lung test set, highlighting that stain variability is only one of multiple factors contributing to cross-dataset performance degradation. The MIL-GRL domain adaptation experiment was performed to address this challenge by encouraging the model to learn domain-invariant features. However, this approach did not yield substantial improvements.

Finally, training models directly on the IPD-Lung dataset demonstrated the highest potential for high classification performance. Restricting training to expert-annotated regions of interest (RoIs) significantly improved classification performance on squamous carcinoma subtype class without compromising performance on adenocarcinoma subtype class. This indicates that targeted annotation of challenging subtypes can be an effective strategy to enhance MIL model performance in datasets with limited or imbalanced samples.

Overall, these findings following key points:

1. MIL-based models are highly effective for WSI-level subtype classification when trained and evaluated within the same domain.

2. Domain shifts between datasets pose significant challenges that are not fully addressed by simple stain normalization or GRL-based adaptation.

3. Targeted expert annotations of RoIs can substantially improve model performance for underrepresented or difficult subtypes, providing a practical pathway for enhancing classification in new datasets.

Conclusion and Future Work

This thesis presents a comprehensive investigation of deep learning-based computational pathology across multiple organs and cohort-specific datasets. In the first study, we demonstrated that patch-level CNNs trained on TCGA WSIs capture features that generalize across organs, revealing shared tumor morphology. Cross-organ inference, t-SNE visualization, GradCAM attention overlap, and nuclei-level morphological analysis collectively confirmed that deep models can extract discriminative and transferable features, providing computational evidence of conserved cancer histology across different organs.

The second study focused on lung cancer in the Indian population via the IPD-Lung dataset. MIL-based models trained on TCGA-Lung performed well within the same dataset but exhibited substantial performance degradation on IPD-Lung, particularly for squamous carcinoma, highlighting the impact of cohort-specific factors such as scanner variability, tissue preparation, and staining. Stain normalization (Macenko) and MIL-GRL domain adaptation provided limited improvements, whereas training directly on IPD-Lung significantly enhanced performance. Notably, restricting training to pathologist-annotated regions of interest (RoIs) yielded the highest gains for the squamous carcinoma subtype, demonstrating that targeted expert annotation is an effective strategy for improving MIL model performance for challenging or underrepresented subtypes.

Overall, these studies underscore that deep learning models in computational pathology are sensitive not only to intrinsic tissue features but also to cohort- and dataset-specific factors. Careful dataset curation, expert annotation, and cohort-aware training strategies are crucial for building reliable and generalizable WSI-level classification models.

Several avenues for future research emerge from this work. First, extending cross-organ analyses to additional cancer types and incorporating multi-modal data (e.g., genomics and immunohistochemistry) could deepen understanding of conserved tumor morphology. Second, advanced domain adaptation techniques, including self-supervised and adversarial feature alignment methods, should be explored to mitigate performance drops across cohorts more effectively than simple stain normalization or GRL-

based approaches. Third, semi-automated or AI-assisted RoI annotation tools could reduce reliance on exhaustive expert labeling, enabling more scalable improvements for subtypes such as squamous carcinoma in small and imbalanced datasets. Finally, studies evaluating how models trained on curated datasets perform in real-world clinical settings would help bridge the gap between computational pathology research and clinical deployment, ensuring models remain robust and generalizable across diverse populations.

List of Publications

1. Ashish Menon[†], **Piyush Singh**[†], P. K. Vinod, and C. V. Jawahar, "Exploring Histological Similarities Across Cancers From a Deep Learning Perspective," *Frontiers in Oncology*, vol. 12, 2022.

2. Ashish Menon, **Piyush Singh**, P. K. Vinod, and C. V. Jawahar, "Interactive Learning for Assisting Whole Slide Image Annotation," in *Proceedings of the Asian Conference on Pattern Recognition*, 2021.

[†] Joint first authors (equal contribution).

Bibliography

[1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 74, pp. 229 – 263, 2024. 1

[2] K. Singh, A. Grover, and K. Dhanasekaran, "Unveiling the cancer epidemic in india: a glimpse into globocan 2022 and past patterns," *The Lancet Regional Health - Southeast Asia*, vol. 34, 2025. 4, 25

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105. 5

[4] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, pp. 211 – 252, 2014. 5

[5] A. Qayyum, S. M. Anwar, M. Majid, M. Awais, and M. R. Alnowami, "Medical image analysis using convolutional neural networks: A review," *Journal of Medical Systems*, vol. 42, pp. 1–13, 2017. 5

[6] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, 2021. 5

[7] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proceedings of the ieee conference on computer vision and pattern recognition*, 2016, pp. 2424–2433. 5

[8] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, 2016. 5, 6

[9] S. Tabibu, P. Vinod, and C. Jawahar, "Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019. 5, 6, 8, 17

[10] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature medicine*, vol. 24, no. 10, pp. 1559–1567, 2018. 5, 6, 8

[11] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *ArXiv*, vol. abs/1606.05718, 2016. 5, 6

[12] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, *et al.*, "Detecting cancer metastases on gigapixel pathology images," *arXiv preprint arXiv:1703.02442*, 2017. 5

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826. 6

[14] S. Kwok, "Multiclass classification of breast cancer in whole-slide images," in *International Conference on Image Analysis and Recognition*, 2018. 6

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9. 6

[16] Y. Fu, A. W. Jung, R. V. Torne, S. Gonzalez, H. Vöhringer, A. Shmatko, L. R. Yates, M. Jimenez-Linan, L. Moore, and M. Gerstung, "Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis," *Nature Cancer*, vol. 1, no. 8, pp. 800–810, 2020. 6

[17] A. Cheerla and O. Gevaert, "Deep learning with multimodal representation for pancancer prognosis prediction," *Bioinformatics*, vol. 35, no. 14, pp. i446–i454, 2019. 6

[18] J. Noorbakhsh, S. Farahmand, A. F. pour, S. Namburi, D. Caruana, D. Rimm, M. Soltanieh-ha, K. Zarringhalam, and J. H. Chuang, "Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images," *bioRxiv*, 2020. 6, 14

[19] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The cancer genome atlas (tcga): an immeasurable source of knowledge," *Contemporary oncology*, vol. 19, no. 1A, p. A68, 2015. 6, 27

[20] L. A. Cooper, E. G. Demicco, J. H. Saltz, R. T. Powell, A. Rao, and A. J. Lazar, "Pancancer insights from the cancer genome atlas: the pathologist's perspective," *The Journal of pathology*, vol. 244, no. 5, pp. 512–524, 2018. 6

[21] D. Komura and S. Ishikawa, "Machine learning methods for histopathological image analysis," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 34–42, 2018. 8

[22] G. J. S. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017. 8

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. 8

[24] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis." *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017. 8

[25] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:196194314 9

[26] N. Pandey, V. Lanke, and P. K. Vinod, "Network-based metabolic characterization of renal cell carcinoma," *Scientific Reports*, vol. 10, 2020. 14

[27] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, pp. 336–359, 2019. 16

[28] X. Zhan, J. Cheng, Z. Huang, Z. Han, B. R. Helm, X. Liu, J. Zhang, T.-F. Wang, D. Ni, and K. Huang, "Correlation analysis of histopathology and proteogenomics data for breast cancer*," *Molecular & Cellular Proteomics*, vol. 18, pp. S37 – S51, 2019. 17

[29] J. Cheng, J. Zhang, Y. Han, X. Wang, X. Ye, Y. Meng, A. V. Parwani, Z. Han, Q. Feng, and K. Huang, "Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis." *Cancer research*, vol. 77 21, pp. e91–e100, 2017. 17

[30] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. N. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009. 17

[31] H. A. Phoulady, D. B. Goldgof, L. O. Hall, and P. R. Mouton, "Nucleus segmentation in histology images with hierarchical multilevel thresholding," in *SPIE Medical Imaging*, 2016. 20

[32] S. Doyle, S. C. Agner, A. Madabhushi, M. D. Feldman, and J. E. Tomaszeweski, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 496–499, 2008. 20

[33] V. Noronha, R. Pinninti, V. M. Patil, A. D. Joshi, and K. Prabhash, "Lung cancer in the indian subcontinent," *South Asian Journal of Cancer*, vol. 5, pp. 95 – 103, 2016. 25

[34] L. M. Solis, C. Behrens, M. G. Raso, H. Y. Lin, H. Kadara, P. Yuan, H. Galindo, X. Tang, J. J. Lee, N. Kalhor, I. I. Wistuba, and C. A. Moran, "Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome," *Cancer*, vol. 118, 2012. 26

[35] S. Berezowska, M. Maillard, M. Keyter, and B. Bisig, "Pulmonary squamous cell carcinoma and lymphoepithelial carcinoma – morphology, molecular characteristics and differential diagnosis," *Histopathology*, vol. 84, 2023. 26

[36] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *International Conference on Machine Learning*, 2018. 28

[37] M. Y. Lu, D. F. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, "Data-efficient and weakly supervised computational pathology on whole-slide images," *Nature Biomedical Engineering*, vol. 5, no. 6, pp. 555–570, 2021. 29

[38] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1107–1110, 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:15008471 31

[39] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, pp. 34–41, 2001. [Online]. Available: https://api.semanticscholar.org/CorpusID:14088925 31

[40] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1962–1971, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:6994762 31

[41] Y. Ganin and V. S. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International Conference on Machine Learning*, 2014. 32

[42] P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. J. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, J. A. James, M. Salto-Tellez, and P. Hamilton, "Qupath: Open source software for digital pathology image analysis," *Scientific Reports*, vol. 7, 2017. 32